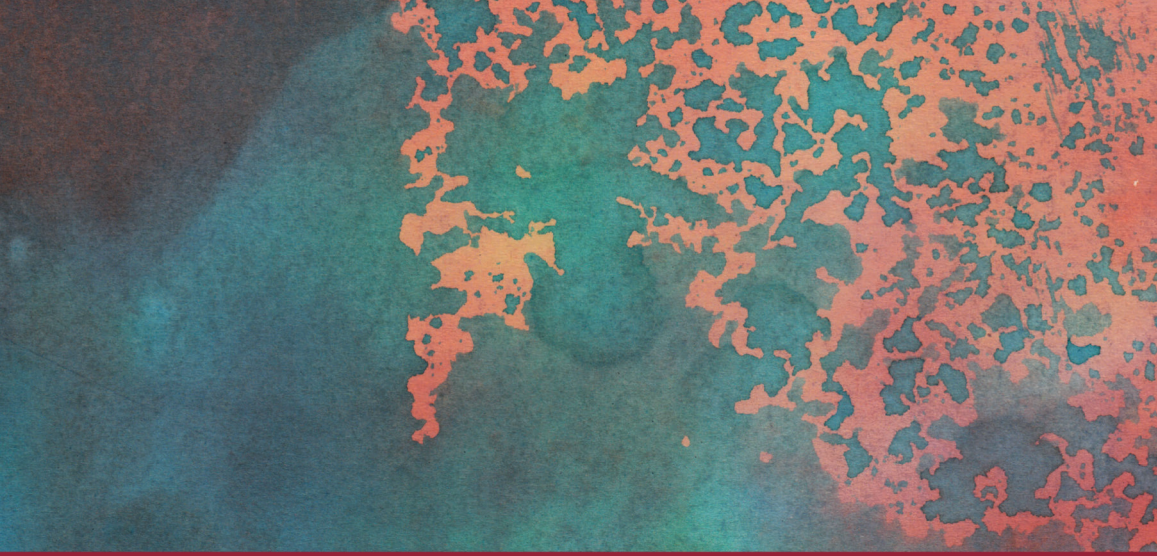
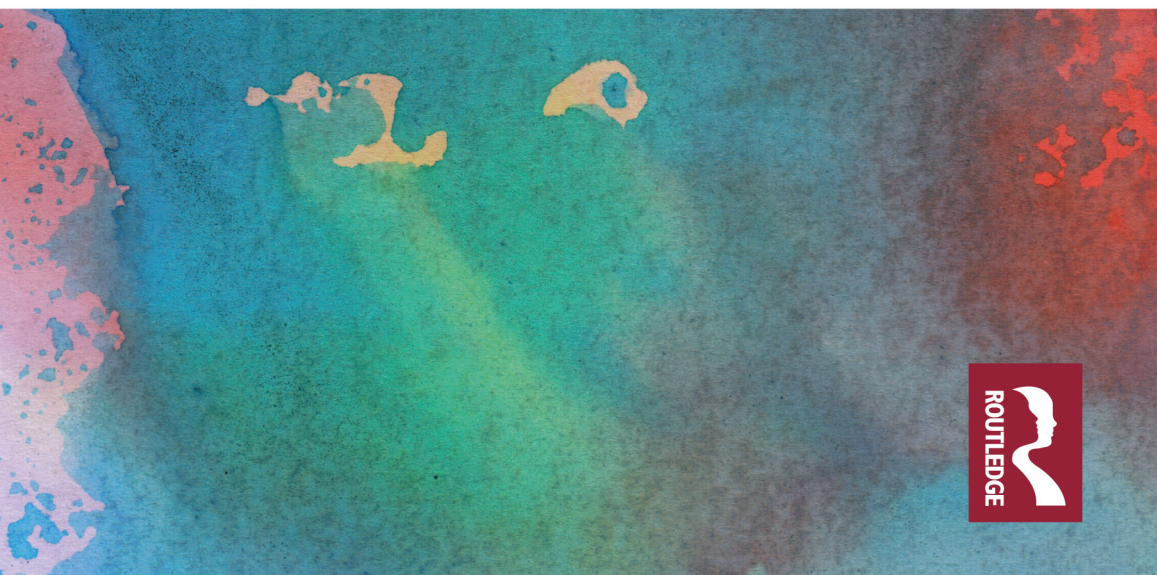


Routledge Studies in Applied Linguistics

VOCABULARY AND THE FOUR SKILLS

PEDAGOGY, PRACTICE, AND IMPLICATIONS FOR TEACHING VOCABULARY

Edited by
Jon Clenton and Paul Booth



ROUTLEDGE

Vocabulary and the Four Skills

This edited volume provides a single coherent overview of vocabulary teaching and learning in relation to each of the four skills (reading, writing, listening, speaking).

Each of the four sections presents a skill area with two chapters presented by two leading experts in the field, relating recent advances in the field to the extent that each skill area relates differently to vocabulary and how this informs pedagogy and policy. The book opens with a summary of recent advances in the field of vocabulary, and closes by drawing conclusions from the skill areas covered.

The chapters respond to emerging vocabulary research trends that indicate that lexical acquisition needs to be treated differently according to the skill area. The editors have chosen chapters to respond to recent research advances and to highlight practical and pedagogical application in a single coherent volume.

Jon Clenton is Associate Professor at Hiroshima University, Japan. His main research interests include the assessment of vocabulary knowledge, L2 vocabulary development in terms of bilingual models, second language acquisition, word association studies, lexical processing, and L2 measurement tools. He has examined several vocabulary measurement tools, with focus on attempts to isolate the construct of productive vocabulary knowledge.

Paul Booth is Senior Lecturer at Kingston University, London, and teaches both undergraduate and postgraduate courses. His research interests include second language vocabulary, especially individual differences, lexical diversity, and frequency. He has also researched situated learning and language teacher development.

Routledge Studies in Applied Linguistics

The Social Lives of Study Abroad
Understanding Second Language Learners' Experiences through Social Network Analysis and Conversation Analysis
Atsushi Hasegawa

Defining and Assessing Lexical Proficiency
Agnieszka Leńko-Szymańska

Global Perspectives on Project-based Language Learning, Teaching, and Assessment
Key Approaches, Technology Tools, and Frameworks
Edited by Gulbahar H. Beckett and Tammy Slater

Referring in a Second Language
Studies on Reference to Person in a Multilingual World
Edited by Jonothan Ryan and Peter Crosthwaite

Aspects of Language Development in an Intensive English Program
Alan Juffs

First Language Influences on Multilingual Lexicons
Edited by Paul Booth and Jon Clenton

Corpus Analysis in Different Genres
Academic Discourse and Learner Corpora
Edited by María Luisa Carrió-Pastor

Vocabulary and the Four Skills
Pedagogy, Practice, and Implications for Teaching Vocabulary
Edited by Jon Clenton and Paul Booth

For more information about this series, please visit: https://www.routledge.com/Routledge-Studies-in-Applied-Linguistics/book-series/RSAL

Vocabulary and the Four Skills

Pedagogy, Practice, and Implications for Teaching Vocabulary

Edited by
Jon Clenton and Paul Booth

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2021
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

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

Library of Congress Cataloging-in-Publication Data
Names: Clenton, Jon, editor. | Booth, Paul (Linguist), editor.
Title: Vocabulary and the four skills: pedagogy, practice, and implications for teaching vocabulary / edited by Jon Clenton and Paul Booth.
Description: First. | New York: Routledge, 2020. |
Series: Routledge studies in applied linguistics |
Includes bibliographical references and index.
Identifiers: LCCN 2020004835 (print) | LCCN 2020004836 (ebook) |
ISBN 9780367249977 (hardback) | ISBN 9780367503284 (paperback) |
ISBN 9780429285400 (ebook)
Subjects: LCSH: Vocabulary–Study and teaching. |
Language and languages–Study and teaching. |
Second language acquisition–Study and teaching.
Classification: LCC P53.9 .V6325 2020 (print) |
LCC P53.9 (ebook) | DDC 418.0071–dc23
LC record available at https://lccn.loc.gov/2020004835
LC ebook record available at https://lccn.loc.gov/2020004836

ISBN: 978-0-367-24997-7 (hbk)
ISBN: 978-0-429-28540-0 (ebk)

Typeset in Times New Roman
by Newgen Publishing UK

Contents

Figures

Tables

Contributors

Paul Booth is Senior Lecturer at Kingston University, London, and teaches both undergraduate and postgraduate courses. His research interests include second language vocabulary, especially individual differences, lexical diversity and frequency. He has also researched situated learning and language teacher development.

Jon Clenton is Associate Professor at Hiroshima University, Japan. His main research interests include the assessment of vocabulary knowledge, L2 vocabulary development in terms of bilingual models, second language acquisition, word association studies, lexical processing, and L2 measurement tools. He has examined several vocabulary measurement tools, with focus on attempts to isolate the construct of productive vocabulary knowledge.

Dion Clingwall is an Associate Professor at the Prefectural University of Hiroshima, and a PhD student investigating the relationship between second language oral ability and vocabulary. His research interests include fluency, pronunciation, and ESP.

Averil Coxhead is Associate Professor and teaches undergraduate and postgraduate courses in Applied Linguistics and TESOL in the School of Linguistics and Applied Language Studies, Victoria University of Wellington, Aotearoa/New Zealand. She is the author of *Vocabulary and English for Specific Purposes research* (2018; Routledge) and co-author of *English for vocational purposes: Language use in trades education* (2020; Routledge).

Nivja J. de Jong is Associate Professor in Second Language Acquisition and Pedagogy at the University of Leiden, the Netherlands. In 2002, she received her PhD at the Max Planck Institute for psycholinguistics and Radboud University Nijmegen on the topic of lexical access within psycholinguistics. Her research focuses on second language (L2) speaking, L2 learning, and language testing. Currently, she supervises MA and PhD students on diverse topics in the field of L2 learning at Leiden University.

Irina Elgort is a Senior Lecturer in Higher Education at Victoria University of Wellington. She teaches on the Master's Programme in Applied Linguistics

and TESOL. Her research interests include second language vocabulary, reading and writing, the bilingual lexicon, and CALL. She conducts empirical research using behavioural response data, eye-movements, ERPs and computational text analysis. Her recent publications appeared in *Studies in Second Language Acquisition*, *Applied Linguistics*, *Language Learning*, *Language Testing*, *Language Learning and Technology* and the *CALICO Journal*.

Simon Fraser is a Professor at the Institute for Foreign Language Research and Education, Hiroshima University. His research interests are vocabulary related, and include the lexical analysis of specialized texts and the modelling of lexical networks. He is currently working on the compilation and analysis of medical English corpora, and the development of medical English materials and word lists. He is the co-author of several lexically-based textbooks for Japanese EFL learners.

Suzanne Graham is Professor of Language and Education at the University of Reading, in the UK. Her research focuses on the teaching and learning of languages in schools, from primary school through to university. She is particularly interested in language comprehension, how learners go about making sense of what they hear or read in the language, and how teachers might facilitate that process.

Jingyi Huang is Assistant Teacher at Tong Liao Vocational College, China. Her main research interests include second language teaching, teacher assessments of vocabulary knowledge, and reading comprehension.

Ahmed Masrai is Assistant Professor of Applied Linguistics in the Department of Languages at King Abdulaziz Military Academy, Saudi Arabia. His research and teaching interests include second language acquisition, lexical studies, and vocabulary knowledge testing. He has published in a number of journals, including *Journal of English for Academic Purposes*, *Applied Linguistics Review*, *Language Learning Journal*, *International Journal of Listening*, and *Sage Open*.

James Milton is Professor of Applied Linguistics at Swansea University, Wales, UK. A long-term interest in measuring vocabulary breadth and establishing normative data for learning progress has led to extensive publications including *Modelling and assessing vocabulary knowledge* (2007, with Michael Daller and Jeanine Treffers-Daller) and *Measuring second language vocabulary acquisition* (2009).

Kazuya Saito is Associate Professor in Second Language Acquisition at University College London, UK. His research interests include how second language learners develop various dimensions of their speech in naturalistic settings; how their learning patterns can be related to individual variability in experiential, cognitive, and sociopsychological profiles; and how instruction can help optimize the complex process and production of L2 speech acquisition in classroom contexts.

Jeanine Treffers-Daller is Professor of Multilingualism at the University of Reading. She has published widely on the measurement of language ability (including language dominance) in bilinguals and second language learners, with a specific focus on vocabulary. She is on the Editorial Board of *Bilingualism, Language and Cognition*, as well as the *International Journal of Bilingualism*.

Takumi Uchihara is currently completing a PhD in Applied Linguistics at the University of Western Ontario, Canada. His research includes studies of second language speaking, pronunciation, vocabulary learning, and meta-analysis. He is currently working on a research project investigating the effects of repeated encounters, different modes of input, and acoustic variability on L2 vocabulary and pronunciation learning.

Pengchong Zhang is a Lecturer in Second Language Learning at the University of Reading. He teaches graduate level courses in second language learning theories and research methods in applied linguistics. He has researched the development of L2 vocabulary knowledge and listening proficiency. His main research interests relate to classroom-based L2 language teaching and learning, especially how L2 vocabulary knowledge and listening proficiency can be developed in the classroom context.

Acknowledgements

We should begin by expressing our sincerest thanks to everyone who wrote their papers for this edited volume. The volume began as a humble idea at the ALTAANZ conference in New Zealand and took off from there, with the sincere patience and hard work of both contributors and the Routledge publishing team. There are, of course, others whose work is not included in the volume but to whom we are very grateful. In this specific regard, a special thanks goes to Paul Meara, Tess Fitzpatrick, and Clarissa Wilks for their inspiration and guidance, not least as advisers at the beginning of the editors' own vocabulary journey, but also for their generous time and influence both of which can certainly be felt throughout this volume. We would also like to convey a special thanks to both Katie and Samantha at Routledge. Their guidance throughout the publishing process has been nothing short of excellent. Finally, we would also like to thank James Milton, Paul Nation, Averil Coxhead, and Jeanine Treffers-Daller for their kind and supportive advice.

Part I

Introduction

1 Introduction

Vocabulary and the four skills – current issues and future concerns

Jon Clenton and Paul Booth

This book is an edited volume of recent studies relating vocabulary knowledge and the four skills of reading, writing, listening, speaking. In addition to providing a compendium for vocabulary researchers, the appeal will extend to postgraduate students. Vocabulary as a course on graduate programs around the world is well supported by many theoretical and practical volumes. Such volumes present broad aspects of vocabulary in relation to, for instance, teaching (e.g., Schmitt, 2000), or specific studies to a research audience (e.g., Milton & Fitzpatrick, 2014; or Nation, 2001). The focus of the current volume, however, is different to these other books. The current volume will show that vocabulary knowledge depends on some extent on the skill area, because individuals with strengths in listening vocabulary knowledge may not demonstrate an equivalent knowledge of vocabulary in, for instance, their spoken vocabulary knowledge. As the nature of vocabulary research expands its reaches in many different arenas, the current volume represents an important central resource of recent developments to address these important concerns. On the basis that no current single volume exists that presents vocabulary knowledge and the four skills as an integrated whole, this book fulfils this need. Each chapter presents the very latest advances in the field of vocabulary research, with each chapter including recognized vocabulary experts in their respective field.

First, we begin with a brief and broad overview of vocabulary, highlighting earlier advances in considering vocabulary as a fundamental component of language. We highlight the increasing concern that what we once thought constitutes vocabulary knowledge might not be quite so straightforward. This early beginning sets the stage for the volume by serving to demonstrate that vocabulary should be treated differently depending on the skill under consideration.

That vocabulary knowledge should be considered in a more detailed manner is nothing new. A number of papers over the past twenty to thirty years (e.g., Fitzpatrick and Clenton, 2017; Meara, 1996; Nation, 2001, 2013;

Read, 2000; Webb, 2005, 2007) show that the construct includes numerous aspects not exclusive to grammar, collocations, and use, to name but a few. Given the very many different aspects of knowledge under consideration, vocabulary research then considered dimensions in order to incorporate such detail. One such example stems from Daller, Milton, and Treffers-Daller's (2007: 8) work, in their 'lexical space: dimension of word knowledge and ability' to incorporate aspects of breadth, depth, and fluency. Dimensions could also be extended to consider the extent to which vocabulary items might be known productively (written or spoken) or receptively (heard or read), with such items on a continuum. Dimensions, however, might not exclusively explain vocabulary knowledge. Meara's (2007) fascinating paper suggestive of a network of vocabulary items indicates that vocabulary knowledge might be far more multifaceted than once first considered.

Our specific focus in the current volume is to respond to recent advances in vocabulary research, and to suggest that vocabulary knowledge must be treated differently depending on the skill area. To make this point, consider Milton (2010) who compared the potential relationships between two versions of a receptive vocabulary knowledge task (in written, and aural form) with tasks from a generic language test (IELTS) that elicits knowledge of the four skills. Milton's study indicates that individual skill areas are sensitive to how they are elicited (on the basis that aural receptive measures were found to predict speaking task scores). Taken in isolation this specific study might not justify an entire volume, but our view is that a number of such studies (e.g., Adolphs & Schmitt, 2003; Elgort, 2017b; Kremmel and Schmitt, 2016; Milton, 2010; Milton, Wade, and Hopkins, 2010; Nation & Meara, 2010; Staehr, 2008; Uchihara and Clenton, 2018) support our view that vocabulary knowledge is inconsistent across the four skills.

Second, this introduction summarizes the chapters to follow, highlighting the practical threads that form the backbone of the book. We begin by first outlining the organization.

The structure of the book is organized according to receptive and productive skill sections. The first two sections explore the receptive skills, beginning with listening, and then reading. The second two sections explore the productive skills, beginning with speaking, and then writing. Each of the four 'skill' sections include four chapters devoted to each skill. The first chapter in each section discusses current research, discusses existing tools, and considers current practices. Two chapters then follow, with each relating the specific skill under enquiry to recent advances in vocabulary research. The final chapter in each section explores future research, considers potential tools, and practices. Each final chapter, we hope, provides a useful springboard for future research by listing a series of potential research questions.

Vocabulary and listening

In Chapter 2, Suzanne Graham and Pengchong Zhang begin the listening section by outlining current research tools and practices. Their chapter considers the extent to which vocabulary can be acquired through listening, covering types of listening activity, learner variables, various means (e.g., via television/video viewing), and the specific features of item-variables.

Chapter 3 by Pengchong Zhang and Suzanne Graham investigates the extent to which some words learned through listening might be more difficult or easier to learn. Their study explores the learning of a small set of words (43) from listening by Chinese second language high school learners. They explore a range of different factors that might influence how well the words are acquired.

These potentially influencing factors include five from Goldschneider & DeKeyser (2001) who suggest that the successful learning of L2 grammatical morphemes is largely determined by: perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency. Additionally, and adding a further five potentially influencing factors, Zhang and Graham examine the extent to which learning word forms might be strongly influenced by perceptual salience, morphophonological regularity, frequency, and whether semantic complexity and syntactic category determine the learnability of word meanings. Zhang and Graham add the extent to which the L1 influences L2 vocabulary learning, and whether classroom presentation might also influence learning.

The chapter reports on a study in which participants completed a listening task, which was then followed by a treatment in which half of their participants received either an L1 or L2 focus on the same vocabulary. Their analysis then explored the factors that influence how well their forty-three words were acquired. Zhang and Graham report a range of different findings. In brief, they report that nouns and adjectives were easier to acquire than verbs, that words with concrete meanings were easier to acquire than words with abstract meanings, that those words with equivalent L1 translations were generally easier to acquire than those words with direct L1 translations.

Zhang and Graham contend that vocabulary learning through listening is an under-researched area. Their findings are encouraging because they report potential implications for the order in which different types of vocabulary items are presented to learners and the amount and nature of teaching focus each type may require. They close by highlighting the importance for research to explore the factors that might influence the learning of vocabulary through such input, as reported in their chapter. Their findings are of relevance to both researchers and practitioners interested in the listening classroom for L2 vocabulary teaching and learning.

Chapter 4 by James Milton reports that the way the lexicon and listening skills interact is not so well investigated, and the results of such research are equivocal. Milton contrasts such relationships with those between the lexicon

along with reading and writing skills that he suggests are well researched, with the common view that there is quite a strong relationship between size and performance.

In considering research, Milton highlights the lack of agreement between different studies. For instance, Kelly (1991) suggests vocabulary knowledge is the main obstacle to successful listening comprehension, which contrasts with Bonk (2000) who suggests that it is not and that good comprehension can be obtained with a comparatively modest vocabulary. Milton points to Stæhr's (2008) paper, in situating listening and vocabulary compared to reading and writing and vocabulary, who demonstrates that the correlation of listening comprehension scores with vocab size is generally smaller than with reading and writing.

Milton suggests that such differences might result from methodological issues, referring to an earlier co-authored paper to highlight such potential issues. He suggests that Milton, Wade & Hopkins (2010) are able to get strong correlations with listening, comparable to those with reading and writing, where the vocabulary size test used matches the skill it is being compared with. The study also showed that receptive measures of vocabulary size, based on orthographic tests, correlated well with written skills. Furthermore, respective measures of vocab size, based on aural tests, correlated well with aural skills. Milton suggests that most studies do not make such a distinction. Milton highlights that in drawing on the dual route model of comprehension Milton et al. suggest the lexicon, in literate L2 learners, has two halves: a phonological and an orthographic half. Milton highlights that this study, Milton et al. (2010), and other studies, further suggest that a characteristic of most L2 language learners is to grow their L2 lexicons disproportionately and to develop their orthographic half faster than their phonological half. The chapter goes on to discuss how such a process might make sense if much of the lexicon is developed from extensive reading where good phonological models of new words that are encountered are never provided. Moreover, Milton shows that such a development might be an efficient way of developing the reading and writing skills that academic study and formal exams favour. Milton et al. also demonstrate that large-scale tests, such as the IELTS test, are heavily dependent on written vocab knowledge and are much less reliant on phonological knowledge.

The chapter discusses the notion that such an unbalanced lexical development, with a comparatively small phonological vocabulary size, ought to make the task of listening comprehension more difficult and less successful. Milton closes by highlighting that where such a skill is required then extensive listening exposure might be analogous to the extensive reading exposure shown to drive uptake in orthographic vocabulary size and speed of word processing (e.g., Masrai and Milton, 2018).

For Chapter 5, James Milton and Ahmed Masrai close the section on vocabulary and listening by considering future research, tools, and practices. Their chapter begins by highlighting the differences between vocabulary reading and

vocabulary listening, pointing to the fact that eye-tracking studies reveal how learners deal with individual words but that the spoken word is far more elusive for aural comprehension. They divide their chapter into four areas (the spoken word and storage; the spoken word and processing; tests and research methods for understanding the spoken word in the lexicon; and learning words from listening) for future research. They close by describing that the field of vocabulary and listening is 'a highly fertile direction for future work'.

Vocabulary and reading

In Chapter 6, Jeanine Treffers Daller surveys current vocabulary research, tools, and practices related to vocabulary and reading. The chapter emphasizes the central place vocabulary takes in relation to developing reading skills for learners, researchers, and practitioners. The chapter highlights the variety of tests that are available, pointing to the recent addition of bilingual and L2 learner tests. The chapter also refers readers to online sources where a number of these tests are available in the public domain.

Chapter 7 by Irina Elgort address the issue of building vocabulary knowledge from and for reading, with a particular focus on lexical quality. As a key goal of any language learning programme is to help students quickly build their target language lexicon (Nation, 2001), Elgot notes that vocabulary research suggests, in order to take advantage of the wealth of language input available electronically and in print, a high proportion of the running words in text (95–98%) needs to be known. In English, for example, readers need at least 8000–9000-word families (Nation, 2006), in order to read unsimplified texts with understanding and further develop their target language lexicon from reading. Elgort explains that the goal here is larger than increasing the number of words learners are familiar with (i.e., their vocabulary size); it is also about improving the quality of knowledge (including the development of robust lexical-semantic networks and fluency of access to word knowledge in real language use). Poor quality of L2 word knowledge is likely to be an impediment to continuous lexical development. This is because learners need to accurately, fluently, and effortlessly access contextually-relevant word meanings in reading and listening, in order to convert input into intake. Enter Lexical Paradox outlined in Cobb (2007) – in order to gain new lexical knowledge from reading, language learners need to bring sufficient lexical knowledge to reading. In particular for students whose target language lexicons are being formed, by and large, in the context of a foreign language classroom (Jiang, 2000), a research-led understanding of the kinds of learning and instructional activities that promote high quality of lexical knowledge is critical.

In her chapter, Elgort, considers L2 word knowledge from the perspective of the Lexical Quality Hypothesis framework (Perfetti, 2007; Perfetti & Hart, 2001, 2002), which interprets lexical quality in terms of formal (orthographic and phonological) and lexical semantic representations, and their

interrelationship. Elgort suggests how this framework can guide the selection of vocabulary learning treatments that contribute to the development of lexical quality and inform measures of lexical development. She refers to research on deliberate word learning as a means of delivering a qualitative boost in a relatively short time. Elgort also discusses research into supplementary learning activities that can be used to optimize vocabulary learning from reading.

Elgort looks back at the findings of the L2 vocabulary learning research she has conducted in collaboration with colleagues from applied linguistics, cognitive psychology, and language education, with a view to translating these findings into recommendations for teachers and learners. Specifically, she considers the role of deliberate learning (such as paired-associate learning using flashcards) in vocabulary development (Elgort, 2011; Elgort & Piasecki, 2014; Nakata, 2008; Nakata & Webb, 2016). She then considers contextual word learning during reading and discusses learner, text, and word variables that affect lexical quality development (Elgort & Warren, 2014; Elgort, Brysbaert, Stevens, and Van Assche, 2017). She also refers to the studies that investigate the effects of instructional and learning treatments on contextual learning from reading, and summarizes how these treatments affect lexical quality. A series of such studies (published and in-progress) investigating effects of different approaches to form-focused and meaning-focused elaboration inform the discussion in this section of the chapter (Boutorwick, 2017; Elgort, 2017a; Elgort, Candry, Eyckmans, Boutorwick, and Brysbaert, 2016; Elgort, Beliaeva, Boers, and Demecheleer in preparation; Toomer & Elgort, in revision). Elgort concludes her chapter with a summary of research-based recommendations for improving the lexical quality of contextual word learning.

In Chapter 8, Jeanine Treffers-Daller and Jingyi Huang report on an investigation into the validity of the Test for English Majors 4 (TEM4) as a measure of reading comprehension and vocabulary knowledge. They report on a study in which they analyse correlations between the TEM4 with widely used tests of vocabulary knowledge and reading comprehension among university-level students of English in China. Their study is based on the responses of sixty students, pursuing a second year English Major in north China, who completed the bilingual Mandarin-English version of the Vocabulary Size Test (Nation & Beglar, 2007) for vocabulary size, the Vocabulary Knowledge Scale (Brown, 2008), modified from Wesche and Paribakht (1996) for assessing depth of vocabulary knowledge, and the TEM-4. Thirty participants were randomly selected from the sixty students and tested using the *York assessment of reading for comprehension (secondary)* (YARC, Stothard, Hulme, Clarke, Barmby, and Snowling, (2010)). Treffers-Daller and Huang highlight that the YARC was originally developed for secondary school children aged 11–16 in the UK and is yet to be used with adult L2-learners of English.

Treffers-Daller and Huang's three key findings show that (i) the reading comprehension part of the TEM-4 did not really measure reading comprehension as it did not correlate with the different components of the YARC;

(ii) there were moderate correlations between the TEM4, the VST, and the VKS, indicating that the TEM-4 tapped into different dimensions of vocabulary knowledge; and (iii) there were modest correlations between the VST, the VKS, and different components of the YARC Secondary Test: the VKS correlated more strongly with comprehension measures of the YARC whilst the VST was correlated with students' decoding skill as measured with the Single Word Reading Test (a component of the YARC).

Treffers-Daller and Huang suggest their results provide empirical evidence supporting the importance of depth of vocabulary in reading comprehension. Their chapter concludes that the validity of the TEM-4 as a measure of reading comprehension is questionable given that the test appears to measure vocabulary knowledge instead of reading comprehension. They end their chapter with reflections on the suitability of the YARC secondary for use with adult L2 learners of English.

In Chapter 9, Irina Elgort considers vocabulary reading, future research, tools, and practices. Ending this section on reading, the chapter considers the construct of 'lexical quality' required for fluent reading. The chapter briefly discusses lexical quality and proposes three related L2 contextual vocabulary learning projects.

Vocabulary and speaking

In Chapter 10, the first chapter in the speaking section, Takumi Uchihara considers vocabulary and speaking in terms of current research, tools, and practices. The chapter shows that this emerging area of vocabulary research can be examined from different perspectives to include both human rating as well as objective tools. The chapter concludes by highlighting the need for detail on the specific aspects of L2 oral proficiency and their relation with vocabulary knowledge.

In Chapter 11 Jon Clenton, Nivja J. De Jong, Dion Clingwall, and Simon Fraser present a small-scale study in which they identify potential relationships between specific vocabulary tasks, previously employed speaking fluency tasks (De Jong, Steinel, Florijn, Schoonen, and Hulstijn, 2013; De Jong, Groenhout, Schoonen, and Hulstijn, 2015), and 'vocabulary skills'. They use tasks that have not been used together before based on the lower proficiency of their first language participant group. Their study is unique, as it represents a first approach to examining the relationship between vocabulary skills (e.g., automaticity retrieval), vocabulary knowledge, and aspects of fluency.

For Clenton et al., the primary aim of their chapter is to elucidate the specific vocabulary knowledge required by using speaking tasks at the specific proficiency level of their participants (pre-intermediate). They base their vocabulary investigation on recent papers that indicate that: (i) the specific vocabulary knowledge captured by different tasks varies according to proficiency; (ii) vocabulary knowledge is multifaceted, to the extent that different tasks appear to elicit quite different vocabulary knowledge; and (iii) vocabulary knowledge development depends on proficiency. Their chapter responds to developments

in fluency research that suggest that oral ability varies according to task (e.g., DeJong, 2016; Tavakoli, 2016). They therefore investigate this claim and present a multifaceted approach to their investigation. They partially replicate an earlier fluency study (De Jong et al., 2013), using different vocabulary tasks. Clenton et al. suggest that the choice of vocabulary task should reflect the lexical resource of the specific participant group investigated. They reflect on a recent 'vocabulary task capture model' (Fitzpatrick & Clenton, 2017), which they reconsider in light of earlier papers on fluency. Earlier papers on fluency (De Jong et al., 2013, 2015; Uchihara & Saito, 2018) have employed specific productive vocabulary tasks such as the Productive Levels Test (PVLT; Laufer & Nation, 1999), or Lex30 (Meara and Fitzpatrick, 2000) Clenton et al. present a 'revised vocabulary task capture map model' (Clenton et al., 2019), based on the earlier model (Fitzpatrick & Clenton, 2017), in order to explore the task differences between these two widely cited productive vocabulary knowledge tasks. Clenton et al., therefore, use this approach in their investigation, in order to account for their use of a specific productive vocabulary knowledge task (Lex30), but also to explore various findings related to the vocabulary resource in their discussion. One specific finding suggests that participants with a limited vocabulary resource might reproduce vocabulary in various tasks; they suggest that this occurs less often with increases in proficiency. Part of their study includes a comparison of the spoken output of their participant group with output from the Lex30 task. They report comparisons using the Academic Spoken Word List (ASWL; Dang, Coxhead, and Webb, 2017). They also explore the various 'vocabulary skills' of their participant group. They suggest that delays in speech, or delays in response, relate to the vocabulary knowledge available to their specific participant group. In discussing such findings, they encourage follow-up studies to explore the extent to which their findings can be replicated with different proficiency levels and with participants from different L1 backgrounds. Their 'vocabulary skills' findings, they suggest, provide a foundation from which to explore the extent to which speed and automaticity of retrieval is level dependent.

In Chapter 12, Uchihara et al consider the potential relationships between productive vocabulary and second language oral ability. They begin by highlighting the important role vocabulary knowledge plays in second language (L2) proficiency and development, pointing to research investigating the relationship between vocabulary and L2 proficiency supporting the long-standing view that vocabulary serves as a proxy for communicative language ability (Meara, 1996). They show that a growing body of research within lexical studies relates vocabulary knowledge to a range of proficiency indicators: overall proficiency benchmarks (e.g., Common European Framework of Reference for Languages (CEFR) levels; Milton, 2010), in-house placement tests (e.g., Harrington & Carey, 2009); standardized language proficiency examinations (e.g., International English Language Testing System (IELTS); Milton, Wade, & Hopkins, 2010, or Test of English as a Foreign Language (TOEFL); Qian, 2002). They highlight the lack of research

designed to investigate the relationship between speaking and vocabulary. For specific aspects of linguistic proficiency, they suggest, research has tended to investigate potential relationships between reading and vocabulary (e.g., Laufer & Levitzky-Aviad, 2017). Their chapter is designed to redress this imbalance by drawing on instruments, and a recent framework devised to assess multiple L2 oral ability dimensions (Crossley, Salsbury, and McNamara, 2017; Saito, Trofimovich, and Isaacs, 2017; Trofimovich & Isaacs, 2012); they explore the extent to which productive vocabulary knowledge correlates with aspects of L2 oral ability including global (comprehensibility), temporal (speed, breakdown fluency), and lexical (appropriateness, variation, sophistication) features.

They report on a study in which their participants, with varying degrees of L2 proficiency, completed a productive vocabulary task (Lex30; Meara & Fitzpatrick, 2000) and a speaking task (suitcase story; Derwing & Munro, 2009). Productive vocabulary test scores were calculated in two ways: raw scores and percentage scores (Fitzpatrick & Meara, 2004; Fitzpatrick & Clenton, 2010; Meara & Fitzpatrick, 2000), the former representing a construct related to fluency (i.e., speed of production) as well as knowledge of infrequent words, the latter being more closely related to lexical knowledge with fluency controlled (Clenton, 2010; Uchihara & Saito, 2016). Speaking task data were submitted for listener judgements and for a range of linguistic analyses. Thirteen L1 English raters were recruited to rate subject speech according to perceived comprehensibility (1 = easy to understand, 9 = hard to understand) (Derwing & Munro, 2015). To measure a variety of linguistic features of oral ability, the transcribed texts of the spoken data were analysed in terms of temporal (articulation rate, silent pause ratio, filled pause ratio) and lexical (appropriateness, diversity and sophistication) dimensions.

Uchihara et al.'s results show that Lex30 raw scores were associated with various aspects of oral proficiency including comprehensibility, fluency (articulation rate, silent pause ratio), and lexical richness (lexical diversity and sophistication). Lex30 percentage scores were only correlated with lexical diversity. They suggest their findings indicate that the two approaches to scoring productive vocabulary may show varying relationships between productive vocabulary knowledge and oral ability, with data suggesting that any definition of 'productive vocabulary knowledge' should not exclusively be limited to a frequency-based operationalization (i.e., Lex30 percentage scores), but a multifaceted construct to include speed of production (i.e., Lex30 raw scores).

Uchihara et al. suggest that their findings offer several implications for vocabulary L2 teaching and assessment. These include their view that L2 teachers should not only teach infrequent words, but also focus on fluency development in production (Nation, 2006). They contend that teachers should focus on increasing low-frequency word knowledge for L2 learners intent on gaining better control of different words in speech,. Uchihara et al.'s data suggest that an additional focus on lexical fluency might positively influence

broader aspects of oral ability. Learners could benefit from activities requiring oral production of known words under increasing time pressure (e.g., 4/3/2 task; Thai & Boers, 2016). Their findings also demonstrate the potential usefulness of the productive vocabulary test as an assessment tool by way of a broad estimate of learners' L2 oral ability. For diagnostic purposes, teachers can administer a vocabulary task (e.g., Lex30) at regular intervals and potentially use data to provide a broad indication of their oral ability progress. The validity of such an attempt remains to be confirmed, but its feasibility is pedagogically appealing given the time taken to administer Lex30 (compared to the relatively time-consuming collection of speech score ratings).

In Chapter 13, the final chapter in the speaking section, Jon Clenton considers vocabulary and speaking in terms of future research, tools, and practice. The chapter suggests the need for several specific research questions, based on recent trends. These include the need to consider multiple aspects of vocabulary knowledge and their relation to spoken output, as well as the potential formulation of an implicational scale of vocabulary knowledge and vocabulary skills.

Vocabulary and writing

In Chapter 14, the first chapter in the writing section, Paul Booth considers current research, tools, and practices. The chapter focuses on measures of lexical sophistication, and suggests a number of means to determine this measure. These include frequency profiles, P-Lex software, N-grams, intrinsic measures such as lexical diversity and TTR (Type Token Ratio), and external and internal measures of lexical sophistication,

In Chapter 15, Averil Coxhead explores the need for specialized vocabulary in writing, and the specific benefit for English Language Teaching (ELT) to investigate outside its own field. Coxhead begins by showing that the vocabulary used in writing in a second language can be a source of anxiety and difficulty for language learners (Coxhead, 2011), and she highlights that there is much to know about a word in order to use it in writing (Nation, 2013). Learners may resort to using high frequency words that they know well rather than taking a risk with lesser known vocabulary. Coxhead suggests that learners might find it difficult to gauge an audience or register for writing, or that they might simply lack the background knowledge of a topic and therefore the vocabulary required to write about it (Coxhead, 2011). She highlights that it is well known that learners know or can recognize more words than they use in English: see Malmström, Pecorari, and Gustafsson, 2016).

Coxhead's chapter shows that recent research has focused on specialized or technical vocabulary in an attempt to support learners and teachers in English for Academic Purposes (EAP) (see Coxhead, 2016; Gardner & Davies, 2014) or English for Specific Purposes (ESP) through developing word lists (Nation, Coxhead, Chung, and Quero, 2016), for example, as possible shortcuts to the vocabulary that these learners need in their studies or in their professional lives (see Coxhead, 2018). Typically, technical vocabulary might be expected

to occur mostly inside a field or be known by people who have studied or worked in that field (Chung & Nation, 2003). That said, she suggests, there are also everyday words that can have specialized meanings in a particular context such as *host* or *string* in Computer Science (see Coxhead, 2018). Furthermore, new research into the technical vocabulary of welding, for example, suggests that more than 30% of a written text in that trade could be technical (Coxhead, McLaughlin, and Reid, *under review*).

Having highlighted these various fields, the focus of Coxhead's chapter is to investigate writing in a field outside English Language Teaching (ELT) in order to suggest that ELT and ESP might be inspired to take on or adapt this technique. While the location of Coxhead's research is courses in carpentry at a polytechnic in Aotearoa/New Zealand, the research sits squarely in research and teaching in English for Specific Purposes, because some of the students in these courses are second or foreign language speakers of English. A mandated writing task for all learners in carpentry is a builder's diary. These diaries contain regular accounts of the classwork and building site work of the carpentry students in a course where they build a house over a year. The diaries are assessed as part of the course and are modeled on diaries that builders keep in their everyday professional work. The diaries include pictures and diagrams, as well as short passages of writing. See Parkinson, Mackay, and Demecheleer, 2017 for more on the diaries and Parkinson et al. (2017a) for more on the Language in Trades Education (LATTE) project overall.

Coxhead's chapter reports on interviews with students about their use of diaries for learning and keeping track of technical vocabulary in the course of their studies over a year. The chapter also draws on a word list of carpentry which was developed using learning materials and teacher talk in classrooms and building sites at the polytechnic (Coxhead, Demecheleer, and McLaughlin, 2016). Corpus linguistics techniques were used to analyse the vocabulary used in the student diaries. Coxhead's analysis shows that the students value their diaries a great deal and use them strategically for tackling the large amounts of technical vocabulary they encounter in their studies. From the corpus analysis, she observes that students tend to use more technical vocabulary in the later parts of their course and some learners used more technical vocabulary than others. Comparisons with a corpus of professional writing in carpentry show higher amounts of technical writing in this corpus compared to the student diary corpus, but not in every case. A particular point to note is the development of spelling knowledge of difficult technical terms such as scotia and joist. The idea of using these diaries has been taken on by other courses at the polytechnic because they foster writing skills, allow students to demonstrate their knowledge, and provide a powerful learning tool for tracking and learning vocabulary. Coxhead's chapter ends with practical suggestions on how the builder's diary might be adapted to ESP courses in a range of contexts and in secondary and university level education.

In Chapter 16, Paul Booth explores Lexical development paths in relation to academic writing, discussing how L2 lexical richness develops over time. Booth builds on lexical frequency profile studies as a quantitative measure

of the lexical frequency of words (Bell, 2009; Laufer 1994a 1994b, 1998; Laufer and Nation 1995). Such studies show that lexical development is not always linear and so researchers must consider individual differences to the development patterns of lexical richness. He suggests that one way to understand differences in lexical production in writing is to explore language aptitude, namely memory and analysis, earlier associated with language learning (Harley and Hart 1997). A framework for learning style, memory, and analysis, first introduced by Skehan (1998), is explored in this chapter as a means to investigate lexical richness of academic writing over time.

Booth reports on a lexical production study that is longitudinal at the beginning and the end of one university semester. Booth analyses university student discursive essays at time 1 and time 2 to analyse lexical richness, using Web Vocabprofile (v2.6) to analyse the student texts. The participants were tested for memory and analysis, and grouped according to high (IELTS 6.0 or equivalent) or low (below 6.0 IELTS) proficiency. Running a range of statistical analyses, Booth investigates whether there were lexical richness gains within the two proficiency groups. The chapter reports on gains that Booth suggests have implications for L2 vocabulary teaching, to the extent that specific grammar should be taught to second language learners when aiming to acquire new words. Booth explores the extent to which context plays a part in the learning of writing, and suggests that a data-driven approach might help when deducing the word meaning in context (Boulton, 2009).

Booth's chapter indicates how learners can apply different approaches to language learning, suggesting that for lexis to develop learners should focus on how words are used in context. Booth adds that a plateau effect might impact production, to the extent that writers could consider how to make their writing more complex.

In Chapter 17, the final chapter in the writing section, Averil Coxhead outlines vocabulary and writing – future research, tools, and practices. The chapter explores some of the key factors in exploring vocabulary use in writing to include corpus-based approaches, evaluating pedagogy, and to find out more about L2 writer intent, belief, and practice. The chapter also suggests tools to examine vocabulary use in writing, and includes five potential research questions for future investigations.

Chapter 18 concludes this volume, and collates the various threads presented throughout. We end with a section on how we began, by highlighting that individual skill areas (e.g., speaking, reading) very much influence the vocabulary knowledge and use. We hope this edited volume provides a springboard for future research, and for researchers and practitioners to further disentangle and unpack this essentially multifaceted area of concern.

References

Adolphs, S., & Schmitt., N. (2003). Lexical coverage of spoken discourse, *Applied Linguistics*, *24(*4) (1 December), 425–438, doi.org/10.1093/applin/24.4.425

Bell, H. (2009). The messy little details: A longitudinal case study of the emerging lexicon. In Fitzpatrick, T., and Barfield, A. (eds) *Lexical processing in second language learners* (pp. 111–140). Bristol: Multilingual Matters.

Bogaards, P. (1996). Lexicon and grammar in second language learning. In Jordens, P., and Lalleman, J. A. (eds) *Investigating second language acquisition* (pp. 357–379). Berlin: Mouton de Gruyter.

Bonk, W. J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening Comprehension*, *14,* 14–31.

Boulton, A. (2009). Styles d'apprentissage et consultation de corpus. XXXIe Congrès de l'APLIUT: Le dictionnaire dans tous ses etats – Dictionnaires et enseignement des langues. Dijon, France: IUT de Dijon, 11–13 June.

Boutorwick, T. J. (2017). Vocabulary development through reading: A comparison of approaches. Unpublished PhD Thesis. Victoria University of Wellington.

Brown, L. (2008). Using mobile learning to teach reading to ninth-grade students. Ph.D dissertation Capella University, USA. Available Online at http://gradworks.umi.com/3330949.pdf.

Chung, T., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, *15*(2), 103–116.

Clenton, J. (2010). Investigating the construct of productive vocabulary knowledge with Lex30. Unpublished doctoral dissertation. *University of Swansea, UK*.

Clenton, J., Elmetaher, H., & Uchihara, T. (2019). Capturing and quantifying individual differences in bilingual learners: a task-based study. Paper presented at Capturing and quantifying individual differences in bilingualism workshop. Arctic University of Norway, Tromsø.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology*, *11*, 38–63.

Coxhead, A. (2011). What is the exactly word in English? Investigating second language vocabulary use in writing. *English Australia*, *27*(1), 3–17.

Coxhead, A. (2016). Acquiring academic and discipline specific vocabulary. In Hyland, K., & Shaw, P. (eds) *Routledge handbook of English for academic purposes*, pp. 177–190. London: Routledge.

Coxhead, A. (2018). *Vocabulary and English for specific purposes research: Quantitative and qualitative perspectives*. London: Routledge.

Coxhead, A., Demecheleer, M., & McLaughlin, E. (2016). The technical vocabulary of carpentry: Loads, lists and bearings. *TESOLANZ Journal*, *24*, 38–71.

Coxhead, A., McLaughlin, E., & Reid, A. (under review). The development and application of a specialised word list: The case of fabrication.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, 35(5), 570–590.

Daller, H., Milton, J., & Treffers-Daller, J. (2007). Editors' introduction: conventions, terminology and an overview of the book. In Daller, H., Milton, J., & Treffers-Daller, J. (eds) *Modelling and assessing vocabulary knowledge* (pp. 1–32). Cambridge: Cambridge University Press.

Dang, T. N. Y, Coxhead, A., & Webb, S (2017) The academic spoken word list. *Language Learning*, *67*(4), 959–997. ISSN 0023-8333

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching, 54*(2), 113–132.

De Jong, N., & Mora, J. (2017). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, 1–13. doi:10.1017/S0272263117000389

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916.

De Jong, N.H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015), Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior, *Applied Psycholinguistics, 36*(2), 223–243. doi: 10.1017/S0142716413000210

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476–490.

Derwing, T. M., & Munro, M. J. (2015). Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research. Amsterdam: John Benjamins.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*(2), 367–413.

Elgort, I. (2017a). Incorrect inferences and contextual word learning in English as a second language. *Journal of the European Second Language Association*, *1*(1), 1–11.

Elgort, I. (2017b). Technology-mediated second language vocabulary development: A review of trends in research methodology. *CALICO Journal* (Sheffield), *35(*1) (December), 1–29. ISSN 2056–9017.

Elgort, I., & Piasecki, A. (2014). The effect of a bilingual learning mode on the establishment of lexical-semantic representations in the L2. *Bilingualism: Language and Cognition*, *17(3),* 572–588.

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*(2), 365–414.

Elgort, I., Beliaeva, N., Boers, F., & Demecheleer, M. (in preparation). Trial-and-error and errorless treatments in contextual word learning.

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2017). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, 1–26. doi:10.1017/S0272263117000109.

Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and ERP indicators of learning. *Language, Cognition and Neuroscience*, *30*(5), 506–528.

Elgort, I., Candry, S., Eyckmans, J., Boutorwick, T. J., & Brysbaert, M. (2016). Contextual word learning with form-focused and meaning-focused elaboration. *Applied Linguistics*, amw029, https://doi.org/10.1093/applin/amw029.

Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, *27*, 537–554.

Fitzpatrick, T. and Clenton, J. (2017). Making sense of learner performance on tests of productive vocabulary knowledge. *TESOL Quarterly, 51*, 844–867. doi:10.1002/tesq.356

Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics*, *1*, 55–74.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the 'natural order of L2 morpheme acquisition' in English: A meta-analysis of multiple determinants. *Language Learning*, *15*(1), 1–50. http://dx.doi.org/10.1111/1467–9922.00147

Harrington, M., & Carey, M. (2009). The on-line Yes/No test as a placement tool. System, 37, 614–626.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19*, 370–400.

IELTS Partners (2009–2017). IELTS Australia Pty Ltd (solely owned by IDP Education Pty Ltd) and Cambridge Assessment English. www.ielts.org/about-the-test/test-format-in-detail.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research, 4*(5)*,* 900–913. doi:10.4304/jltr.4.5.900-913

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47–77. https://doi.org/10.1093/applin/21.1.47

Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language learners. *IRAL*, *29*, 135–149.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, *4*(5)*,* 900–913. doi:10.4304/jltr.4.5.900-913

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, *13*(4), 377–392.

Laufer, B. (1994a). The lexical profile of second language writing: Does it change over time? *RELC Journal, 25* 2), 21–33.

Laufer, B. (1994b). Beyond 2000: A measure of productive lexicon in a second language. In Eubank, L., Selinker, L., & Sharwood Smith, M. (eds) *The current state of interlanguage* (pp. 265–272). Amsterdam/Philadelphia: John Benjamins Publishing.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, *19*(2), 255–271.

Laufer, B., & Levitzky-Aviad, T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *Modern Language Journal*, *101*(4), 729–741.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production, *Applied Linguistics*, *16*(3), 307–322.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. doi:10.1177/026553229901600103.

Malmström, H., Pecorari, D., & Gustafsson, M. (2016). Coverage and development of academic vocabulary in assessment texts in English Medium Instruction. In Göpferich, S., & Neumann, I. (eds) *Assessing and developing academic and professional writing skills* (pp. 45–69). New York: Peter Lang.

Masrai, Ahmed, & Milton, James. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes, 31*, 44–57.

McLean, S., & Kramer, B. (2015). The creation of a new vocabulary levels test. *Shiken*, *19*, 1–11.

Meara, P. (1996). The dimensions of lexical competence. In Brown, G., Malmkjaer, K., & Williams, J. (eds) *Performance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.

Meara, P. M. (2007). Growing a vocabulary. *EUROSLA Yearbook*, *7(* 1) (Jan.), 49–65.

Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, *28*(1), 19–30. doi: 10.1016/S0346-251X(99)00058-5.

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In Vedder, I., Bartning, I., & Martin, M. (eds) *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211–232). Second Language Acquisition and Testing in Europe Monograph Series 1.

Milton, J., & Fitzpatrick, T. (2014). *Dimensions of vocabulary knowledge.* Basingstoke: Palgrave Macmillan.

Milton J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In Chacón-Beltrán, R., Abello-Contesse, C., & Torreblanca-López, M. (eds) *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.

Nakata, T. (2008). English vocabulary learning with word lists, word cards, and computers: Implications from cognitive psychology research for optimal spaced learning. *ReCALL*, 20(1), 3–20.

Nakata, T., & Webb, S. A. (2016). Vocabulary learning exercises: Evaluating a selection of exercises commonly featured in language learning materials. In Tomlinson, B. (ed.), *Second language acquisition research and materials development for language learning* (pp. 123–138). Abingdon, Oxon: Taylor & Francis.

Nation, I. S. P. (2001). *Learning vocabulary in another language* (1st ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*, 59–82.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press

Nation, I. S. P., & Beglar, D. (2007) A vocabulary size test. *Language Teacher*, *31*(7), 9–13.

Nation, P., & Meara, P. (2010). Vocabulary. In Schmitt, N. (ed.), *An introduction to applied linguistics* (2nd ed., pp. 34–53, chapter 3, p. 49). London: Hodder Education.

Nation, P., Coxhead, A., Chung, M., & Quero, B. (2016). Specialized word lists. In Nation, I. S. P. (ed.) *Making and using word lists for language learning and testing* (pp. 145–151). Amsterdam: John Benjamins.

Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, *11*(1), 9–29. https://doi.org/10.18806/tesl.v11i1.623

Parkinson, J., Coxhead, A., Demecheleer, M., Mackay, J., Matautia, L, McLaughlin, E., & Tuamoheloa, F. (2017a). *The language in the trades education project*. Wellington: Ako Aotearoa.

Parkinson, J., Mackay, J., & Demecheleer, M. (2017b). Situated learning in acquisition of a workplace genre. *Vocations and Learning*. First view: 30 August 2017.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–83.

Perfetti, C. A., & Hart, L. (2001). The lexical bases of comprehension skill. In D. Gorfien (ed.) *On the consequences of meaning selection* (pp. 67–86). Washington, DC: American Psychological Association.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In Vehoeven, L., Elbro, C., & Reitsma, P. (eds) *Precursors of functional literacy* (pp. 189–213). Amsterdam: John Benjamins.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. Language *L*earning, 52(3), 513–536.

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *37*(2), 217–240. doi:10.1017/S0142716414000502.

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439–462.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, New York: Cambridge University Press.

Skehan, P. (1998). *A cognitive approach to language leaning*. Oxford: Oxford University Press.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*, 139–152.

Stothard, S. E., Hulme, C., Clarke, P., Barmby, P., & Snowling, M. J. (2010). *YARC York assessment of reading for comprehension (secondary)*. [Book]. London: GL Assessment

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 133–150.

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50, 369–393. doi:10.1002/tesq.232

Tian, L. (2011). *Teacher codeswitching in a communicative EFL context: Measuring the effects on vocabulary learning*. Oxford: VDM.

Toomer, M., & Elgort, I. (under revision). The effects of different learning conditions on implicit and explicit knowledge of English collocations.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916.

Uchihara, T., & Clenton, J. (2018). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*. https://doi.org/10.1177/1362168818799371

Uchihara, T. & Saito, K. (2018). Exploring the relationship between productive vocabulary knowledge and second language oral ability, *Language Learning Journal*. doi:10.1080/09571736.2016.1191527.

Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Language Assessment Quarterly*, *9*(2), 172–185. doi:10.1080/15434303.2011.625579.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*, 33–52.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge depth vs. breadth. *Canadian Modern Language Review*, *53*, 13–39.

Part II
Listening

2 Vocabulary and listening

Current research, tools, and practices

Suzanne Graham and Pengchong Zhang

To what extent can vocabulary can be acquired through listening? This question is less widely researched than it is for reading, but a common thread through the literature is that levels of vocabulary learning from listening are typically lower than for reading (Brown, Waring, and Donkaewbua, 2008; Vidal, 2011). That view is changing, however, with a recent metanalysis by de Vos, Schriefers, Nivard, and Lemhöfer (2018) finding a large effect size for aural-based vocabulary learning from which they conclude that such learning "may be more effective than has previously been thought" (p. 930). Vocabulary gains from listening (online videos) were also found to be comparable to those from reading (online blogs) in a recent study of EFL learners by Arndt and, Woore (2018). Indeed, the video group scored more highly on meaning recognition and on the recall of grammatical function and meaning. Differences were however not statistically significant.

These findings suggest that listening develops different kinds of vocabulary knowledge compared to reading. As van Zeeland and, Schmitt (2013) argue, "more fine-grained and earlier acquired aspects of vocabulary knowledge" (p. 609) may be developed through listening, such as grammatical recognition, requiring measurement tools that tap into these different forms of knowledge. In their own study they used a vocabulary dimensions framework, assessing learning through tests of form recognition, grammar recognition, and meaning recall. Listening to different types of spoken input led to vocabulary gains ranging from 29% (short-term learning) to 19% (longer-term learning).

Alongside looking at the kind of vocabulary knowledge that develops from listening and how best to measure that, a growing area of interest is who learns what under what conditions – in other words, what is the role of task type, learner variables, different forms of support, and item variables? We consider each of those interrelated areas in what follows.

Type of listening activity

Particular types of listening may lead to greater vocabulary gains, according to de Vos et al. (2018). They grouped studies into four categories: audio input only (e.g., audiobooks); audio-visual input (audio accompanied by pictures or

video); audio plus visual support in the form of task materials that involved interaction (e.g., asking questions of the speaker); and audio plus visual support in the form of task materials that involved a meaning-focused task but no interaction with another person. They found no difference in learning gains between audio and audio-visual input, but learning was greater from interactive rather than non-interactive tasks.

Learner variables

The importance of learner proficiency was indicated in Vidal (2011), who although finding lower levels of vocabulary retention on average when students listened to academic lectures rather than reading academic texts, the higher learners' proficiency level was, the smaller the difference between gains from reading and from listening. Greater learning gains through listening for higher proficiency learners (prior vocabulary knowledge and/or listening skills) have also been found in many recent studies (Peters & Webb, 2018; Pujadas & Muñoz, 2019; Suárez & Gesa, 2019) but not in all (Rodgers, 2013). Learner proficiency may also interact with the type of the support provided, an issue we return to below in relation to video captioning.

Types of support

The ephemerality of aural input may account for challenges in learning vocabulary from listening for many learners. Therefore, various studies have explored the impact of different forms of support or text enhancement to lessen those challenges, with some also looking at how far their effect varies according to proficiency. These include looking at the impact of captions, images, and teacher explanations, before or after the aural input, for listening and also video/TV viewing.

Images – television/video viewing

The ubiquity of television/video viewing in out-of-class contexts makes this an important form of aural input for researchers to explore. Vocabulary gains have been found in a growing number of studies considering different contexts: for example, those involving young learners and short clips (d'Ydewalle & Van de Poel, 1999); intermediate university learners of English watching ten full-length episodes of a TV program (Rodgers, 2013); adult university learners of English watching an hour-long documentary about economics (Peters & Webb, 2018); formulaic sequences as well as single words (Puimège & Peters, 2019).

Although de Vos et al. (2018) found that there was no difference in vocabulary learning comparing audio and audio-visual input within their metanalysis, other studies lend weight to the claim that the support afforded by visual images enhances vocabulary learning. Sydorenko (2010), exploring

learning through video both with and without captions, found that learners reported finding images most helpful for learning new words and working out their meaning. Visual images may however be more effective in certain circumstances than in others; for example, when they are in close temporal proximity to the occurrence of the word they represent (Peters, 2019).

Captions or subtitles

The presence of L2 captions with video is considered to lead to greater vocabulary gains compared with just watching or with L1 subtitles (Vanderplank, 2016). A meta-analysis (Montero Perez, Van Den Noortgate, and Desmet, 2013) found a large effect size for captioning compared with no captioning across a range of proficiency levels. Contrastingly, Suárez and Gesa (2019), studying high school and university learners, reported that captions only benefited the former, indicating that having the additional support of the written form of the aural input was especially important for learners with lower listening proficiency and vocabulary knowledge. Adding to the complexity of this question, several caption studies confirm the positive relationship between higher levels of prior vocabulary knowledge and vocabulary gains outlined earlier (Peters, Heynen, and PuimÃgë, 2016; Peters & Webb, 2018; Pujadas & Muñoz, 2019).

Teachers' pre- or post-listening vocabulary explanations

In an early study showing that vocabulary can be learned incidentally through listening to stories (Elley, 1989), one of the factors to which increased gains were attributed was teacher explanations of word meanings as learners listened. Since then, a small but growing group of studies has explored the effect of teacher explanations of vocabulary before or after listening and, within some studies, whether the effectiveness varies according to the nature of the explanations.

Two recent studies have considered the effect of pre-teaching vocabulary, both finding that it aided learning (Montero Perez, 2019; Pujadas & Muñoz, 2019). Eye tracking, an increasingly important research tool in the field, indicated that pre-teaching did not influence eye fixations and time spent on the pseudo-words in the captions, or on the images, but pre-learned pseudo-words were skipped over more often than unknown pseudo-words in the captions (Montero Perez, 2019). This suggests perhaps that pre-teaching leads learners to concentrate more on the aural form of the words rather than needing to rely on images and captions for support.

Giving vocabulary explanations after aural input in the form of listening passages has also been found to lead to greater vocabulary gains than listening without such explanations, across four studies concerning university (Lee & Levine, 2020; Tian & Macaro, 2012) and high school learners (Hennebry et al., 2013; Zhang & Graham, 2019). All four studies are discussed in detail in the next chapter.

Item-variables

Vocabulary gains through aural input have been found to relate to properties of the target vocabulary. For example, Puimège and Peters (2019), focusing on the incidental acquisition of formulaic language, found that more concrete items were more likely to be learnt than less concrete ones, and adjective-noun combinations better learnt than verb-particle combinations (see Zhang & Graham, this volume). Cognateness has also been found to be an influencing factor (Peters & Webb, 2018; Vidal, 2011).

Finally, with regards to frequency of occurrence of vocabulary items in aural input, there is a rather mixed picture. While several studies have reported lower frequency effects than are associated with reading (Brown et al., 2008; van Zeeland & Schmitt, 2013; Vidal, 2011), a systematic review by Uchihara et al. (2019) found a correlation between frequency and learning gains of .39 for listening, only slightly below that for reading (r = .41). By contrast, vocabulary gains in listening supported by visual images was found to relate only weakly to frequency (.21). Yet, more recent studies of video viewing and hence involving images, such as Peters and Webb (2018) and Peters (2019), do report a frequency effect. This contradictory picture suggests a need for further investigation into repetition effects for aural input.

Conclusion

Overall, this review suggests that more vocabulary is learned through listening than has been thought in the past. That is no trivial conclusion, given that learners today may be more likely to access spoken input in the form of TV/video, or music, than to read in the L2 (Arndt & Woore, 2018). Who learns what under what circumstances remains however a question to be explored further.

References

Arndt, H. L., & Woore. R. (2018). Vocabulary learning from watching YouTube videos and reading blog posts. *Language Learning and Technology*, *22*(1), 124–142.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading a Foreign Language*, *20*(2), 136–163.

de Vos, J. F., Schriefers, H., Nivard, M. G., & Lemhöfer, K. (2018). A meta-analysis and meta-regression of incidental second language word learning from spoken input. *Language Learning*, *68*(4), 906–941.

d'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, *28*(3), 227–244.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*, 174–187.

Hennebry, M., Rogers, V., Macaro, E., & Murphy, V. (2013). Direct teaching of vocabulary after listening: is it worth the effort and what method is best? *Language Learning Journal*, *45*(3), 282–300.

Lee, J. H., & Levine, G. S. (2020). The effects of instructor language choice on second language vocabulary learning and listening comprehension. *Language Teaching Research*, *24*, 250–272. https://doi.org/10.1177/1362168818770910

Montero Perez, M. (2019). Pre-learning vocabulary before viewing captioned video: an eye-tracking study. *Language Learning Journal*, *47*, 460–478.

Montero Perez, M., Van Den Noortgate, W., & Desmet. P. (2013). Captioned video for L2 listening and vocabulary learning: a meta-analysis. *System*, *41*, 720–739.

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audio-visual input. *TESOL Quarterly*. *53*, 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 55–577.

Peters. E., Heynen. E., & Puimège. E. (2016). Learning vocabulary through audio-visual input: the differential effect of L1 subtitles and captions. *System*, *63*, 134–148.

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *Language Learning Journal*, *47*, 424–438.

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: a study of L2 vocabulary learning by adolescents. *Language Learning Journal*, *47*, 479–496.

Rodgers, M. P. H. (2013). *English language learning through viewing television: an investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions*. Unpublished doctoral dissertation. Wellington, NZ: Victoria University.

Suárez, M. M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference? *Language Learning Journal*, *47*, 497–517.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning and Technology*, *14*, 50–73.

Tian, L., & Macaro, E. (2012). Comparing the effect of teacher codeswitching with English-only explanations on the vocabulary acquisition of Chinese university students: a lexical focus-on-form study. *Language Teaching Research*, *16*(3), 367–391.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: a meta-analysis of correlational studies. *Language Learning*, *69*, 559–599.

Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching*. London: Palgrave Macmillan.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: a dimensions approach. *System*, *41*, 609–624.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*(1), 219–258.

Zhang, P., & Graham, S. (2019). Vocabulary learning through listening: Comparing L2 explanations, teacher codeswitching, contrastive focus-on form and incidental learning. *Language Teaching Research*. https://doi.org/10.1177/1362168819829022

3 Vocabulary learning through listening

Which words are easier or more difficult to learn and why?

Pengchong Zhang and Suzanne Graham

Introduction

As highlighted in the previous chapter, while research into incidental learning through listening has increased in recent years, we know less about the nature of such learning than is the case for reading. One area that has received recent increased attention concerns lexical focus on form (Laufer, 2005) in relation to aural input, whereby learners' attention is drawn to features of the input to help them 'notice' them (Schmidt, 1990). Four recent studies (Hennebry et al., 2013; Lee & Levine, 2020; Tian & Macaro, 2012; Zhang & Graham, 2019) have investigated the extent to which lexical focus on form after aural input through different types of teacher explanations leads to vocabulary learning. Within those studies, however, there has not been any consideration of whether the effectiveness of the instruction varies according to different types of vocabulary items.

In this chapter we return to the intervention considered in Zhang and Graham (2019), exploring the impact of three types of aural lexical focus-on-form vocabulary explanations (L2-only, codeswitched, and contrastive focus-on-form) but in relation to the learning and retention of words across different word classes and of different concreteness levels.

Literature review

While previous research has considered the extent to which rates of learning differ according to word class and whether the word is concrete or abstract, the impact of different types of instruction on such learning has rarely been considered. Generally speaking, across L1 and L2 learning, nouns are learnt more easily than adjectives, which are then learnt more easily than verbs (Ellis & Beaton, 1993). There is some variation across L1s regarding the proportion of verbs to nouns that are learnt in early childhood language acquisition; in pro-drop languages such as Korean, for example, there is a higher proportion of verb learning compared with other, non pro-drop languages, perhaps because the absence of pronouns and the verb taking final position makes

verbs more salient (Kim, McGregor, and Thompson, 2000). In most languages, however, noun learning tends to precede verb learning because in the former there is a more "transparent semantic mapping of the object-reference terms to the perceptual world" (Kim et al., 2000: 227, as cited in Graham & Santos, 2013). Further aspects of difficulty in verb learning include their morphological complexity and greater difficulty in mapping verbs to actions than nouns to objects. Additionally, understanding a verb in an utterance requires the processing of a number of features at once such as tense, subject, object, and so forth (Graham & Santos, 2013; Maguire, Hirsh-Pasek, and Golinkoff, 2006). As Graham and Santos (2013) found, these demands are likely to be much greater still in an L2, and particularly in the context of aural input. They themselves found that across both a listening recall and a transcription task, intermediate learners of French identified and recalled a lower proportion of verbs compared with nouns, with a tendency to focus more on nouns especially on the task with less time for language processing and heavier cognitive load, namely the recall task.

In terms of word learning, one of the few studies to explore the learnability of words of different grammatical categories within the context of aural input was conducted by van Zeeland and Schmitt (2013). University-level learners with a variety of L1s listened to four passages in English from different genres, containing 24 target items in the form of nonwords. Target items consisted of an equal number of nouns, verbs, and adjectives, and were also equally divided between concrete and abstract words. Using a knowledge dimensions approach, in which levels of form and grammatical recognition were assessed as well as meaning recall, van Zeeland and Schmitt found that overall learning was greatest for nouns, followed by verbs, and then by adjectives.

Additionally, on all three knowledge dimensions (form, grammatical form recognition, meaning recall), concrete words were learnt significantly better than abstract words in van Zeeland and Schmitt (2013), perhaps reflecting the difference found for verb and noun learning. According to Maguire et al. (2006), verbs possess lower levels of "individuability" and "imageability" (p. 23) than nouns. As such they tend to be less easily associated with a mental image and are less concrete and more abstract. More recently, in a study of the acquisition of formulaic language through aural input in TV viewing, Puimège and Peters (2019a) found that more concrete items were more likely to be learnt than less concrete ones. The highest learning gains were for adjective-noun combinations, the lowest for verb-particle combinations, adding weight to evidence that verbs and abstract words are harder to learn. By contrast, Puimège and Peters (2019b), suggested that part of speech was not a significant predictor for either meaning recall or meaning recognition of the English words.

There is a larger body of research into the relative learnability of abstract and concrete terms than there is for grammatical word classes. Again, for

both L1 and L2 contexts, concrete words are generally learnt more easily than abstract ones, across a number of different test types (Ellis & Beaton, 1993; Ding, Liu, and Yang, 2017; van Hell & Candia Mahn, 1997). For example, in a study of 10–12-year-old learners' English-L2 vocabulary knowledge, Puimège and Peters (2019b) assessed the role of concreteness. Their results indicated that concreteness was a significant predictor of learners' passive vocabulary knowledge (meaning recognition), although not of meaning recall.

Two main positions have been taken to explain the effect of concreteness. First, dual coding theory (Paivio, 1986) holds that in semantic memory, concepts can be processed by two different systems, a verbal-based system which processes linguistic information, and an imagery-based system which processes nonverbal information. Both systems come into play for concrete words but for abstract words, the connection is made with the verbal-based system alone. For L2 words, while there is a verbal representation in both the L1 and the L2 verbal systems, for concrete words only there is also a non-verbal image common across both languages, meaning that there is thus greater anchorage for concrete words (Paivio & Desrochers, 1980).

The second position, the context availability hypothesis (Schwanenflugel & Shoben, 1983), suggests that it is the difference in availability of contextual information that makes concrete words easier to learn than abstract ones. Concrete words tend to be easier to place into a context, and tend to be associated with a narrower, more predictable set of contexts and circumstances than is the case for abstract terms. Summarizing Schwanenflugel, Akin, and Luh (1992), Ding et al. (2017) consider that it is the 'looser' representation of contextual information associated with abstract words that makes them harder to learn, because of the variety of different contexts in which they can occur in comparison to concrete words. Abstract words are thus less clearly linked with "associated contextual information in memory" (p. 97). Furthermore, and important for the study we report here, the context availability hypothesis suggests that the learnability of concrete and abstract terms can be equalized by making relevant contextual information available for both types of words.

This suggestion was investigated by van Hell and de Groot (1998), using words which had been previously rated by participants on concreteness and context availability (i.e., on a seven-point scale, participants indicated how easy or difficult it was to suggest a particular context or circumstance in which the word might appear). They found that on L1 and L2 lexical decision tasks, and for translation tasks, scores were higher for concrete words. When concrete and abstract words were matched for context availability ratings, however, there was no longer a significant effect for concreteness.

This has implications for how abstract words are presented to learners, and the type of additional information that needs to be presented with them to increase their learnability. As outlined above, a small but growing number of studies have explored the extent to which vocabulary learning through listening can be facilitated through post-listening explanations given by the teacher, typically contrasting the effectiveness of explanations in the L2,

and codeswitched or L1 explanations (Hennebry et al., 2013; Lee & Levine, 2020; Tian & Macaro, 2012). Findings across these three studies suggest the superiority of L1 explanations, but not consistently so; for example, Tian and Macaro (2012) found an advantage for codeswitching (CS) for short-term learning among university learners of English but not for long-term learning as assessed through delayed post-tests administered two to seven weeks after each instructional session. By contrast, Hennebry et al. (2013) found a longer-term CS advantage (between one and four weeks after the vocabulary instruction) for high-school learners of French. In the most recent study, Lee and Levine (2020) report similar levels of vocabulary retention for explanations given in the L2 or the L1 for advanced ESL learners, but a significant advantage for intermediate learners for the L1 explanations condition.

One potential weakness of codeswitched, L1 explanations may be the amount of information that is conveyed about a given item, if the codeswitching only amounts to what Laufer and Girsai (2008) refer to as "bilingual glosses which simply state the meaning of L2 words". They compare that approach with contrastive form-focused instruction "which leads to learners' understanding of the similarities and differences between their L1 and L2 in terms of individual words and the overall lexical system", through the explicit highlighting of such differences (p. 696). In Zhang and Graham (2019) we found an advantage for short and long-term vocabulary learning from post-listening explanations that included such contrastive focus on form (CFoF), compared with codeswitched (CS), L2, and no explanations. Additional contrastive information may be especially useful for abstract words and verbs. L1 translations may be advantageous for concrete terms and nouns, facilitating the formation of a direct conceptual link. According to Kroll and Stewart's (1994) revised hierarchical model of bilingual memory, L2 words need to be attached to their L1 translation before access to the underlying mental concepts occurs (i.e., at lower proficiency levels). However, direct conceptual links between L2 words and conceptual representations can also be created when higher proficiency levels are reached, although again this may be easier with nouns and with verbs. Adjectives may fall somewhere in between verbs and nouns in terms of concreteness and contextual availability.

The present study

The data collected for the present study were from a larger intervention study, which compared the vocabulary learning through listening of Chinese high-school EFL learners who received three types of vocabulary explanations (L2, CS, and CFoF, forming three treatment groups) and those who did not receive any input enhancement (no explanation group, NE). It also investigated whether the listening proficiency of these four groups was improved through the intervention. Findings indicated that the three treatment groups learnt and retained significantly more vocabulary than the NE group did. In addition, the CFoF approach showed a significant advantage over the L2 and

CS approaches for both short-term and long-term vocabulary improvement. Regarding listening however, the NE group had significantly larger pre to post gains than the CFoF did (all as reported in Zhang & Graham, 2019).

The present study considered the three treatment groups alone and adopted a different perspective on their learning, investigating the extent to which the impact of each type of vocabulary explanations varied across different word classes and for words of different concreteness levels. We therefore had two research questions:

1. To what extent does the impact of the three types of post-listening vocabulary explanations (target language explanations; teacher codeswitching; and contrastive Focus-on-Form) vary for words of different concreteness levels, for short-term and long-term vocabulary learning?
2. To what extent does the impact of the three types of explanations vary for words across different word classes, for short-term and long-term vocabulary learning?

Method

Sampling and baseline measurement

Four classes of 137 first-year high school EFL learners participated in the present study. Following Tian and Macaro (2012), we adopted a quasi-experimental design, randomly assigning these four intact classes to three treatment groups (L2, CS, and CFoF) and one NE group (whose data were not considered for the present analysis, because we wanted to focus on the effects of explanations leaving 114 learners). These learners had seven years' EFL learning experience (four years in primary school and three years in secondary school), and therefore were between A2-B1 level (Common European Framework of Reference for Languages) for English language proficiency.

To further measure learners' English language proficiency as well as to assess any partial knowledge they had of the target words for the classroom intervention, we administered a baseline general vocabulary knowledge test (GEVT) with all four groups. The GEVT was in two parts. The first was based on the aural vocabulary levels test (AVLT) developed by McLean, Kramer, and Beglar (2015). We decided to include only 100 items from that test, measuring knowledge of up to the 3000 English word frequency level and academic word list, which piloting with learners of a similar level showed to be appropriate. Participants first heard the researcher reading out an English word together with an example sentence for that word. They were then required to select from the four given options matching the correct Chinese meaning to that word. Taking the same format as the AVLT, the second part of the GEVT was the vocabulary pre-test, which included 60 items, assessing participants' partial-knowledge of the target lexical items used for the intervention. The 100 items for the AVLT and the 60 items for the vocabulary test were then intermingled in order to reduce possible priming effects of having a pre-test.

Intervention procedures

The classroom intervention started three weeks after the participants completed the baseline GEVT test. The three treatment groups received six intervention sessions respectively, one each week, delivered by the first author. Each intervention session started with a listening comprehension task, followed by the researcher providing vocabulary explanations for 10 target lexical items (the majority of which were single words) which appeared in the listening passage. A vocabulary post-test assessing the short-term learning of these target items was administered at the end of each session. From the third intervention session inclusive onwards, an additional vocabulary delayed post-test measuring the long-term retention of the vocabulary items taught two weeks before was administered together with the post-test for the current session. In the following two weeks after the final intervention session, we had two additional sessions with the participants (one each week), which enabled us to administer the delayed post-tests for the fifth and the sixth intervention sessions and therefore ensured an equal two-week delay between each post-test and delayed post-test. For further details of the research design, see Zhang and Graham (2019).

Listening passages and target lexical items

The six listening passages used for the intervention were initially selected from a standard EFL textbook. They were then modified to ensure that the topics were relevant to the participants and that they had an equal length of approximately 250 words each (Tian, 2011). Six passages were recorded by three English L1 speakers, controlling the speech rate at between 150–190 words/min (Brindley & Slatyer, 2002; Tauroza & Allison, 1990). Lexical coverage for these passages was controlled at a minimum 95% level of what the high school English curriculum and the textbook authors indicated participants would know. Three listening comprehension questions were designed for each passage, involving one global question asking for the general idea of the passage and two local questions focusing on detailed information from the passage. Altogether 43 single words (seven to eight for each passage) and 17 collocations (two to three for each passage, not considered in the present analysis) were selected as target items from these passages. The 43 single words included 18 nouns, 13 verbs, and 12 adjectives.

Three types of vocabulary explanations

The content of the intervention sessions was similar across the three treatment groups, apart from the vocabulary explanations given by the researcher. For the L2 and CS groups, participants were first told the part of speech of the target word, followed by the meaning of the word, in either English or Chinese respectively. Both groups were then given a sentence

exemplifying the target word and were required to understand the sentence using the meaning of the word provided. The CFoF group, similar to the CS group, were told the part of speech and L1 meaning of the target word. Instead of having a sentence exemplifying the word, they were provided with additional cross-linguistic information about how the word functions in Chinese and in English. Examples of each type of vocabulary explanation are given below:

> 'Yesterday, another student and I, representing our university's student ***association***, went to the Capital International Airport to meet this year's international students.'
>
> **L2 explanation**: Here, association is a noun which means 'club or society where a group of people working together'. Therefore, in the listening passage, student association means student club or society. Another example for this word can be 'we have a costume play association in our school'.
>
> **CS explanation**: Here, association is a noun which means '协会, 社团'. Therefore, in the listening passage, student association means student 协会 or student 社团. Another example for this word can be 'we have a costume play association in our school'.
>
> **CFoF explanation**: Here, association is a noun. In the structure 'n. + ***association***', it means '协会, 社团'. Therefore, student association in the listening passage means student 协会 or student 社团. However, it has a different meaning '联合' when used as '***association*** with sb./sth.'.

Hence in the CS explanations, the L2 to L1 translation potentially provided a direct conceptual link with the concept conveyed by the word, but offered relatively sparse contextual information about it, except for in the example sentence. The latter was also given in the L2 explanations, but the L2 explanation offered potentially fuller, if less direct, contextual information about the word. Finally, in the CFoF condition, the initial L2 to L1 translation again potentially provided a direct conceptual link with the concept conveyed by the word, which was then supplemented by a fuller amount of contextual information about the word in so far as it compared and contrasted how the L1 and L2 expressed the same concept.

Vocabulary post-test and delayed post-test

An aural form of vocabulary meaning recall test, modified from the one employed in Tian and Macaro (2012), was used to measure any learning of the target words through the intervention. Altogether there were six post-tests and six delayed post-tests as outlined in the intervention procedures. Similar to the GEVT, these tests involved the researcher reading out the target lexical item and an example sentence for that item. Participants were then asked to write down the meaning of the lexical item they heard, either in Chinese or in English.

Data analysis

Our data analysis involved undertaking logistic regression tests using generalized linear mixed-effects models (GLMMs), 'lmerTest' Package (Kuznetsova, Brockhoff, and Christensen, 2017), in R computing environment (R Core Team, 2018). All models included random intercepts for participant (137 participants) and item (43 words, coded 1 if correct while 0 if wrong). Both by-participant and by-item random slopes were fit using a maximal random effects structure (Barr et al., 2013). In cases where a full random effects structure model did not converge, we first took out the interactions between random slopes, and then gradually removed random slopes which accounted for the least variance until a converged model was obtained. Further simplification of both random and fixed effects structures was made on the converged model, following the backward selection procedure, in order to obtain the final model with best fit of the data. Effect sizes were calculated as odds ratios, but in cases where the predictor only involved categorical variables, odds ratios were translated into Cohen's *d*, and the following rules applied to interpret the latter: between-groups contrasts, small = .4, medium = .7, large = 1.0; within-groups/pre-post contrasts, small = .6, medium = 1.0, large = 1.4 (Plonsky & Oswald, 2014).

Items were categorized as either verbs, nouns, or adjectives. We then used Brysbaert et al.'s (2014) concreteness index [rating from 1 (abstract) to 5 (concrete)] to indicate the concreteness level of each item (M = 2.96, SD = 1.09, Min = 1.43, Max = 4.96).

Findings

In order to address our two research questions, exploring the impact of the intervention on the learning and retention of the target words across different word classes and at different concreteness levels, descriptive statistics for the three vocabulary tests were first calculated within each group for each word class (Table 3.1).

Subsequently, a first model, Model A, was built by running a series of logistic regression tests using GLMMs, which included four fixed effects: Group (L2 vs. CS vs. CFoF); Time (1. pre-test vs. 2. post-test vs. 3. delayed post-test); Conc (concreteness index); Type (noun vs. verb vs. adjective). Three-way interactions involving the fixed effect of Time and Group were also added to this model (i.e., Time × Group × Conc; Time × Group × Type). For the three categorical fixed effects, L2 was set as the baseline level for Group, and Time 1 for Time and Verb for Type. The random effects structure for the final converged model included random slopes for both Participants and Items, by-Participant random slopes for Time, and by-Item random slopes for Time.

The two R^2 values ($R^2_{marginal}$ = 0.30; $R^2_{conditional}$ = 0.64) indicated that Model A represented a relatively good-fit to the data, with 30% of the variance in the dependent variable explained by the fixed effects and the 64% of the variance in the dependent variable explained by both the fixed effects and random effects. The baseline level for the fixed effects of Type and Group was relevelled

Table 3.1 Descriptive statistics for the vocabulary tests by word classes within each treatment group (%)

Group (N)	*Word class*		*Pre-test*	*Post-test*	*Delayed post-test*
L2 (35)	Noun	Mean (*SD*)	30.48 (11.54)	51.11 (16.10)	38.10 (18.43)
		Min – Max	5.56–50.00	22.22–88.89	5.56–72.22
	Verb	Mean (*SD*)	22.20 (10.18)	40.22 (16.70)	21.76 (15.10)
		Min – Max	7.69–46.15	7.69–76.92	0.00–61.54
	Adjective	Mean (*SD*)	15.71 (9.85)	44.05 (14.93)	18.81 (11.67)
		Min – Max	0.00–33.33	16.67–66.67	0.00–50.00
CS (36)	Noun	Mean (*SD*)	34.92 (14.61)	77.14 (13.53)	33.65 (17.57)
		Min – Max	5.56–61.11	50.00–100.00	5.56–83.33
	Verb	Mean (*SD*)	18.46 (9.91)	63.52 (18.41)	23.30 (15.88)
		Min – Max	0.00–38.46	23.08–92.31	0.00–53.85
	Adjective	Mean (*SD*)	20.00 (14.03)	70.71 (13.61)	18.10 (14.07)
		Min – Max	0.00–50.00	41.67–91.67	0.00–50.00
CFoF (33)	Noun	Mean (*SD*)	19.44 (10.07)	86.93 (12.30)	58.85 (20.20)
		Min – Max	0.00–55.56	55.56–100.00	16.67–94.44
	Verb	Mean (*SD*)	12.67 (7.06)	71.72 (14.39)	38.24 (16.34)
		Min – Max	0.00–23.08	38.46–92.31	7.69–76.92
	Adjective	Mean (*SD*)	7.11 (7.98)	83.58 (14.58)	32.35 (16.51)
		Min – Max	0.00–25.00	50.00–100.00	0.00–75.00

to obtain all contrasts of interest. Results from Model A indicated that there were Time × Group × Concreteness and Time × Group × Type three-way interactions (Table 3.2), suggesting that learners from the different treatment conditions progressed differently in terms of learning words of different word classes and at different concreteness levels. The effect plots for the two sets of three-way interactions are given in Figure 3.1 and 3.2 respectively.

Regarding the Time × Group × Conc interactions, two contrasts were found, one between the CS and L2 groups, the other between the CS and CFoF groups. The odds ratio for the CS/L2 contrast (Line2, Table 3.2) suggested that with one unit increase in the concreteness level (from more abstract to more concrete), the odds of learners in the L2 group correctly recalling the meaning of the target word from the pre-test to the delayed post-test were 1.45 (1/0.69) times greater than for those in the CS group. Similarly, with one unit increase in the concreteness level, the odds of learners in the CFoF group correctly recalling the meaning of the target word from the pre-test to the post-test were 1.45 times greater than for those in the CS group (Line 5, Table3. 2). That is, while the CS approach was more beneficial for learning more abstract words compared with the L2 and CFoF approaches, L2 and CFoF explanations were more beneficial than codeswitching for learning more concrete words. No significant CFoF/L2 contrasts were confirmed (Lines 3–4, Table 3.2), meaning learners in these two groups progressed similarly in learning the words at different concreteness levels.

Table 3.2 Results for the three-way interactions in Model A

	Fixed effects	*β (SE)*	*95% CI for Odds Ratio*		
			Lower	*Odds Ratio*	*Upper*
1	$Time_{Time2 \rightarrow Time1} \times Group_{CS \rightarrow L2} \times Conc$	-0.22 (0.14)	0.61	0.80	1.06
2	$Time_{Time3 \rightarrow Time1} \times Group_{CS \rightarrow L2} \times Conc$	-0.37 (0.14)**	0.52	0.69	0.91
3	$Time_{Time2 \rightarrow Time1} \times Group_{CFoF \rightarrow L2} \times Conc$	0.15 (0.16)	0.86	1.17	1.59
4	$Time_{Time3 \rightarrow Time1} \times Group_{CFoF \rightarrow L2} \times Conc$	-0.25 (0.15)	0.58	0.78	1.04
5	$Time_{Time2 \rightarrow Time1} \times Group_{CFoF \rightarrow CS} \times Conc$	0.37 (0.16)*	1.07	1.45	1.97
6	$Time_{Time3 \rightarrow Time1} \times Group_{CFoF \rightarrow CS} \times Conc$	0.12 (0.15)	0.84	1.13	1.51
7	$Time_{Time2 \rightarrow Time1} \times Group_{CS \rightarrow L2} \times Type_{Adj \rightarrow Verb}$	-0.52 (0.39)	0.27	0.59	1.29
8	$Time_{Time3 \rightarrow Time1} \times Group_{CS \rightarrow L2} \times Type_{Adj \rightarrow Verb}$	-1.36 (0.42)**	0.11	0.26	0.58
9	$Time_{Time2 \rightarrow Time1} \times Group_{CS \rightarrow L2} \times Type_{Noun \rightarrow Verb}$	-0.12 (0.33)	0.47	0.89	1.68
10	$Time_{Time3 \rightarrow Time1} \times Group_{CS \rightarrow L2} \times Type_{Noun \rightarrow Verb}$	-1.17 (0.33)***	0.16	0.31	0.59
11	$Time_{Time2 \rightarrow Time1} \times Group_{CS \rightarrow L2} \times Type_{Noun \rightarrow Adj}$	0.40 (0.41)	0.67	1.50	3.32
12	$Time_{Time3 \rightarrow Time1} \times Group_{CS \rightarrow L2} \times Type_{Noun \rightarrow Adj}$	0.19 (0.41)	0.54	1.21	2.70
13	$Time_{Time2 \rightarrow Time1} \times Group_{CFoF \rightarrow L2} \times Type_{Adj \rightarrow Verb}$	0.43 (0.46)	0.63	1.54	3.78
14	$Time_{Time3 \rightarrow Time1} \times Group_{CFoF \rightarrow L2} \times Type_{Adj \rightarrow Verb}$	-0.45 (0.45)	0.26	0.64	1.55
15	$Time_{Time2 \rightarrow Time1} \times Group_{CFoF \rightarrow L2} \times Type_{Noun \rightarrow Verb}$	-0.21 (0.37)	0.39	0.81	1.67
16	$Time_{Time3 \rightarrow Time1} \times Group_{CFoF \rightarrow L2} \times Type_{Noun \rightarrow Verb}$	-0.18 (0.35)	0.42	0.83	1.66
17	$Time_{Time2 \rightarrow Time1} \times Group_{CFoF \rightarrow L2} \times Type_{Noun \rightarrow Adj}$	-0.64 (0.47)	0.21	0.53	1.33
18	$Time_{Time3 \rightarrow Time1} \times Group_{CFoF \rightarrow L2} \times Type_{Noun \rightarrow Adj}$	0.27 (0.44)	0.55	1.31	3.11
19	$Time_{Time2 \rightarrow Time1} \times Group_{CFoF \rightarrow CS} \times Type_{Adj \rightarrow Verb}$	0.95 (0.45)*	1.07	2.59	6.26
20	$Time_{Time3 \rightarrow Time1} \times Group_{CFoF \rightarrow CS} \times Type_{Adj \rightarrow Verb}$	0.91 (0.45)*	1.04	2.49	5.98
21	$Time_{Time2 \rightarrow Time1} \times Group_{CFoF \rightarrow CS} \times Type_{Noun \rightarrow Verb}$	-0.09 (0.37)	0.44	0.91	1.89
22	$Time_{Time3 \rightarrow Time1} \times Group_{CFoF \rightarrow CS} \times Type_{Noun \rightarrow Verb}$	0.99 (0.35)**	1.36	2.69	5.32
23	$Time_{Time2 \rightarrow Time1} \times Group_{CFoF \rightarrow CS} \times Type_{Noun \rightarrow Adj}$	-1.04 (0.47)*	0.14	0.35	0.89
24	$Time_{Time3 \rightarrow Time1} \times Group_{CFoF \rightarrow CS} \times Type_{Noun \rightarrow Adj}$	0.08 (0.44)	0.46	1.08	2.54

Note. * – $p < .05$, ** – $p < .01$, *** – $p < .001$; CI = confidence interval.

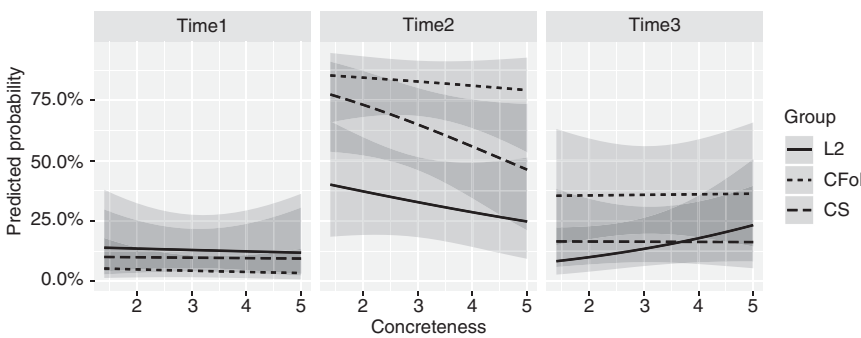


Figure 3.1 Time × Group × Concreteness effect plot.

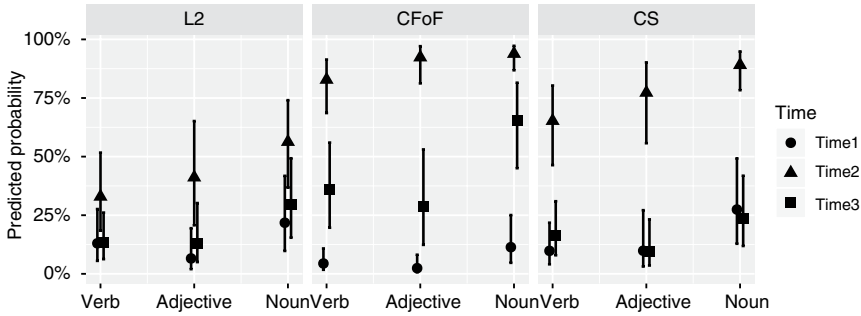


Figure 3.2 Time × Group × Type effect plot.

Interpreting the Time × Group × Type interactions, two significant CS/L2 and four significant CS/CFoF contrasts were confirmed. The first CS/L2 contrast (Line 8, Table 3.2) suggested that for long-term vocabulary retention, the L2 approach showed a significant advantage over the CS approach for learning the target adjectives, while the CS approach was more beneficial for learning the verbs compared with the L2 approach. There was a medium effect size for this contrast. The second CS/L2 contrast (Line 10, Table 3.2) indicated that comparing the long-term learning of nouns and verbs, the L2 approach was significantly better than the CS approach for learning the nouns, while the CS approach was superior to the L2 approach for learning the verbs. A small effect size was found for this contrast.

The first two CS/CFoF contrasts (Lines 19–20, Table 3.2) suggested that considering both short-term and long-term learning for the adjectives and verbs, the CFoF approach was significantly better than the CS approach for learning the adjectives while the CS approach was more beneficial than the CFoF approach for learning the verbs. In addition, the third CS/CFoF contrast (Line 22, Table 3.2) suggested that for the long-term learning comparing nouns and verbs, the CS approach was more beneficial for learning the verbs and the CFoF approach was significantly better for the nouns. The final CS/CFoF contrast (Line 23, Table 3.2) indicated a short-term advantage of the CFoF over the CS approach for learning the adjectives and therefore a corresponding short-term advantage for the CS over the CFoF approach for the nouns. All the four contrasts showed small effect sizes. No significant CFoF/L2 contrasts were confirmed (Lines 13–18, Table 3.2), indicating that for learning words in different classes, learners in these two groups showed similar improvements.

Finally, as we were also interested in whether the learning of different word classes or words at different concreteness level differed across the three treatment conditions, a second model was built, but only included two two-way interactions (i.e., Time × Type; Time × Conc). Model selection of the converged model indicated while the random effects structure was retained (random intercepts for Participants and for Items, by-Item random slopes for

Time, and by-Participant random slopes for Time), the Time × Type interaction could be removed from the model ($X^2(4) = 4.65$, $p = .32$). This simplified model was therefore named Model B for further interpretation. Model B represented a good-fit to the data ($R^2_{marginal} = 0.23$; $R^2_{conditional} = 0.64$). Results showed that there was a main effect of $Type_{Noun \rightarrow Verb}$ ($\beta = 1.09$, $SE = 0.45$, $p = .02$, $d = .60$, 95%CI[1.22, 7.19]), indicating that regardless of the treatment condition and the test time point, the acquisition of the nouns was significantly better than that of the verbs, with a small effect size. Model B results, however, did not show any significant $Time_{Time2 \rightarrow Time1}$ × Conc or $Time_{Time3 \rightarrow Time1}$ × Conc interactions, meaning that when the learning outcomes were averaged across the three treatment groups, the effect of word concreteness did not exist.

Discussion

We begin with a summary of our main findings.

- Regarding learning across the three treatment groups, concrete and abstract words were acquired and retained to a similar degree.
- The two three-way Time × Group × Conc interactions from the effect plot (Figure 3.1) suggested that:
    a. For short-term learning, CFoF seemed to be the most balanced teaching approach for words at different concreteness levels, followed by the L2 approach. The CS approach was the most helpful for the abstract words.
    b. For long-term learning, both CFoF and CS seemed to be balanced teaching approaches, yet the L2 approach was the most beneficial for the concrete words.
- Regarding short-term learning, the CFoF approach was better than the CS approach for learning the adjectives, yet the CS approach was more beneficial than the CFoF approach for the verbs and nouns.
- For long-term learning, the CFoF and L2 approaches were more beneficial than the CS approach for the target nouns and adjectives, while the CS approach was more helpful than the CFoF and L2 approaches for the retention of the target verbs.

We now interpret these findings in relation to the research questions and the theoretical framework discussed earlier.

Our first research question investigated to what extent the impact of the three types of explanations varied for words of different concreteness levels, for short-term and long-term vocabulary learning. When exploring the vocabulary learning across the three treatment groups, we did not find a significant effect of concreteness levels, meaning that concrete words and abstract words were acquired and retained similarly. This differs from what was found by van Zeeland and Schmitt (2013), most likely because we had another predictor, Group (i.e., the three intervention conditions), within our

study design. The three conditions provided contextual information, albeit at different levels, which overall may have eradicated the effect of concreteness, as would be suggested by the Context Availability Hypothesis (CAH) and as was found by van Hell and de Groot (1998).

Further exploration of the two three-way Time × Group × Conc interactions confirmed in the model results gives a clearer picture. The first interaction was found between the CS and CFoF groups indicating that, for short-term learning, the CFoF approach was more beneficial for learning the concrete words and the CS approach was better for learning the abstract words. Three issues should be noted here. First, as observed from the effect plot in Figure 3.1, in both approaches the abstract words were learnt better than the concrete words, especially in the CS condition. Second, this advantage of the CS approach over the CFoF approach for learning abstract words disappeared at the delayed post-test, with both approaches showing equal benefits for learning words of different concreteness levels. Third, short-term and long-term learning of both the concrete and the abstract words was greater for the CFoF group compared with the CS group overall.

These somewhat complicated findings suggested that although the CS approach was more beneficial for learning the abstract words, the CFoF approach seemed to be a more balanced approach for learning words of different concreteness levels. Learners in that group received both the L1 translation of the concrete and abstract words and then additional contextualizing information in the form of cross-linguistic information. This conforms firstly to the CAH (Schwanenflugel & Shoben, 1983), whereby if sufficient contextual information about both concrete and abstract words is provided, the advantage for learning concrete items over abstract ones may disappear (Ding et al., 2017). Since learners in the CFoF were provided with cross-linguistic information about the target words, arguably a type of enhanced contextual information, they showed similar progress for learning the concrete words and abstract words in the intervention. As they also received direct L1 translations of the words, however, they enjoyed similar advantages to those of the CS group in so far as a conceptual link (non-verbal representation) was built for the target items with the mental lexicon as per the dual coding theory and Kroll and Stewart (1994).

Our other findings further suggest that enhancing contextual information irons out concreteness effects for initial learning, but that this equalizing effect diminishes in the long term; the findings are hence only partially in line with the CAH. Support for this argument arises from the second three-way interaction that was obtained between the CS and L2 groups. This interaction suggested that, for long-term learning, both CFoF and CS seemed to be balanced teaching approaches, yet more concrete words were retained significantly better by the learners from the L2 group while the CS approach was significantly superior for helping learners retain the abstract words.

These findings thus suggest that, for long-term learning, the dual coding theory (Paivio & Desrochers, 1980; Paivio, 1986) seems to hold true, whereby

concrete words establish both verbal and non-verbal representations across the two languages yet there are only verbal representations available for abstract words. By providing the L1 meaning of the target words using the CS approach, a conceptual link (non-verbal representation) was built for the target items with the mental lexicon (Kroll & Stewart, 1994). The acquisition of the abstract words was therefore brought up to the level of the concrete words for learners in the CS group. This did not happen for learners in the L2 group by contrast, for whom the target language explanations did not facilitate the establishment of this conceptual link and for whom thus the concrete words were learnt more easily.

Turning to our second research question, exploring to what extent the impact of the intervention varied for the learning and retention of the target words across different word classes, again unlike previous research (Graham & Santos, 2013; van Zeeland & Schmitt, 2013) which indicated nouns were more easily learnt than verbs, we did not find any significant Time × Type two-way interactions, meaning that when looking at the data from the three treatment conditions as a whole, the learning and retention of the target words was similar across the three word classes. A series of three-way Time × Group × Type interactions was however confirmed, indicating that the effect of word classes was significantly dependent on the predictor Group. Similar to what we found for the Time × Group × Concreteness interactions, the contrasts obtained were between the CS and L2 groups and between the CS and CFoF groups.

Two significant CS/L2 contrasts were obtained for long-term vocabulary retention, one between verbs and adjectives and the other between verbs and nouns. On the one hand, the CS approach showed significant advantages for learning the verbs. Conversely, nouns and adjectives were retained better when they were delivered using the L2 approach. One possible explanation may be that the meanings of the verbs was easier to convey in the L1 than in the target language, as direct translations are likely to give a clearer and fuller sense of the concept, leading to the superior learning of these words for the CS group. By contrast, the meanings of the target nouns and adjectives were less difficult to convey in the target language. They were therefore retained better by learners from the L2 group. These findings would concur with those of Lee and Levine (2020), whereby intermediate level learners in their study seemed to have experienced difficulties in understanding L2-only vocabulary explanations, especially for words which are difficult to convey in the target language (i.e., verbs), and therefore benefited significantly more from the L1 vocabulary explanations.

Regarding the learning differences of different word classes between the CS and CFoF groups, our findings in general indicated that the CFoF approach was significantly superior for learning the adjectives and nouns compared with the verbs while the CS approach correspondently was better for the learning of the target verbs than the adjectives and nouns. As discussed above, verbs conveying more abstract concepts tend to benefit more from direct L1

translation as was provided for the CS group, enabling a directly conceptual link to be built between the word and the mental lexicon (Kroll & Stewart, 1994). The target adjectives and nouns, however, were 'easier' words with higher levels of concreteness and contextual availability (Schwanenflugel & Shoben, 1983) than the verbs; their learning may have been facilitated by the additional cross-linguistic information in terms of how to use them, and were therefore retained better under the CFoF approach.

Conclusions and pedagogical implications

The present study explored the impact of three types of vocabulary explanations on 114 high school EFL learners' learning and retention of 43 words across different word classes and of different concreteness levels. We found that providing additional lexical focus-on-form vocabulary explanations overall may have eradicated the effect of word concreteness, balancing the uptake rates between the concrete and abstract words. In addition, our findings suggested that, on the one hand, the CFoF approach seemed to be a balanced approach for teaching words at different concreteness levels. On the other hand, abstract words were learnt better than concrete words under the CS approach while the learning of concrete words benefited more than the abstract words from the L2 explanations. Moreover, compared with the CS approach, the L2 and CFoF approaches were better for teaching nouns and adjectives. The verbs, however, were learnt better under the CS approach than under the L2 and CFoF approaches.

From a pedagogical point of view, these somewhat complicated findings suggest that second language teachers need to carefully plan and select the appropriate type of vocabulary explanation for teaching words across different word classes and of different concreteness levels. Although using the L2 approach may be helpful for learning concrete words, it needs to be modified, for example, combining it with codeswitched L1 explanations to maximize the learning of abstract words for which the concepts are relatively difficult to acquire through the L2 approach. Additionally, in order to improve the learning of the verbs, CFoF and L2 explanations may need to be further simplified to make them more accessible. To facilitate the learning of nouns and adjectives however, CS explanations may need to be supplemented with additional L2 or cross-linguistic explanations.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, *19*(4), 369–394. https://doi.org/10.1191/0265532202lt236oa

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading a Foreign Language, 20*(2), 136–163.

Ding, J., Liu, W., & Yang, Y. (2017). The influence of concreteness of concepts on the integration of novel words into the semantic network. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02111

Ellis, N., & Beaton. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*(4), 559–617. http://dx.doi.org/10.1111/j.1467–1770.1993.tb00627.x

Graham, S., & Santos, D. (2013). Selective listening in L2 learners of French. *Language Awareness*, *22*(1), 56–75. https://doi.org/10.1080/09658416.2011.652634

van Hell, J., & Candia Mahn, A. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experiences and inexperienced learners. *Language Learning*, *47*(3), 507546. https:// doi.org/10.1111/0023-8333.00018

van Hell, J., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association.

Bilingualism: *Language and Cognition*, *1*(3), 193–211. https:// doi.org/ 10.1017/S1366728998000352

Hennebry, M., Rogers, V., Macaro, E., & Murphy, V. (2013). Direct teaching of vocabulary after listening: Is it worth the effort and what method is best? *Language Learning Journal*, *45*(3), 282–300. https://doi.org/10.1080/09571736.2013.849751

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In Robinson, P. (ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.

Kim, M., McGregor, K. K., & Thompson, C. K. (2000). Early lexical development in English- and Korean-speaking children: Language-general and language-specific patterns. *Journal of Child Language, 27*(2), 225–254. https://doi.org/10.1017/S0305000900004104

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*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kuznetsove, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laufer, B. (2005). Focus on form in second language vocabulary learning. *Eurosla Yearbook, 5*(1), 223–250. http://dx.doi.org/10.1075/eurosla.5.11lau

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694–716. https://doi.org/10.1093/applin/amn018

Lee, J. H., & Levine, G. S. (2020). The effects of instructor language choice on second language vocabulary learning and listening comprehension. *Language Teaching Research*. *24*, 250–272. https://doi.org/10.1177/1362168818770910

Maguire, M. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2006). A unified theory of word learning: Putting verb acquisition in context. In Hirsh-Pasek, K., & Golinkoff, R. M. (eds) *Action meets word: How children learn verbs* (pp. 364–391). New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195170009.003.0015

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research, 19*(6), 741–760. https://doi.org/10.1177/1362168814567889

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology, 34*(4), 388–399. https://doi.org/10.1037/h0081101

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Puimège, E., & Peters, E. (2019a). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *Language Learning Journal*, *47*, 424–438. https://doi.org/10.1080/09571736.2019.1638630

Puimège, E., & Peters, E. (2019b). Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning*, *64*, 944–977. https://doi.org/10.1111/lang.12364

R Development Core Team. (2018). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 82–102. https://doi.org/10.1037/0278-7393.9.1.82

Schwanenflugel, P. J., Akin, C., & Luh, W. M. (1992). Context availability and the recall of abstract and concrete words. *Memory and Cognition*, *20*, 96–104. https://doi.org/10.3758/BF03208259

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics*, *11*(1), 90–105. https://doi.org/10.1093/applin/11.1.90

Tian, L. (2011). *Teacher codeswitching in a communicative EFL context: Measuring the effects on vocabulary learning*. Oxford: VDM.

Tian, L., & Macaro, E. (2012). Comparing the effect of teacher codeswitching with English-only explanations on the vocabulary acquisition of Chinese university students: A Lexical Focus-on-Form study. *Language Teaching Research, 16*(3), 367–391. https://doi.org/10.1177%2F1362168812436909

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609–624. https://doi.org/10.1016/j.system.2013.07.012

Zhang, P., & Graham, S. (2019). Vocabulary learning through listening: Comparing L2 explanations, teacher codeswitching, contrastive focus-on-form and incidental learning. *Language Teaching Research*. https://doi.org/10.1177/1362168819829022

4 Vocabulary and listening

James Milton and Ahmed Masrai

Introduction

The link between vocabulary knowledge in a second language and the learner's ability to function in that second language is now well-established. In simple terms, and for most learners up to the highest levels of performance, the relationship is thought of as linear where the more words a learner knows then the more they are likely to understand and better they are likely to communicate in their second language. So, in an early test of vocabulary size (*Eurocentres Vocabulary Size Test*; Meara & Jones, 1990), created as a placement test, the authors point to the way scores on the test allow learners to be placed in classes of the right language level and with other learners at the same level. But they also point to the way, it seems, that vocabulary predicts language performance in some skills better than others. Scores on the *Eurocentres Vocabulary Size Test* predict performance in test of grammatical knowledge, reading comprehension, and writing particularly well. They do not, it seems, predict the oral/aural skills of speaking and listening so well. Stæhr's (2008) paper appears to illustrate this weaker relationship in his study of the relationship between scores on Nation's (2001) Vocabulary Levels Test, used as a proxy for vocabulary size, and school test scores of reading, writing, and listening. The correlations he notes are high for writing and particularly reading but only a moderate, though still statistically significant, correlation is noted with the listening scores.

There are reasons for thinking that vocabulary knowledge may have a different relationship with these two different types of skill: the oral/aural and the written skills. Milton (2009) points to the way users of English tend to use a greater variety of words, and sometimes different words, in writing than they do in speech. The pronoun *I*, for example, is more frequent in speech than in most forms of writing, and we tend to use the most frequent words in English even more frequently in speech than in writing. In principle, therefore, it is possible to reach the 95% and 98% coverage of texts, associated with fluency, with fewer words in speech than in writing. Both Nation (2006) and Laufer & Ravenhorst-Kalovski (2010) suggest knowledge of the most frequent 8,000 or even 9,000 words in English are needed to handle written texts with fluency.

In a skill such as listening, therefore, the relationship need not be linear and learners may, for most purposes, acquire all the words they need at something far less than 9,000 words. There are historic figures that suggest such a number might be very small (Schonell, Middleton, and Shaw, 1956) but more recent, and probably more useful, figures suggest about 3,000 words for general speaking and listening (Adolphs & Schmitt, 2003) and maybe up to 6,000 or 7,000 for other activities, such watching a film in English (Nation, 2006).

Our understanding of the way vocabulary knowledge interacts with listening fluency is coloured by the quantity of research which addresses this issue. As Vandergrift (2007) reflects, L2 listening has received relatively little research attention. This absence of a good research base extends, van Zeeland & Schmitt (2013) suggest, to incidental vocabulary learning through listening and the details of the relationship this vocabulary has to the listening process. If the relationship between vocabulary and communicative fluency is different for written and for aural skills, this raises the question why this should be so. This could be a processing issue, a product of the way the spoken and written words are managed differently by the brain. Alternatively, this may be a knowledge issue, that the knowledge we have of written forms may be significantly different from our knowledge of aural forms. Aural forms may be stored in a different place or in a different way.

Listening and processing issues

The research we have acknowledges the idea that the processing of spoken language will have to be significantly different from that of written language. Van Zeeland & Schmitt (2013) point, for example, to the way that fast processing is needed for listening. In reading, the speed with which words are processed is controlled by the reader but the listener has no such control. In reading, the reader can process the language faster or slower as need demands. For the language learner, however, spoken language will likely be delivered at a pace where the words cannot be recognized and processed, and meaning can very rapidly be lost. Listening is also unidirectional and transient. The spoken word is delivered, and if meaning is lost, there is little recourse to revisit what was said or slow it down. Once the spoken word is uttered, it is gone. The written word, by contrast can be read multiple times if needed. The general conclusion of this is that for L2 learners, communication failures are much more likely in listening than in writing, because of the nature of the medium and the processing this entails.

Van Zeeland & Schmitt (2013) also point to the absence of word boundaries in listening. In writing, and in most languages, individual words are separated by spaces and this is a major asset in comprehension. These word boundaries do not exist in normal speech, and to introduce them would be highly artificial. Words, even if they are known by sound, may be missed or misidentified in the flow of language where other sounds and syllables precede and follow immediately. To these difficulties might be added the inconvenient

way word pronunciation can vary according to context. In English, some words have both citation and weak forms of pronunciation and the forms interchange according to the phonetic context. The process of speech also allows words to change pronunciation according to style or speed of delivery. Words may change their phonological form as processes of assimilation and elision amend or even remove phonemes from the citation form of the word.

The extra difficulty in listening, therefore, comes from the need to process at speed, but also the need to recognize words from a multiplicity of different, and possibly partial, forms. Perhaps it is not surprising, then, that the lexicon associated with the spoken word is generally smaller than that of the written word. The listener has a better chance of recognizing words if there is a smaller number of words to recognize.

Listening and vocabulary knowledge issues

There is, we think, no question that these issues of the different medium and the processing issues which are required will contribute to some of the difficulties observed in the process of listening as distinct from the process of reading the written word. However, there is another explanation for thinking that the relationship between vocabulary and the skill of listening particularly appears as it does. This explanation can also explain the observations made by Stæhr (2008) and Meara & Jones (1990) that vocabulary size appears not to inter-relate with comprehension in listening as strongly as it does with the written skills. The issue may be a knowledge problem, and this suggests that the written and spoken word forms may be stored differently and the scale of knowledge may be significantly different according to the word form. This idea does not hypothesize a different relationship between vocabulary and the skill of listening in a foreign language in the sense that it continues to tie the skill of listening tightly to vocabulary knowledge and quantity of words known.

This explanation emerges from an attempt to model the mental lexicon and make its structure clear, and the choice of test used to characterize word knowledge in the lexicon. Milton et al. (2010) note that while we all recognize that there is more to word knowledge than just the ability to identify the written form of a word, we tend to use only written tests to measure vocabulary size and broader vocabulary knowledge. While the use of multiple tests, to test a range of aspects of vocabulary knowledge, is considered part of good practice in the field (Nation, 2007), for practical reasons this is not always done. Assessments that are able to characterize learners' knowledge of the phonological form of words, rather than the written form, are hard to find in the literature. However, a number of studies, starting with Milton et al.'s (2010) paper, have now suggested that strong correlations between vocabulary size and listening, equivalent to those found with the written skills, can be found where these phonological tests are used.

To understand the relationship between the skill of listening and vocabulary knowledge it is probably useful to revisit how we understand word

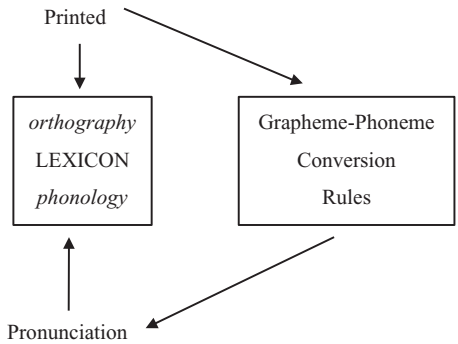


Figure 4.1 Coultart and Rastle's dual route model (1994).

knowledge in the mental lexicon. Successive attempts to characterize the complexity of word knowledge have resulted in ever longer lists of what it means to know a word. Milton & Fitzpatrick (2014) develop Nation's (2001) table of what it means to know a word, to explain this in some depth, but at the heart of understanding the lexicon is to understand the distinction between knowledge of the written form of words and knowledge of the phonological form of words. Coultart & Rastle's dual route model (1994), Figure 4.1, pictures the lexicon as having two halves and different skills such as reading and listening will access different halves of the lexicon. In reading the most efficient route to reach word meaning is to access words in written form in the orthographic half of the lexicon. Words stored in this form will not be useful for the skill of listening where the search for meaning will involve access to the sound of words in the phonological half of the lexicon. They also illustrate that there is an alternative route to access the meaning of written words where, if the words are sounded out via grapheme-phoneme conversion rules so you can hear them, you can also read using knowledge of words in phonological form.

Coulthart & Rastle's characterization of the reading process draws attention to the idea that the two halves of the lexicon may not match each other precisely, and that there will be instances where a written word form is not known and that an alternative route to meaning is possible. However, there seems to be an assumption, where we talk about the number of words needed for communicative fluency in English, that the two halves will map onto each other pretty well. Words that you can read and understand, you can also hear and understand; each written form will have a corresponding phonological form. We tend to assume, therefore, that tests of vocabulary delivered in written format will deliver scores which meaningfully, and usefully, represent knowledge of the whole lexicon. Research, however, suggests that in L2 learners this may not be the case. Milton et al. (2010) and Milton & Riordan (2006) demonstrate that learners from most L1 backgrounds tend

to grow an unbalanced lexicon, where the orthographic half of the lexicon is bigger than the phonological half. This imbalance is not static. The better these learners become in overall performance, the bigger the written half is in relation to the phonological half. High performing and fluent users of English as an L2, therefore, recognize a lot of words by sight but appear to have only hazy ideas of how many of these words are pronounced. It is not entirely clear why they do this. It might be a functional outcome of the way you need more words to read for meaning than you do to listen for meaning well. Equally, it might be a product of the way learning occurs. Learners, it is thought, gain a lot of their vocabulary in an L2 from reading and so never encounter the aural forms of these words. There is a very strong tradition that lexicons are built this way as in Krashen (1989), and Bright & McGregor (1970). An outcome of a lexicon constructed this way is that there are likely to be higher correlations between written vocabulary size measures and the written skills than with the oral skills. It is likely to mean, too, higher correlations between the phonological vocabulary size measures and the oral skills than with the written skills. This is what is observed in Milton et al.'s (2010: 91) study, where written and oral vocabulary size measures (X_Lex and A_Lex respectively) are correlated with IELTS sub-scores. This is shown in Table 4.1.

It is interesting from the perspective of the listening skills that medium to high correlations are noted with both written and aural vocabulary measures and this is explained by the way the IELTS listening sub-test requires the reader both to read the questions and listen for the answers, activating knowledge in both halves of the lexicon.

Viewed in this light, vocabulary knowledge interacts with the skill of listening in the same way that it interacts with the written skills. Better vocabulary knowledge leads to higher levels of listening comprehension, and the relationship appears linear, but it is knowledge of the spoken form of words that is relevant in understanding this relationship. Where knowledge of this form of words is compromised for some reason, meaning the phonological half of the lexicon is small, then the skill of listening will also be compromised. Where this is observed, and the oral skills of learners are important, then teachers will want to undertake remedial action.

The Coultart & Rastle (1994) model of the lexicon draws attention to the way the skill of listening can interact with the process of reading. The reason for this may not be immediately obvious, but there are instances where the

Table 4.1 Correlations between vocabulary size scores and IELTS scores

	A_Lex	*Reading*	*Listening*	*Writing*	*Speaking*	*Overall*
X_Lex	.46*	.70**	.48**	.76**	.35**	.68**
A_Lex		.22*	.67**	.44*	.71***	.55****

* = sig to 0.05 level, ** = sig to 0.01 level

orthographic half of the lexicon is undeveloped leading to this alternative route for decoding written text. While we have said that most learners demonstrate an unbalanced lexicon where the orthographic half is larger than the phonological half, it is thought that for some learners the reverse is true. Among L1 Arabic speakers, for example, it has been noted that they appear to possess a balanced lexicon where scores for written vocabulary size match scores for phonological vocabulary size, but it is thought this is a product of an unbalanced lexicon in favour of the phonological half. Alhazmi & Milton (2019) in an eye-tracking study argue that these learners favour reading by phonological decoding. Words on the page are read by sounding them out and it is only by listening to the sounds of the words that access to meaning can be made. This leads Masrai & Milton (2018) to suggest this is a product of a lexicon which lacks good representations of words in orthographic form and where knowledge is primarily held in phonological form. Scores on the two different vocabulary tests match each other because by vocalizing the written tests, so the tests' words can be heard, these learners turn both tests into listening tests. Evidence to support this can be seen in Alhazmi & Milton (2016) who repeat Milton et al.'s (2010) study with L1 Arabic speakers and produce the correlations seen in Table 4.2.

It can be observed in this data that the L1 Arabic speakers, who are the subjects in this study, possess a statistically significant correlation between their vocabulary knowledge and reading only where vocabulary knowledge is measured using a phonological test. The strong correlation between written vocabulary knowledge and reading comprehension, observed in other learners, is absent.

The conclusion to be drawn from this is that a test of the written form of words may not in all cases make a good representation of the whole of the lexicon and may not access knowledge in the form in which words will need to be known for aural skills such as speaking and listening. Where vocabulary knowledge is explicitly considered, separately, from both a phonological and orthographic standpoint, listening skills, at least as characterized by the IELTS listening sub-test, emerge as an interaction of knowledge in both halves of the lexicon. Regression shows a combination of aural and written vocabulary size tests predicts listening skill scores best. This reflects the way the test requires both reading and listening skills for successful completion.

Table 4.2 Correlations between IELTS sub-components and scores from X_Lex and A_lex (Alhazmi and Milton, 2016: 33)

	A_Lex	*Reading*	*Listening*	*Writing*	*Speaking*	*Overall*
X_Lex	0.669**	0.354	0.711**	0.289	0.419*	0.590**
A_Lex		0.476*	0.772**	0.199	0.581**	0.698**

* = sig to 0.05 level, ** = sig to 0.01 level

Removed from the confusion of reading elements in listening tests, it is possible to envisage a more straightforward relationship between vocabularies. Once vocabulary knowledge is considered in its aural form, then the expected relationship emerges; vocabulary knowledge correlates very strongly with listening skill performance, at a level equivalent to that observed between vocabulary and reading and writing, where vocabulary is measured by written tests.

This interpretation of the structure of the mental lexicon, and the nature of word knowledge required for the different communicative skills, makes sense too of the difficulties in reading observed with L1 Arabic learners of L2 English, where reading skills are best predicted by measures of aural vocabulary knowledge. For these readers, reading requires hearing the language since visual representations of words are undeveloped or insufficiently developed to permit processing through the written half of the lexicon.

The learning of vocabulary through listening

This discussion of knowledge of words in their phonological form raises issues about the learning of such vocabulary. For good aural representations of words to exist in the lexicon, it might be thought, good models of these forms are needed, presumably gained from listening practice. The assumption that vocabulary is learned from reading is well established but it is less well established that vocabulary can be learned with equal efficiency from extensive listening.

The studies that exist generally conclude that vocabulary can be learned from purely aural input, but that uptake is generally less than that obtained from written input. Vidal (2011), for example, measured the uptake from the two different forms of input and concluded not just that uptake from reading was better than uptake from listening, but also more repetitions were needed for uptake from listening. This is tentatively attributed to the need for fast processing, noted above, in order for words to be recognized and some kind of word form committed to memory. Van Zeeland & Schmitt (2013) likewise noted this need for greater repetition, when compared to learning from reading, if words are to be learned aurally. Vidal's conclusion also replicates that of Brown et al (2008) whose study produced similar results. Not all studies are quite so unequivocal, and Chandry, Deconinck, and Eyckmans (2018) study suggested that while immediate recall was marginally better in written recall than oral recall, it was an effect that disappeared within a short time. Little & Kobayashi (2011), in their study, found both reading-based and listening-based interventions were equally effective in promoting vocabulary learning.

Pigada & Schmitt (2006) and Pellicer-Sánchez & Schmitt (2010) note a sequence in the development of vocabulary knowledge where knowledge of form and a word's part of speech tend to be acquired before a link between form and meaning are established. However, it is not clear whether the way these different dimensions of word knowledge are learned is also conditioned

by whether the word is known in written or aural form. Other studies investigate the effects of different forms of intervention designed to promote vocabulary development and enhance listening comprehensions. Zhang & Graham (2019), for example, conclude that a listening-based focus on form interventions may not be effective in improving vocabulary, while Bulut & Karasakaloğlu (2017) conclude that that active listening training positively contributed to vocabulary development.

These studies are based on assessing the impact of formal interventions in trying to decide whether listening-based vocabulary exposure alone is an effective way to develop a usefully sized lexicon in a foreign language. There is an idea, suggested above, that much of a learner's lexicon comes from less formal sources, through reading, and there appears to be less published research that attempts to quantify vocabulary uptake from equivalent, informal aural input. Milton's (2008) paper, for example, does include an examination of vocabulary uptake from listening to songs, for example, but the listening activity is supplemented by written input also. The combination of reading supplemented by aural input is thought to be a strength and may help explain the success of the activity reported. The remaining part of this chapter, therefore, reports an attempt to add to this small literature base, and to estimate the vocabulary uptake that can occur from informal language input that is entirely aural.

Investigating acquisition of the aural form of words

General aim and research questions

The broad aim of the study presented in this chapter is to investigate the effect of an extensive listening strategy on lexical uptake. It will use the case study technique demonstrated in Horst & Meara (1999), and used in Masrai & Milton's (2018), which investigated whether an extensive reading strategy could be shown to impact on lexical acquisition, and the speed and process of L2 reading, in some measurable way.

In order to address the broad aim of the current study, a number of more specific research questions are asked.

Before the listening intervention

- Do the learners display the balanced, and unusual, lexicons noted in Milton & Hopkins (2006), and Milton & Riordan (2006)?

After the listening intervention

- Do the participants grow the phonographic side of their lexicons at the rates noted in other studies?
- Has the participant's lexicon developed as an unbalanced lexicon favouring the phonological side?

- Can the listening comprehension of participants be shown to have improved?

The participants

The participants were 27 native Arabic speaking students at an academy in Saudi Arabia, studying English for their courses. These students were invited to watch movies and news supplemented by subtitles for 20 weeks. The news was listened to in English with sub-titles in Arabic to allow comprehension of the English where this was needed. As in Masrai & Milton's (2018) study, the students received about 2.5 classroom hours per week of English instruction for their courses and they should have received on average about 1000 hours of English classroom input before the intervention for this study. It is thoughts they are at, roughly, intermediate level of knowledge and performance (Masrai & Milton, 2012). The materials were chosen by the learners based on their preference, and they were given diaries to record details of their listening, including type and length of the materials, repetitions of the same material, and dictionary use.

Instruments and procedure

The learner took three pre-tests prior to the intervention. These were:

- Aural Lex (A_Lex) (Milton & Hopkins, 2005),
- X_Lex (Meara & Milton, 2003), and
- an IELTS style listening comprehension paper (Milton et al., 2001).

A_Lex is a computerized Yes/No test that measures phonological vocabulary knowledge of L2 learners in the most 5,000 frequent words of English. The test includes 100 real English items and 20 pseudowords that sound similar to the real items. The inclusion of the pseudowords is used to adjust for guesswork if practised by a test-taker.

The X_Lex is also a Yes/No test which provides estimates of L2 learners' orthographic vocabulary size of the most 5,000 frequent words in English. Similar to the A_Lex, the X_Lex comprises 100 real English words and 20 pseudowords. Like A_Lex, X_Lex is administered electronically and provides a final adjusted score, out of 5,000, automatically. To reach an estimate of the total score on the X_Lex, first the raw score of the pseudowords is multiplied by 250 and the raw score of real words is multiplied by 50. Second, the total raw score of pseudowords is deducted from the total raw score of real words to provide a final adjusted vocabulary score of the test-taker.

The listening comprehension test is an IELTS style test with 40 written questions and answers to be found by listening to a series of recorded passages.

The informal vocabulary learning materials which the informant was instructed to use are movies and news that are supplemented by Arabic subtitles.

Table 4.3 Tests results before the intervention

	N	*Mean*	*SD*
Pre-test A-Lex	24	1943.75	862.17
Pre-test X-Lex	24	2920.83	856.80
Pre-test IELTS	24	10.13	6.18

Results and discussion

Following the intervention, three sets of results proved incomplete or unreliable and these were removed from the data set. Results are drawn from the remaining 24 subjects, therefore. Results from the tests administered before the intervention are shown in Table 4.3.

These results suggest that learners are approximately at B1 level (Meara & Milton, 2003) and are therefore lower intermediate.

Masrai & Milton (2018) suggest lexical uptake from classroom input is about 2.5 words from contact hour prior to the intervention with, approximately, a balanced lexicon. The results from this study suggest uptake is approximately the same although it appears words are learned more readily in their written form, at a rate of closer to three words per contact hour, and less readily in aural form, a rate of just under two words per contact hour. This appears to confirm the observation of previous studies, comparing uptake in written and aural form, that uptake tends to be greater in written form. The result suggests also that the learners in this study do not display the balanced lexicon found in other studies (e.g., Milton & Hopkins, 2006; Milton & Riordan, 2006; Alhazmi & Milton, 2016; and Masrai & Milton, 2018) but have the unbalanced lexicons more similar to the non-L1 Arabic speaking learners described in those studies. The difference in the mean scores on X_Lex and A_Lex is statistically significant ($t = 6.45$, sig < .001).

The listening intervention resulted in, on average, about 50 hours of listening to movies and news in English, very similar in scale to the reading intervention described in Masrai & Milton (2018). There is, however, considerable variation in the amount of time spent in this activity as the standard deviation and maximum and minimum figures in Table 4.4 show.

After the intervention, the scores on the tests used increased and the results are shown in Table 4.5.

These results suggest that lexical uptake has occurred and that there has been an improvement in listening comprehension test scores. These results are shown in Table 4.6.

These results suggest that both sides of the lexicon have increased. The intervention is not the sole source of input and it seems likely that some vocabulary will have been gained from classroom input. Of the gains noted in Table 4.6, and following Masrai & Milton's (2018) procedures, it is thought

Table 4.4 Time spent in the intervention activity

	N	*Max*	*Min*	*Mean*	*SD*
Listening hours	24	102	9	51.77	36.39

Table 4.5 Test results after the intervention

	N	*Mean*	*SD*
Post-test A-Lex	24	2297.92	989.23
Post-test X-Lex	24	3325.00	819.60
Post-test IELTS	24	14.67	7.34

Table 4.6 Growth in lexical test scores

	N	*Mean*	*SD*
A-Lex growth	24	354.17	223.08
X-Lex growth	24	447.92	523.47

about 100 words will have been learned aurally and about 150 words will have been learned in written form, from normal classroom input. Uptake from the intervention, therefore, appears to be of the order of approximately 250 additional words learned aurally, and 300 additional words learned orthographically. The difference between the mean scores for vocabulary gains is not statistically significant (t = .815, sig = .424). These figures suggest a rate of uptake of about five or six words per hour spent in the intervention, a number far closer to the gains associated with classroom instruction (Milton & Meara, 1998) than those associated with the kind of informal activity used here (as in, e.g., Milton, 2008, and Masrai & Milton, 2018).

It appears, therefore, that an informal listening intervention is capable of promoting lexical knowledge in phonological form, albeit relatively slowly. This, then, fits with the kind of uptake noted in more formal vocabulary learning interventions (as in, e.g., Little & Kobayashi, 2011), where uptake of words in written form appeared to outstrip uptake in phonological form. If it appears strange that in this study there appears to have been uptake of orthographic vocabulary where no obvious extra input was proved by the intervention, this might be explained by the characteristics of the subjects. Learners from an L1 Arabic background, it is suggested, are particularly prone to phonological decoding and this means that, unlike most other learners, vocabulary knowledge in phonological form will match vocabulary knowledge in written form.

Table 4.7 Correlations between intervention time and growth in phonological vocabulary and listening comprehensions scores

	A-Lex growth	*Listening scores*
Intervention time	.569**	.536**

** = sig to 0.01 level

Usually, where there is a growth in vocabulary knowledge, there tends to be a benefit in the performance of language skills. In this study, the participants improved their scores on a listening comprehension task by about 40% on average. The difference between the pre- and post- listening comprehension test mean scores is statistically significant ($t = 7.10$, sig < .001), so it would seem that with the growth of vocabulary, albeit less than was expected, comes improved listening comprehension.

Support for the idea that it is the nature of the intervention that is causing this improvement comes from the correlations between the time spend in the intervention and the growth in both phonological vocabulary size and listening comprehension test scores and these results are summarized in Table 4.7.

In both cases the correlations are moderate and are statistically significant. It appears, therefore, that for teachers seeking to improve their learners' listening skills an informal intervention, targeted at developing the learners' phonological vocabulary, will be beneficial. This confirms the relationship, which is acknowledged in other skill areas, that there is a close link between the size and nature of the lexicon, and performance in communicative language skills. However, these results probably confirm too that it is rather harder to promote the growth of vocabulary for listening than it is for the skill of reading. There are several possible explanations for this. One is the difference in the medium which makes the recognition of unknown words a much harder task than it is for the written form of the word. This, in turn, may make the quality and nature of the intervention crucial for success. The level of the learners in this study, which is low intermediate means that the chosen form of listening input may have been very difficult for many of the subjects. Interventions may have to be controlled for lexical level, speed of delivery, and overall difficulty to allow such learners sufficient overall comprehension for optimal learning to take place. Slow uptake from listening sources such as the news may also be connected to the quality of real-time translation which occurs in the sub-titling of news programmes. Our observation is that it appears this was sometimes poor (something noted also among studies in Montero-Perez et al. 2013), and this may have affected the quality of the learning that took place. Finally, there is the issue of the L1 background of the learners in this study, different to that in other studies. Alhazmi & Milton (2019) argue that the nature of lexical knowledge among these learners is likely to be both qualitatively and quantitatively different from the learners in other studies leading to different language process, at least in reading. We

have little understanding as to whether this will have any impact on the skill of listening.

Conclusions

The observation was made at the outset of this chapter that this area is less well researched and understood than is the relationship between vocabulary and other skills, particularly reading. The medium itself makes a big difference even if the fundamental relationship between vocabulary and comprehension, where more is better, stays the same. The transitory nature of the spoken medium is likely to make both vocabulary comprehension and vocabulary learning harder. It also makes it less susceptible to easy measurement and observation than the analogous skill of reading. Reading can use techniques such as eye-tracking to better understand the process of word access and comprehension but nothing equivalent for listening exists. The processes used for the skill of listening are likely to be significantly different in relation to how words are recognized and meaning retrieved for comprehension to occur. It is also argued that the nature of the vocabulary knowledge itself, words held in phonological rather than orthographic form, may significantly affect the relationship between vocabulary knowledge and the skill of listening. This is something we are still grappling to understand fully. An outcome of these differences, however, appears to be that the process of teaching vocabulary and using interventions to grow vocabulary to enhance the skill of listening is significantly more difficult than for reading. Studies repeatedly report this, and the study in this chapter is no exception. Understanding how to best promote vocabulary for the skill of listening is a subject that still appears to be in its infancy.

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, *24*(4), 425–438.

Alhazmi, K., & Milton, J. (2016). Phonological vocabulary size, orthographic vocabulary size, and EFL reading ability among native Arabic speakers. *Journal of Applied Linguistics*, *30*, 26–43.

Alhazmi, K., & Milton, J. (2019). Examining 'vowel blindness' among native Arabic speakers reading English from the perspective of eye-tracking. *System*, *80*, 235–245.

Bright, J. A., & McGregor, G. P. (1970). *Teaching English as a second language.* London: Longman.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136–163.

Bulut, B., & Karasakaloğlu, K. (2017). Benefiting from listening in vocabulary development. *Journal of Education and Training Studies*, *5*(12), 99–109.

Chandry, S., Deconinck, J., & Eyckmans, J. (2018). Written repetition vs. oral repetition: Which is more conducive to L2 vocabulary learning? *Journal of the European Second Language Association*, *2*(1), 72–82.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1197–1211.

Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Review/La Revue canadiennne des langues vivantes*, *56(*2) (December), 308–328

Krashen, S. (1989). We acquire vocabulary and spelling via reading: Additional evidence for the input hypothesis. *Modern Language Journal*, *73*(4), 440–464.

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign language*, *22*(1), 15–30.

Little, A., & Kobayashi, K. (2011). Comparing listening and vocabulary enhancement activities for vocabulary acquisition and retention. *Malaysian Journal of ELT Research*, *7*(1), 1–63.

Masrai, A., & Milton, J. (2012). EFL vocabulary knowledge of university students in Saudi Arabia. *Perspectives*, *19*(3), 13–19.

Masrai, A., & Milton, J. (2018). The role of informal learning activities in improving L2 lexical access and acquisition in L1 Arabic speakers learning EFL. *Language Learning Journal*, *46*(5), 594–604.

Meara, P., & Jones, G. (1990). *Eurocentre's vocabulary size test: User's guide.* Zurich: Eurocentres.

Meara, P., & Milton, J. (2003). *X_Lex: The Swansea Levels Test.* Newbury: Express.

Milton, J. (2008). Vocabulary uptake from informal learning tasks. *Language Learning Journal*, *36*(2), 227–238.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Milton, J., Bell, H., & Neville, P. (2001). *IELTS practice tests.* Newbury: Express.

Milton, J., & Fitzpatrick, T. (eds) (2014). *Dimensions of vocabulary knowledge*. Basingstoke: Palgrave.

Milton, J., & Hopkins, N. (2005). *Aural lex*. Swansea: Swansea University.

Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners. *Canadian Modern Language Review*, *63(*1) (September), 127–147.

Milton, J., & Meara, P. (1998). Are the British really bad at learning foreign languages? *Language Learning Journal*, *18*, 68–76.

Milton, J,. & Riordan, O. (2006). Level and script effects in the phonological and orthographic vocabulary size of Arabic and Farsi speakers. In Davidson, P., Coombe, C., Lloyd, D., & Palfreyman, D. (eds) *Teaching and learning vocabulary in another language* (pp. 122–133). UAE: TESOL Arabia.

Milton J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In Chacón-Beltrán, R., Abello-Contesse, C., & Torreblanca-López, M. (eds) *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.

Montero-Perez, M., van den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A met-analysis. *System*, *21*(3), 720–739.

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

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63(*1) (September), 59–82.

Nation, I. S. P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In Daller, H., Milton, J., & Treffers-Daller, J. (eds) *Modelling and assessing vocabulary knowledge* (pp. 3–43). Cambridge: Cambridge University Press.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*, 31–55.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*, 1–28.

Schonell, F. J., Middleton, I. G., & Shaw, B. A. (1956). *A study of the oral vocabulary of adults*. Brisbane: University of Queensland Press.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*(2), 139–152.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*(3), 191–210.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*, 609–624.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*, 219–258.

Zhang, P., & Graham, S. (2019). Vocabulary learning through listening: Comparing L2 explanations, teacher codeswitching, contrastive focus-on-form and incidental learning. *Language Teaching Research*. Ahead of print.

5 Vocabulary and listening

Future research, tools, and practices

James Milton and Ahmed Masrai

Introduction

Papers that report research into the lexicon and the skill of listening repeatedly reflect that this is an area that is less well researched than the other skills. Chapter 4 has already noted that the listening process is inherently more difficult to monitor and assess than, say, reading and writing. We have less understanding of the way words are stored in the lexicon for listening although we do know that the listener must tolerate greater variety of word forms than would be the case for writing. The spoken word is ephemeral, so it is more difficult to monitor the processes by which the spoken word is accessed for comprehension. We lack the tools for its assessment. Techniques such as eye-tracking can reveal much about the way the reading process handles both individual words and more extensive texts to reach comprehension, but we have few such tools that can illuminate the listening process. Perhaps it is not surprising, therefore, that we understand the teaching and learning process for words in sound form less well than for those in written form. We know it seems more difficult, and to take longer, than for the teaching of words in written form, but we seem ill-equipped to find methods that are supported by good learning theory to teach the words for listening more effectively and efficiently. For researchers, this is an opportunity to make headway in an area where few others have ventured systematically. This brief chapter will suggest where such research more usefully be directed, and this can be summarised under these four headings:

1. Spoken word knowledge and storage,
2. Spoken word processing,
3. Tests and research methods for understanding the spoken word in the lexicon; and
4. Learning words for listening.

Spoken word knowledge and storage

There is currently a consensus that the lexicon is multifaceted and knowledge of the spoken form of words will be one of those facets (as in Nation's table,

'What is involved in knowing a word?' 2001: 27). However, research which characterizes and quantifies the lexicon, in terms of knowledge of spoken form, is comparatively rare.

Tests which are able to convincingly count words known and available for listening are few and are not widely used in comparison to written lexical size measures. Given what we know about the importance of developing sufficient volumes of words to enable smooth communication, being able to count such words, and assess where learners stand in relation to learners' communicative goals, ought to be of importance to both researchers and teachers. Nation's (2006) seminal work ascribing a figure of 6,000 to 7,000 words as a requirement of reaching 98% coverage of a film script, is unsupported by research telling us whether these kinds of figures for knowledge can be replicated in an actual listening task, as in an IELTS listening sub-test, for example.

Tests which quantify the number of words recognized in aural form might usefully be extended into the relationship between knowledge of the phonological forms of words and aspects of depth of knowledge. It is assumed that, in both lexical dimensions, knowledge in written form translates into equivalent knowledge in aural form. However, this may not be the case. In a series of studies, Milton has suggested that the aural or spoken side of the L2 lexicon may be very different from the orthographic side where written words are stored, and that the nature of this difference may be governed by the learner's L1 (e.g., Alhazmi & Milton, 2016; Milton & Riordan, 2006; Milton et al. 2010). This may have considerable implications for the way we attempt to teach word knowledge which might vary according to the L1, and orthographic background, of the learner. In short, we lack a convincing model of the phonological side of the lexicon from which good theories of learning, processing, and teaching can emerge, and work which elucidates this must be useful.

Spoken word processing

It is interesting to note that consideration of lexical processing still thinks first and foremost of strings or sounds and letters which are serially accessed (e.g., Schütze, 2017). Research into the reading process tells us that things are much more sophisticated than this, and much more rapid than this allows. Not all letters in words are fixated on, and not all words are fixated on, in the process of reading, although much of this information may be taken in peripherally. The saccadic process where the eye jumps over words is well accepted and implies that meaning is anticipated in the reading process. The gaps left in this process are filled in by the reader from knowledge of language and of what makes sense of the text – we assume a text will make sense.

The degree to which this model of processing can be translated to listening is unclear. The speed with which the spoken word must be processed when listening implies that every spoken sound cannot be processed to create meaning. So, what is being sampled? The text we listen to does not, usually, introduce spaces between words to help word identification. The spoken word is further prone to variation not just through accent but by the processes of

sentence stress, by assimilation, and by elision. Research has yet to convincingly tell what sounds the listener is picking up on in order to create word and text meaning. We need research to tell us how the brain manages this kind of variety in listening. This has implications for the teaching of words for communication in listening. Is it sufficient to teach a single citation form of a word to learners, relying on the learner's ability to derive varied forms, or would learners' listening performance be enhanced by deliberate and explicit teaching of word variation in listening form? At the heart of the absence of convincing and accepted research in this area is the absence of a good methodology for approaching it and this is the subject of the next section.

Tests and research methods for understanding the spoken word in the lexicon

The previous two sections have highlighted the gap in our understanding of the lexicon and have also drawn attention to a reason why this gap exists, which is the absence of good tools for working with the lexicon for listening. Even in terms of quantifying this half of the lexicon, probably only Milton & Hopkins (2005) have attempted to create a test of phonological word recognition with the intention that any scores which emerge can be meaningfully compared with an equivalent written word recognition test. They can do this because the tests they use are very minimal and consider only the passive recognition of individual words in isolation, and the methodology is therefore very approachable. It would be useful to extend this idea to other aspects of word knowledge to better characterize the lexicon for listening and contrast it with the orthographic half.

Finding tools to illuminate the process of words for listening is probably even more challenging. While methods like eye-tracking have been able to give us insight into the reading process, it is not clear how the processes involved in listening can be approached in an equivalent manner. Presumably, we need methods that involve gating spoken text, controlling the sounds and words available for access in listening, to allow us to model the listening process. Just as we lack a convincing model of the phonological side of the lexicon, we also lack a convincing model of lexical processing in listening and this is an area future research could usefully address.

The kind of research that involves testing tool creation is one where researchers need patience, persistence, and imagination. It often involves the creation and application of multiple test forms before useful and insightful approaches emerge, and then much honing and refinement before a test can be demonstrated to be reliable and valid in what it attempts to assess. Systematic research, perhaps with multiple research students, is likely to yield results here.

Learning words for listening

Extensive reading has been found to be very useful for learning words in written form, and a similar approach with listening might reveal a notable

vocabulary gain with words stored in appropriate form to promote the skill of listening best. Compared to the wide extensive reading literature, only limited research has been conducted with extensive listening. Viewing DVDs, listening to news, listening while reading are just some of the many forms of aural input which might be used as ways for developing the phonological side of the lexicon. Milton et al. (2010) rightly noted that knowing words in written modality does not necessarily mean these lexical items are known aurally, and vice versa. So, developing a systematic model of sound-form mapping might serve the theory of word learning for listening.

There are some empirical studies of extensive viewing which provide indications that L2 learners, who were engaged in this informal activity, improve their ability to perceive and parse L2 text (Renandya, 2012; Wang & Renandya, 2012), increase their listening fluency (Chang & Millett, 2014), and enhance their overall listening comprehension skills (Onoda, 2014). All of these beneficial outcomes might be connected to the observations made by Masrai (2019) who found that extensive viewing significantly improved L2 learners' aural vocabulary gain and listening comprehension. Prior phonological vocabulary knowledge was also found to contribute to the learning of new words in aural modality. Further work to evaluate the impact of all of these avenues of listening intervention would be welcome.

Conclusions

This, brief, overview of current practices and suggests, draws attention to the wide range of opportunities for research into the phonological lexicon and the skill of listening. The four areas highlighted here are by no means the only avenues for research but we do think that work which elaborates the lexicon for listening, and how this relates to the skill of listening, is likely to be a highly fertile direction for future work.

References

Alhazmi, K., & Milton, J. (2016.) Phonological vocabulary size, orthographic vocabulary size, and EFL reading ability among native Arabic speakers. *Journal of Applied Linguistics*, *30*, 26–43.

Chang, A. C., & Millett, S. (2014). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal*, *68*, 31–40.

Masrai, A. (2019). Can L2 phonological vocabulary knowledge and listening comprehension be developed through extensive movie viewing? The case of Arab EFL learners. *International Journal of Listening* (ahead of print).

Milton, J., & Hopkins, N. (2005). *Aural Lex.* Swansea: Swansea University

Milton, J., & Riordan, O. (2006). Level and script effects in the phonological and orthographic vocabulary size of Arabic and Farsi speakers. In Davidson, P., Coombe, C., Lloyd, D., & Palfreyman, D. (eds) *Teaching and learning vocabulary in another language* (pp. 122–133). UAE: TESOL Arabia.

Milton J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In Chacón-Beltrán, R., Abello-Contesse, C. &

Torreblanca-López, M. (eds) *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.

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

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63(*1) (September 2006), 59–82.

Onoda, S. (2014). Investigating effects of extensive listening on listening skill development in EFL classes. *Journal of Extensive Reading in Foreign Languages*, *1*, 43–55.

Renandya, W. A. (2012). The tape is too fast. *Modern English Teacher*, *21*, 5–9.

Schütze, U. (2017). *Language learning and the brain*. Cambridge: Cambridge University Press.

Wang, L., & Renandya, W. A. (2012). Effective approaches to teaching listening: Chinese EFL teachers' perspectives. *Journal of Asia TEFL*, *9*, 79–111.

Part III
Reading

6 Vocabulary and reading

Current research, tools, and practices

Jeanine Treffers-Daller

Developing vocabulary knowledge is crucially important for monolinguals as well as second language learners because vocabulary is the key determinant of reading skills, and these are, in turn, of vital importance for academic success, both in children (Bleses et al., 2016) and adults (Trenkic & Warmington, 2017). As discussed in detail in Elgort (in Chapter 7), there is a bidirectional relationship between vocabulary and reading, which means that learners who know many words find it easier to read, and these learners are therefore in a better position to learn new words from reading too. Conversely, children with smaller vocabularies, such as immigrant children who are learning through the medium of their L2, have often been found to underperform in reading and other subjects by comparison with their monolingual peers (Højen et al., 2019; Hutchinson, 2018), even though their performance varied depending on variables, such as children's regional backgrounds and socio-economic status.

Studying the relationship between vocabulary and reading is particularly relevant for researchers and practitioners interested in understanding why many L2 learners struggle to read authentic texts independently, and who want to support students in developing their reading skills. Many educated L2 learners know only between 1,000- and 4,000-word families (Laufer, 2000), which is not nearly enough to independently read a novel or a newspaper article, as Nation (2006) has shown that between 8,000 and 9,000 word families are needed for this reading task.

Research into the relationship between reading and vocabulary knowledge has focused on both directions of the bidirectional relationship between these two constructs. On the one hand, studies have concentrated on reading as the dependent variable and investigated the contribution of vocabulary knowledge to different components of reading. On the other hand, there is a growing body of research which concentrates on vocabulary as the dependent variable and looks at the ability of learners to learn new words from reading, a process which is generally referred to as *incidental learning* (Swanborn & DeGlopper, 1999). As incidental vocabulary learning from reading is covered in detail in Elgort's contributions to this volume, in this introduction I will focus on the contribution of vocabulary to reading comprehension, and in particular on different aspects of the construct of vocabulary and its measurement. Further

analyses of the construct of reading comprehension are offered in Chapter 8 by Treffers-Daller and Huang.

Many researchers in the field have investigated which percentage of words a reader needs to know in order to understand a text. Establishing an exact cut-off point is difficult, because, as Schmitt, Jiang, and Grabe (2011) show, the level of coverage depends on the level of comprehension that is required. If a score of 70% on a post test is required, a coverage of 98–99% is needed, but if more of the text needs to be understood, the reader may need to be familiar with all the words. However, there is considerable evidence that it is not only important how *many* words a person knows (vocabulary size or breadth) but also how *well* words are known (vocabulary depth or lexical quality). Binder et al. (2017) found that both depth and size were related to reading comprehension, but only vocabulary size explained unique variance in comprehension in their study. Other authors (e.g., Qian, 2005; Treffers-Daller & Huang, Chapter 8) found the opposite, namely that vocabulary depth is the key variable which explains reading comprehension. There is therefore no consensus about this issue, which is probably due to the fact that the relationship between size and depth depends on how both constructs are conceptualized and measured (Schmitt, 2014). While it is generally assumed that size and depth are related, there is also considerable evidence that these are separate constructs (Tseng & Schmitt, 2008). As Schmitt (2014) points out, there are many different dimensions to depth (knowledge of collocations, polysemous forms, derivations, etc.) and it is hardly possible for one test to cover the many interpretations of this construct. Another issue is that many tests of vocabulary depth have hardly been validated and therefore Schmitt (2014) calls for the development of better measures of lexical quality.

The previous section has already revealed that knowing a word involves far more than knowing the link between the form and the meaning of an item. The most widely used model of the construct of vocabulary knowledge is the one proposed by Nation (2001, 2013) who distinguishes between knowledge about the form, meaning, and use of words. There are many different aspects to these three dimensions, each of which can be known productively and/or receptively. In addition, there is the dimension of fluency, that is the degree of automaticity with which a person can access different aspects of word knowledge (Laufer & Nation, 2001; Segalowitz, 2010). It is hardly possible to measure all the different dimensions of vocabulary knowledge with one test, which means that different tests are needed to tap into each.

A widely used typology of vocabulary tests is the one introduced by Laufer and Goldstein (2004), who distinguish between four different degrees of meaning knowledge. First of all, a distinction is made between *active recall*, which means that learners need to supply the L2 target word form and *passive recall*, which refers to learners supplying the L1 translation equivalent of the L2 word. *Active recognition*, by contrast, involves giving the learners a prompt in the L1 and asking them to provide the L2 target by choosing from four (or

more) options. Finally, in *passive recognition*, the L2 target word is given as a prompt, and the learner chooses the L1 translation equivalent from four options. Laufer and Goldstein (2004) demonstrate that the order of difficulty of the different degrees of knowledge (from most to least difficult) is: active recall, passive recall, active recognition, and passive recognition. The importance of the distinction between these degrees of knowledge was also confirmed by González-Fernández and Schmitt (2019), who showed that recognition knowledge is generally acquired before recall knowledge. As pointed out by Schmitt (2014) most vocabulary tests assess meaning recognition (passive recognition), and there are risks of overestimating a person's vocabulary size due to the effects of guessing (Gyllstadt, Vilkaitė, and Schmitt, 2015). However, in a recent study in which the explanatory power of recognition versus recall for reading comprehension was analysed, Laufer and Aviad-Levitzky (2017) concluded that both types of test correlated equally strongly with reading comprehension.

A key problem with testing vocabulary as well as reading among L2 learners (and bilinguals) is that the available tests have generally been developed for monolinguals. As pointed out by Schmitt, Nation, and Kremmel (2019), tests that were developed for one group, in a specific context, are not necessarily suitable for other groups in other contexts. L2 learners' experience with the target language, including the age at which they started learning it, is usually very different from that of monolinguals, which means that existing tests, as well as norms that are associated with these, are often not suitable for L2 learners and bilinguals (Mueller Gathercole, 2013a, b).

A positive development is that, in recent years, there is an increasing interest in the development of new tests and adaptation of existing tests for bilinguals and L2 learners. Mueller Gathercole, Thomas, and Hughes (2008), for example, developed a Welsh picture vocabulary test with associated norms for Welsh-English bilingual children, and Goriot et al. (2018) examined the suitability of the Peabody picture vocabulary task (Dunn & Dunn, 2007) among children learning English as an L2 in the Netherlands. It remains very difficult to develop norms for bilinguals and L2 learners because the circumstances under which they learn, the amounts of exposure to the language, and the quality of the exposure varies so much. As Goriot et al. (2018) point out, scores on tests from learners with different L1s can hardly be compared due to L1 effects: there are many cognates between Dutch and English but not between Chinese and English, which makes the learning burden (Nation, 2001) very different for both groups of learners.

Larson (2017) made a very important contribution to our understanding of the effect of the L1 on learners' vocabulary sizes, by making the Vocabulary Size Test (VST) (Nation & Beglar, 2007) available through an online platform to learners of English across the world. The results show important differences in vocabulary sizes between thousands of learners at different ages and speaking 44 different L1s. The fact that nine different bilingual versions of the VST have been made freely available to researchers and teachers through Paul

Nation's website (www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-tests) is also very important for the field, and illustrates the move towards the development of tests that are suitable for non-native speakers.

Another major contribution to the field was made by Tom Cobb's Compleat Lexical Tutor (www.lextutor.ca), where a wide range of widely used vocabulary tools and tests have been made freely available to researchers and practitioners, which illustrates the importance of technological innovation in research into vocabulary.

A limitation of the current overview is that it has focused solely on the importance of vocabulary for reading, and not on other variables impacting reading. Of course, in order to arrive at a good comprehension of a text, readers need to do more than just understand the words. As demonstrated by Schmitt et al. (2011), even learners who know all the words in a text do not always understand it completely. In addition to vocabulary knowledge, variables such as phonological and morphological awareness, grammatical knowledge, contextual or prior knowledge, and inferencing skills (that is the ability to go beyond explicitly stated ideas and to build a mental representation of what a text is about), impact on reading comprehension (Khalifa & Weir, 2009). While providing further information about all these variables is beyond the scope of the current introduction, which focused on the relationship between vocabulary and reading, it is clear that a good understanding of the process of reading is only possible if a wider range of variables is considered.[1]

Note

1 I am very grateful to Jon Clenton and Paul Booth for inviting me to write this introductory chapter and to Irina Elgort for comments on a draft. All remaining errors are mine.

References

Binder, K. S., Cote, N. G., Lee, C., Bessette, E., & Vu, H. (2017). Beyond breadth: The contributions of vocabulary depth to reading comprehension among skilled readers. *Journal of Research in Reading*, *40*(3), 333–343. doi:10.1111/1467–9817.12069.

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*(6), 1461–1476. doi: 10.1017/S0142716416000060

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 Manual*. Bloomington, IN: NCS Pearson.

González-Fernández, B., & Schmitt, N. (2019). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, *40*(2), 329–354.

Goriot, C., van Hout, R., Broersma, M., Lobo, V., McQueen, M. J., & Unsworth, S. (2018). Using the Peabody Picture vocabulary test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism*, *1*–23. doi:10.1080/13670050.2018.1494131.

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *International Journal of Applied Linguistics, 166(*2), 278–306. doi 10.1075/itl.166.2.04gyl.

Højen, A., Bleses, D., Jensen, P., & Dale, Ph. (2019). Patterns of educational achievement among groups of immigrant children in Denmark emerge already in preschool second-language and preliteracy skills. *Applied Psycholinguistics, 40*(4), 853–875. doi:10.1017/S0142716418000814.

Hutchinson, J. (2018). *Educational outcomes of children with English as an additional language*. Cambridge: Bell Foundation.

Khalifa, H., & Weir, C. J. (2009). *Examining reading. Research and practice in assessing second language reading*. Cambridge: Cambridge University Press.

Larson, M. (2017). *Thresholds, text coverage, vocabulary size, and reading comprehension in applied linguistics*. Published PhD thesis. Victoria University of Wellington.

Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of 'involvement'. *Selected Papers from AILA '99 Tokyo* (pp. 47–62). Tokyo: Waseda University Press.

Laufer, B., and Aviad-Levitzky, T. (2017). What type of vocabulary knowledge predicts reading comprehension:Word meaning recall or word meaning recognition? *Modern Language Journal*, *101*(4), 729–741.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength and computer adaptiveness. *Language Learning*, *54*(3), 399–436.

Laufer, B. & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? EUROSLA Yearbook *1*(1): 7–28. DOI: 10.1075/eurosla.1.05lau

Mueller Gathercole, V. (2013a). *Issues in the assessment of bilinguals*. Bristol: Multilingual Matters.

Mueller Gathercole, V. (2013b). *Solutions for the assessment of bilinguals*. Bristol: Multilingual Matters.

Mueller Gathercole, V., Thomas, E. M., & Hughes, E. (2008). Designing a normed receptive vocabulary test for bilingual populations: A model from Welsh. *International Journal of Bilingual Education and Bilingualism*, *11*(6), 678–720.

Nation, P. (2001, 2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *63*(1), 59–82. DOI: 10.3138/cmlr.63.1.59.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Qian, D. D. (2005). Demystifying lexical inferencing: The role of aspects of vocabulary knowledge. *TESL Canada Journal*, *22*(2), 34–54.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64(*4), 913–951.

Schmitt, N., Jiang X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, *95*(1), 26–43.

Schmitt, N., Nation, P., & Kremmel, B. (2019). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching, 53*(1)*,* 109–120. doi:10.1017/S0261444819000326.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Swanborn, M. S. L., & DeGlopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, *69*(3), 261–285.

Trenkic, D., & Warmington, M. (2017). Language and literacy skills of home and international university students: How different are they, and does it matter? *Bilingualism: Language and Cognition*, *22*(2), 349–365.

Tseng, W.-T., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A structural equation modelling approach. *Language Learning*, *58*(2), 357–400.

7 Building vocabulary knowledge from and for reading

Improving lexical quality

Irina Elgort

Introduction

Building a lexicon is an important language learning goal. Whether in the first (L1) or second (L2) language, a lexicon is the pressure point of comprehension. Although there is a wealth of language input available electronically and in print, understanding and learning from this input is possible only if the majority of the running words is known. In English, Laufer and Ravenhorst-Kalovski (2010) suggested that an L2 vocabulary of at least 5,000–6,000-word families[1] is needed to achieve about 95% lexical text coverage, which is likely to be minimally sufficient for basic reading comprehension (i.e., for a 55% score on a reading comprehension test). A 98% lexical text coverage needed to read unsimplified English books and to learn new words from reading would require the knowledge of about 8,000–9,000-word families (Nation, 2006). Nation (2014) roughly calculated that, in order to achieve this vocabulary size in English, learners would need to read about 25 authentic novels, averaging about 120,000 word tokens a novel. Nation (2014) estimated that to learn the 2nd to 9th 1000 words at the rate of 1000 words a year, reading at a moderate speed of 200 words per minute, learners would have to read initially 21 minutes a week for forty weeks for the first year and then increase the amount of reading each year to eventually reach over 6 hours a week for forty weeks in the eighth year. This amount of reading is not impossible but would require a high degree of dedication on the part of the learner, a sympathetic language learning curriculum, and availability of sufficient numbers and variety of graded readers or simplified texts, at a number of frequency levels. When reading in a second language, the situation is further complicated by what is known as the *Lexical Paradox* (Cobb, 2007) – language learners need to bring sufficient lexical knowledge *to* reading in order to gain new lexical knowledge *from* reading. In this context, an important question is whether there are ways to facilitate contextual word learning in a second language using some supplementary instructional interventions or learning activities.

Increasing vocabulary size is not the only goal of lexical development; building an L2 lexicon is also about improving the quality of lexical knowledge. Poor lexical quality is an impediment to L2 processing and use, and it slows down further lexical development. High-quality lexical representations

are particularly important in reading because print may be less effective in activating lexical representations than the spoken language (Braze et al., 2007). Achieving high-quality lexical knowledge is particularly challenging in a *foreign* language learning context, where the language classroom is often the main (or, even, the only) source of target language input, and for speakers of orthographically distant L1s, such as Chinese speakers learning English (Jiang, 2000). Language teachers and teaching and learning material developers, therefore, need to be familiar with research-informed learning and instructional methods that promote high quality of lexical knowledge. But, before proposing some ideas on how to augment contextual word learning, let us consider what we mean by *lexical quality*, especially in the context of reading.

Lexical quality

At the word level, the *Lexical Quality* framework introduced by Perfetti & Hart (2001, 2002) and further developed in Perfetti (2007) interprets the quality of word knowledge in terms of three core interrelated components: *linguistic form* (phonology and morpho-syntax), *literacy form* (orthography), and *meaning*. Robust word identity needed for fluent reading comprehension is represented by the binding of these constituents. As a consequence of having precise word-specific formal representations, good readers are able to efficiently and accurately decode words presented visually during reading. Importantly, beyond decoding, good reading comprehension is underpinned by fast and dynamic processes of form–meaning mapping and word-to-text integration that rely on the online availability of robust and flexible semantic representations needed to access contextually appropriate word meanings during reading (Perfetti & Stafura, 2014). Because lexical quality arises through learning, skilled readers are more effective word learners. At any point in time, skilled readers will have some high-quality lexical representations and also some low-quality word representations (e.g., for lower frequency words, such as *evanescence*); however, they are able to continuously gather and add new information from input (across all knowledge constituents) and thus improve their mental representations of less familiar words (Perfetti & Hart, 2002). Their high-quality orthographic and phonological representations facilitate further contextual word learning, including the development of more integrated and flexible lexical semantic representations of new and known words. Conversely, for poor readers, whose formal lexical representations are less robust, contextual word learning is more difficult (be it in their first or second language). For L2 readers, especially those whose L1 is orthographically distant from the target language, slow development of high-quality formal-lexical representations could be a bottleneck in contextual word learning from reading. Furthermore, in their first language, children and adults tend to read at an appropriate age and lexical coverage level, thus, being able to take advantage of repeated encounters with less familiar words in diverse engaging contexts. In the second language, because access to

engaging, age-appropriate, proficiency-appropriate reading materials is much less common, the development of lexical quality from reading is likely to be slower and less efficient.

Contextual word learning

Reading is an important source of increasing the size and quality of lexical knowledge, especially for lower-frequency words that are much less likely to be encountered in day-to-day interactions or even when watching film and television programmes. So, what are the mechanisms underpinning contextual word learning from reading? When an unfamiliar word is encountered in a connected text, it may be processed (and, hence, learned) either overtly or covertly. In the case of overt learning, the novel word form is noticed as 'unfamiliar' in the input. Skilled readers, whose orthographic representations are highly fine-tuned, may recognize an unfamiliar word form straight away (during the first pass). Less skilled L1 readers and L2 readers whose orthographic representation are less fine-tuned may not notice an unfamiliar word on the first pass (Laufer, 2003). For instance, they may mistake a low-frequency word, *goad*, for its high-frequency orthographic neighbour *goal*; or *succor* for *soccer* (Elgort et al., 2018; Perfetti, 2007). However, if the word does not fit in the ongoing meaning construction at the sentences or text level, the reader is likely to revisit the word, re-examine its form, and then try to deliberately derive its meaning from context, especially if this word is important for the understanding the text. Contextual word learning may also take place covertly, through implicit learning. Via its cooccurrence with known words in context, the novel word form may become associated with certain semantic features of these known words, without the reader having to make explicit meaning inferences. The same reader may engage in both overt and covert contextual learning at different times and under different circumstances. Because these different types of learning are supported by different memory structures they result in two different kinds of word knowledge (Reber, 2008, 2013; Ullman & Lovelett, 2018); explicit noticing and deliberate efforts to derive meaning are likely to result in encoding information into declarative memory, leading to explicit, declarative word knowledge; implicit learning is likely to result in encoding into nondeclarative memory leading to nondeclarative, tacit word knowledge. The nondeclarative memory is necessarily involved in fluent language processing, so gaining nondeclarative knowledge is an important goal of language learning in general and L2 lexical development, more specifically.

In alphabetic languages, a theoretical framework known as *Self-teaching hypothesis* (Share, 1995) details a mechanism by which children learn orthographic word forms and become skilled word readers. This process involves sounding out written words (i.e., translating print to sound) in the course of reading; this decoding process affords focus on letters and their sequences, which facilitates the development of orthographic knowledge. A recent L2

study with English children acquiring the target language in an immersion French context in Canada suggests that the self-teaching mechanism is also available in the L2 (Chung et al., 2019). However, it is unclear whether this exact mechanism can also support contextual word learning by adults in a *foreign* language context, particularly for distant pairs of languages. Nevertheless, a conjecture that attention to the spelling of novel L2 words is helpful for their initial encoding and, consequently, learning seems logical. Initial evidence in support of this conjecture is reported by Elgort et al. (2016) in a contextual word learning study with Chinese and Dutch learners of English, discussed below.

Importantly, lexical quality needed for fluent reading comprehension depends not only on precise and fast processing of the form but also low-effort online access to the contextually appropriate meaning. In understanding mechanisms underpinning contextual learning of meaning, I refer to two influential theoretical frameworks: (1) the *Instance-based Framework* (Bolger et al., 2008; Reichle & Perfetti, 2003) and (2) *parallel distributed processing* (PDP) models of lexical semantic memory (Seidenberg & McClelland, 1989).

The Instance-based Framework proposes that word knowledge gained during reading is initially encoded in episodic memory. The framework predicts that each encounter with a new word in context leaves an episodic memory trace of the word and its context. After multiple encounters with the same word in diverse supportive contexts, the semantic features that overlap between individual encounters are reinforced and those that do not are discarded. The reinforced semantic features are eventually abstracted from specific contextual episodes, forming a core meaning of the word that can be accessed not only in supportive but also in neutral (unsupportive) contexts, or in isolation.

Computationally, the process of word learning has been modelled using the so-called distributed models of lexical semantic knowledge. In distributed connectionist architectures, linguistic knowledge is represented as patterns of activation of interconnected nodes in neural networks (McClelland, Rumelhart, and Hinton, 1986). The PDP models are especially well-geared to account for interactions between orthographic, phonological, and meaning representations, as well as for the processes related specifically to semantic representations. These models suggest that word meanings are not represented as distinct units but are instead composed of multiple micro-features which are reused as kind of building blocks to create semantic representations of multiple words (e.g., Cree, McRae, and McNorgan, 1999; Hinton & Shallice, 1991; Masson, 1991, 1995; McRae, de Sa, and Seidenberg, 1997). The semantic (or meaning) domain units also influence, and can be influenced by, contextual factors, arising from syntactic, thematic, and pragmatic constraints. Another relevant feature of PDP models is a built-in learning mechanism that can account for contextual word learning. Distributed models maintain that every new encounter with a word is effectively a learning instance, which strengthens connection weights between processing units of the distributed

activation pattern corresponding to the word. The learning of a word is understood as the acquisition of connection strengths that allow a network of these units to quickly and effortlessly fall into the right activation pattern (McClelland et al., 1986). Thus, distributed models are able to account for the incremental frequency-driven nature of contextual word learning (Ellis, 2002; Hulstijn, 2002).

In contextual learning, as mentioned earlier, cooccurrence in the text with known words plays a critical role in the covert development of lexical semantic representations. With multiple contextual encounters with a word in diverse meaningful contexts, the semantic (and thematic) features that are consistently activated for the novel word-form, over time, form its lexical semantic representation. By sharing semantic features and clusters of features with other words, a new word is integrated into the lexical semantic networks of individual readers (Mason, 1995; McRae et al., 1997; Plaut & Booth, 2000). Words that are better integrated into lexical semantic memory will have more overlapping features (in neurological terms, patterns of activation) with other words, and will therefore be recognized and accessed faster in real language use.

This integration process is easier to achieve when existing lexical semantic networks are rich, interconnected, and robust, such as those of most adult L1 readers and high-proficiency L2 readers. At lower L2 proficiencies, readers have fewer and weaker L2 lexical semantic connections, which would explain why unassisted contextual word learning is less effective for them. In the L1 literacy literature, this 'the rich get richer' phenomenon is known as the *Matthew effect* (Stanovich, 1986; Walberg & Tsai, 1983). Frishkoff, Perfetti, and Collins-Thompso (2008) and Frishkoff et al. (2011), for example, found that vocabulary gains in learning from context were larger for higher skilled readers. The *Matthew effect* applies also to L2 contextual word learning (Ferrel Tekmen & Daloğlu, 2006; Horst, Cobb, and Meara 1998; Pulido, 2007). Similarly, in an L2 contextual word learning study, Elgort et al. (2015) found that lower-proficiency ESL students established less robust semantic representations compared to higher-proficiency students, and their access to these semantic representations was less automatic, even though the lexical coverage of the reading materials was controlled with the lower-proficiency participants in mind. Because, in this study, there was no difference in the working memory between higher- and lower-proficiency participants, and their comprehension of the reading materials was about the same, the difference in the quality of their semantic knowledge was squarely attributed to the different contextual learning potential at the two proficiency levels, likely linked to the quality of their existing L2 lexical semantic knowledge. This suggests that, for lower-proficiency L2 learners, unassisted contextual word learning (even from level appropriate texts, such as graded readers) may not be sufficient to establish high-quality lexical semantic representations and, therefore, deliberate learning methods are needed, at the very least, as a supplementary approach (Cobb, 2007, 2016).

Contextual word learning from reading in English as a second or foreign language

Notwithstanding the challenges that lower proficiency language learners may have, contextual word learning is an important (if not the key) form of increasing target language vocabulary beyond the first 5,000-word families and improving lexical quality of partially known vocabulary. In this section, I review findings from unassisted L2 contextual word learning studies that used long authentic or authentic-like texts, in order to reveal vocabulary learning patterns observed in naturalistic contextual word learning.

In the first study of its kind on vocabulary acquisition from reading, Saragi et al. (1978) used an authentic L1 novel, Anthony Burgess's *A clockwork orange* (1972; about 60,000 words), that included 90 novel *nadsad* words – a made-up language based on Russian used by Burgess in the novel. Their findings based on the meaning recognition test were very optimistic – over 75% of the novel words were learned if they occurred about 10 times in the novel. However, because participants in this study were reading in their L1 (English), contextual learning of new words would have been easier for them, compared to an L2 reading situation. Indeed, the findings of the majority of later studies with L2 readers are less optimistic.

Since Saragi's et al. (1978) pioneering study, there have been a number of L2 contextual word learning studies, but only a few of them investigated vocabulary development from unassisted reading. A key drawback of some of these studies evaluating vocabulary learning from reading, as pointed out by Nation and Webb (2011), was that they only used a single vocabulary test that was not sensitive to partial learning or the learning of different aspects of word knowledge. An important step forward in measuring outcomes of contextual word learning from longer texts was made by Waring and Takaki (2003) and Pellicer-Sanchez and Schmitt (2010), who used multiple tests to evaluate the learning of different aspects of word knowledge. Waring and Takaki (2003) measured form and meaning recognition and recall, and used immediate and delayed post-tests. In their study with 15 Japanese speakers these were exposed to 25 English pseudowords in a 400 headword English graded reader, *A Little Prince*. They found that, immediately after reading, recognition of word form was at about 60%, recognition of meaning at about 40%, and meaning recall (translation) was at 18%, with the outcomes on all aspects decreasing significantly after three months to 34%, 24%, 3.6%, respectively. Thus, fairly modest gains occurred in the knowledge of meaning, but higher gains were observed in the receptive knowledge of form. Pellicer-Sanchez and Schmitt (2010) assessed recognition and recall of spelling, meaning, and grammatical class of novel words from a Nigerian language, Ibo, after reading an L2 (English) novel (67,000 words). They concluded that at least 8–12 encounters with a word were needed for measurable learning to be observed. However, even these important contextual word learning studies only measured declarative,

explicit word knowledge but did not assess how nondeclarative knowledge of the target words developed from reading.

The development of both declarative and nondeclarative aspects of word knowledge was considered in a contextual word learning study conducted by Elgort and Warren (2014). The study measured learning outcomes for 48 English pseudowords encountered between 8 and 40 times in reading an authentic expository text (about 40,000 words), and considered the effects of and interactions between participant, item, and text variables. Elgort and Warren (2014) found that form-meaning mapping (measured by a meaning generation task) was, on average, at 20% after multiple encounters with novel words in reading but nondeclarative (tacit) knowledge (measured by form and semantic priming) was observed only for participants who started learning English earlier in their life. In addition, better tacit knowledge of lexical semantic representations was established for the readers who understood the text better and used vocabulary learning strategies. The findings related to the development of tacit word knowledge from reading seem to reaffirm the existence of the Matthew effect in L2 contextual word learning: early learners and those with better reading comprehension skills were also better able to learn new words and improve the quality of their lexical knowledge through reading-only. This suggests that, for many second and foreign language learners, lexical quality established from reading only (at least in the early stages) may not be sufficient for real online language use. Elgort and Warren (2014) also found that the acquisition of nondeclarative lexical knowledge was affected by different and fewer participant, item, and text predictors than the acquisition of declarative knowledge. An important finding of this study was that even the most powerful predictors of learning, such as frequency of occurrence in the text, do not tell the full story and do not operate in isolation; the positive effect of rate of occurrence, for example, was modulated by participants' ability to understand the text and their lexical proficiency. Better text comprehenders and readers with estimated greater vocabulary size were better able to take advantage of multiple occurrences of the novel vocabulary during reading.

Although Elgort and Warren's (2014) study provided a number of insights into different aspects of vocabulary knowledge gained from reading and variables that affect them, they measured learning outcomes *after* reading. In order to observe how contextual word learning progresses in real time, Elgort et al. (2018) conducted an eye-tracking study with Dutch speakers reading the first part (just over 12,000 words) of the book used in Elgort and Warren (2014), while their eye movements were being recorded. The data on 14 unfamiliar words and 9 matched familiar control words were used to compare how the readers' eye movements changed over time. The eye-movement data showed that the word form became familiar relatively quickly: the first-fixation durations on the novel and familiar words became very similar within the first five to seven occurrences in the text (see Pellicer-Sánchez, 2015, for a similar finding). However, the learning of meaning progressed at a much

slower rate, with the estimated go-past time (an eye-movement measure indicative of the ease of word-to-text integration) being slower for the novel words than that recorded for the known words, even after multiple contextual encounters. Also, when contextually learned words were presented in neutral contexts on the post-test, the gains observed by the end of the treatment phase were mostly gone; with the exception of first-fixation duration, all other eye-movement measures on the contextually-learned words were significantly slower than those on the known words. Finally, the newly learned words were processed slower in neutral contexts, in the post-test, than on the final encounter in the reading text (in the treatment phase). This also suggests that their lexical semantic representations were not yet abstracted from context; in other words, their lexical quality was still relatively poor. This confirms the conclusion from Elgort and Warren (2014) that L2 contextual word learning is a slow and incremental process.

Two other contextual word learning studies, in which L1 and L2 participants read a continuous text while their eye-movements were recorded have been recently conducted by Godfroid (2018) and Pellicer-Sánchez (2015). In Godfroid's study, participants read five chapters of an authentic English novel (9,000 words) containing 29 Dari words (used as learning targets) over two days. Pellicer-Sánchez used a short story (2,300 words) written specifically for her study, with six embedded nonwords and six controls (known words). Similar to Elgort et al. (2018), both studies observed a decrease in reading times on the novel words over the course of reading. Godfroid found that reading times on the novel words decreased at a similar rate for L1 and L2 readers, while Pellicer-Sánchez (2015) reported faster learning rates for L1 than L2 readers. Godfroid also found that additional contextual exposure (number of encounters with a word during reading) positively affected gains in declarative word knowledge of form and meaning. Both Pellicer-Sánchez and Godfroid also reported that longer total reading times on individual target words during reading were associated with higher score on the offline post-tests of the knowledge of meaning (but not form). This finding suggests that more deliberate engagement with the words during reading (as indicated by longer total reading times) may lead to higher gains in the learner's declarative knowledge of meaning.

Alternative approaches to contextual word learning

In addition to unassisted contextual word learning from level appropriate authentic or quasi-authentic texts, two alternative approaches used to overcome the lexical paradox in L2 reading are *reading-while-listening* and *narrow reading*. In reading-while-listening, learners usually listen to the recorded text of a level-appropriate book (often, a graded reader) and follow the recoding along, while reading the text in the book. This approach exposes learners to aligned auditory and visual input, which has the potential to facilitate the establishment of connections between written and spoken word

representations, increasing the chances of developing better quality lexical representation, compared to reading-only. Indeed, a number of reading-while-listening studies found that providing an audio recording of the text supports contextual vocabulary learning (Brown et al., 2008; Horst et al., 1998; Webb & Chang, 2015; Webb et al., 2013). In these studies, similar to reading-only studies, the Matthew effect is also present: readers' existing vocabulary knowledge is an important predictor of their ability to learn new L2 words from reading-while-listening (Webb & Chang, 2015).

The second approach, *narrow reading*, deals with the challenges of contextual word learning in a second language by encouraging students to read multiple books by the same author and within the same book series (e.g., Cho & Krashen, 1994, 1995; Krashen, 1981). Potential advantages of narrow reading include higher repetition rates of individual words, phrases, and syntactic structures in which they occur, due to the author's personal writing style; accumulation of background knowledge on the topic of reading that makes reading easier to understand; and learner motivation, provided they are able to choose the author, series, or topic for their own narrow reading programme (Krashen, 2004; McQuillan, 2016; Schmitt & Carter, 2000). Although the conjecture regarding positive effects of narrow reading on contextual word learning makes sense, there are no empirical studies, to my knowledge, that evaluate lexical development as a direct consequence of narrow reading and, thus, the vocabulary building potential of narrow reading remains yet to be confirmed.

One limitation of the *reading-while-listening* and *narrow reading* studies conducted so far is that they only measure gains in declarative vocabulary knowledge, often deploying offline multiple-choice tests of form and meaning recognition and recall. These measures do not tell us much about the development of lexical quality as a consequence of these two contextual word learning approaches. Is vocabulary encountered in reading-while-listening and narrow reading likely to be available online during reading, without the need for the reader to divert their cognitive resources to lower-level word recognition? Will the word knowledge gained be sufficient to support fluent reading with understanding? Future reading-while-listening and narrow reading studies will need to add online and nondeclarative knowledge measures to their design, in order to answer these questions.

Having reviewed research into unassisted contextual word learning, I shall now consider whether adding deliberate instructional and/or learning treatments into the mix can be used to improve lexical quality for L2 readers.

Optimizing contextual word learning for L2 readers

Since improving lexical quality is an important end goal of L2 lexical development and it may be more difficult to attain it in the second than first language from reading only, the question is whether it may be possible to use supplementary instructional or learning interventions in order to make L2

contextual word learning more effective? This question was considered in Laufer's (2009) article that provided a research timeline of L2 vocabulary acquisition from language input and from form-focused activities, covering a period of 1982 to 2008. Laufer (2009) concluded that 'a growing number of empirical studies suggest that input together with engaging word-focused activities and frequent rehearsals are likely to yield the best results' (pp. 341–342). A similar message is communicated by Schmitt (2008) who suggests that 'perhaps the most effective way of improving incidental learning is by reinforcing it afterwards with intentional learning tasks' (p. 352). Schmitt points out that increasing the amount of engagement with lexical items maximizes learning. In line with this, recent studies outlined below suggest that deliberate learning activities and procedures which draw participants' attention to various aspects of vocabulary knowledge (in addition to contextual exposure in reading) are beneficial for the development of declarative lexical knowledge. The picture is far less clear, however, as far as nondeclarative knowledge is concerned.

One approach that is known to result in high-quality lexical representations, affecting both declarative and nondeclarative knowledge, is using L2 flashcards (Elgort, 2011) – a paired associate learning approach, in which a word's form and meaning (presented as a simple L2 definition, defining example of use, and/or picture) are practised in a way that facilitates receptive and productive retrieval (see also Mondria & Mondria-de Vries, 1994; Nation, 2013). Elgort (2011) showed that after a week's spaced training regime (about four to five practice sessions), 48 novel vocabulary items were not only known explicitly but also exhibited good quality of lexical representations, as shown by the learners' ability to retrieve these items fluently in a series of online behavioural priming tasks. The L2 words, learned using flashcards, were shown to be integrated into the L2 formal-lexical and lexical semantic networks of the learners. These results show that deliberate learning through repetition and retrieval can result in high-quality lexical knowledge that is on a par with known L1 words. The results of Elgort's (2011) study were partially replicated by Elgort and Piasecki (2014) that investigated the use of L2-L1 flashcards (instead of L2-L2 flashcards) with German-English bilinguals. Similar to the participants in the original L2-L2 flashcard study, the bilinguals in this study developed robust knowledge of form (tested in a form-priming experiment); however, only more proficient English speakers developed robust knowledge of meaning. The lower-proficiency participants developed weaker semantic knowledge that resulted in an inhibition (rather than a facilitation) effect in a semantic priming task. This finding suggests that flashcards that do *not* encourage explicit links between the L2 word form and its L1 translation equivalent are a more effective learning method, compared to L2-L1 flashcards. Such flashcards can be used in conjunction with learning words from reading, in order to deliver a qualitative boost to the knowledge of lexical items learned contextually, from input. Therefore, language teachers should train learners to create their own flashcards (either paper-based or

electronic), explain how best to use them, and encourage them to create cards for unfamiliar words they encounter in reading.

Another quick and relatively easy way to facilitate the development of lexical quality in contextual word learning is to fine-tune formal-lexical representations during initial learning stages. Although recent eye-movement studies suggest that the form of the word becomes familiar to the reader relatively quickly (Elgort et al., 2018; Pellicer-Sánchez, 2015), form-meaning mapping takes longer and could become a bottleneck in achieving high-quality lexical representations for contextually learned words. The Lexical Quality Framework holds that a more precise representation of the orthographic form is a catalyst in word learning because it helps create a stronger association between the form of a novel word and its contextually inferred meaning. The importance of the knowledge of form has been highlighted by a number of vocabulary researchers in applied linguistics (e.g., Bogaards, 2001; Laufer, 1988, 2005; Nation, 2013). Hill and Laufer (2003), for example, showed superiority of word form oriented, compared to general meaning oriented, post-reading tasks on immediate and delayed meaning generation post-tests. In order to examine the effect of fine-tuning formal-lexical (orthographic) representations in conjunction with contextual word learning, Elgort et al. (2016) compared the effect of form-focused elaboration (operationalized as word writing) with that of meaning-focused elaboration (deliberately deriving meaning from context) on the quality of resulting word knowledge, with two groups of L2 readers: native speakers of Chinese (a language orthographically distant from English) and native speakers of Dutch (a language orthographically close to English). The authors predicted that Chinese speakers were likely to benefit more from form-focused elaboration because of the challenges in establishing precise orthographic representations in English – an orthographically distant L2 (Hamada & Koda 2008; Koda 1997). What they found was that both Chinese and Dutch readers benefited more from the word-writing activity than from the meaning inference activity, in terms of developing their declarative knowledge of form and meaning. Furthermore, the form-focused procedure also improved lexical representations of these words: for Chinese speakers, it improved the precision of lexical representations (measured by accuracy of lexical decisions) while, for Dutch speakers, who were able to establish precise orthographic representations in both learning conditions, it improved the fluency of lexical processing (as indicated by faster response times). Thus, simply writing down unfamiliar words encountered in reading led to an improvement in their lexical quality. This is something L2 readers can easily do themselves if they know about the benefits of such a procedure. However, the study findings need to be interpreted with caution because novel lexical items were presented in the sentence (rather than long text) context and the reading materials were designed to be supportive of the contextual learning of meaning. The form-focused elaboration approach may have been particularly effective under these conditions, compared to the meaning-focused elaboration approach, because the latter did not add much to the already

favourable contextual learning condition, while the former did. Therefore, future studies will need to verify this writing advantage when unfamiliar vocabulary occurs in less informative contexts, in long authentic texts.

Another supplementary learning technique that seems to positively affect outcomes of contextual word learning is giving readers access to the meanings of unfamiliar word encountered in reading through glosses and dictionaries (Hulstijn, Hollander, and Greidanus, 1996; Knight, 1994; Luppescu & Day 1993). This approach is particularly effective when readers actively derive the meaning of unfamiliar words from context (Fraser, 1999; Ko, 1995). In a recent study, Elgort et al. (2019) found that, when contextual meaning inferences about an unfamiliar word were *followed* by a dictionary-type definition of that word after reading a passage (about 150 words in length), L2 readers gained not only explicit knowledge of its meaning but were also somewhat better able to access this meaning online, when reading this word in a new context. This suggests that, in line with the Instance-based framework (Reichle & Perfetti, 2003; Bolger et al., 2008), processing new L2 words in informative contexts with subsequent access to their definitions creates favourable conditions for establishing their lexical semantic representations. The study by Elgort et al. (2019) was the first to show that this approach is better than presenting definitions prior to reading (i.e., reducing the likelihood of inference errors at the point of encoding). The authors conjectured that presenting a definition first may change the way an unfamiliar word is processed when it is encountered during reading, potentially, negatively affecting its learning. The superior learning from definitions presented *after* reading may be the result of a closer engagement with the context in an attempt to make a meaning inference and the post-reading verification process associated with reviewing definitions of the novel words after a contextual inference has been made.

Interestingly, in line with previous studies (Carpenter et al., 2012; Hulstijn, 1992), incorrect meaning inferences during reading negatively affected the declarative knowledge of the contextually learned words but they did not affect their nondeclarative knowledge, measured as reading times in the self-paced reading task. This dissociation of the effect of incorrect contextual meaning inferences on declarative and nondeclarative knowledge of meaning was also reported in Elgort (2017), where nondeclarative knowledge was measures using masked mixed-modality repetition priming task. These results suggest that making a meaning inference during reading should be encouraged (even if readers initially get the meaning wrong), as long as this inference is soon followed by the correct definition, e.g., in the form of a dictionary look-up.

Conclusion

In this chapter, I have considered how L2 vocabulary knowledge develops from reading. Going beyond increasing vocabulary size through learning new words, the focus of the chapter was on the development of lexical quality. Also, distinct from most of the L2 vocabulary literature, this chapter looked

into the development of nondeclarative (as well as declarative) word knowledge. Although the L2 vocabulary research is currently thriving, there are few contextual word learning studies that examine how lexical quality and nondeclarative knowledge develop over time from reading and how they may be affected by supplementary vocabulary-focused instructional and learning activities.

The studies reviewed in this chapter suggest that unassisted contextual word learning will be most effective for higher-proficiency L2 learners and early bilinguals, who already have decent-size L2 vocabularies, good-quality L2 lexical semantic knowledge, and are good at reading comprehension. For such L2 readers, the most important ingredients of successful contextual word learning are rate of occurrence in context, quality, and diversity of contextual encounters, and how much reading they are willing to do. They need access to engaging, level-appropriate reading materials that meet these criteria, encouragement from teachers and peers, and motivation to read at the right pace to meet their vocabulary learning goals (Nation, 2014).

For less skilled L2 readers, learners with smaller vocabularies, less developed L2 lexical semantic networks, and lower L2 proficiency, and for older foreign language learners, contextual word learning from reading needs to be supplemented with deliberate form-focused and meaning-focused learning activities of the kind described by Nation (2007) as language-focused learning. Activities that bring learner attention to the word form (in particular, writing down unfamiliar words encountered in reading) and encourage deliberate contextual meaning inferences followed by feedback seem to be particularly effective for the development of lexical quality and both declarative and nondeclarative lexical knowledge. Learners should also be strongly encouraged to make flashcards for the key words they encounter during reading and use them for regular retrieval practice in both form→meaning and meaning→form directions.

At this stage, the main gap in the applied linguistics literature on L2 vocabulary teaching, learning, and acquisition, is an almost complete absence of empirical studies investigating the development nondeclarative lexical knowledge. Clinging to the more traditional view of what it means to know a word, the majority of L2 vocabulary studies use measures of declarative knowledge of form, meaning, and use. However, these measures of offline, explicit knowledge are not informative of the real time lexical processing that needs to happen in fluent reading, i.e., ballistic and precise visual word recognition and resource-light and flexible access to contextually appropriate meaning. We need more evidence to understand how L2 readers are able to continue fine-tuning lexical quality of partially known words through reading, and how they use context in real time to resolve semantic ambiguity and continue acquiring new meanings of known words (Rodd, 2019). Such evidence is crucial in making informed decisions about optimizing conditions for effective contextual word learning for classroom-based teaching and out-of-class learning.

Note

1 Word family is a unit of counting used in word frequency lists organized in levels (or bands) of 1,000 headwords. List one contains the most-frequent 1,000-word families of English, based on the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). The lists are used to measure the vocabulary load of reading texts (Nation, 2006) and estimate receptive vocabulary knowledge of languages users (Nation & Beglar, 2007). A word family consists of a headword (e.g., cancel) and its inflected forms (cancelled, cancelling), as well as derivative forms that share a common meaning with the headword (cancellation, cancellations).

References

Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *23*(3), 321–343. doi:10.1017/S0272263101003011

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Contextual variation and definitions in learning the meaning of words. *Discourse Processes*, *45,* 122–159. https://doi.org/10.1080/01638530701792826

Braze, D., Tabor, W., Shankweiler, D., & Mencl, W. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities*, *40*(3), 226.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136–163.

Carpenter, S. K., Sachs, R. E., Martin, B., Schmidt, K., & Looft, R. (2012). Learning new vocabulary in German: The effects of inferring word meanings, type of feedback, and time of test. *Psychonomic Bulletin and Review*, *19*, 81–86.

Chang, A. (2019). Effects of narrow reading and listening on L2 vocabulary learning: Multiple dimensions. *Studies in Second Language Acquisition*, *41*(4), 1–26. https://doi.org/10.1017/S0272263119000032.

Cho, K. S., and Krashen, S. (1994). Acquisition of vocabulary from the Sweet Valley Kids series: Adult ESL acquisition. *Journal of Reading*, *37*(8), 662–667.

Cho, K. S., & Krashen, S. (1995). Becoming a dragon: Progress in English as second language through narrow free voluntary reading. *California Reader, 29*(1), 9–10.

Chung, S. C., Chen, X., Commissaire, E., Krenca, K., & Deacon, S. H. (2019). Testing the self-teaching hypothesis in second language reading, *Writing Systems Research.* https://doi.org/10.1080/17586801.2019.1625478.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology*, *11*, 38–63.

Cobb, T. (2016) Numbers or numerology? A response to Nation (2014) and McQuillan (2016). *Reading in a Foreign Language, 28*(2), 299–304.

Cree, G., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, *23*, 371–414.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*(2), 367–413.

Elgort, I. (2017). Incorrect inferences and contextual word learning in English as a second language. *Journal of the European Second Language Association, 1*(1), 1–11.

Elgort, I., Beliaeva, N., & Boers, F. (2020). Contextual word learning in the first and second language: Definition placement and inference error effects on declarative and nondeclarative knowledge. *Studies in Second Language Acquisition*, *42*(1), 7–32. doi:10.1017/S0272263119000561

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, *40*(2), 341–366. doi:10.1017/S0272263117000109

Elgort, I., Candry, S., Eyckmans, J., Boutorwick, T. J., & Brysbaert, M. (2016). Contextual word learning with form-focused and meaning-focused elaboration. Applied Linguistics, amw029, https://doi.org/10.1093/applin/amw029

Elgort, I., Perfetti, C.A., Rickles, B., & Stafura, J.Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and ERP indicators of learning. *Language, Cognition and Neuroscience*, *30*(5), 506–528.

Elgort, I., & Piasecki, A. (2014). The effect of a bilingual learning mode on the establishment of lexical-semantic representations in the L2. *Bilingualism: Language and Cognition*, *17*(*3*), 572–588.

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables, *Language Learning*, *64*(2), 365–414.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.

Ferrel Tekmen, E. A., & Daloğlu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, *39*(2), 220–243.

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, *21*, 225–241.

Frishkoff, G. A., Collins-Thompson, K., Perfetti, C., & Callan, J. (2008). Measuring incremental changes in word knowledge: Experimental validation and implications for learning and assessment. *Behavioral Research Methods*, *40*, 907–925.

Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2011). Predicting robust vocabulary growth from measures of incremental learning. *Scientific Studies of Reading*, *15*(1), 71–91.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., … Yoon, H. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584. doi:10.1017/S1366728917000219

Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning*, *58*, 1–31.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *Iral*, *41*, 87–106.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*, 74–95.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11,* 207–223.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In: Arnaud, P. J. L., & Béjoint, H. (eds), *Vocabulary and applied linguistics* (pp. 113–125). London: Macmillan. doi: https://doi.org/10.1007/978-1-349-12396-4_11

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, *18*, 193–223.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal*, *80*, 327–339.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47–77.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal*, *78*, 285–299.

Ko, M. H. (1995). Glossing in incidental and intentional learning of foreign language vocabulary and reading. *University of Hawaii Working Papers in ESL*, *13*, 49–94.

Koda, K. (1997). Orthographic knowledge in L2 lexical processing: A crosslinguistic perspective. In Coady, J., and Huckin, T. (eds) *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 35–52). Cambridge: Cambridge University Press.

Krashen, S. D. (1981). The case for narrow reading. *TESOL News*, *12*, 23.

Krashen, S. (2004). The case for narrow reading. *Language Magazine*, *3*(5), 17–19.

Laufer, B. (1988). The concept of 'synforms' (similar lexical forms) in vocabulary acquisition. *Language and Education*, *2*(2), 113–132.

Laufer, B. (2003). Vocabulary acquisition in a second language. Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*(4), 567–587. http://dx.doi.org/10.3138/cmlr.59.4.567

Laufer, B. (2005). Focus on form in second language vocabulary acquisition. In S. H. Foster-Cohen, M. P. Garcia-Mayo & J. Cenoz (eds), *EUROSLA Yearbook 5* (pp. 223−250). Amsterdam: John Benjamins.

Laufer, B. (2009) Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching*, *42*, 341–354, https://doi.org/10.1017/S0261444809005771

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*, 15–30.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*, 647–672. https://doi.org/10.111/j.1467-9922.2010.00621.x

Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning*, *43*, 263–287.

Masson, M. E. J. (1991). A distributed memory model of context effects in word identification. In Besner, D., and Humphreys, G. W. (eds) *Basic processes in reading: Visual word recognition* (pp. 233–263). Hillsdale, NJ: Erlbaum.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 3–23.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In Rumelhart, D. E . McClelland, J. L., & he PDP Research Group (eds), *Parallel distributed processing: Explorations in the microstructure of cognition* (vol. 1, pp.3–44). Cambridge, MA: MIT Press.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (1986). Parallel distributed processing: Explorations in the microstructure of cognition (vol. 2). Cambridge, MA: MIT Press.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997) On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.

McQuillan, J. (2016). What can readers read after graded readers? Reading in a Foreign *Language*, *28*(1), 63–78.

Mondria, J. A., & Mondria-de Vries, S. (1994). Efficiently memorizing words with the help of word cards and 'hand computer': Theory and applications. *System*, *22*(1), 47–57. https://doi.org/10.1016/0346-251x(94)90039-6

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Rowley, MA: Newbury House.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*, 59–82.

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*(1), 2–13. http://dx.doi.org/10.2167/illt039.0

Nation, I. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Nation, I. S. P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*(2), 1–16.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(1), 9–13.

Nation, I., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.

Pellicer-Sánchez, A. (2015). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*(1), 97–130.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*(1), 31–55.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383.

Perfetti, C. A, & Hart, L. (2001). The lexical bases of comprehension skill. In Gorfien, D. (ed.), *On the consequences of meaning selection* (pp. 67–86). Washington, DC: American Psychological Association.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In Verhoeven, L., Elbro, C. & Reitsma, P. (eds) *Precursors of functional literacy* (vol. 11, pp. 189–213). Amsterdam: John Benjamins.

Perfetti, P., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review*, *107*, 786–823.

Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *57*, Suppl. 1, 155–199.

Reber, P. J. (2008). Cognitive neuroscience of declarative and nondeclarative memory. In Aaron, S., Benjamin, A. S., De Belle, J. S., Etnyre, B., & Polk, T. A. (eds) *Advances in Psychology*, *139* (pp. 113–123). Amsterdam: North-Holland.

Reber, P. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia*, *51*, 2026–2042. https://doi.org/10.1016/j.neuropsychologia.2013.06.019

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading*, *7*, 219–237.

Rodd, J. M. (2019). Settling into semantic space: An ambiguity-focused account of word-meaning access. 30 January. https://doi.org/10.31234/osf.io/75m4k

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Schmitt, N., & Carter, R. (2000). The lexical advantages of narrow reading for second language learners. *TESOL Journal*, *9*(1), 4–9.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151–218.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research ma*nual. Basingstoke: Palgrave Macmillan.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–400.

Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, *34*, 39–65. https://doi.org/10.1177/0267658316675195

Walberg, H. J., & Tsai, S.-L. (1983). Matthew effects in education. *American Educational Research Journal*, *20*, 359–373.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? Reading in a Foreign *Language*, *15*, 130–163.

Webb, S., & Chang, A. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, *37*(4), 651–675. https://doi.org/10.1017/S0272263114000606

8 Measuring reading and vocabulary with the Test for English Majors Band 4

A concurrent validity study

Jeanine Treffers-Daller and Jingyi Huang

Introduction

Testing students' English language proficiency is an enormous undertaking in China. Every year no less than 18 million students take the College English Test as part of their undergraduate studies (Yu & Jin, 2014). The current project focuses on another widely used test, namely the Test for English Majors, Band IV (TEM-4), for which the number of test takers has soared to 270,000 between 1992 and 2015 (Xu & Liu, 2018). The TEM-4 is a criterion-referenced English language test for university undergraduates majoring in English Language and Literature in China and aimed at testing a wide range of aspects of students' English proficiency levels, as well as their knowledge of the content of the National College English Teaching Syllabus for English Majors. According to Jin and Fan (2011), the TEM-4 is considered to be a reliable and valid test. At the same time, they mention a study by Chen (2009), who notes there are issues with construct underrepresentation and construct-irrelevant variance, both of which are key aspects of the construct relevance of a test (Messick, 1995). Construct underrepresentation refers to situations where important dimensions of a construct are not included in a test. Construct-irrelevant variance, by contrast, is found in situations where a test is too broad and contains variance that is associated with other constructs. An example of how these issues affect a test of reading comprehension can be found in Ready, Chaudry, Schatz, and Strazullo (2012) study of the Nelson-Denny reading comprehension test (Brown, Fishco, and Hanna 1993). The authors found that many test items could be answered correctly by testees who had not seen the reading passage (i.e. passageless administration). This means that the scores obtained did not really reflect reading comprehension of the passage. Instead, they were associated with general intelligence, levels of vocabulary and knowledge and broad reading skills. In other words, there was a considerable amount of construct-irrelevant variance in the data set. As pointed out by Weir (2005: 18), it is very important to ensure that the construct we are eliciting with a test is indeed the construct we aim to measure. Khalifa and Weir's (2009) model of reading presents a detailed overview of the different processes involved in reading. One of these is inferencing, that is the ability of readers to go beyond explicitly stated ideas and to build a mental representation of what a text is about (Khalifa & Weir, 2009: 50). Graesser, McNamara, and Louwerse (2003)

also mention the importance of measuring readers' inferencing skills. A reading test which only measures understanding of literal meanings as found in the text but does not assess respondents' ability to infer meanings that are not explicitly mentioned in the text is therefore likely to underrepresent the construct (see also the Methods section for further discussion).

From the literature available to researchers in Western Europe it is not clear to what extent the validity issues sketched above been investigated in any detail for the TEM-4. In the 1990s Zhou, Weir, and Green (1998) carried out a three-year validation study of the TEM-4 and the TEM-8 published by the TEM Test Centre in Shanghai. Zhou et al. found that the TEM tests (both Band IV and Band VIII) were 'reasonably valid and reliable tests' (p. 63). However, they also note that the concurrent validity of the version of the TEM-4 (from 1995) was not high, as correlations between scores on the different components of the TEM-4 and the corresponding scores on the Test of English for Educational Purposes (TEEP), developed at the University of Reading[1] (UK), were relatively low (r = 0.4037). The authors suggest that one of the reasons for the low correlations might be participants' lack of familiarity with the open-ended format of the TEEP questions. The reading component of the TEM-4 has changed considerably since the publication of Zhou et al.'s validation study in that the test no longer distinguishes between careful reading and speed reading. Therefore, conclusions from this report may not hold anymore for more recent versions of the TEM-4. Jin and Fan (2011) report that the average test reliability was good between 2008 and 2010 but also note that there are still very few published validation studies of the test: It is not clear whether, for example, a passageless administration of the TEM-4 has been attempted or whether any concurrent validity studies have been carried out. They therefore call for further validation studies of the test.

The current study sets out to evaluate the concurrent validity of the vocabulary and reading components of the TEM-4 by correlating respondents' scores on this test with those on widely used tests of vocabulary size and depth as well as a test of reading comprehension, the York Assessment of Reading Comprehension Secondary (from now on the YARC, Snowling et al., 2009). We are of course aware that tests are not suitable for all learners in all contexts (Schmitt, Nation, and Kremmel, 2019). A potential issue with the YARC Secondary is that it was developed for students in the UK (both L1 and L2 users of English) who receive English input in their daily lives. We therefore also look in detail at the suitability of the test for the target group of L2 learners in China.

In the current study we will, first of all, investigate to what extent the reading and vocabulary components of the TEM-4 correlate with widely used tests of vocabulary and reading. Second, we will look at the relative contribution of size and depth of vocabulary knowledge to explaining reading comprehension as measured with the TEM-4 and the YARC.

The structure of the current paper is as follows. First, we present the construct of reading and the ways in which this is measured with different tests.

The next section focuses on vocabulary and its measurement. After this, we sketch the aims and research questions that have guided the current study and the methods chosen to answer our questions. We then present the results and a discussion, and finish with a section which offers a summary and a conclusion.

Reading comprehension: the construct and its measurement

Under the Simple View of Reading (Gough & Tunmer, 1986), which is the most widely used model of reading ability, decoding is one of the two key dimensions of reading, the other one being linguistic comprehension. Decoding refers to readers' ability to recognize words, that is to make a link between the printed word and the appropriate entry in the mental lexicon. Reading comprehension, by contrast, is defined as the ability to understand written language. More specifically it refers to readers' ability to use lexical (semantic) information and to derive sentence and discourse level interpretations from it (Gough & Tunmer, 1986). Reading comprehension is thus different from linguistic comprehension, in that the former relates to written and the latter to aural language. While Gough and Tunmer recognize that there are many aspects to understanding a text, the two dimensions of decoding and linguistic comprehension are the essential ones without which no reading can take place.

In her discussion of the construct of reading, Snow (2002: 11) elaborates on the notion of reading comprehension, which she defines as 'the process of simultaneously extracting and constructing meaning through interaction and involvement with written language'. To be able to construe meaning in this way, at sentence as well as textual levels, the reader needs to know about the domain and the topic, have the necessary linguistic and discourse knowledge, and rely on cognitive capacities (e.g., attention, memory, critical analytic ability, inferencing, visualization ability).

Reading fluency, that is the 'ability to read rapidly with ease and accuracy, and to read with appropriate expression and phrasing' (Grabe, 2009: 291), is another variable which has been found to correlate strongly with reading comprehension. Grabe (2010) suggests that readers who read fast and have very efficient word recognition skills are generally able to integrate information from different sources and construe text-level interpretations, even under time pressure. Indeed, the available research on L2 reading fluency indicates that word reading fluency and passage reading fluency impact on reading comprehension. Conversely, a lack of reading fluency is also a reliable predictor of reading comprehension difficulties (Stanovich, 1991).

Among the linguistic variables that are relevant for reading, vocabulary has often been found to be a key predictor (Laufer & Ravenhorst-Kalovski, 2010). Since Stanovich's (1986) seminal publication on the Matthew effect in reading, it has been known that there is a reciprocal relationship between vocabulary knowledge and reading: readers who know more words are better

readers, and better readers can learn new words from reading more easily. The relationship between reading and vocabulary has therefore been the focus of a wide range of studies. For the purposes of this chapter we will limit the presentation of the available literature to studies which focus on adult L2 learners of English, as these are the target group for the current study.

According to Nation and Waring (1997) adult learners of English often have vocabularies smaller than 5,000 words, despite having learned English for several years. However, for reading a newspaper or a novel, however, around 8,000–9,000 words are needed (Nation, 2006). This means that both decoding and reading comprehension are likely to be more difficult for this group than for monolinguals. Clearly it is not just the size of a person's vocabulary that matters but also how well words are known (vocabulary depth). Qian (2005) found that depth of vocabulary knowledge contributes more to reading comprehension than readers' vocabulary size, although for Binder et al. (2017) only vocabulary size explained unique variance in reading fluency. As the authors point out, it is quite challenging to measure vocabulary depth and the battery used in the study may not have been sufficient to tap this construct successfully.

Before looking in more detail at the construct of vocabulary and how this can be measured, a few words must be said about the measurement of different components of reading. As Ready et al. (2012) point out, there are not many reading tests for adults, and even fewer that specifically target adult L2 learners. One of the tests for adult native speakers is the National Adult Reading Test (NART, Nelson, 1982). This test assesses word recognition and familiarity of words, and consists of 50 words of increasing difficulty, all of which have irregular grapheme-phoneme correspondences. However, this test is unlikely to be suitable for L2 learners who have small vocabularies because a Vocabprofile analysis of the items, provided by Tom Cobb's Lextutor (https://www.lextutor.ca/vp/), shows that 82% of the words in the NART belong to frequency levels lower than 5k, and the test includes words from frequency layers up to 20k. L2 learners are therefore likely to know only very few of these items.

The YARC Secondary is a comprehensive test of reading, based on the Simple View of Reading (Gough & Tunmer, 1986). The test was developed for 11 to 16-year-old students in the UK (including non-native speakers of English). A sample of 89 students for whom English was an Additional Language was included in the standardization sample. As might be expected, scores for non-native speakers were lower than those for native speakers (see https://www.gl-assessment.co.uk/support/yarc-support/). The reliability information as provided in the manual shows that Cronbach's alpha varied from .85 to .90 for most components, except for the summarization part, where reliability ranged from .65 to .74.

The reading passages are accessible for readers with smaller vocabularies because they contain very few words beyond the 5k level (further discussion below). According to Stothard (2010), the reading comprehension questions

include a range of inference questions that can be used to assess predictive, evaluative, knowledge-based, and cohesive inference. As a detailed analysis of these different kinds of inferencing is beyond the scope of the current chapter, the reader is referred to Bowyer-Crane and Snowling (2005) for an overview. In the YARC, reading comprehension is not only assessed with comprehension questions, but also with a summarization task (see Yu, 2008, for a discussion of summarization to assess reading comprehension).

Whether or not the YARC Secondary is also suitable for university students of English in China is an empirical question. The TEM-4 contains a reading component which consists of four to five different reading texts and understanding of these texts is measured with multiple choice questions. However, it is not clear which model of reading underpins the test, and it seems to only target reading comprehension, as measures of word recognition or fluency are not included. Although inferencing skills are mentioned in the 2015 test specification,[2] most of the reading comprehension questions in the TEM-4 test paper used in the current study, and later TEM-4 papers from 2015 and 2016 which we have seen, appear to mainly assess literal information.

Before explaining the specific objectives of the current study, we first briefly present two concepts which, as we have seen in the opening section, are key to reading comprehension, namely vocabulary size and vocabulary depth.

Vocabulary size and vocabulary depth: the constructs and their measurement

Most researchers in the field would agree that it is not only important for readers to know a large number of words, but how well they know these words (the depth of their knowledge) is relevant too. The most widely used model of vocabulary knowledge is that of Nation (2013) who proposes there are three basic components to vocabulary knowledge, namely form, meaning and use, each of which can be known to different degrees both receptively (passively) or productively (actively). There is a wide range of possible options for vocabulary tests for L2 learners, depending on whether active or passive recognition or active or passive recall is measured (see Laufer & Goldstein, 2004), even though many widely used vocabulary tests have not sufficiently been validated (Schmitt et al., 2019).

Nation and Beglar's (2007) Vocabulary Size Test (VST) is a widely used free test of both first language and second language learners' written receptive vocabulary size, that is the vocabulary size needed for reading in English. As pointed out by Gyllstadt, Vikaitė, and Schmitt (2015), there is a clear risk that the test overestimates the vocabulary learners know, as is often the case with multiple choice format. Further information about the validity of the test can be obtained from Beglar (2010).

Testing vocabulary depth is even more complex than testing vocabulary size. A widely used format is the Vocabulary Knowledge Scale (Paribakht &Wesche, 1996), which is a self-report form on which respondents indicate

on a five-point scale how well they know a particular word. While there are obvious disadvantages to using self-report, the format has been widely used, also because users can include items they want to focus on. In addition, other tests of vocabulary depth, such as the Word Associates Test (Read, 1993), which taps into collocational knowledge as well as antonyms and synonyms, can be very complex and unsuitable for learners with relatively small vocabularies.

Measuring vocabulary knowledge is very important for studies of reading, because vocabulary is key determinant of reading comprehension (Laufer & Ravenhorst-Kalovski, 2010). In a comprehensive study of over 600 learners of English from eight different countries, Schmitt, Jiang, and Grabe (2011) found that there is no specific vocabulary threshold for understanding text. Rather, it was the case that students with higher scores on different vocabulary measures could demonstrate more in-depth understanding of the texts in the study. Establishing an exact threshold is also complicated because the results depend on the degree of comprehension that is required: if only 60% needs to be understood, then a coverage of 95% is probably sufficient. However, most teachers would probably want their students to understand more of the text. The authors therefore suggest that if 70% comprehension is required, a 98–9% coverage is needed. Importantly, they also found that even students who knew all the words did not always get full marks on the comprehension task. This is likely due to the fact that non-native speakers may not be familiar with the genre or the wider context or lack cultural information that is needed to comprehend a text.

In the Chinese context, according to figures from the College English Curriculum Requirements of the Ministry of Education from 2007, reported in Zhao, Wang, Coniam, and Xie (2017), at the Basic Level, students should know 4,795 words and 700 phrases and expressions; at the Intermediate level it is 6,395 words, and 1,200 phrases and expressions; and at the Advanced level it is 7,675 words and 1,870 phrases and expressions. However, actual vocabulary knowledge of students is often much more limited, even among students who are majoring in English. In a recent study among second year non-English major students in China in which Nation and Beglar's (2007) Vocabulary Size Test was used to measure vocabulary size, Wang and Treffers-Daller (2017) found that the students knew on average just under 3,000 words, which is far less than the 5,000 words they are required to know according to the syllabus. These figures may even be inflated as according to Gyllstad et al. (2015), vocabulary sizes as measured with the VST overestimate students' knowledge by up to 26%. Other sources do indeed report lower actual vocabulary sizes for Chinese university students. In a study among English majors and non-English majors, in which Schmitt, Schmitt, and Clapham (2001) vocabulary levels test was used to assess students' vocabulary knowledge, Zhang (2009) found students knew 2,156 words receptively and 859 productively. While a detailed overview of the vocabulary knowledge of Chinese students is beyond the scope of the current study, these studies suggest that Chinese university

students do not always know enough words for independent reading of authentic materials such as texts from newspapers or novels.

The current study hopes to contribute to a further understanding of these issues, as will be explained in the next section.

The current study

The aim of the current project is, first of all, to evaluate the concurrent validity of the reading and vocabulary components of the TEM-4. Investigating concurrent validity entails investigating whether the data collected from one instrument correlate highly with the data collected from another instrument which purportedly measures the same construct (Cohen, Manion and Morrison 2018: 258). We will assume the TEM=4 to be a valid test of reading if there are positive and significant correlations between the results of this test and different components of the YARC Secondary. The vocabulary component of the TEM-4 will be assessed in a similar way against two widely used vocabulary tests. It is important to note here that vocabulary as well as grammar are assessed together in one component of the TEM-4. Therefore, this component does not assess only one construct, but two. This is not necessarily a problem, however, as many researchers assume with Halliday (1994: 14) that 'grammar and vocabulary are merely different ends of the same continuum'.

Next, as some studies have shown that vocabulary depth is more important than vocabulary size for reading comprehension, while other studies found the opposite, the second objective is to investigate to what extent these two dimensions of vocabulary knowledge can explain unique variance in reading comprehension in our study.

The following two research questions have guided our investigation:

RQ1: To what extent do the reading and vocabulary components of the TEM-4 tap into the constructs they are intended to measure?

RQ2: To what extent do vocabulary size and vocabulary depth explain unique variance in reading comprehension as measured with the TEM-4 and the YARC Secondary?

Methods

The participants in this study were 60 second-year English Major (Education) undergraduate students who were studying at a university in the North of China. The students' ages ranged from 18 to 22, and their first language was Mandarin. A brief questionnaire revealed that the students rarely spoke English after class and communicated only a few times a year with (near-) native speakers. All 60 students took the following three tests: the TEM-4, the VST, and the VKS, and a subsample of 30 students were also administered the YARC (see further down in this section).

The TEM-4 contains six different parts: dictation, listening, cloze, vocabulary, reading, and writing, each of which is described in detail in Jin and Fan (2011). For the purposes of the current study it is important to know that the vocabulary and grammar component of the version of the TEM-4 we used (a past paper from 2011) consisted of 30 multiple choice test for which respondents ed to recognize the meaning of a target word out of four options. Half a point was given for each correctly answered question, which means the maximum score was 15. The reading comprehension part of the TEM-4 contained four passages with 20 multiple-choice questions in total. Each of the four passages had a different theme (see Appendix A for an example). An analysis of the vocabulary in the texts revealed that 95% coverage of the texts was achieved at the 6k level. This means that the texts were probably relatively difficult for the students, although some rare words were translated with glosses in the text. Reliability of the individual components could not be computed as only total scores for each component were provided by the school. Students' results for the TEM-4 were obtained from the school administration.

The YARC consists of three parts: in Part 1 decoding is measured, in Part 2 reading comprehension, and in part 3 reading fluency. For Part 1, the Single Word Reading Test (SWRT), students need to read 70 single words aloud. One point is awarded for each word read correctly. An analysis with Vocabprofile showed that in the first half of the test all words except one (*yawned*) belong to the highest three frequency levels. The second half, however, contained words in frequency levels up to 13k. While this means that the test is easier than the NART (Nelson, 1982), which contains words up to the 20k level, the second half of the SWRT is likely to be difficult for students.

Part 2 assesses Reading Comprehension. For this part there is a choice of two levels for the reading fluency passages (Level 1 and Level 2). For this study, Level 1 was chosen because a pre-test revealed this was the more appropriate one for the target group. As we expected the students to have relatively low levels of vocabulary, we chose two passages from Level 1: the School Boy (fiction) and Honey for You, Honey for me (non-fiction). A Vocabprofile analysis of these two stories revealed that 95% coverage was reached at K4. These texts were therefore likely to be a little easier for the students than the reading texts of the TEM-4, at least as far as the vocabulary is concerned. Analyses of the readability of the texts confirm this. We used the Flesch Reading Ease score, which is based on the number of words per sentence and the number of syllables per word, and found a score of 60.1 for the TEM-4 texts, while the YARC texts obtained 79.9. As texts for which a score of 30 is given are considered difficult and those which obtain 70 easy (Stajner, Evans, Orăsan and Mitkov 2012), these results suggest that the YARC texts were easier than the TEM-4 ones. The results of the Flesch Kincaid readability scores, which are a simplified version of the Flesch Reading Ease score, and indicate US grade levels, point in the same direction: 8.5 for the TEM-4 and 5.5 for the YARC. According to the data provided by Stajner et al. (2012), this means that the TEM-4 readability scores are closer to those for news texts and the

YARC ones closer to fictional texts. Both the lexical analysis and the readability indices therefore show that the YARC texts were simpler than the TEM-4 texts. As one reviewer points out, readability indices present only one aspect of the difficulty levels of a text: the comprehension questions for a text may be easy for a difficult text and vice versa. However, comparing the difficulty of the questions is unfortunately beyond the scope of the current study. For each text students had to respond to thirteen comprehension questions, each worth one point, and for the summarization part of each story eight and nine points could be obtained. In addition, they were required to summarize the texts.

Part 3 assessed reading fluency. The passage students read contained 137 words, and one point was awarded for each word read correctly (reading fluency). The time needed to read the passage (reading rate) was also recorded.

As the YARC Secondary had to be administered on a one-to-one basis, and it was not feasible to test all students with the available means, 30 of the 60 students were randomly selected and administered the YARC Secondary Test. Each tutor assessed ten students at different times over a period of two months. For the analysis, we did not make use of the ability scores because the students' ages were higher than those of the group for which the test was developed, and the students were classroom learners of English with relatively little contact with day-to-day English. The ability scores and their associated norms as found in the manual are therefore unlikely to be appropriate for the sample in the current study.

As the test was intended for UK-based students, we were interested in obtaining teachers' and students' opinions about the YARC too. Therefore interviews were held with a small sample of students and the classroom teachers involved in the study.

Both vocabulary tests that were used in this study are widely used with adult L2 learners. The bilingual Mandarin-English version of the VST was chosen to ensure students were able to understand the answer options. Considering the students' relatively low vocabulary levels, only the first eight levels of the fourteen levels in the VST were used in this research. Thus, the maximum score was 80. Students' vocabulary sizes (word families) were computed by multiplying the scores with 100, as suggested in Nation's (2012) test specification for the VST. As the reliability of the VST was a little low (Cronbach's Alpha = .671), we decided to leave out the third level of the VST, which led to an improvement of the reliability (Cronbach's alpha = .712). The VST was administered to all participants during class time. Students also filled in a brief background questionnaire about their language learning history and personal background.

Brown's (2008) slightly simplified version of the VKS was used to assess vocabulary depth. No points were given when students ticked level 1 ('I don't know this word') or level 2 ('I have seen this word but I don't know what it means'), because recognition of the form of the word was not assessed at level 2, and therefore information provided by students could not be verified. The

difference between 'I think this word means X' (level 3) and 'I know this word and it means X' (level 4) was considered to be indicative of students' confidence rather than their actual degree of knowledge, and therefore level 3 was not used. One point was given for receptive knowledge (ability to translate the word) and one for productive knowledge of a word (ability to use the word in a sentence), even if there was a spelling or grammar error in the answer. In total 20 words randomly selected from the 1K until the 8K levels of the VST were included in the VKS (see Appendix B). The maximum number of points that could be obtained was therefore 40. Participants were all given the VKS in class at the same time.

Before carrying out any further analyses, we investigated whether the scores on the different tests were normally distributed. No significant differences were found with the normal distribution for any of the test results. No floor or ceiling effects were found.

Results

Here, we will first give an overview of the descriptive results for all tests. This will include an analysis of the suitability of the YARC Secondary for adult Chinese L1 learners of English. Subsequently the correlations between the different components of the TEM-4, the YARC, and the vocabulary tests will be discussed (RQ1) and, finally, we will look into the dimensions of vocabulary knowledge which can predict reading comprehension as measured with the TEM-4 and the YARC (RQ2).

Descriptive results

Students' total mean scores on the TEM-4 were 63.6 (SD 10.5), with a minimum of 37 and a maximum of 85. This means that, on average, students obtained a pass mark for the test, as scores between 60 and 69 are a 'pass' (Jin & Fan, 2011). However, one-third of the students in this group obtained a mark below 60. Students' English language levels are therefore likely to be relatively low. For vocabulary and grammar the mean score was 8.88 (SD 3.1), which means that students answered 59% of the questions correctly. For the reading component, the mean score was 13.81 (SD 3.31). As 69% of the answers were correct, for this part of the TEM-4 students therefore obtained slightly better scores.

The results for the VST show that students obtained a mean score of 41.5 (SD 7.3) out of 70 items (without level 3 which had to be deleted for reasons of reliability). The minimum score was 25 and the maximum 63. The total number of word families known by students was therefore on average around 4,000, although ten students had vocabularies smaller than 4,000 word families. Figure 8.1 reveals that students' performance decreased at the lower frequency levels, so that at the K6 and K7 levels they knew only half of the items, and at the K8 level they were just above chance level.

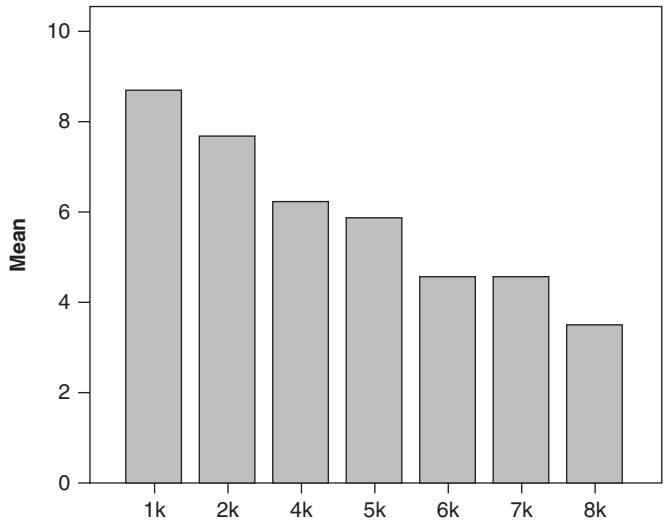


Figure 8.1 VST results.

Table 8.1 Results (raw mean scores) from the YARC Secondary

	Minimum	*Maximum*	*Raw mean (%)*	*SD*
SWRT	25	50	37.37 (53.39)	7.33
Fluency Accuracy score	128	137	133.10 (97.15)	2.43
Fluency Time	52	83	65.87	7.38
Reading Comprehension	4.00	18.00	10.73 (41.27)	3.25
Summarization	4.00	16.00	8.67 (51)	2.47

SWRT = mean raw scores on the Single Word Reading Test (maximum 70)
Fluency Accuracy Score = mean raw scores on the Reading Fluency Items (maximum 137)
Fluency time: number of seconds needed to read the words
Reading Comprehension = mean raw scores on two passages (maximum 26)
Summarization = mean raw scores of two passage summaries (maximum 17)

For the VKS, the mean score was 20.1 (SD 4.4) and the minimum and maximum scores were 7 and 36. This means that the students knew on average on half of the items in the test, and 42% of the students knew less than half of the items.

The results for the different components of the YARC are given in Table 8.1. The scores reveal that students obtained around 50% on most components, except for reading fluency, where they obtained almost full marks. The lowest scores were given for the reading comprehension part.

As the Vocabprofile analyses and the readability indices indicate that the texts were relatively easy, certainly by comparison with the TEM-4 texts, the

texts are unlikely to have been too difficult for the students, except for the ones who knew less than 4,000 word families. Instead, the reading comprehension part may have been particularly difficult for students because they were not familiar with inferential questions. The reading comprehension component of the TEM-4 version we used contained mainly questions which assessed comprehension of information which had been provided in the text, and inferential skills were hardly assessed. In addition, students may not have been familiar with the question format, as the comprehension questions were open questions rather than multiple choice. The students' results are also low by comparison with the raw scores for secondary school pupils in the UK. In a large-scale study among students in state schools in the UK, in which 8.2% of students were known to have English as an Additional Language, Stothard et al. (2010) report that on the SWRT year 7 students (N =178) obtained mean scores of 47.88 (SD 9.20), with values ranging from 18 to 67. For the other components only the standardized scores are reported, so that a comparison with our sample is not possible.

That students were struggling was confirmed in interviews held with teachers and the students after the completion of the tests. Teachers reported that most students were able to accurately read the first 40 words of the SWRT, but struggled with the last 30 words. This is not surprising as we had already seen that the first 35 words on the list belonged to the highest three frequency levels, but in the second half words up to 13k were included. While this component was considered difficult, the teachers confirmed that the reading fluency test was relatively easy for the students, which was also clear from the high scores on this part of the test.

Students who were interviewed mainly pointed to problems with listening comprehension, but the second most common problem the students encountered related to their grammar mistakes and limited vocabulary knowledge. While one student reported not knowing some keywords in the sentence, which made it difficult to understand the meaning of the entire sentence, the comments of another student pointed in the direction of her problems with integrating the information at sentence-level despite knowing the meaning of the words: 'I know most of words, but I [am] still not sure [of] the meaning of the sentence'.

In light of the above, it is likely that the YARC Secondary was probably rather difficult for the students in the current sample for students with low vocabulary levels and because of their lack of familiarity with inferential questions. However, the vocabulary in the stories was not too difficult, as the Vocabprofile analyses have shown. In fact, the vocabulary in the stories was simpler than those in the TEM-4 and the readability indices confirmed this too.

Correlations between the TEM-4, the vocabulary tests, and the YARC Secondary

The first aim of our study was to investigate to what extent the reading and vocabulary components of the TEM-4 tap into the constructs they are

intended to measure. To enable us to answer this question we will first analyse the correlations between the different tests.

Table 8.2 gives an overview of all Pearson correlations between the variables. It reveals that the TEM-4 (total scores) correlates most strongly with the VKS (.521**), and the VST (.420**), but among the variables associated with the YARC only reading rate correlates significantly with the TEM-4 (-.371*). For the reading component of the TEM-4 the same picture emerges: significant correlations are again found only with the vocabulary tests (both around .352**), although these are slightly less strong than the correlations with the overall TEM-4 scores. Interestingly, there is also a moderate correlation between the reading component of the TEM-4 and the grammar and vocabulary component of this test (.369*). As might be expected, the latter also correlates with both vocabulary tests. Again the correlations are slightly stronger with the VKS (.354**) than with the VST (.270*). Finally, there are some correlations between the different components of the YARC, which are less relevant for the aims of the current study.

In summary, these results show that the reading component of the TEM-4 is strongly related to students' vocabulary knowledge, and in particular to vocabulary depth. The absence of significant correlations between the YARC variables and the reading comprehension component of the TEM-4 is worrying, and makes the reader wonder if the TEM-4 reading comprehension component really taps into this construct. As one reviewer points out, it is also possible that the TEM-4 reading component measures different aspects of reading than the YARC. The fact that the *overall* scores on the TEM-4 (rather than the reading component on its own) correlated moderately but significantly with reading rate means that those who obtained higher scores on the TEM-4 are faster readers than those who obtained lower scores. In other words, the TEM-4 does indeed tap into one of the dimensions of reading that the YARC measures too, namely reading rate. However, the TEM-4 reading component, which is labelled 'reading comprehension', targets the same construct as the YARC Secondary reading comprehension task. The lack of correlations between these two tests therefore raises questions regarding the construct validity of the TEM-4 reading comprehension component.

As for the dimensions of vocabulary that are most relevant for reading, in the current study it appears to be the case that vocabulary depth is more important than vocabulary size as the TEM-4 correlates more strongly with the VKS than with the VST. In addition, only the VKS correlates significantly with different dimensions of the YARC (reading fluency and reading comprehension). It is possible of course that the scores on the VST are not representative of students' actual vocabulary knowledge because of the guessing factor that might lead to an overestimation of students' vocabulary size (Gyllstad et al., 2015). Because the multiple-choice format is not used in the VKS, it might present a more valid picture of their vocabulary knowledge. In the next section we will look into whether both measures make a unique contribution to reading comprehension.

Table 8.2 Correlations between TEM-4, VST, VKS, and YARC Secondary

	TEM-4 Reading	*TEM Grammar and Vocab*	*VST*	*VKS*	*YARC SWRT*	*YARC Fluency*	*YARC Rate*	*YARC Comprehension*	*YARC Summari-sation*
TEM-4 Total	.690**	.738**	.420**	.521**	0.295	0.265	-.371*	0.071	0.211
TEM-4 reading		.369**	.352**	.355**	0.086	0.173	-0.136	0.107	0.012
TEM-4 GrammarVocab			.270*	.354**	0.287	0.141	-0.298	0.040	0.049
VST				.510**	0.340	0.287	-0.219	0.285	0.047
VKS					0.255	.386*	-0.200	.379*	0.157
YARC SWRT						.696**	-0.236	0.311	0.301
YARC Fluency							-.375*	.406*	.369*
YARC rate								0.013	-0.273
YARC comprehension									0.152
YARC summarisation									

* = correlation significant at $p < .05$; ** correlation significant at $p < .01$

The contribution of the VST and the VKS to reading comprehension as measured with the TEM-4 and the YARC Secondary

The second objective of the current study was to establish what extent the two vocabulary tests explain unique variance in reading comprehension as measured with the TEM-4 and the YARC. We first ran a hierarchical regression analysis with the TEM-4 reading component as the dependent variable, and the VKS and the VST as predictor variables. In the first model we entered the VKS in the first step and the VST in the second step. The β value for the VKS was .237 and for the VST it was .231. The overall model was significant ($F(2,57) = 5.65$, $p < .006$) and the VIF and tolerance values were within acceptable ranges. Together these variables explained 16.6% of the variance in reading. The changes in R^2 were .126 for the VKS and .040 for the VST, which means the VST explained unique variance in reading over and above the contribution of the VKS. When the order of the entry of the predictors was reversed, the model was virtually identical: the β value for the VST was .231 and for the VKS it was .237. The changes in R^2 were .124 for the VST and .041 for the VKS.

As the TEM-4 contains a grammar and vocabulary component too, we also attempted a model which used this variable in addition to the VKS as a predictor for the reading component of the TEM-4. This model turned out to be significant ($F(2,57) = 6.848$, $p = 0.002$). Again multicollinearity values were within acceptable limits. The addition of this variable led to an increase in explained variance (19.4%), with β values of 0.256 for the VKS and 0.278 for the TEM-4 Grammar and vocabulary predictor. That this variable explained additional variance in the model may in part be explained by the fact that it covers not only vocabulary but also grammar. After entering the TEM-4 Grammar and vocabulary Scores, the VST was no longer a significant predictor in this model.

Thus, on the basis of the models presented above, we can conclude that vocabulary size and vocabulary depth both explain unique variance in reading as measured with the TEM-4. As the total explained variance is relatively low, we wondered whether other variables, such as reading rate or reading fluency, as measured with the YARC, would explain additional variance but that was not the case.

We subsequently ran regression models with the YARC fluency score as the dependent variable. Recall that we did not compute the ability scores as the group differed in age from the group for which the test was created. The VKS turned out to be the only significant predictor of fluency ($F(1,28) = 4.89$, $p = 0.35$). It explained 14.9% of the variance in fluency ($\beta = .386$). The VST was not a significant predictor of fluency, neither on its own, nor in combination with the VKS. Adding reading rate (or the grammar and vocabulary component of the TEM-4) to the model did not bring about a significant change in R^2.

When the reading comprehension score of the YARC was used as the dependent variable, the result was very similar. Again the VKS was the only

significant predictor of reading comprehension (F(1,28(= 4.70, $p = 0.039$). The VKS explained 14.4% of the variance in reading comprehension ($\beta = .379$). Adding reading fluency to the model brought the total explained variance to 24.1% (F (2,27) = 4.29, $p = 0.024$), because a further 9% of variance was explained by reading fluency. The addition of any other variables could not improve the model.

The results of the regression analyses confirm not only that vocabulary is an important predictor of reading ability as measured by both tests, but also suggest that the TEM-4 and the YARC measure different aspects of reading ability. For reading as measured with the TEM-4, vocabulary size and vocabulary depth play an approximately equally important role, but for the YARC it is only vocabulary depth that matters and not vocabulary size. The findings from the regression models based on the YARC therefore confirm the findings of Qian (2005) who emphasized the importance of vocabulary depth for reading comprehension. In addition, reading fluency was a significant predictor of reading comprehension as measured with the YARC. The model based on the YARC data therefore supports the view of Grabe (2010) that reading fluency is strongly linked to reading comprehension.

A limitation of the current study was that we used a 2011 version of the TEM-4 as this was the only one available to us. A limitation of the analyses was that other variables which are known to affect reading comprehension, such as phonological or morphological awareness, working memory, and non-verbal cognitive abilities were not measured. If these had been included, more variance could have been explained.

Summary and conclusion

The current chapter set out to investigate the validity of the TEM-4 reading and vocabulary components in a study among 60 English major students between the ages of 18 and 22 from a university in Northern China. To investigate the concurrent validity of the test, students also took two vocabulary tests, the Vocabulary Size Test (Nation & Beglar, 2007) and a modified version of the Vocabulary Knowledge Scale (Brown, 2008). In addition, 30 students were administered the YARC reading comprehension Secondary (Snowling et al., 2009). We found that the YARC Secondary was rather difficult for the students in China, but this was not so much related to their smaller vocabularies (as the vocabulary and the readability of the texts showed they were easier than those in the TEM-4). Instead it is likely that students were not familiar with inferential questions and the test format of some of the other components. Nevertheless, the results of the YARC Secondary should not be completely discarded: first of all, for one component (reading fluency), the results were very promising, and there were no floor or ceiling effects in any of the components. Second, there were moderate correlations between reading comprehension and reading fluency, as predicted by Grabe (2010). Finally, there were moderate correlations between reading comprehension

and the VKS, which confirms findings of Qian (2005). The existence of these correlations lends support to the assumption that the YARC Secondary did provide useful information about students' abilities, despite the fact that they found some parts rather difficult.

There were also moderate to strong correlations between the reading component of the TEM-4 on the one hand, and vocabulary size as measured with the VST and vocabulary depth as measured with the VKS on the other hand. It was rather unexpected that the TEM-4 reading component did not correlate with any of the components of the YARC Secondary. This pattern of correlations leads us to the conclusion that the TEM-4 reading component mainly taps into vocabulary knowledge, and there is little evidence that this component measures dimensions of reading as distinguished in the YARC Secondary.

In the final part of our study we looked at whether or not vocabulary size or vocabulary depth explained unique variance in reading as measured with both tests. We found that the VST, the VKS, and the TEM-4 grammar and vocabulary component all explained unique variance in the TEM-4 reading scores. In total, almost 20% of the variance could be explained.

For the YARC Secondary a different model emerged. The VST was not found to be a significant predictor of the YARC Secondary reading fluency component, nor of the reading comprehension component. The VKS, by contrast, was a significant predictor of both fluency and comprehension, and explained on its own between 14 and 15% of the variance. The addition of reading fluency to the model meant that an additional 10% of the variance in comprehension could be explained.

Overall we conclude that, if the reading component of most recent versions of the TEM is similar to the reading component of the TEM-4 we used (which dated from 2011), it is in need of an overhaul. The construct validity of the test should be improved as it is not clear on which model of reading it is built, which is of crucial import in the process of test development (Weir, 2005). The fact that students' inferencing ability is hardly assessed in the test means that it taps virtually only into students' ability to provide answers to literal meanings that are found in the texts. While it is therefore likely that the test underrepresents the construct of reading, a possible way forward in this would be to obtain further clarification about the meaning of inferencing and its role in the TEM-4. The low scores obtained by students on the test are worrying: these could be due to the texts containing many low frequency items and to the complexity of the texts as measured with the readability indices, as our analyses have revealed. However, as the questions are all multiple choice, it is likely that scores are inflated because of guessing. The use of alternative formats, such as a gap filling task which is not based on multiple choice, can help to reduce construct-irrelevant variance (in particular students' ability to guess). In this context it is important to note that the VKS, which does not make use of multiple choice questions, turned out to be a stronger predictor of students' reading ability than the VST, the scores of which may also have

been in part the result of students' guessing (see Gyllstad et al., 2015). While for the current group the YARC worked reasonably well, this does not mean that it can be used with any group of adult L2 learners. Prior to using the test with adult L2 learners, students' vocabulary levels should be carefully checked against the vocabulary that is used in the texts. In addition, staff and students would need to receive training in formulating and answering inferential questions. We hope that the findings of the current study have provided useful information for test developers in China and that it can inform further discussions on the ways in which reading is measured in the TEM-4, and other tests that are widely used in China.

Notes

1 We are very grateful to Rita Green for having provided us with a copy of this report, and to Anthony Zhang and Changqing Zheng for sending us the final version.
2 We are very grateful to Guoxing Yu for providing us with the information regarding the inferencing skills in the test specification for the TEM-4 (2015).

References

Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101–118.

Binder, K. S., Cote, N. G., Lee, C., Bessette, E., & Vu, H. (2017). Beyond breadth: The contributions of vocabulary depth to reading comprehension among skilled readers. *Journal of Research in Reading*, *40*(3), 333–343. doi:10.1111/1467–9817.12069.

Bowyer-Crane, C., & Snowling, M. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology, 75*(2), 189–201.

Brown, D. (2008). Using a modified version of the vocabulary knowledge scale to aid vocabulary development. *The Language Teacher, 32*(12). No page numbers.

Brown, J. A., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny reading test: Manual for scoring and interpretation*, Forms G & H. Rolling Meadows, IL: Riverside Publishing.

Chen, X. (2009). *The validation study of the objective items in TEM4*. Shanghai: Fudan University Press.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education*. Oxford: Routledge.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.

Grabe, W. (2010). Fluency in reading: Thirty-five years later. *Reading in a Foreign Language, 22*(1), 71–83.

Graesser, A., McNamara, D., & Louwerse, M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In Sweet, A. P., & Snow, C. E. (eds) *Rethinking reading comprehension* (pp. 82–98). New York: Guilford Press.

Gyllstad, H., Vilkaitė, L,. & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats. Issues with guessing and sampling rates. *International Journal of Applied Linguistics, 166(*2), 278–306. doi 10.1075/itl.166.2.04gyl

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Arnold.

Jin, Y., & Fan, J. (2011). Test for English Majors (TEM) in China. *Language Testing, 28*(4), 589–596.

Khalifa, H., & Weir, C.J. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge: Cambridge University Press.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436. doi: 10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons'responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741–749. doi: 10.1037/0003-066X.50.9.741

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, *63*(1), 59–82. doi: 10.3138/cmlr.63.1.59

Nation, I. S. P. (2012). The Vocabulary Size Test. www.wgtn.ac.nz/lals/about/staff/Publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf

Nation, P. (2013). *Learning Vocabulary in Another Language*, 2nd ed. Cambridge: Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In Schmitt, N., and McCarthy, M. (eds) *Vocabulary: Description, acquisition and pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.

Nelson, H. E. (1982). *National Adult Reading Test (NART): Test manual.* Windsor: NFER-Nelson.

Paribakht, T. S., & Wesche, M. (1996). Enhancing vocabulary acquisition through reading: A hierarchy of text-related exercise types. *Canadian Modern Language Review, 52*(2), 155–178.

Qian, D. D. (2005). Demystifying lexical inferencing: The role of aspects of vocabulary knowledge. *TESL Canada Journal, 22*(2), 34–54.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, *10*(3), 355–371.

Ready, R. E., Chaudhry, M. F., Schatz, K. C., and Strazullo, S. (2012). 'Passageless' administration of the Nelson-Denny Reading Comprehension Test: Associations with IQ and reading skills. *Journal of Learning Disabilities, 46*(4), 377–384. doi: 10.1177/0022219412468160

Schmitt, N., Jiang, X., & Grabe, W. P. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Learning Journal, 95*(1), 26–43.

Schmitt, N., Nation, P., & Kremmel, B. (2019). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching* (first view). https://doi.org/10.1017/S0261444819000326.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88.

Snow, C. (2002). Reading for *u*nderstanding. *Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Corporation.

Snowling, M. J., Stothard, S. E., Clarke . P., Bowyer-Crane C., Harrington A., Truelove, E., and Hulme, C. (2009). *York assessment of reading for comprehension*. London: GL Assessment.

Stajner, S., Evans, R., Orăsan, C., & Mitkov, R. (2012). What can readability measures really tell us about text complexity? In Rello, L., and Saggion, H. (eds) *Proceedings of the LREC'12 Workshop: Natural language processing for improving textual accessibility* (NLP4ITA). Istanbul: European Language Resources Association (ELRA).

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–406.

Stanovich, K. E. (1991). Word recognition: Changing perspectives. In Barr, R., Kamil, M. L. Mosenthal, P., & Pearson, P. D. (eds) *Handbook of reading research* (vol. 2, pp. 418–452). New York: Longman.

Stothard, S. E. (2010). Identifying reading difficulties in the secondary school years. *Literacy Today*, *64*, 34–35.

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System, 65*, 139–150. doi: 10.1016/j.system.2016.12.0

Weir, C. (2005). *Language testing and validation: An evidence-based approach*. Houndmills: Palgrave Macmillan.

Xu, Q., & Liu, J. (2018). *A study on the washback effects of the Test for English Majors (TEM)*. New York: Springer.

Yu, G. (2008). Reading to summarize in English and Chinese: A tale of two languages? *Language Testing, 25*(4), 521–551.

Yu, G. & Jin, Y. (2014). English language assessment in China: Policies, practices and impacts. *Assessment in Education: Principles, Policy and Practice*, *21*(3), 245–250. doi:10.1080/0969594X.2014.937936

Zhang, B. (2009). FL Vocabulary learning of undergraduate English Majors in Western China: Perspective, strategy use and vocabulary size. *English Language Teaching, 2*(3), 178–185.

Zhao, W., Wang, B., Coniam, D. & Xie, B. (2017). Calibrating the CEFR against the China Standards of English for College English vocabulary education in China. *Language Testing in Asia*, *7*(5). DOI 10.1186/s40468-017-0036-1.

Zhou, S., Weir, C. J., and Green, R. (1998). *The Test for English Majors Validation Project*. Shanghai: Foreign Language Education Press.

APPENDIX A

TEM-4 TEXT B

I know when the snow melts and the first robins (知更鸟) come to call, when the laughter of children returns to the parks and playgrounds, something wonderful is about to happen.

Spring cleaning.

I'll admit *spring cleaning is a difficult notion for modern families to grasp*. Today's busy families hardly have time to load the dishwasher, much less clean

the doormat. Asking the family to spend the weekend collecting winter dog piles from the melting snow in the backyard is like announcing there will be no more Wi-Fi. It interrupts the natural order.

'Honey, what say we spend the weekend beating the rugs, sorting through the boxes in the basement and painting our bedroom a nice lemony yellow?' I say.

'Can we at least wait until the NBA matches are over?' my husband answers.

But I tell my family, *spring cleaning can't wait*. The temperature has risen just enough to melt snow but not enough for Little League practice to start. Some flowers are peeking out of the thawing ground, but there is no lawn to seed, nor garden to tend. Newly wakened from our winter's hibernation (冬眠), yet still needing extra blankets at night, we open our windows to the first fresh air floating on the breeze and all of the natural world demanding 'Awake and be clean!'

Biologists offer a theory about this primal impulse to clean out every drawer and closet in the house at spring's first light, which has to do with melatonin, the sleepytime hormone (激素) our bodies produce when it's dark. When spring's light comes, the melatonin diminishes, and suddenly we are awakened to the dusty, virus-filled house we've been hibernating in for four months.

I tell my family about the science and psychology of a good healthy cleaning at spring's arrival. I speak to them about life's greatest rewards waiting in the removal of soap scum from the bathtub, which hasn't been properly cleaned since the first snowfall.

'I'll do it,' says the eldest child, a 21-year-old college student who lives at home.

'You will? Wow!' I exclaim.

Maybe after all these years, he's finally grasped the concept. Maybe he's expressing his rightful position as eldest child and role model. Or maybe he's going to Florida for a break in a couple of weeks and he's being nice to me who is the financial-aid officer.

No matter. Seeing my adult son willingly cleaning that dirty bathtub gives me hope for the future of his 12-year-old brother who, instead of working, is found to be sleeping in the seat of the window he is supposed to be cleaning.

'Awake and be clean!' I say.

86. According to the passage, '*…spring cleaning is difficult notion for modern families to grasp*' means that spring cleaning

A. is no longer an easy practice to understand.
B. is no longer part of modern family life.
C. requires more family members to be involved.
D. calls for more complicated skills and knowledge.

87. Which of the following is LEAST likely to be included in family spring cleaning?

A. Beating the rugs.
B. Cleaning the window.
C. Restoring Wi-Fi services.
D. Cleaning the backyard.

88. Why does the author say '*spring cleaning can't wait*'?

A. Because there will be more activities when it gets warmer.
B. Because the air is fresher and the breeze is lighter.
C. Because the whole family is full of energy at spring time.
D. Because the snow is melting and the ground is thawing.

89. Which of the following interpretations of the biologists' theory about melatonin is INCORRECT?

A. The production of melatonin in our bodies varies at different times.
B. Melatonin is more likely to cause sleepiness in our bodies.
C. The reduction of melatonin will cause wakefulness in our bodies.
D. The amount of melatonin remains constant in our bodies.

90. Which of the following can best sum up the author's overall reaction to her adult son's positive response to spring cleaning?

A. Surprised and skeptical.
B. Elated and hesitant.
C. Relieved and optimistic.
D. Optimistic and hesitant

APPENDIX B: WORDS RANDOMLY SELECTED FROM THE VST

accessory
allege
compost
compound
deficit
devious
drawer
drive
hallmark
haunt
jug
latter
maintain
olive

soldier
standard
strangle
threshold
upset
yoghurt

9 Vocabulary and reading

Future research, tools, and practices

Irina Elgort

In this brief chapter, I take another look at the construct of *lexical quality* that is required influent reading and put forward ideas for contextual vocabulary learning research projects that can be carried out by teacher-practitioners and applied linguistics researchers alike.

The *Lexical Quality Hypothesis* (Perfetti, 2007; Perfetti & Hart, 2001) proposes that accurate and fluent access to lexical knowledge during reading depends on the tight binding between orthographic, phonological, and semantic representations and that this binding is, in turn, predicated on the quality of each of these component representations. In L1 acquisition, high lexical quality can be achieved through continuous fine-tuning of lexical representations during exposure to written input, partly because many words are already familiar to the reader in their spoken form. Because L2 acquisition often starts later in life, readers' existing lexical knowledge (whether spoken or written) is limited in both quality and quantity, and the amount and quality of input tend to be inferior to those in the L1, the development of lexical quality from input is slower and less effective. Therefore, L2 contextual vocabulary learning needs to be augmented by instructional and learning activities that facilitate the fine-tuning of L2 orthographic and semantic representations needed in reading. This is not to say that deliberate vocabulary learning is sufficient; L2 lexical quality likely arises from the combination of contextual vocabulary learning, which affords implicit learning from input, and deliberate learning activities, aimed at developing more precise knowledge of the written form and more robust form-meaning mapping.

Three approaches to augmenting contextual word learning, presented in Chapter 7, that had a positive effect on the development of lexical quality were: handwriting, dictionary looking up, and flashcards. These approaches can be combined in practice. Let us say, a language learner is reading some articles or short stories of interest. When she comes across an unfamiliar word, she writes it down in a vocabulary notebook; when she sees this word again in the text, she tries to infer its meaning from context. At the end of the reading session, she looks up the meaning of each of the words in her notebook (and checks their pronunciation if she is using an online dictionary). She then creates flashcards for each of the words. Since learning new words is

something language learners are generally keen on, this process can be mostly driven by intrinsic motivation. The effectiveness of this approach could be evaluated in a longitudinal in-situ study by a language teacher or an applied linguistics researcher. When measuring lexical quality of the contextually learned vocabulary in such a study, it is important to use tasks that probe *nondeclarative* knowledge of the orthographic and semantic representations.

In formal teaching, one way of augmenting contextual word learning is by adding post-reading activities, such as book reports, small-group discussions, worksheets, and vocabulary notebooks (Beglar et al., 2012; Horst, 2005; Yamashita, 2008). Recently, Boutorwick, Macalister, and Elgort (2019) examined vocabulary learning under two extensive reading conditions: a traditional reading-only method and a reading-plus method, in which L2 readers engaged in a *Say-it* activity (Macalister, 2014) after each graded reader. In this activity, learners form triads and take turns in discussing characters and events from the story they've read, using a set of prompts. Vocabulary development was measured using a pre- and post-reading word association task, in which participants provided up to five associations for 60 target words (20 in high-, mid-, and low-frequency bands each) that occurred in the graded readers. The semantic relevance (distance) of the associations provided by the participants was established using the Latent Semantic Analysis (LSA) approach, a statistical method that represents a word's meaning as a sum of all of contexts in which it does and does not occur (Landauer & Dumais, 1997). The difference between the LSA similarity indices calculated in the pre- and post-reading word association tasks were compared for the reading-only and reading-plus treatments, in order to determine which treatment results in higher-quality meaning associations. Boutorwick et al. (2019) found that, although gains in the knowledge of meaning were about the same in the two conditions, when an analysis was conducted on the words that had received attention in the Say-it activity, the reading-plus treatment resulted in significantly greater gains, compared to the reading-only treatment, for the mid-frequency vocabulary (i.e., words just outside of the readers' L2 vocabulary knowledge). This suggests that learner-learner interactions in which they discuss recently read stories may be helpful in promoting higher quality meaning representations of contextually learned words, as long as discussion prompts are designed to encourage the use of mid- and low-frequency words that occur in the books. There is no evidence at this stage, however, that such post-reading activities promote the development of nondeclarative knowledge. This is a task for future research studies.

Beyond individual words, contextual learning of *formulaic language* (a cover term for different types of multiword expressions) presents an even greater challenge for L2 readers. Even high-proficiency L2 learners have trouble with formulaic language, particularly in production (e.g., Laufer & Waldman, 2011; Levitzky-Aviad & Laufer, 2013). One of the main obstacles to the acquisition of multiword expressions from reading is their lack of salience in the written input. A number of learning interventions have been

proposed, in order to draw learners' attention to multiword expressions, increasing learners' chances to notice them explicitly in the written input. This is commonly done by means of typographic enhancement, such as bolding or underlining of the target phrases in the input, thus, creating an externally induced focus on the multiword expressions. Typographic enhancement has been shown to improve explicit, declarative knowledge of target expressions (Boers et al., 2006; Boers et al., 2017; Choi, 2017; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016). However, this approach may interfere with contextual learning of the non-enhanced vocabulary that occurs alongside highlighted expressions (Boers et al., 2017; Choi, 2017) and it does not appear to improve the nondeclarative knowledge of these expressions (Sonbul & Schmitt, 2013; Toomer & Elgort, 2019). Because readers overtly allocate attention to typographically-enhanced multiword expressions, they are likely to encode this knowledge into their declarative memory, gaining explicit knowledge of the target expressions but not their tacit, nondeclarative knowledge that arises as a consequence of implicit (covert) contextual learning in the course of continuous (uninterrupted) reading of a connected text (Reber, 2008, 2013; Ullman & Lovelett, 2018).

An alternative approach that does seem to positively affect the development of nondeclarative knowledge of formulaic language is text *enrichment*, i.e., increasing the likelihood of repeated encounters with target multiword expressions in supportive contexts. Similar to contextual learning of individual words, frequency of encounters with multiword expressions is one of the strongest factors in their contextual learning and development of nondeclarative, tacit knowledge that eventually leads to their faster processing during reading (Conklin & Schmitt, 2008). In a recent study, Toomer and Elgort (2019) found that increasing the number of encounters with the target L2 lexical collocations (to nine instances over two days) resulted in a positive collocational priming effect indicative of the development of their nondeclarative knowledge. Because the likelihood of high density of occurrence of the same multiword expression in reading texts or language textbooks is not very high (Pellicer-Sánchez & Boers, 2019), these findings suggest that teachers, textbook developers, and graded reader writers could create favourable conditions for the acquisition of L2 collocations by intentionally increasing their rate of occurrence in L2 teaching and learning materials. To test this hypothesis and extend the initial evidence from Toomer and Elgort's (2019) to longer, less contrived L2 reading texts, future research should evaluate the effect of enriching graded readers (or similar level-appropriate texts) with repeated instanced of multiword expressions (e.g., lexical and grammatical collocations) and trace the development of L2 readers' nondeclarative collocational knowledge over time.

In summary, in this brief chapter, I have proposed three L2 contextual vocabulary learning research projects focused on the development of lexical quality: (1) augmenting contextual learning from reading with an intentional learning procedure comprising handwriting, contextual meaning inferences

followed by dictionary lookup, and the use of flashcards; (2) supplementing extensive reading with post-reading activities that necessitate productive use of contextually learned vocabulary; and (3) enriching L2 reading materials with repeated instances of multiword expressions. The effectiveness of each of these interventions must be evaluated using measures of not only declarative but also nondeclarative knowledge of form and meaning, such as tests that require online access to lexical knowledge, under time pressure and under conditions that minimize the use of explicit task/test strategies.

References

Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese university EFL learners' reading rates. *Language Learning*, *62*(3), 665–703. https://doi.org/10.1111/j.1467-9922.2011.00651.x

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, *27*(2), 448–469. https://doi.org/10.1111/ijal.1214

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, *10*(3), 245–261. https://doi.org/10.1191/1362168806lr195oa

Boutorwick, T. J., Macalister, J., & Elgort, I. (2019). Two approaches to extensive reading and their effects on second language vocabulary development. *Reading in a Foreign Language*, *331*(2), 15–172.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3), 403–426.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review*, *61*(3), 355–382. https://doi.org/10.3138/cmlr.61.3.355

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*, 647–672. https://doi.org/10.111/j.1467-9922.2010.00621.x

Levitzky-Aviad, T., & Laufer, B. (2013). Lexical properties in the writing of foreign language learners over eight years of study: Single words and collocations. *EUROSLA Monograph Series*, *2*, 127–148.

Macalister, J. (2014). The say-it activity. *Modern English Teacher*, *23*(1), 29–32.

Pellicer-Sánchez, A., & Boers, F. (2019). Pedagogical approaches to the teaching and learning of formulaic language. In Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (eds) *Understanding formulaic language* (chapter 8). New York: Routledge.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383.

Perfetti, C. A., & Hart, L. (2001). The lexical bases of comprehension skill. In Gorfien, D. (ed.), *On the consequences of meaning selection* (pp. 67–86). Washington, DC: American Psychological Association.

Reber, P. J. (2008). Cognitive neuroscience of declarative and nondeclarative memory. In Aaron, S., Benjamin, A. S., De Belle, J. S., Etnyre, B., & Polk, T. A. (eds) *Advances in Psychology*, *139* (pp. 113–123). Amsterdam: North-Holland.

Reber, P. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia*, *51*, 2026–2042. https://doi.org/10.1016/j.neuropsychologia.2013.06.019

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*(2). https://doi.org/10.1111/ijal.12092

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121–159.

Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, *69*(2), 405–439. https://doi.org/10.1111/lang.12335

Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, *34*, 39–65. https://doi.org/10.1177/0267658316675195

Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System*, *36*, 661–672. https://doi.org/10.1016/j.system.2008.04.003

Part IV
Speaking

10 Vocabulary and speaking

Current research, tools, and practices

Takumi Uchihara

The relationship between vocabulary and speaking has received recent attention in L2 research (e.g., Koizumi & In'nami, 2013; Kyle & Crossley, 2015), the area which used to be neglected in comparison to, for example, the relationship between vocabulary and reading (Uchihara & Saito, 2019). Researchers generally take two approaches when investigating the vocabulary-speaking link. The first approach is to elicit speech samples through oral tasks (e.g., oral narrative), assess the samples holistically (e.g., native judgements of communicative adequacy), score the same samples lexically (e.g., number of infrequent words), and examine the extent to which multiple lexical measures predict general speaking proficiency using multivariate statistical analysis (e.g., Kyle & Crossley, 2015; Saito et al., 2016). The second approach, unlike the first, elicits vocabulary and speaking data separately, and therefore the two elicited samples are not dependent. Focusing on studies taking the second approach, this chapter reviews research on vocabulary and four aspects of L2 speech – fluency, lexical richness, pronunciation, and global features – and introduces tools commonly used for measuring oral proficiency in this field. Due to space limitation, this chapter will focus exclusively on reviewing and discussing speaking measurement (for those who are interested in tools for measuring vocabulary knowledge, refer to Read, 2000; Schmitt, 2010).

Current research on vocabulary and speaking

Oral fluency is probably the most extensively researched aspect of L2 speech in relation to vocabulary knowledge (e.g., De Jong et al., 2013; Hilton, 2008; Koizumi & In'nami, 2013; Uchihara & Saito, 2019; Chapter 12, this volume). From a theoretical standpoint, the speech production model posits that speaking is lexically driven, to the extent that learners with rich lexicons are hypothesized to retrieve lemmas efficiently, making their overall speech production fast (Kormos, 2006). Research supports this view, as studies consistently report medium-to-large correlations between vocabulary knowledge and fluency – particularly, with speed fluency (e.g., articulation rate, mean length of run), r = .34 to .67 (De Jong et al., 2013; Hilton, 2008; Uchihara & Saito, 2019).

Exploration of lexical richness in spoken responses is another way of evaluating L2 speaking proficiency. Lexical richness is commonly defined as lexical sophistication and diversity (see Kyle, 2019 and Read, 2000 for discussion of the construct), and the former is often measured using word frequency information and the latter is measured with a simple type token ratio or advanced variants of it (Kyle, 2019). In principle, learners producing more low frequency words and fewer repetitions are more proficient than those who do the opposite. Recent studies examine the relationship between vocabulary knowledge (size test scores) and use (lexical richness) in order to test the hypothesis that learners with rich lexical knowledge show lexically rich language use in speech (Uchihara & Clenton, 2018; Chapter 12, this volume). Their findings show significant correlations between the two, but a closer examination of their data suggest a complexity of the relationship, indicating that speakers with larger vocabulary sizes might not necessarily produce lexical richer words.

Pronunciation, perhaps receiving the least attention in vocabulary research, is also one of the important aspects of speaking proficiency. Although our understanding of the vocabulary-pronunciation link is limited, research has begun to shed some light on the role that vocabulary plays in phonological development. Bundgaard-Nielsen et al. (2011) suggest that learners with larger vocabulary sizes have finer-tuned phonological representations, enabling more accurate phonological perception. In Uchihara and Saito (2019), however, vocabulary knowledge did not seem relevant to the ability to pronounce the L2 in a target-like manner. More research is needed to advance our understanding of how vocabulary knowledge relates to pronunciation ability at both perception and production levels.

Finally, some studies examine vocabulary knowledge and global aspects of L2 speech, in addition to specific oral aspects (e.g., fluency, lexical richness, and pronunciation). The global construct includes perceived comprehensibility (i.e., ease of understanding; Saito et al., 2016) and communicative/functional adequacy (i.e., success of task achievement; De Jong et al., 2012). Research indicates some indirect relationship between vocabulary knowledge and comprehensibility (Uchihara & Saito, 2019; Chapter 12, this volume), whereas De Jong et al. (2012) suggest that learners with large vocabulary sizes are more likely to complete oral tasks successfully.

Tools and practices for measuring speaking proficiency

Given the many different approaches and tools to assess L2 speaking proficiency, this section focuses mainly on the tools and practices that are commonly employed in studies examining the relationship between vocabulary knowledge and speaking. Approaches to assessing L2 speech are broadly divided into the following two: human rating and objective measurement employing acoustic and corpus-based analysis tools.

Human rating

One typical approach to assessing L2 speech is to employ listener judgements. After speech samples are elicited through oral tasks (e.g., picture narrative, interview), trained or untrained raters listen to each of the elicited samples and rate these based on numerical scale points while referring to holistic or analytic language descriptors. By way of an example of holistic rating, in De Jong et al. (2012), four native speaking non-expert raters were selected to evaluate speech data on the communicative/functional adequacy of the oral responses. Recruiting untrained raters was an important decision in their study because the researchers did not want their raters to pay special attention to specific linguistic errors (e.g., lexical and grammatical errors). As for analytic rating, on the other hand, raters are encouraged to attend to specific linguistic features. For instance, in Uchihara and Saito (2019), five native speaking raters were trained to refer to a fluency descriptor stating various kinds of temporal information, including the number of filled/silent pauses and repetition. In Uchihara and Clenton (2018), three native speaking expert raters were instructed to focus on lexical use of L2 speakers in reference to the vocabulary component of the IELTS speaking band descriptors containing various vocabulary-related information.

Objective measures

Another approach to L2 speech assessment is quantifying linguistic features in question by means of, for example, counting the number of lexical, grammatical, and phonological errors. This approach provides insight into learners' L2 use objectively, rather than relying on human rating. For instance, De Jong et al. (2013) used acoustic analysis tools, PRAAT (Boersma & Weenik, 2013), and calculated number of silent pauses, total duration of speaking time, and total duration of pausing time for the purpose of assessing oral fluency. For measuring lexical sophistication, Uchihara and Clenton (2018) used a corpus-based lexical analysis tool, TAALES (Kyle & Crossley, 2015), in order to calculate average frequency scores given to individual words used in oral responses per speaker. For measuring lexical diversity, Uchihara, Saito, and Clenton (Chapter 12, this volume) adopted the measure of textual lexical diversity (MTLD) automatically produced by a text analysis tool, Coh-Mertix (McNamara et al., 2014).

Conclusion

This chapter has provided a brief overview of research on vocabulary and speaking, and introduced existing tools and practices adopted for measuring speaking proficiency in this area of research. Readers should note that the information provided in this chapter is far from exhaustive and many additional speaking measures are available to examine the vocabulary-speaking

relationship, such as syntactic complexity (e.g., number of clauses per speech unit; Koizumi & In'nami, 2013), discourse competence (e.g., number of cohesive devices; Saito et al., 2017), and pronunciation accuracy (e.g., vowel and consonant errors; Suzuki & Kormos, 2019). In order to advance our understanding of the relationship between vocabulary and speaking, it is important to investigate the relative contribution of learners' vocabulary knowledge to different aspects of L2 oral proficiency.

Further reading and useful information

Kyle, K. (2019). Measuring lexical richness. In Webb, S. (ed.) *The Routledge handbook of vocabulary studies* (pp. 454–476). London: Routledge.

This chapter provides a review of lexical richness measures focusing specifically on lexical sophistication and diversity with examples and introduction of analysis tools. For instance, the author introduced the Tool for the Automated Analysis of Lexical Sophistication (TAALES) and the Tool for Automatic Analysis of Lexical Diversity (TAALED), both tools available at Kristopher Kyle's website (www.kristopherkyle.com).

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439–462.

This article provides insight into multifaceted aspects of oral proficiency measured globally (e.g., comprehensibility) and analytically (e.g., temporal, lexical, and phonological features), using various speech analysis techniques, such as human rating, acoustic, and corpus-based measures.

References

Boersma, P., & Weenink, D. (2013). *PRAAT: Doing phonetics by computer* (Version 5.3.39). Available: www.praat.org

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*(1), 51–67.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*, 5–34.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, *36*(2), 153–166.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, *4*(5), 900–913.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kyle, K. (2019). Measuring lexical richness. In Webb, S. (ed.) *The Routledge handbook of vocabulary studies* (pp. 454–476). London: Routledge.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757–786.

McNamara, D. S., Grasesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. New York: Cambridge University Press.

Read, J. (2000). *Assessing vocabulary.* Cambridge: Cambridge University Press.

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439–462.

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations. *Studies in Second Language Acquisition*, *38*(4), 677–701.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan.

Suzuki, S., & Kormos, J. (2019). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 251. doi:10.1017/S0272263119000627

Uchihara, T., & Clenton, J. (2018). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*.

Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *Language Learning Journal*, *47*(1), 64–75.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, *17*(1), 65–83.

11 Investigating the extent to which vocabulary knowledge and skills can predict aspects of fluency for a small group of pre-intermediate Japanese L1 users of English (L2)

Jon Clenton, Nivja H. de Jong, Dion Clingwall and Simon Fraser

Introduction

The words second language speakers choose to use when speaking may have consequences for their speaking fluency (e.g. Seifart et al., 2018). A number of studies (e.g. De Jong et al., 2013; De Jong & Mora, 2017; Miralpeix & Muñoz, 2018; Milton et al., 2010; Segalowitz & Freed, 2004; Uchihara & Saito, 2016) explore the ways in which the relation between vocabulary knowledge and fluent speech can be evaluated objectively. Such evaluation is important because of the variety and volume of second language speakers, especially of English, whose fluency needs to be assessed, with the importance vocabulary plays in such assessment being absolutely central: 'while without grammar very little can be conveyed, without vocabulary nothing can be conveyed' (Wilkins, 1972: 111–12). Measures of vocabulary knowledge and fluency provide stakeholders, such as those involved in research, pedagogy, and assessment, with essential information to discriminate between users of second languages and their respective proficiency levels. Much research, therefore, is designed to explore the specific features necessary to distinguish between second language users with different levels of language ability. The study we report here adds to this body of research by examining the relationship between the vocabulary knowledge of pre-intermediate Japanese learners of English and their oral fluency.

In this chapter, then, we present a small-scale study in which we employ various vocabulary knowledge tasks as well as fluency elicitation tasks. We compare the results from a number of elicitation tasks not conventionally employed together in the hope not only that this combination of tasks is better suited to the users whose second language we measure, but also that the findings are informative in our investigation of the ways in which vocabulary knowledge relates to aspects of second language fluency.

Before we turn to describing how we understand the term fluency, we briefly outline how we approach the vocabulary knowledge investigated in

this study. As we go on to show, a number of fluency papers (e.g. De Jong et al., 2013) report significant relationships between vocabulary knowledge and aspects of fluency. Such papers elicit learner knowledge from vocabulary tasks alongside fluency tasks and then report that, for instance, learners with specific vocabulary knowledge of X items consistently demonstrate particular aspects of fluency. In this study, we add an additional component to our investigation, one, we believe, that is both novel and unique. In addition to adopting the same approach as referenced above in the current chapter by reporting relationships between vocabulary knowledge and aspects of fluency, we compare the vocabulary used in response both to the vocabulary tasks and the fluency elicitation tasks. We attempt to go to the heart of the vocabulary knowledge of our subject population, and we refer to a recent approach (Fitzpatrick & Clenton, 2017) that supports this investigation.

Before we begin by measuring vocabulary and fluency, we need to detail what is currently meant by the construct of 'vocabulary knowledge'. The construct of vocabulary knowledge is far from straightforward, as Fitzpatrick and Clenton (2017: 844–5) point out, because the 'simplicity' of vocabulary task scores is inconsistent with the multitude of interpretations possible to the extent that 'subtle and informed interpretation is required'. Fitzpatrick and Clenton raise several related concerns when highlighting such complexity. The first pertains to vocabulary measures (e.g. the Vocabulary Levels Test (VLT), Nation, 1983; the Productive Levels Test (PVLT), Laufer & Nation, 1999) that conventionally base elicitation on the assumption that the difficulty of vocabulary items generally relates to their frequency of occurrence in corpora. Second, they highlight the commonly shared view (e.g. Nation 2001, 2013; Read, 2000; Webb, 2009) that vocabulary knowledge is 'multidimensional'. Fitzpatrick (2007) makes this point clearly when comparing three vocabulary tasks according to a revised version of Nation's (1990) 'aspects of word knowledge', indicating that despite being designed to elicit the same, all three tasks test different aspects of the construct. Fitzpatrick and Clenton suggest that Nation's (2001, 2013) 'aspects of word knowledge' is important in this discussion, because it attempts to list the multiple aspects of word knowledge. Fitzpatrick and Clenton add that the complexity of the construct includes a range of factors including those related to the way words are 'organized in the mental lexicon (Meara, 1996), and related to this, speed and, ultimately automaticity of retrieval' (Qian, 2002: 846). The current chapter reports an attempted analysis of the final two of this 'list' of factors: speed and automaticity of retrieval. We return to this specific question in our research questions below. A further concern relates to a distinction commonly made when discussing vocabulary knowledge, of that between productive and receptive knowledge.

'Productive' and 'receptive' are widely-used terms that appear to have gathered currency within the field of vocabulary research. Such terms, however, might need reconsidering in light of suggestions (e.g. Fitzpatrick, 2010; Fitzpatrick & Clenton, 2017) that elicitation relates to the aspects of

vocabulary knowledge measured. Productive vocabulary tasks (e.g. the PVLT; Lex30, Meara & Fitzpatrick; G_Lex, Fitzpatrick & Clenton, 2017; The Lexical Frequency Profile, Laufer & Nation; 1995; The Vocabulary Knowledge Scale, Paribakht & Wesche, 1996) broadly relate to spoken or written output, and receptive vocabulary tasks (e.g. the EVST, Meara & Jones, 1987; the VLT; XY_Lex, Meara & Miralpeix, 2016) broadly relate to reading and listening.

Fluency and vocabulary knowledge

For this study we consider fluency in the narrow sense as opposed to broad sense (Lennon, 1990). While the broad sense of fluency appears to relate to overall or global proficiency, the narrow sense of fluency (for diagnostic purposes) 'refer(s) to one, presumably isolatable, component of oral proficiency' (p. 389). Within narrow fluency, fluency is often measured as a component of speech with multiple aspects, referring to quick and perhaps smooth delivery of speech with or without filled or unfilled pauses, repetitions, and repairs. Some researchers, such as Skehan (2003) and Tavakoli and Skehan (2005), suggest fluency can therefore be measured according to three main characteristics: (i) breakdown fluency (referring to how often speech 'breaks down', or the number of pauses); (ii) speed fluency (referring to the speed of speech between these pauses, therefore articulation rate); and, (iii) repair fluency (referring to the number of times a speaker recognizes and repairs speech). The Common European Framework of Reference (CEFR; Council of Europe, 2001), for instance, refers to proficient users (e.g. C2) as being able to express themselves 'spontaneously, very fluently, differentiating finer shades of meaning even in more complex situations' (p. 5); independent users (e.g. B2) as being able to 'interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party' (p. 5); and basic users (e.g. A2) as being able to 'make themselves understood in short contributions, even though pauses, false starts and reformulations are very evident' (p. 31).

Several papers on fluency (e.g. De Jong et al., 2013; Segalowitz & Freed, 2004; Uchihara & Saito, 2016) have shown strong and significant correlations between fluency measures and productive vocabulary knowledge. De Jong et al. (2013) explore fluency according to several fluency characteristics and report strong and significant correlations between fluency and a newly constructed Dutch version of a sentence completion task (The Productive Vocabulary Levels Test (PVLT), Laufer & Nation, 1999). Their 'intermediate to advanced level' proficiency participants (learners of Dutch as an L2) responded to a variety of tasks, with the study designed to explore linguistic skills and speaking fluency. Fluency was measured from speaking performances in which participants were required to respond to eight computer-administered, semi-spontaneous speaking tasks ranging in terms of complexity, formality, and discourse type. De Jong et al. report a number of limited strength but nevertheless significant (and negative) correlations between productive vocabulary

knowledge and the different fluency measures (silent pauses (r=-0.39), filled pauses (r=-0.33), corrections (r=-0.43), repetitions (r=-0.24), and mean syllable duration (r=-0.58)): participants with higher vocabulary scores tend to produce fewer hesitations, pauses, and a lower mean duration of syllables. More recently, in their study of a participant group whose proficiency was widely varied, Uchihara and Saito (2016) found that fluency, as measured by 'optimal speech rate' (Saito et al. 2015, 2016), moderately predicts productive vocabulary task (Lex30) scores (r=0.34). They compared their fluency analyses with Lex30 task scores on the basis that scores have been shown to be 'representative of [each subject's] productive mental lexicon' (Fitzpatrick & Clenton, 2010: 548). Saito et al.'s oral ability measures required participants to respond to a timed picture description task. Their raters then judged optimal speech rate according to temporal information such as proportion of un/filled pauses and mean length of pauses (Derwing et al., 2004). Taken together, these two studies (De Jong et al., 2013; Uchihara and Saito, 2016) suggest that fluency relates to productive vocabulary knowledge and more broadly to vocabulary skills. The reported fluency measures involved pausing and syllable duration (articulation rate inversed), as well as speech rate. The vocabulary measures consisted of productive vocabulary measures, lexical access speed, and lexical access efficiency. The fluency studies reported in this section appear to broadly reflect a relationship between fluency and productive vocabulary knowledge. Fluent speech appears to correlate with vocabulary task scores to the extent that a second language speaker with a higher vocabulary score hesitates and pauses less and produces a lower mean syllable duration (i.e., a higher articulation rate).

Comparisons between aspects of fluency and receptive vocabulary measures (e.g. De Jong & Mora, 2017; Miralpeix & Muñoz, 2018; Milton et al., 2010), however, are somewhat less consistent. De Jong and Mora (2017) used three of the same speaking tasks as those from earlier studies (De Jong et al., 2013, 2015) and compared data with XY_Lex vocabulary size measures (Meara & Miralpeix, 2016). Their upper-intermediate to advanced proficiency subject vocabulary size scores (M= 6144, *Range* =3350–8200) were shown to correlate moderately significantly (r=-.311) with one aspect of fluency (mean syllable duration) but not with other fluency measures. Miralpeix and Muñoz (2018) investigated the relationships between vocabulary size (using XY_Lex vocabulary measures) with reading, writing, listening, and speaking measures. Their upper-intermediate proficiency subject vocabulary size scores (M=5127, *Range*=2500–7200) were shown to correlate moderately significantly with oral fluency (r=.485). In a comparison of two versions of Yes/No (aural and written) tasks, Milton et al. (2010) compared the vocabulary size (M=2844) with the IELTS tasks designed to elicit knowledge of the four skills (reading, writing, listening, and speaking). While Milton et al. did not find a significant correlation between the written form of the Yes/No task and the speaking scores, their study found a significant correlation between the aural form of the Yes/No task (r=.71) and the

speaking scores. These findings indicate that tasks designed to elicit a specific skill are sensitive to the mode of elicitation, which is supported by the fact that the correlations between X_Lex and reading and writing scores were very similar (r=.70 and r=.76).

The construct of productive vocabulary knowledge

In their discussion of *productive* vocabulary knowledge tasks, Fitzpatrick and Clenton (2017) point to different elicitation tasks eliciting different mean proportions of infrequent items. They suggest that mean score differences relate to different tasks not tapping into the same qualities of word knowledge and therefore '[not sampling] the learner lexicon in the same way' (p. 858). Fitzpatrick and Clenton (2017) have devised a 'Vocabulary Test Capture Model' (see pp. 859–61 for details) in which they adapt a model (the Vocabulary Knowledge Scale, VKS; Paribakht & Wesche, 1993, 1997) originally designed to rank learner knowledge of individual items. The vertical dimension of this scale relates to the nature of the task, to the extent that words are produced in response to a specific task if learner knowledge relates to the four levels. Accordingly, items produced in response to the Lex30 task, which elicits single word responses, might relate to learner knowledge of all of the four levels. Word knowledge would likely be populated with highly frequent items for lower proficiency learners, with an emerging lexicon exhibited by progression through the vertical levels. The horizontal dimension relates to what learners have the capacity to produce in response to each elicitation task. Lex30 activates a different semantic field for each of its 30 cues. On the basis that this chapter discusses fluency studies that have used both Lex30 and the PVLT as their productive vocabulary measures, we compare the PVLT with Lex30 in this section. The findings we present here are from a recent study (Clenton, Elmetaher, and Uchihara, 2019) which reports the different proportions of infrequent items each task elicits (n=107; Lex30 score = 18.41 (SD –10.66); PVLT score = 12.41 (SD –5.83) and show that scores on the two tasks correlate moderately significantly (r = .575, p <.01), to the extent that the capture map might better explain differences between these tasks. Therefore, a task such as the PVLT with its 18 elicitation sentence gaps over its five levels might indicate a less broad capture zone, in contrast to Lex30. Fitzpatrick and Clenton (2017) also suggest that their long arrows (see Figure 11.1) indicate the multiple dip activation events by which responses to tasks such as Lex30 require learners to repeatedly return to the 'same subset of lexical resource, pulling out consecutive items that are closely related' (p. 862); this 'same subset of lexical resource' is not available to PVLT task takers. Correct responses to the PVLT might indicate that participants can demonstrate semantic as well as grammatical mastery of their vocabulary knowledge, indicated by levels 3 and 4 in the model. Figure 11.1 shows a revised vocabulary test capture model, serving to highlight task differences and reasons behind those differences.

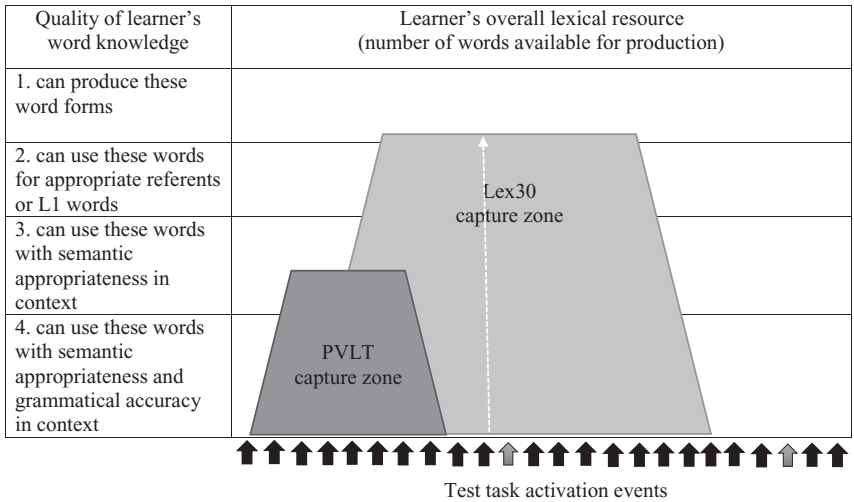


Figure 11.1 A revised vocabulary test capture model: Lex30 and the PVLT.

The study

Aims and research questions

Our main aim is to explore the potential relationships between the knowledge elicited from a productive vocabulary knowledge task and the aspects of fluency elicited from speaking (fluency) tasks. Our second aim is to compare the vocabulary produced in response to the productive vocabulary knowledge task with the vocabulary produced in response to the speaking (fluency) tasks. We also intend to explore findings from earlier papers on fluency in two additional respects, by: (i) comparing receptive knowledge with aspects of fluency; and (ii) exploring the speed and retrieval automaticity, and so including response latency and response duration measures in picture naming tasks in the investigation. The current study, therefore, focuses on the following four questions:

1. Can productive vocabulary knowledge task scores predict aspects of speaking fluency?
2. Can receptive vocabulary knowledge task scores predict aspects of speaking fluency?
3. To what extent do vocabulary skill measures (e.g. response latency and response duration in picture naming tasks) predict aspects of fluency?
4. Is there an overlap between vocabulary used in response to the productive vocabulary task and the vocabulary used in the speaking fluency task?

Methodology

Participants

The participants in the study were 30 pre-intermediate undergraduate adult L1 Japanese learners of English (*M* age = 19, *SD* = 1.3) with an average of 6.5 years' experience of learning English in a school environment; learners had received L2 English instruction for approximately three to four hours a week from L1 Japanese teachers in Japan. They did not use English regularly outside of the learning context. Their X_Lex scores (*M*= 4048, *Range*=2400–4800) also indicated they were of a pre-intermediate proficiency.

Speaking tasks

We chose three speaking tasks from those employed in De Jong et al. (2013), which varied in terms of their task demands: a formal descriptive task (describing a crime scene to a policeman); a formal persuasive task (responding in a town hall meeting to whether a new casino should be built next to an elementary school); and an informal persuasive task (responding to a view on climate change). All tasks were completed on a personal computer. All participants were required to prepare a response and then speak the response aloud. All outputs were recorded. The recordings were subsequently transcribed and analysed using PRAAT (Boersma & Weenink, 2005). The participants were instructed to complete the tasks themselves, and to follow the directions presented on the computer screen. Each task began by presenting participants with a detailed explanation of the situation. Participants were asked to imagine they were speaking for the situation presented. Participants then had a 30-second period within which to prepare their response, indicated by a colour time bar at the bottom of the screen. At the beginning of each task, this coloured bar indicated a time period of two minutes, with the approaching deadline indicated by changing colours, requiring participants to provide their response within the given time.

To measure fluency, using PRAAT (Boersma & Weenink, 2005), syllables were counted manually for all participants. The threshold for a silent pause was set to 350 ms (as De Jong, 2012), and silent pauses were measured manually. All instances of sounds uttered such as *ehh, uhh, mm*, and *umm* were indicated and counted as filled pauses. Similarly, repetitions and repairs were counted manually. All measures were collated over the three tasks. Subsequently, articulation rate was calculated per second of speaking time (total time minus total silent pausing time). Following De Jong and Mora (2017), for all fluency measures indicating hesitations, the total counts were normalized per second total speaking time. Finally, mean silent pause durations for each participant were calculated.

Vocabulary skills tasks

Picture naming: measuring lexical retrieval speed

The same task was used as in De Jong et al. (2013) and De Jong and Mora (2017). From the picture set produced by Snodgrass and Vanderwart (1980), we selected 35 pictures of items all participants were expected to know (i.e. these were highly frequent items). E-Prime was used to present the pictures, one by one. Before the experiment proper commenced, participants were familiarized with the pictures and their names. In this first round, a fixation cross was presented in the middle of the screen for 1000 ms, after which a picture appeared in the centre of the screen, and after yet another 2000 ms, its name was presented underneath the picture. Participants would press the space bar to proceed to the next picture. In the second round, after familiarization, participants were instructed to name (i.e. speak out and name) the pictures as fast and as accurately as possible. In this second round, first, a fixation cross was presented in the middle of the screen for 1500 ms. Then the picture appeared, which was presented for 2000 ms. After the picture, a blank screen followed for 500 ms. The pictures were presented in a random order identical for all participants (but in a different order from the first familiarization round). The time between the appearance of the picture and the beginning of the response was measured manually with the use of PRAAT. Per participant, the mean of all correct responses was used as the measure of lexical retrieval.

Delayed picture naming task: measuring speed of articulation

The materials and apparatus were the same as the ones used for the lexical retrieval measure (picture naming). Following the same picture naming procedure, participants carried out the picture naming task once more. This time, however, they were asked to prepare their response to naming a picture but wait with the actual naming of the picture until the cue was given. A fixation cross was presented in the middle of the screen for 500 ms. Then the picture appeared and remained on the screen for 2000 ms. After 2000 ms, the participant heard a short beep, and a green frame appeared on the screen around the picture. The beep together with the green frame formed the cue for participants to give their response. The picture (with the green frame) remained on the screen for another 1000 ms, during which time the participants responded. The pictures were presented in a random order identical for all participants, but in a different order from the procedures for familiarization and lexical retrieval speed. The experimenter noted incorrect responses and other deviations from the intended responses. Response latency was measured as the latency between the auditory cue and the beginning of the response. Response duration was measured as the duration of the response, i.e. the latency between the beginning and the end of the response.

Both duration measures were measured manually with the use of PRAAT. For each participant, the mean of all correct responses was calculated for both response latency and response duration.

Vocabulary knowledge tasks

Unlike the earlier De Jong et al. (2013) study, which used a Dutch version of Laufer and Nation's Productive Levels Test, we decided to use Lex30 as our productive vocabulary task. We chose Lex30 for four main reasons: (i) Lex30 task scores have been shown to relate to fluency measures (Uchihara & Saito, 2016); (ii) the scores are 'more aligned to the ability to "use" words compared to the PVLT' (Clenton et al., 2019); (iii) we felt the task would better relate to the pre-intermediate proficiency level of our participants (compared, for example, to use of the PVLT in other advanced participant populations (e.g. De Jong et al., 2013); and (iv) to explore the extent to which the vocabulary produced in response to the Lex30 task would match the vocabulary produced in response to the speaking task. Lex30 was created by Meara and Fitzpatrick (2000) in response to issues with other existing productive measures (i.e., PVLT, LFP) at the time of publication. Lex30 has since been used in a wide variety of different papers (Clenton, 2010; Fitzpatrick & Clenton, 2010, 2017; Fitzpatrick & Meara, 2004; Jiménez Catalán & Moreno Espinosa, 2005; Uchihara & Saito, 2016; Walters, 2012). The task requires participants to respond with up to four words to each of the 30 Lex30 cues. Each set of Lex30 responses, a potential 120 items, is processed by correcting misspellings, lemmatizing according to Bauer and Nation's (1993) criteria, and profiling online according to frequency using the Web VP Classic (www.lextutor.ca/vp/eng/). Following the original Lex30 procedure (Meara & Fitzpatrick, 2000), responses contributed to a Lex30 score if they fell outside the first 1000 frequency band and were not proper nouns.

We also used a receptive vocabulary task. A number of studies exploring the relationship between second language fluency and vocabulary knowledge have used receptive measures (e.g. De Jong & Mora, 2017; Milton et al., 2010). We measured the receptive vocabulary of our participants using X_Lex (Meara and Milton, 2003). X_Lex is a computer-based test in which participants are required to respond to whether (120) words presented one at a time are known or unknown. Word knowledge is tested for items from the 1,000-frequency band to the 5,000-frequency band. X_Lex includes pseudo words, and scores are adjusted when such items are identified as genuine.

As well as measuring vocabulary using the productive vocabulary knowledge tasks, we wanted to explore whether this data correlated with the vocabulary used in response to the three speaking tasks. We therefore transcribed the vocabulary produced by participants in response to the three scenario tasks. Corpora generated from our speech data were treated in the same way as in the standard Lex30 task. With the concern that any comparison between Lex30 (written) data and scenario description (spoken) data

is not without difficulties, we turn to an earlier paper (Fitzpatrick & Clenton, 2010). Fitzpatrick and Clenton (p. 546) compared two formats of Lex30, a written and a spoken format. A paired t-test analysis ($t = .751$, $p = .457$) indicated that the means between the two tasks did not significantly differ. However, they report that a correlation analysis between the two task scores ($r = 0.391$, $p < .01$) was significant but weak, and might have been explained by their participants reluctantly having to respond to their classroom teacher. We base our comparison between the speaking fluency task data and Lex30 data on Fitzpatrick and Clenton's findings, but of course note Fitzpatrick and Clenton's (2010) warning that '(w)e should not assume, then, that the sample of vocabulary produced by a test taker in written mode will exactly mirror that which they produce in spoken mode' (p. 547). Based on this assumption, that the two modes may not exactly mirror one another, we tentatively compared our Lex30 data on the basis that our participants' written responses might approximately reflect their spoken responses. An additional issue we needed to address was which corpus to use in order to make this comparison. Lex30 conventionally uses corpora based on writing. Rather than comparing the vocabulary produced by our participant population with written corpora, we wanted to process data using a spoken word list. We used Dang et al.'s (2017) Academic Spoken Word List (ASWL), on the basis that 'there is a clear-cut difference between the linguistic features of academic speech and academic writing' (p. 978) and, the ASWL 'represented (as closely as possible) the academic speech that EAP learners from a wide range of academic disciplines are likely to encounter in their academic study in English-medium events' (p. 968).

Results

To determine the extent to which measures of vocabulary knowledge and vocabulary skills predict fluency variables in speaking, our dependent variables were: Silent pause duration between ASU (transcriptions were broken down into analysis of speech units[1]), Silent pause duration within ASU, Number of silent pauses per second, Number of filled pauses per second, Number of repetitions per second, Number of corrections per second (all, per second speaking time), and Mean syllable duration. The vocabulary measures used as predictor variables were two measures for vocabulary knowledge: Lex30-score (raw score), and X_Lex-score (corrected score); and three measures for vocabulary skills: LRS (Response Latency – picture naming), RL (Response Latency – delayed picture naming), and RD (Response duration – delayed picture naming). Regarding the extent to which vocabulary *use* in speaking can be predicted by measures of vocabulary knowledge and skills, we use vocabulary as the dependent variable.

Table 11.1 shows the descriptive statistics of all dependent variables as measured from the speaking fluency performances (all fluency variables). Table 11.2 shows the descriptive statistics of all predictor variables (three vocabulary knowledge and three (timed) vocabulary skills measures).

Table 11.1 Descriptive statistics of all dependent variables as measured from the speaking performances (all fluency variables)

Fluency variables	*Mean*	*SD*
Silent pause duration between ASU (ms)	2565.9	1120.34
Silent pause duration within ASU (ms)	1759.2	439.55
Number of silent pauses per second	0.88	0.28
Number of filled pauses per second	0.12	0.14
Number of repetitions per second	0.04	0.01
Number of corrections per second	0.015	0.007
Mean syllable duration (ms)	388	78.08

Table 11.2 Descriptive statistics of all predictor variables (vocabulary knowledge and (timed) vocabulary skills measures)

Vocabulary knowledge	*Mean*	*SD*
Lex30 raw score	40.06	10.43
X_Lex score	4048	476
LRS: Response Latency – picture naming	513.4	141.63
RL: Response Latency – delayed picture naming	749.6	104.30
RD: Response Duration – delayed picture naming	516.1	22.86

Correlations between vocabulary knowledge and skills with fluency measures

Table 11.3 shows the bivariate correlations between the fluency measures, on the one hand, and the vocabulary knowledge and skills measures, on the other. As can be seen in the table, for two measures of fluency, significant correlations with vocabulary knowledge and skills were found. For number of silent pauses per second, the higher the participants scored on the Lex30 task, the fewer pauses were found in their speech samples. At the same time, shorter latencies in the delayed picture task were associated with few silent pauses. Finally, the latency measure in the delayed picture naming task was negatively related to mean syllable durations in the speech samples: participants with short latencies tended to speak slower (with longer syllable durations).

Correlations between productive vocabulary knowledge and vocabulary use

Table 11.4 shows the bivariate correlations between the vocabulary used in the speaking fluency performances on the one hand, and the productive vocabulary knowledge measure (Lex30), on the other. As can be seen from Table 11.4, for the vocabulary used in the speaking task, significant correlations with productive vocabulary knowledge were found. For words used from the level 2 and level 4 of the Academic Spoken Word List (ASWL), the higher participants

Table 11.3 Correlations between vocabulary knowledge and skills with fluency measures (N = 30)

	Silent pause duration between ASU	*Silent pause duration within ASU*	*Number of silent pauses per second*	*Number of filled pauses per second*	*Number of repetitions per second*	*Number of corrections per second*	*Mean syllable duration*
Vocabulary knowledge							
Lex30-score	-.06	-.15	-.39*	-.17	.22	-.02	.12
X_Lex-score	.24	.18	-.16	-.14	.13	.12	.02
Vocabulary skills							
LRS: Response Latency – picture naming	-.04	.08	.31	.18	.08	.28	-.34*
RL: Response Latency – delayed picture naming	.27	.16	.37*	-.12	-.03	-.22	-.44*
RD: Response duration – delayed picture naming	.04	.09	-.21	-.21	-.15	-.09	-.22

*: $p < 0.05$

Table 11.4 Correlations between productive vocabulary knowledge and vocabulary use ($N = 30$)

	Speaking fluency task vocabulary (ASWL levels)				
	Level 1	*Level 2*	*Level 3*	*Level 4*	*Level 0*
Lex30-score	0.341	0.389*	0.234	.0.154	0.400*

*: $p < 0.05$

had scored on the Lex30 task, the more words from this band were found in the speaking performances.

Discussion

The current study was designed to further investigate the extent to which vocabulary knowledge and skills can predict aspects of fluency using several tasks. We have reported on an experiment in which the participants carried out three speaking tasks, and responded to tasks designed to capture their vocabulary skills (picture naming to measure lexical retrieval speed and delayed picture naming task to measure articulation speed), as well as two vocabulary tasks (a productive vocabulary task (Lex30; Meara & Fitzpatrick, 2000), and a test of vocabulary size (X_Lex; Meara & Milton, 2003). We also included an analysis in which the vocabulary used in response to the speaking fluency tasks was correlated with the vocabulary knowledge and skills measures. We can now respond to each of our four research questions.

We first asked whether productive vocabulary knowledge task scores predict aspects of speaking fluency. In broad terms, the findings from the current study are to some extent consistent with earlier fluency studies (e.g. De Jong et al., 2013; De Jong & Mora, 2017). The current study, while using different productive vocabulary knowledge measures, supports De Jong et al.'s (2013) finding that a higher vocabulary score correlates negatively and significantly with the number of silent pauses (Lex30). Regarding this specific correlation, we suggest it relates to Lex30 tapping into aspects of fluent speech to the extent that our pre-intermediate participants potentially used a similar set of highly frequent items from the same or similar frequency bands for the written and fluency tasks. In using Lex30, the current study supports Clenton et al.'s (2019) suggestion that it appears more aligned to the ability to use the words than other productive vocabulary knowledge tasks. This implication we feel is borne out by the significant correlations between the vocabulary used in response to the speaking fluency task and the Lex30 score (Table 11.4), on the basis that participants' lexical resource appears to be shown both in response to Lex30 and to the speaking fluency task. We suggest, however, that at higher levels of proficiency such overlap might not exist to this same extent between

productive vocabulary knowledge task corpora and speaking fluency task corpora, because of the vocabulary size of highly proficient users. While we appreciate that our finding might be exclusive to the proficiency of the participants in the current study, we suggest that this interpretation is important because it appears that the Lex30 task might tap the vocabulary knowledge available to such proficiency groups. Clenton et al. (2019) suggest that some aspects of vocabulary acquisition might lag others to the extent that certain aspects of vocabulary knowledge (e.g. form, which we believe Lex30 accesses) come before others (e.g. semantic, and grammatical knowledge, which the PVLT accesses). We also sense that the current study confirms that Lex30 scores predict aspects of fluency at a pre-intermediate level of proficiency, at least for the specific participants examined here in the current study. However, we suggest that future studies explore suggestions (e.g. Webb & Chang, 2012; Zhang & Lu, 2013) that aspects of vocabulary knowledge develop inconsistently with increases in proficiency. We propose that for studies involving higher-level learners a test such as the PVLT (alongside other productive vocabulary tasks such as Lex30) might help to inform the extent to which the quality of vocabulary knowledge develops with increases in proficiency.

Our second research question was designed to explore the findings from earlier papers on fluency (e.g. De Jong & Mora 2017) that found a significant correlation between receptive vocabulary knowledge task scores and one aspect of speaking fluency. The current study, however, did not find any significant correlations between receptive vocabulary knowledge task scores and the various aspects of speaking fluency. We refer readers to the discussions of our first and second research questions in this case, because we believe that the lack of correlations with the receptive vocabulary measures might relate to the specific proficiency level of our participant group and that this might relate to differences in developing lexicons. Previous fluency related studies (e.g. De Jong et al., 2013; De Jong & Mora, 2017; Miralpeix & Muñoz, 2018) have tended to examine more proficient participants. Such higher-level participants might have developed a receptive vocabulary resource which, we suspect, while not only being larger than that of the pre-intermediate participants that were the focus of the current study might also be more closely related to their productive vocabulary knowledge. The lack of any significant correlation between Lex30 and X_Lex (r = 0.371) might support this finding and runs somewhat counter to earlier Lex30 studies (e.g. Fitzpatrick & Clenton, 2010; Fitzpatrick & Clenton, 2017) that tend to show significant correlations between the receptive and productive vocabulary measures. We suggest that follow-up studies explore this specific finding with perhaps leaners of different (lower and higher) proficiency participants.

Our third research question was designed to explore the extent to which vocabulary skill measures (e.g. response latency and response duration) predict aspects of fluency. This specific question investigates Qian's (2002) suggestion that vocabulary knowledge relates to speed and automaticity of retrieval. The findings here all relate to the timed picture naming tasks in

which participants were required to name pre-primed pictures presented on a screen. Our investigation showed three significant correlations. First, it yielded a significant moderate correlation between response latency-delayed picture naming and the number of silent pauses per second in the speaking tasks ($r = 0.37$, $p < 0.05$). In other words, participants who were slower in their response in naming pictures tended to use more silent pauses in their speaking performances. Second, there was also a significant correlation between response latencies in delayed picture naming and mean syllable duration ($r = -0.44$, $p < 0.05$). This negative correlation is counterintuitive, in that fast picture-naming speed is related to a slow articulation rate (long syllable duration). The findings we report here are different to those reported in De Jong et al (2013), who found ten significant relations (with n=179), the largest being .32. We speculate that such differences may relate to the different participant proficiency levels and the sample sizes. Accordingly, we suggest that the three findings we report in this chapter are worthy of further examination in additional studies to determine whether aspects of fluency, such automaticity of retrieval and speed of naming, relate differently at different proficiency levels.

Our fourth and final research question asked whether the vocabulary used in response to the productive vocabulary task predicted the vocabulary used in the speaking fluency task. Our findings here show that there is some degree of overlap between responses to the Lex30 task and the speaking fluency task at levels 2 and 0 of the Academic Spoken Word List (ASWL; Deng et al., 2017). This finding, however should be tempered by the comments we presented earlier in our discussion (e.g. Fitzpatrick & Clenton, 2010) that speaking output may not mirror written output. The current study, however, was originally designed to test our first research question, to evaluate the extent to which productive vocabulary knowledge predicts aspects of fluency with perhaps a measure appropriate to the specific proficiency of our participant group. We maintain that this specific finding is, however, worth exploring further and that future such studies could, of course, adopt a spoken Lex30 format in order to test this specific claim. We do suggest, however, that there are potential limitations to this finding that relies on comparing data from the productive vocabulary knowledge task with the speaking fluency task. For our pre-intermediate proficiency participants, we propose that this kind of approach might fit, to the extent that we can observe some degree of overlap. However, with a highly proficient group, we argue that there might only be limited overlap between the productive vocabulary knowledge task and the speaking fluency task. Arguably, because of the limitations of the lexical resource, this approach might only be relevant for lower proficiency levels. We wonder, therefore, up until which proficiency levels this specific approach is relevant. We might suppose, then, that up to a specific proficiency, Lex30 provides a useful indication of the available lexical resource. The extent to which this finding can relate to other proficiencies and to other productive vocabulary tasks, would, we feel, be worthy of further exploration.

Limitations

We acknowledge that, inevitably, there are limitations with the current study, which should not go unreported. The first of these limitations relates to the sample size in the current study. With N = 30, adopting a power of .8 and alpha level of 0.05, we can only expect to find quite large correlations (at least r = 0.49). A second limitation relates to the fact that the current study only explored the vocabulary knowledge of a participant group with the same L1 (Japanese). It is therefore difficult for us to extend the results to other first language groups, because the findings we report here might be limited to L1 Japanese learners. Accordingly, we encourage replications of the current study with different first language populations in order to explore the extent to which our findings represent a potentially bigger picture of the relationships between vocabulary knowledge and fluency. We also propose that other studies consider additional and different vocabulary measures to explore whether different proficiency levels demonstrate greater (or lesser) word knowledge. We suggest that by doing so, such studies might clarify and support the findings we present here. In short, despite its limitations, we believe the current study represents an important development in determining which aspects of vocabulary relate to second language fluency.

Conclusion

This chapter has explored relationships between vocabulary knowledge and fluent speech, but we cannot overextend our findings given the limitations of this small study. We can now report three, albeit tentative, findings. First, based on comparisons of our results with those of earlier studies (e.g. De Jong, 2013) we suggest that relations between vocabulary knowledge and fluent speech may to some extent be proficiency dependent. This can be followed up in future research designed to investigate the potential interaction between proficiency level and the relation between vocabulary knowledge and fluency. Second, there appears to be some degree of overlap between the productive vocabulary used in response to a productive vocabulary task as well as a speaking fluency task. We do not, however, suggest that this finding would be consistent across all proficiency levels, as we discuss above. We suggest, again, that a series of studies of participants at different proficiency levels with the same tools employed in the current study might help shed some light on this finding. Third, we propose that the responses in (delayed) picture naming might relate to vocabulary knowledge in terms of speed and automaticity of retrieval (i.e. in vocabulary skills). We suggest that such measures are interesting and worthy of more research to the extent that different 'vocabulary skills' (such as automaticity of retrieval) relate differently for participants at different proficiency levels.

In order to further vocabulary research within the field of speaking fluency, we urgently need a range of studies to address the issues raised in this chapter. Specifically, we suggest that follow-up studies employ the same fluency and

vocabulary skills tasks as those used here, but we would add that using concurrent productive and receptive vocabulary knowledge tasks at a range of different proficiency levels, and with different first language populations, might shed additional light on our findings. We also encourage research to explore the relationships between an individual's lexical resource, their vocabulary knowledge, and their vocabulary skills. More studies of vocabulary skills are needed to explore the relationships between lexical resource, speed, and automaticity of retrieval.[2]

Notes

1 An AS-unit is 'a single speaker's utterance consisting of an independent clause, or a subclausal unit, together with any subordinate clause(s) associated with either' (Foster, Tonkyn, and Wigglesworth, 2000).

2 The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the JSPS [project number 16K02922].

References

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279. doi:10.1093/ijl/6.4.253

Boersma, P., & Weenink, D. (2016). PRAAT: Doing phonetics by computer [Computer program]. Version 6.0.23. Retrieved from www.praat.org/.

Clenton, J. (2010). *Investigating the construct of productive vocabulary knowledge with Lex30* (Order No. 10797989). Available from ProQuest Dissertations & Theses A&I. (2008472046). Retrieved from https://search.proquest.com/docview/2008472046?accountid=15247

Clenton, J., Elmetaher, H & Uchihara, T. (2019). Capturing and quantifying individual differences in bilingual learners: A task-based study. Paper presented at Capturing and Quantifying Individual Differences in Bilingualism workshop. Tromsø: Arctic University of Norway,.

Dang, T. N. Y, Coxhead, A., & Webb, S (2017). The academic spoken word list. *Language Learning*, *67*(4), 959–997. ISSN 0023-8333

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In Housen, A., Kuiken F. &Vedder, I. (eds) *Dimensions of L2 performance and proficiency* (pp. 120–142). Amsterdam: John Benjamins.

De Jong, N. H. (2016). Fluency in second language assessment. In Tsagari, D., and Banerjee, J. (eds) *Handbook of second language assessment* (pp 203–218). Boston, MA/Berlin: Mouton de Gruyter.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916.

De Jong, N.H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015), Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior, *Applied Psycholinguistics, 36*(2): 223–243. doi: 10.1017/S0142716413000210

De Jong, N., & Mora, J. (2017). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, *1*–13.doi:10.1017/S0272263117000389

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004), Second language fluency: Judgments on different tasks. *Language Learning*, *54,* 655–679. doi:10.1111/j.1467-9922.2004.00282.x

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *SSLA, 24*, 143–188. doi: 10.1017/s0272263102002024.

Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity. In Daller, H.,. Milton, J., & Treffers-Daller, J. (eds) *Modelling and assessing vocabulary knowledge* (pp. 116–132). Cambridge: Cambridge University Press.

Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, *27*(4), 537–554. doi:10.1177/2F0265532209354771

Fitzpatrick, T., & Clenton, J. (2017). Making sense of learner performance on tests of productive vocabulary knowledge. *TESOL Quarterly*, *51*, 844–867. doi:10.1002/tesq.356

Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics*, *1*, 55–74.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range and frequency programs. Available at www.victoria.ac.nz/lals/staff/paul-nation.aspx .

Jimenez Catalan, R., & Moreno Espinosa, S. (2005). Using Lex30 to measure the L2 productive vocabulary of Spanish primary learners of EFL. *Vigo International Journal of Applied Linguistics*, *2*, 27–44.

Kremmel, B., & Schmitt, N. (2017). Vocabulary levels test. In Liontas, J. I. (ed.) *The TESOL Encyclopedia of English Language Teaching*. Hoboken, NJ: Wiley. doi:10.1002/9781118784235.eelt0499

Laufer, B., & Nation, P. (1995). Vocabulary size and use lexical richness in L2 written production. *Applied Linguistics*, *16*, 307–322.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. doi:10.1177/026553229901600103

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387–417.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. ACL-MIT Press series in natural-language processing. Cambridge, MA: MIT Press.

Meara, P. (1996). The dimensions of lexical competence. In Brown, G., Malmkjaer, K., & Williams, J. (eds) *Performance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.

Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, *28*(1), 19–30. doi: 10.1016/S0346-251X(99)00058-5

Meara, P., & Jones G. (1987) Tests of vocabulary size in English as a foreign language. *Polyglot, 8*(1), [24.4].

Meara, P., & Milton, J. (2003). *X_Lex: The Swansea Levels Test*. Newbury: Express.

Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary.* Multilingual Matters.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In Chacón-Beltrán, R., Abello-Contesse, C., & Torreblanca López, M. (eds) *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching, 56*, 1–24.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, *5*, 12–25.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle Publishers.

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

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Paribakht, T., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, *11*(1), 9–29. doi: 10.18806/tesl.v11i1.623.

Paribakht, T. S., & Wesche, M. (1996). Enhancing vocabulary acquisition through reading: A hierarchy of text-related exercise types. Canadian Modern Language Review *52*(2), 155–178.

Paribakht, T., & Wesche, M. (1997). vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. *Second Language Vocabulary Acquisition,* 174–200. doi: 10.1017/cbo9781139524643.013.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*, 513–536. doi:10.1111/1467–9922.00193

Read, J. (2000). *Assessing vocabulary.* Cambridge: Cambridge University Press.

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*. doi: 10.1093/applin/amv047

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical correlates of comprehensibility versus accentedness in second language speech. *Bilingualism: Language and Cognition, 19*(3), 597–609. doi:10.1017/S1366728915000255

Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*(2), 173–199. doi:10.1017/S0272263104262027.

Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., & Bickel, B. (2018). Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences*, *115*(22), 5720–5725.

Skehan, P. (2003). Task-based instruction. Language Teaching, *36*(1), 1–14. doi:10.1017/S026144480200188X

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215. http://dx.doi.org/10.1037/0278-7393.6.2.174

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In Ellis, R. (ed.) *Planning and task performance in a second language* (pp. 239–277). Amsterdam: Benjamins.

Uchihara, T., & Saito, K. (2016). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *Language Learning Journal*. doi:10.1080/09571736.2016.1191527

Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Language Assessment Quarterly*, *9*(2), 172–185. doi:10.1080/15434303.2011.625579

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52. doi:10.1017/S0272263105050023

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65. doi:10.1093/applin/aml048

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, *30(*1), 79–95. doi:https://doi.org/10.1017/S0272263108080042.

Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, *40*(3), 360–376. https://doi.org/10.1177/0033688209343854

Webb, S., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, *43(*1). doi: https://doi.org/10.1177/0033688212439367.

Wilkins, D. A. (1972). *Linguistics in language teaching*. Cambridge: MFT Press

Zhang, X., & Lu, X. (2013). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied Linguistics*, *35*, 283–304. doi: https://doi.org/10.1093/applin/amt014.

12 Re-examining the relationship between productive vocabulary and second language oral ability

Takumi Uchihara, Kazuya Saito and Jon Clenton

Introduction

Vocabulary knowledge has been viewed as one of the most essential elements of second language (L2) proficiency and development (Meara, 1996), as research investigating the relationship between vocabulary and L2 proficiency has developed to support the long-standing view that vocabulary serves as a proxy for communicative language ability (Miralpeix & Muñoz, 2018). A growing body of research in this area relates vocabulary knowledge to overall proficiency benchmarks (e.g., Common European Framework of Reference for Languages (CEFR) levels; Milton, 2010), in-house placement tests (e.g., Harrington & Carey, 2009), and standardized language proficiency examinations (e.g., International English Language Testing System (IELTS); Milton et al., 2010, or Test of English as a Foreign Language (TOEFL); Qian, 2002), and each of the four language skills (e.g., Laufer & Levitzky-Aviad, 2017 for reading; Wang & Treffers-Daller, 2017 for listening; Baba, 2009 for writing; Uchihara & Clenton, 2018 for speaking). This line of research has been largely devoted to investigating the relationship between vocabulary and reading, yet surprisingly little is known about the role of vocabulary in speaking (cf. Uchihara & Clenton, 2018). Under recent frameworks of L2 speech, L2 oral ability is considered multifaceted in nature, since it comprises a range of different skills related to phonological, fluency, and lexicogrammar (Crossley et al., 2015; Saito et al., 2017). To move the research agenda in L2 vocabulary and speech ahead, the current study attempts to explore whether, to what degree, and how L2 learners' productive vocabulary knowledge is associated with global (comprehensibility), temporal (speed, breakdown fluency), and lexical (appropriateness, variation, sophistication) aspects of L2 oral ability, and to then discuss implications for vocabulary assessment and future research.

Productive vocabulary measures

Researchers agree that vocabulary knowledge encompasses a wide array of word knowledge, characterized by knowledge of form, meaning, and use

in receptive and productive dimensions (Nation, 2013), or partial to precise knowledge, shallow to deep knowledge, and receptive to productive knowledge (Henriksen, 1999). Among these dimensions, a receptive vs. productive feature has to date attracted theoretical and empirical research attention (see Pignot-Shahov, 2012) and such a conceptual distinction has also applied to the development of vocabulary testing (Read, 2000): Vocabulary Levels Test (Schmitt, Schmitt, and Clapham 2001) and the Eurocentres Vocabulary Size Test (Meara & Jones, 1990) for receptive (or passive) vocabulary measures, and Productive Vocabulary Levels Test (PVLT; Laufer & Nation, 1999) and Lexical Frequency Profiling (LFP; Laufer & Nation, 1995) for productive (or active) vocabulary measures. In studies exploring the role of vocabulary knowledge as a predictor for L2 proficiency, researchers tend to rely on receptive vocabulary measures instead of productive vocabulary measures. Receptive vocabulary tests are frequently used in many studies (e.g., Baba, 2009; Harrington & Carey, 2009; Milton, 2010; Milton et al., 2010; Miralpeix & Muñoz, 2018; Qian, 2002; Stæhr, 2009; Uchihara & Clenton, 2018; Wang & Treffers-Daller, 2017); yet, by contrast, productive vocabulary tests are less commonly used (De Jong et al., 2012a; Koizumi & In'nami, 2013). This tendency may yield misleading results if any data born out of receptive vocabulary measurement alone are subsequently interpreted as an indication of overall vocabulary knowledge and compared with productive language skills.

Some productive vocabulary tests have been widely used to measure L2 productive vocabulary knowledge. For example, the Productive Vocabulary Levels Test (PVLT) (Laufer & Nation, 1999) elicits productive vocabulary in a controlled manner (i.e., sentence completion task), since the answer key is predetermined and none of the other words produced are awarded a point. The Lexical Frequency Profile (LFP) (Laufer & Nation, 1995) is a free production task, in which test takers write short essays producing texts that are categorized and scored according to frequency levels (e.g., the first and second 1,000 words).

These two productive measures, however, suffer from issues with scoring procedures in addition to issues with task formats (Clenton et al., 2019; Fitzpatrick & Clenton, 2010). Scores from productive vocabulary tasks such as the PVLT and LFP alike appear heavily dependent on the general assumption that word frequency determines language development (i.e., learning occurs in order of high to low frequency words; Schmitt et al., 2001). Yet such an assumption and general principle might not apply to extrapolating productive vocabulary size because words produced in speech or writing, or lexical *use,* may not necessarily represent lexical *knowledge* (Nation & Webb, 2011). An additional concern relates to the seemingly multifaceted nature of the construct of productive vocabulary (Fitzpatrick & Clenton, 2017). In addition to the quantity (i.e., size) and quality (i.e., depth) of word knowledge, the ability to use words fluently has been considered as an integral part of productive vocabulary construct (Daller et al., 2007). Accessibility or automaticity involved in word use reflects the ability to use words in spontaneous communication

(Chapter 11, this volume; McLean et al., 2015), and considering the degree to which a vocabulary task truly mirrors real-life language use is crucial in vocabulary test development (Kremmel & Schmitt, 2016). Among other productive vocabulary tasks (e.g., PVLT), the Lex30 task format (i.e., the test we used in the current study) has been found to relate to L2 learners' proficiency (Meara & Fitzpatrick, 2000; Walters, 2012) and considered as a reasonably valid measure necessary to 'elicit a representative sample of vocabulary from the productive lexicon' (Fitzpatrick & Meara, 2004: 70–1).

One additional strength of the Lex30 task relates to reported dual scoring approaches (i.e., raw scoring and percentage scoring). While the percentage scoring system depends heavily on a frequency index to report scores specifically pertaining to learners' lexical resources, the raw scoring system reflects fluency (i.e., production speech) as well as learners' lexical resources (Clenton, 2010; Chapter 11, this volume; Uchihara & Saito, 2019). Researchers also suggest that certain measures (e.g., PVLT, LFP, translation test) might not encourage test takers to demonstrate 'their full range of lexical knowledge' or appear to tap 'superfluous' knowledge extraneous to the productive vocabulary knowledge construct (Fitzpatrick & Meara, 2004). The Lex30 word-association task is therefore adopted in the current study to respond to these demands and avoid tapping receptive vocabulary knowledge (cf. the sentence completion tasks of the PVLT), as well as to avoid eliciting additional contextual knowledge (cf. the composition completion task of the LFP), and decontextualizing the task to a great extent (cf. L1-to-L2 translation task).

Measuring second language oral ability

Second language oral ability is conceived of as a black box full of complexities, encompassing multi-componential linguistic factors spanning pronunciation, fluency, vocabulary, and grammar (De Jong et al., 2012a). To better understand such complex constructs, researchers have measured oral ability in a variety of different ways, though broadly divided into two approaches: subjective and objective assessment. On a global level, L2 oral ability is operationalized as native listener judgements as per different types of rating rubrics. In particular, L2 speech researchers have extensively examined comprehensibility (i.e., how easily listeners can understand L2 speech) (Derwing & Munro, 2015). Such global judgements have been found to be tied to a range of linguistic factors, such as pronunciation accuracy (Saito et al., 2017), fluency (Suzuki & Kormos, 2019), lexicogrammar (Saito et al., 2016), and language use at a discourse level (Trofimovich & Isaacs, 2012).

In the objective approach, L2 speech scores are yielded by means of quantifying a variable of temporal (e.g., the number of pauses for fluency) and lexical (e.g., appropriateness, diversity, sophistication) properties with acoustic (e.g., PRAAT) and corpus (e.g., Coh-Mertix) tools. Quantified linguistic properties are envisaged as a useful parameter for describing L2 oral proficiency with the focus on Complexity, Accuracy, and Fluency (CAF) of L2 speech

(Suzuki & Kormos, 2019). Each construct has been captured and measured in a variety of ways – e.g., complexity as lexical diversity, lexical sophistication, syntactic complexity, and morphological complexity (Bulté & Housen, 2012; Crossley et al., 2015), accuracy as lexical, morphological, and syntactic accuracy (Crossley et al., 2015; Yuan & Ellis, 2003), and fluency as speed fluency (the number of syllables per second), repair fluency (the number of repetitions and self-corrections), and breakdown fluency (the number of filled and unfilled pauses) (Tavakoli & Skehan, 2005; Tavakoli & Uchihara, 2019).

Motivation for the current study

Despite a growing number of studies examining the effects of lexical richness and accuracy on L2 speech, few researchers (De Jong et al., 2012a; Koizumi & In'nami, 2013; Vermeer, 2000) have specifically focused on the relationship between vocabulary knowledge and L2 speech by measuring both constructs independently. These researchers adopt either an objective or subjective approach to measuring oral ability and then relate it to vocabulary task scores.

In the objective approach, Vermeer (2000) conducted a study in which Dutch L1 and L2 children performed a receptive vocabulary task (similar to the Peabody Picture Vocabulary Test in format), a productive vocabulary (i.e., definition) task, and a spontaneous storytelling task. The elicited speech samples were transcribed and analysed on a series of lexical diversity measures (e.g., type token ratio and the Guiraud index). The results showed that both receptive and productive vocabulary measures were significantly correlated with most of the diversity measures (r = -.19 to .51 and r = -.19 to .53 respectively), yet the relationship appeared no longer present in a group of proficient speakers with vocabulary sizes of 3,000 or more. In the context of novice to intermediate Japanese learners of English, Koizumi and In'nami (2013) found significant correlations between productive vocabulary knowledge (measured with L1-to-L2 translation tasks) and spoken fluency, accuracy, and syntactic complexity in speech.

In the subjective approach, De Jong et al. (2012a) carried out a large-scale study (N = 181) in which intermediate to advanced learners of Dutch performed productive vocabulary tests (i.e., a sentence completion task based on the PVLT from Laufer & Nation, 1999) and eight speaking tasks, the elicited speech was subsequently rated by four untrained judges according to communicative adequacy. The results showed that productive vocabulary knowledge was among the strongest predictors of L2 oral ability.

Another recent study adopting the subjective approach was conducted by Uchihara and Saito (2019). In this study, Japanese EFL undergraduates took a task designed to elicit productive vocabulary word knowledge from a series of lexical cues (i.e., Lex30), and a picture description task designed to elicit speech samples. The results showed a significant correlation between productive vocabulary scores and fluency (i.e., optimal rate of speech delivery) evaluated by five trained native speakers.

Such studies, regardless of the obvious variations in their approaches (being either subjective or objective), appear to converge on the opinion that vocabulary knowledge contributes to L2 oral ability: fluency (Koizumi & In'nami, 2013), lexical richness (Vermeer, 2000), and communicative adequacy (De Jong et al., 2012a). However, it is also noteworthy that all of these studies exclusively focus on a few specific aspects of L2 speech by adopting either subjective or objective approaches. Recently, a growing number of researchers (e.g., Trofimovich & Isaacs, 2012) have begun to emphasize the importance of conceptualizing L2 oral proficiency as a *multifaceted* (rather than monolithic) phenomenon that needs to be assessed from various angles. In this view, such different dimensions of L2 oral proficiency include not only the ability to choose appropriate, diverse, and sophisticated words (Crossley et al., 2015), but also the ability to deliver them with correct pronunciation forms (Derwing & Munro, 2015) at an optimal tempo without too many dysfluencies and pauses (Tavakoli & Skehan, 2005).

Taken together, the current literature review identifies two gaps apparent in the relationship between vocabulary and oral proficiency. First, our understanding of the relationship between vocabulary and L2 proficiency is largely contingent on receptive measures rather than productive measures. Second, virtually no studies have systematically looked at the relationship between L2 learners' productive vocabulary knowledge and multiple dimensions of their oral proficiency. To advance the research agenda on this topic, and following the latest framework in L2 oral proficiency (Crossley et al. 2015; Saito et al., 2017), the current study is designed to explore the extent to which L2 learners' productive vocabulary knowledge can be associated with global (comprehensibility), temporal (breakdown, speed), and lexical (appropriateness, diversity, sophistication) aspects of speech production. Accordingly, this study is guided by the following two research questions:

1. To what extent is productive vocabulary knowledge associated with comprehensibility rating?
2. To what extent is productive vocabulary knowledge associated with objectively measured fluency and lexical aspects of L2 speech?

Method

The current study follows up on the data set presented in another venue (Uchihara, Eguchi, Clenton, Kyle, & Saito, under review) and is designed to pursue a different research purpose by exploring the relationship between vocabulary knowledge and oral proficiency.

Participants

The participants were 40 first-year Japanese students (26 females and 14 males) at a university in Japan. All participants had learned English for six

years starting at Grade 7. Their general proficiency in English spread widely according to Test of English for International Communication (TOEIC) scores, comprising reading and listening sections (M = 697.9, SD = 125.7, range = 515–890), and the Test of English as a Foreign Language (TOEFL) (iBT) test scores (M = 71.7, SD = 13.7, range = 40–96).

Productive vocabulary task (Lex30)

Lex30 is a productive vocabulary measure created by Meara and Fitzpatrick (2000) in response to perceived issues with productive measures (e.g., PVLT and LFP) and widely employed for research purposes (Clenton, 2010, 2015; Fitzpatrick & Clenton, 2010, 2017; Fitzpatrick & Meara, 2004; Uchihara & Saito, 2019; Walters, 2012). We selected this task because productive vocabulary elicited through the Lex30 word association task is viewed as an indication of learners' lexical resource in addition to fluency (i.e., access and retrieval speed of L2 items) (Clenton, 2010). Our initial concern in choice of a vocabulary measure relates to the fact that speakers with a large vocabulary can be disfluent. This 'fluency' (Daller et al., 2007) or 'automaticity' (Meara, 1996) is recognized as a crucial dimension in describing productive vocabulary construct and particularly for this study (i.e., exploring the relationship with L2 speech measures), we consider Lex30 an appropriate measure to compare the vocabulary knowledge it generates with oral ability.

Test administration and scoring

For the Lex30 task, participants write four responses to each of the 30 cue words provided on a sheet of paper within 15 minutes (e.g., *attack* > *game, offense, defense, war*) (see Appendix A in Meara & Fitzpatrick, 2000 for the cue words). Each set of responses (120 items in total) elicited through the task are processed as follows: misspellings are corrected, responses are lemmatized following Bauer and Nation's (1993) criteria, and individual sets of responses are analysed according to the JACET 8000 resource (JACET, 2003) whereby each response item is profiled per frequency levels. All the lemmatized items beyond the first 1,000 frequency words except numbers and proper nouns are considered infrequent and awarded a point. Raw scores are calculated by tallying the infrequent items per participant, and percentage scores calculated as the proportion of infrequent words in relation to the total number of responses.

Highly vs. moderately frequency-based scoring

Following suggestions from earlier studies (Clenton, 2010; Uchihara & Saito, 2019), we considered raw scoring as moderately frequency-based and percentage scoring as highly frequency-based. Percentage scores are heavily dependent on a frequency-based principle, since both the increase in

the number of low-frequency responses and the decrease in the number of high-frequency responses together raise percentage scores (= number of low-frequency items/total number of items produced). Conversely, raw scores are not influenced by the increasing number of high-frequency words, since the scores are counts of low-frequency items. Thus, raw scoring is less dependent on the frequency index.

L2 oral ability measures

The current study elicited spontaneous speech data via a picture narrative task (i.e., a Suitcase Story) following previous L2 speech studies using the same task (e.g., Saito et al., 2016; Trofimovich & Isaacs, 2012) and based on the fact that the format has been extensively employed for research purposes (e.g., Daller et al., 2003; Saito et al., 2017; Trofimovich & Isaacs, 2012; Vermeer, 2000). The task can also differentiate L2 speakers in accordance with differing lexical proficiency, because a picture narrative is 'unforgiving in what needs to be covered' and involves 'unavoidability' when speakers select L2 words to describe the situation during a given task (Skehan, 2009: 517). The elicited speech samples are rated by L1 English speakers using a global rubric (comprehensibility) and submitted to objective speech measures including fluency and lexis.

Speaking task

The participants describe a sequence of eight-frame pictures without any explicit time restriction after spending approximately one-minute familiarizing themselves with the pictures. The story consists of two strangers carrying suitcases identical in appearance, bumping into each other at the corner of a city street, inadvertently exchanging their suitcases, and later discovering their mistake when opening the other's suitcase.

Speech recordings were carried out individually in a sound-proof laboratory at the testing venue (a university), and each elicited speech sample was digitally stored as a WAV file. The total length of each story ranged between 105 and 251 seconds. To ensure that the content of the story was consistent across speakers, the first few seconds were trimmed to remove pauses and false starts from the outset of each recording and 30 seconds were extracted from each sample for subsequent rating sessions. This procedure conformed to the methodology adopted in L2 speech literature (Trofimovich & Isaacs, 2012), and such small samples (i.e., 30 seconds) are reported to reliably elicit listeners' impressionistic ratings of L2 speech (Munro, Derwing, and Burgess 2010).

Subjective measures

We chose comprehensibility (i.e., ease of understanding) as a global construct of L2 oral ability (Derwing & Munro, 2015). The rationale behind the choice

of comprehensibility is informed by the pedagogical value attached to this construct. L2 speech studies suggest that targeting comprehensible speech is pedagogically sensible because attaining native-like accents is extremely difficult (Flege, Munro, and MacKay 1995), and L2 speakers with heavy accents can be highly comprehensible (Derwing & Munro, 2015).

Comprehensibility rating

Thirteen L1 English raters were recruited (10 females and 3 males) from a university in Japan. All of the raters were undergraduate students enrolled on different programs (e.g., one-year exchange program) with a mean age of 21.7 years. Their familiarity with Japanese language was considered as high across raters, as the survey of their language background showed moderately frequent use of Japanese (M = 3.4, 1 = *not at all*, 6 = *very often*) in a range of contexts (e.g., classmates, teachers, clerks) and with at least a three-month period of residence in Japan. Three raters had taken an introductory linguistics course at university, but none had had pronunciation-focused training or practice prior to the current experiment. None of the raters reported hearing difficulties.

To reflect their perception of how easy it was to understand the speaker in our 40 speech samples, we followed the rating method used in earlier L2 speech studies (e.g., Saito et al., 2017; Trofimovich & Isaacs, 2012). The raters first received a practice session, rating three examples based on the rubrics of comprehensibility (Derwing & Munro, 2015). Then the raters listened to all samples in a randomized order using PRAAT software (Boersma & Weenink, 2013). Upon hearing each sample only once, they rated comprehensibility on a 9-point rating scale presented on a computer screen (1 = *easy to understand*, 9 = *hard to understand*).

Objective measures

The 40 oral narratives were transcribed with all orthographic markings of pausing (e.g., *uh, um, oh, ehh*) removed and obvious pronunciation errors fixed (e.g., *the story* for *the stoly*). The transcripts in length ranged between 57 and 208 words. All transcripts were submitted to linguistic analysis in terms of fluency and lexis. Three variables were selected for fluency analysis: articulation rate, filled pause ratio, and silent pause ratio. Four variables were chosen for lexical analysis: appropriateness (i.e., error ratio), diversity, and sophistication (i.e., two different types of corpus used for this analysis).

Fluency analysis

We adopted speech and breakdown fluency from a triad of utterance fluency measures: speed fluency (e.g., speech rate, articulation rate), breakdown

fluency (e.g., length of run, number of pauses, length of pauses), and repair fluency (e.g., self-corrections, false starts, repetitions, hesitations) (Tavakoli & Skehan, 2005). Our choice was based on earlier studies highlighting an important relationship between vocabulary knowledge and these two fluency constructs in particular (De Jong, 2016; Koizumi & In'nami, 2013).

For breakdown fluency, measures of both silent pause (i.e., silent pause ratio) and filled pause (i.e., filled pause ratio) were calculated. Building on De Jong et al.'s (2012b) study, a silence longer than 0.35 seconds was counted as a silence pause and a silent pause ratio was calculated as the percentage of silent pausing time in the total speaking time. In addition to silent pause ratio, we also computed a filled pause ratio. After checking the transcripts as well as speech data and counting the number of filled pauses (e.g., *uh, um*), the total number of filled pauses was divided by the total number of words and then we calculated a filled pause ratio.

For speed fluency, we measured articulation rate (total number of syllables divided by speaking time excluding pauses) as 'a pure speed measure' (De Jong et al., 2012b: 136). Articulation rate is unlikely susceptible to variability in task complexity and appears to reflect 'a task-independent articulatory skill' (p. 125).

Lexical analysis

Lexical appropriateness, diversity and sophistication measures were calculated for lexical analysis (Read, 2000; Saito et al., 2016). First, appropriateness (or lemma errors) was defined based on earlier literature (e.g., Yuan & Ellis, 2003) as the number of contextually and conceptually inappropriate words (e.g., *hit* or *attack* instead of *bump into*) and Japanese substitutions (e.g., トラベルバッグ for *travelling bag*), over the total number of words. All 40 transcripts were initially coded by a trained coder, and then another trained coder recoded 10 randomly selected transcripts (i.e., 25% of all transcripts). The resulting intraclass correlations showed high consistency (r = .97).

Second, we computed lexical diversity as the variation of words in a text. Although lexical diversity is normally defined as the number of different words used by a speaker or writer (e.g., type-token ratio), the reliability of such measures is considered questionable due to its dependency on text length (i.e., the longer the texts, the lower the values). To circumvent this operational problem, we employed a sophisticated measure of lexical diversity, or the Measure of Textual Lexical Diversity (MTLD) (McCarthy & Jarvis, 2010). MTLD, derived through Coh-Metrix (McNamara et al., 2014), involves indices that are mathematically transformed to account for text length so that the computed values can be adequately independent of text-length effect (McCarthy & Jarvis, 2010).

Third, lexical sophistication was defined as the degree of difficulty or rarity of the words produced in speaking or writing (Read, 2000). This view underlies

the empirical assumption that learning occurs in the order of frequency (i.e., high to low frequency words) and sophistication is often measured objectively through corpus-based lexical profiling (e.g., LFP; Laufer & Nation, 1995). As such, we employed a corpus-based frequency measurement. Our sophistication measure involved dividing the total number of words (types, not tokens) beyond 2,000 most frequent words of English (excluding proper nouns) by the total number of word types (Daller, Van Hout, and Treffers-Daller, 2003; Uchihara & Clenton, 2018). As our participants were first language Japanese speakers, we applied the JACET 8000 (JACET, 2003) corpora to this analysis.

Results

Productive vocabulary (Lex30) and L2 oral ability

Table 12.1 shows the means, standard deviations, and ranges of Lex30 raw and percentage scores and L2 oral ability scores. A series of Cronbach's alpha analyses confirmed inter-rater reliability in L2 oral ability ratings. Given that a panel of 13 untrained raters demonstrated a high inter-rater agreement for comprehensibility (α = 0.92), a single average score was calculated for each speaker for comprehensibility.

Pearson correlation analysis was performed among ten L2 speech measures to examine the aspects of oral ability we used in this study (see Table 12.2). As reported by previous L2 speech studies (Trofimovich & Isaacs, 2012), the comprehensibility rating was moderately associated with a wide range of linguistic variables including fluency and lexis.

Table 12.1 Lex30 raw and percentage scores and L2 Speech Measures

	M	*SD*	*Range*
Lex30			
Raw scores	45.3	10.5	22–65
Percentage scores	44.21	7.48	30.9–61.9
Subjective measure			
Comprehensibility	4.8	1.6	1.4–7.3
Objective measures			
Fluency			
Articulation rate	109.12	26.61	50.94–176.20
Filled pause ratio	0.07	0.60	0.00–0.22
Silent pause ratio	0.60	0.21	0.28–1.16
Lexis			
Appropriateness	0.08	0.04	0.00–0.18
Diversity	37.09	10.21	19.20–63.08
Sophistication	0.08	0.03	0.00–0.16

Table 12.2 Intercorrelations among L2 speech measures

Speech Measure	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1. Comprehensibility	1	.75**	.39*	.68**	.38*	-.38*	-.01
2. Articulation rate		1	-.43*	-.71**	-.38*	.32*	.24
3. Filled pause ratio			1	.37*	.26	-.08	-.14
4. Silent pause ratio				1	.50**	-.21	-.32*
5. Appropriateness					1	-.12	.04
6. Diversity						1	.05
7. Frequency							1

Note. * indicates $p < .05$; ** indicates $p < .01$.

With regard to the interrelationship among objectively measured linguistic variables, moderate-to-strong associations were found between speed fluency (articulation rate) and breakdown fluency (filled pause ratio, silent pause ratio). Lexical diversity was significantly correlated with lexical sophistication (JACET), whereas lexical appropriateness did not correlate with either diversity or sophistication.

Overall, the results are in line with previous studies (Trofimovich & Isaacs, 2012; Saito et al., 2016) in that comprehensibility was associated with a wide range of linguistic features, except for lexical sophistication. We observe that the six fluency and vocabulary measures appear to tap into three distinct aspects of L2 oral proficiency – fluency (breakdown, speed), lexical appropriateness, and richness (diversity, sophistication).

Relationships between productive vocabulary and L2 speech measures

To respond to the first and second research questions regarding the relationship between productive vocabulary knowledge and L2 oral ability, Pearson correlation analyses were performed on the Lex30 scores and a subjective measure (comprehensibility) and the six objective measures (articulation rate, filled pause ratio, silent pause ratio, lexical appropriateness, lexical diversity, lexical sophistication). Results show that learners' Lex30 raw scores moderately correlated with the comprehensibility rating ($r = -.35$, $p < .05$), indicating the more productive vocabulary learners have, the more comprehensible they are perceived to be (Table 12.3). Lex30 percentage scores did not significantly correlate with the comprehensibility rating. With respect to objective measures, their Lex30 raw scores significantly but moderately correlated with two fluency variables – articulation rate ($r = .48$, $p < .01$), silent pause ratio ($r = -.43$, $p < .01$), and two lexical variables – diversity ($r = .47$, $p < .01$), sophistication (JACET) ($r = .44$, $p < .01$). In contrast, their Lex30 percentage scores were not associated with any objective measures except lexical diversity ($r = .45$, $p < .01$), comparable to the effect size derived from the correlation between diversity and Lex30 raw scores ($r = .47$).

Table 12.3 Pearson correlations between productive vocabulary and L2 speech measures

	Lex30	
	Raw	*Percentage*
Subjective measure		
Comprehensibility	-.35*	-.23
Objective measures		
Fluency		
Articulation rate	.48**	.31
Filled pause ratio	-.03	.07
Silent pause ratio	-.43**	-.23
Lexis		
Appropriateness	-.18	-.10
Diversity	.47**	.45**
Frequency	.44**	.30

Note. * indicates $p < .05$; ** indicates $p < .01$.

Discussion

In response to the first research question concerning the relationship between productive vocabulary knowledge and L2 speech rating scores (comprehensibility), the results appear supportive of earlier research (De Jong et al., 2012a), suggesting that L2 speakers with a larger productive vocabulary can speak the L2 in a more comprehensible manner. This finding was in accord with L2 speech literature indicating that lexically proficient speakers' speech is perceived as comprehensible (Trofimovich & Isaacs, 2012; Saito, 2019; Saito et al., 2016, 2017). However, the result in this study does not seem to be in line with Uchihara and Saito's (2019) study finding that Lex30 scores did not correlate significantly with comprehensibility ratings but correlated with fluency ratings. Based on this previous finding and the fact that comprehensibility and fluency are highly related (Suzuki & Kormos, 2019), we conducted post hoc partial correlation analyses (i.e., the first one on the correlation between Lex30 raw scores and comprehensibility while the influence of articulation rate is controlled for, and the second one on the correlation between Lex30 raw scores and articulation rate while comprehensibility is controlled for). Articulation rate was selected as a fluency variable in this analysis, given that it showed the largest correlation with Lex30 scores among the three fluency measures. The post hoc results show that the correlation between Lex30 scores and comprehensibility did not reach statistical significance when the effect of fluency was accounted for (pr = -.075, p = .648), whereas the significant correlation between Lex30 scores and articulation rate remained while the effect of comprehensibility was accounted for (pr = .351, p = .028). These findings

support Uchihara and Saito's argument and suggest that (a) productive vocabulary knowledge is closely linked with oral fluency, (b) fluency and comprehensibility are highly related (Suzuki & Kormos, 2019), and (c) consequently learners with larger productive lexicons can speak the L2 fluently, which in turn might make their speech more comprehensible to listeners.

For the second research question concerning the relationship between productive vocabulary knowledge and objective L2 speech measures, results showed important relationships between productive vocabulary and fluency (articulation rate, silent pause ratio) and lexical use (diversity and sophistication), apart from two (filled pause ratio, lemma appropriateness). These findings suggest that L2 speakers with a large productive vocabulary can produce lexically sophisticated and rich language at a faster rate without too many pauses. The following discussion outlines the relationship between productive vocabulary and each of the objectively measured variables, fluency and lexis.

Regarding the relationship between vocabulary knowledge and fluency, we found that productive vocabulary correlated with articulation rate and silent pause ratio. Our findings, along with earlier research (Koizumi & In'nami, 2013), support the view that the speech production process is lexically driven (Kormos, 2006; Levelt, 1989) and speakers with smaller lexicons might experience difficulties in producing L2 words smoothly due to the inefficiency of lexical retrieval process (Tavakoli & Uchihara, 2019; Skehan, 2009). This view might also be explained in light of learner vocabulary knowledge and its effect on any pausing phenomena. Among many possible accounts for such phenomena, attempting to retrieve lower frequency words is reported to cause pauses during speech (De Jong, 2016). In this respect, for speakers with larger productive vocabularies, L2 words required in oral narratives might be within the range of their vocabulary size and therefore they might be adequately capable of selecting and producing semantically appropriate words fluently. Conversely, for low-proficiency speakers, the same words might be perceived as infrequent and unfamiliar due to limited lexical resources, resulting in more frequent and lengthier pauses (see Chapter 11, this volume). The relationship between vocabulary size and breakdown fluency is also supported by a weak but significant correlation between silent pause ratio and lexical sophistication (r = -.32), indicating that L2 speakers who can retrieve low-frequency words successfully tend to speak without too many silent pauses. Interestingly, we did not find any relationship between filled pause ratio and productive vocabulary scores. This finding implies that even lexically advanced L2 speakers might rely on filler use while speaking, the trend attributable to factors external to linguistic systems such as personality, speaking style, or speaking strategies to buy planning time during tasks (Préfontaine & Kormos, 2016).

As for the relationship between productive vocabulary and lexical use in speech, we found that productive vocabulary knowledge significantly correlates with lexical diversity and sophistication. These findings were consistent with Vermeer's (2000) study in that the quality of lexical output (i.e.,

diversity) is broadly determined by L2 speakers' vocabulary knowledge. Another intriguing finding to come from the current study was the absence of a significant correlation between productive vocabulary scores and lexical appropriateness. This specific finding might indicate that, even though speakers with larger productive vocabularies can produce a greater number of different and low-frequency words, their choice of words may not be conceptually and contextually appropriate to describe the situation depicted in the pictures. One possible reason might be that the Lex30 test may not be sufficiently sensitive to capture the ability of L2 speakers to use contextually appropriate words. According to Fitzpatrick and Meara (2004), the construct Lex30 taps into is more closely related to productive vocabulary *recall* (i.e., retrieving forms of target words from memory) than the ability to actually *use* L2 words for communicative purposes (see Read, 2000, for further discussion related to recall and use). To test Fitzpatrick and Meara's hypothesis, Walters (2012) conducted validation studies and found that test takers, especially at lower proficiency levels, failed to use the elicited words appropriately and accurately in their communication. Alternatively, the Lex30 task reflects the ability to recall a variety of associated L2 words prompted through 30 different stimuli or 'a wide range of conceptual fields' (Fitzpatrick & Clenton, 2010: 539), potentially increasing the likelihood of producing lexically diverse language as well as sophisticated L2 words.

On a final note, the result of this study (i.e., Lex30 raw scores correlated with oral fluency measures, whereas percentage scores correlated exclusively with lexical diversity) indicates that a frequency-based approach alone to assess productive lexical knowledge is not satisfactory. In this study, we used two types of scoring procedures to calculate Lex30 scores, raw and percentage scores; the former depends on production speed as well as production of infrequency words, whereas the latter emphasizes production of infrequent words, thus depending on frequency information to a greater extent (Clenton, 2010; Uchihara & Saito, 2019). Our findings suggest that, even with the same word elicitation task, different scoring methods might influence the way that the resultant scores are interpreted to the extent that raw Lex30 scores might give an indication of temporal aspects of oral proficiency, whereas percentage scores might indicate specifically learners' lexical knowledge or lexical recourses in their lexicons.

Conclusion

The current study supports the proposed view of the relationship between productive vocabulary knowledge and L2 oral ability using various aspects of L2 speech measures and a vocabulary measure separately. We believe that the findings from this study provide implications not only for vocabulary assessment in the L2 classroom, but also future directions of researching vocabulary and L2 proficiency. We also acknowledge several limitations. In what follows, we discuss implications, future directions, and limitations.

Implications

Our findings demonstrate the potential usefulness of a productive vocabulary test as an assessment tool to obtain a broad estimate of learners' L2 oral ability in classroom settings. For diagnostic purposes, teachers can administer a vocabulary task (i.e., Lex30) at regular intervals and use such data as a broad indication of students' oral ability with which to monitor their progress in speaking. The validity of such an attempt remains to be seen and needs to be investigated, but the feasibility appears pedagogically appealing. Collecting human-rated speech scores (e.g., comprehensibility) requires a lot of time in order to elicit individual speech samples and to ask first language speakers to rate subsequently elicited speech data. Likewise, linguistic operationalizations (e.g., articulation rate) involve considerable amounts of time and effort for collecting, transcribing, and quantifying speech data. Administering the Lex30 task is, in contrast, time-efficient (i.e., 15 minutes) and user-friendly (http://www.lognostics.co.uk/tools/Lex30/index.htm for a computerized version of Lex30).

Future directions

First, methodology in measuring vocabulary and L2 speech separately deserves earnest consideration in future research. Although the vast majority of studies in this emerging research area compare both lexical and speaking scores derived from the same speech data (e.g., Crossley et al., 2015; Daller et al., 2003; Saito et al., 2016, 2017; Trofimovich & Isaacs, 2012), we want to highlight the concern that such traditional methodology may not serve the purpose of gaining deeper insights regarding the effect of what they have (vocabulary *knowledge*) upon what they actually do (vocabulary *use* in speech). Just as we found some interesting relationships between knowledge (measured by Lex30) and use (e.g., diversity and sophistication vs. appropriateness), this line of future research, by using different vocabulary measures (e.g., a gap fill task: Fitzpatrick & Clenton, 2017), advances our understanding of the relationship between vocabulary knowledge and oral ability from a different perspective.

Another potential future research direction relates to reappraisal in predominantly using frequency-based vocabulary measures. Though earlier Lex30 studies allude to the superiority of percentage scoring system as a highly frequency-based measure by subtracting the fluency factor (i.e., number of infrequent responses produced within 30 seconds, or number of blanks) from the raw scoring system, the current study suggests otherwise. Research evidence has accumulated to pose the question of a by-default mode of dependency on a frequency index alone to extrapolate learners' overall vocabulary knowledge, and it has put forward lexical indices of additional value alongside word frequency such as contextual diversity, multiword units, and psycholinguistic information (Kyle & Crossley, 2015; Saito,

2019). Following this line of reasoning, the Lex30 scores derived through the raw scoring system should not be deemed inferior to percentage scoring, but rather should be regarded as valid as a composite measure of multiple aspects encompassing lexical knowledge and fluency. Similarly, future studies should not equate learners' overall vocabulary knowledge to scores calculated on the frequency basis alone; rather, such studies should be based on a variety of scoring indices.

Limitations

First, we acknowledge that the responses produced in the Lex30 task are by no means transformed into proper estimates of productive vocabulary size. Given such limitations, it is not our intention to make conclusive statements about the relationship between productive vocabulary and speaking on the sole ground of the data in the current chapter. However, we feel that our findings based on Lex30 task usage, in combination with findings from earlier research using different formats of productive vocabulary tasks such as a sentence completion task (De Jong et al., 2012a), L1-to-L2 translation task (Koizumi & In'nami, 2013), and definition task (Vermeer, 2000), have the potential to provide additive insights into our understanding of the role of productive vocabulary in L2 oral ability.

Second, our findings are based on a single task (i.e., a picture narrative) to elicit speech samples. It is important to note that task characteristics such as task complexity, monologic or dialogic modes, and planning time have the potential to significantly impact upon oral performance. We acknowledge that the oral ability elicited via a picture narrative task was not comprehensive but rather limited in scope. Therefore, in subsequent investigations we intend to test the generalizability of this small-scale study (i.e., 40 Japanese first-year university students) with multiple task modalities and contexts (see De Jong et al., 2012a for varieties of task contexts), and with different population groups in the form of a larger-scale study.[1]

Note

1 This work was supported by the Grant-in-Aid for Scientific Research in Japan [grant number 26770202].

References

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing*, *18*(3), 191–208.

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279.

Boersma, P., & Weenink, D. (2013). PRAAT: Doing phonetics by computer (Version 5.3.51). Retrieved from http://www.praat.org/

Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In Housen, A., Kuiken, F., & Vedder, I. (eds) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam: John Benjamins.

Clenton, J. (2010). *Investigating the construct of productive vocabulary knowledge with Lex30*. (Unpublished doctoral dissertation.) University of Swansea, Swansea.

Clenton, J. (2015). Testing the revised hierarchical model: Evidence from word associations. *Bilingualism: Language and Cognition*, *18*(1), 118–125.

Clenton, J., Elmetaher, H., & Uchihara, T. (2019). *Capturing and quantifying individual differences in bilingual learners – A task-based study.* Paper presented at Capturing and Quantifying Individual Differences in Bilingualism conference, Tromso, Norway.

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *35*(5), 570–590.

Daller, H., Milton, J., & Treffers-Daller, J. (2007). Editors' introduction: conventions, terminology and an overview of the book. In Daller, H., Milton, J., & Treffers-Daller, J. (eds) *Modelling and assessing vocabulary knowledge* (pp. 1–32). Cambridge: Cambridge University Press.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, *24*(2), 197–222.

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 113–132.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012a). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*, 5–34.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012b). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In Housen, A., Kuiken, F., & Vedder, I. (eds) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 121–142). Amsterdam: John Benjamins.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.

Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, *27*, 537–554.

Fitzpatrick, T., & Clenton, J. (2017). Making sense of learner performance on tests of productive vocabulary knowledge. *TESOL Quarterly*. Advance online publication. doi: 10.1002/tesq.356

Fitzpatrick, T., & Meara, P. M. (2004). Exploring the validity of a test of productive vocabulary. *Vigo international journal of applied linguistics*, *1*, 55–74.

Flege, J., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*(5), 3125–3134.

Harrington, M., & Carey, M. (2009). The on-line Yes/No test as a placement tool. *System*, *37*, 614–626.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in second language acquisition*, *21*(2), 303–317.

JACET (2003). *JACET list of 8000 basic words*. Tokyo: Japan Association of College English Teachers.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research, 4*(5)*,* 900–913.

Kormos, J. (2006). *Speech production and second language acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, *13*(4), 377–392.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757–786.

Laufer, B., & Levitzky-Aviad, T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *Modern Language Journal*, *101*(4), 729–741.

Laufer, B., & Nation, I. S. P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322.

Laufer, B., & Nation, I. S. P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392.

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, *19*(6), 741–760.

McNamara, D. S., Grasesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. New York: Cambridge University Press.

Meara, P. M. (1996). The dimensions of lexical competence. In Brown, G., Malmkjaer, K., & Williams, J. (eds) *Performance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.

Meara, P. M., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, *28*(1), 19–30.

Meara, P., & Jones, G. (1990). *The Eurocentres Vocabulary Size Tests. 10KA*. Zurich: Eurocentres Learning Service.

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. In Vedder, I., Bartning, I., & Martin, M. (eds) *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211–232). Eurosla Monographs Series 1. Amsterdam: Eurosla.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In Chacón-Beltrán, R., Abello-Contesse, C., & Torreblanca-López, M. (eds) *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching*, *56*(1), 1–24.

Munro, M. J., Derwing, T. M., & Burgess, C. S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, *52*, 626–637.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle Cengage Learning.

Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers*, *4*, 37–45.

Préfontaine, Y., & Kormos, J. (2016). A qualitative analysis of perceptions of fluency in second language French. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 151–169.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*(3), 513–536.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Saito, K. (2019). Multi- or single-word units? The role of collocation use in comprehensible and contextually appropriate second language speech. *Language Learning*.

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439–462.

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations. *Studies in Second Language Acquisition*, *38*(4), 677–701.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*(4), 577–607.

Suzuki, S., & Kormos, J. (2019). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition.*

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In Ellis, R. (ed.), *Planning and task performance in a second language* (pp. 239–273). Amsterdam: John Benjamins.

Tavakoli, P., & Uchihara, T. (2019). To what extent are multiword sequences associated with oral fluency? *Language Learning*.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916.

Uchihara, T., & Clenton, J. (2018). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*.

Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *Language Learning Journal*, *47*(1), 64–75.

Uchihara, T., Eguchi, M., Clenton, J., Kyle, K., & Saito, K. (under review). To what extent is collocation knowledge associated with oral proficiency? *Corpus-based approach to word association*.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, *17*(1), 65–83.

Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Language Assessment Quarterly*, *9*(2), 172–185.

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, *65*, 139–150.

Yuan, F., & Ellis, R. (2003). The effects of pre-task and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1–27.

13 Vocabulary and speaking

Future research, tools, and practice

Jon Clenton

Speaking is often conceptualized as consisting of multiple aspects (e.g., De Jong et al, 2013), but, in such studies, vocabulary is not. Studies designed to compare aspects of speaking with vocabulary knowledge (e.g., De Jong et al., 2013, De Jong and Mora, 2017; Uchihara and Clenton, 2018; Uchihara and Saito 2016) present various aspects of speaking ability (e.g., aspects of fluency include pausing data such as the number of filled or silent pauses an English user produces) in comparison with vocabulary as a single task score. Various attempts to describe vocabulary knowledge (e.g., Daller et al., 2007; Milton and Fitzpatrick, 2014; Nation, 2001, 2013), however, indicate that the construct of vocabulary knowledge is multifaceted (and includes aspects of knowledge that include, for instance, spoken productive knowledge of form, receptive knowledge of collocations, and so on). We therefore propose that future studies in which comparisons are made between speaking and vocabulary knowledge should include multiple aspects of both. While researchers appear, however, to recognize the importance of the need to investigate 'multiple word components' (e.g., González-Fernández and Schmitt, 2019; Webb 2005, 2007), only a handful of studies (e.g., Saito, 2019) explore the extent to which such multiple aspects of lexical knowledge relate to spoken output.

Fitzpatrick and Clenton's (2017) paper represents a potential platform from which to explore a multifaceted approach to investigate the productive vocabulary used in speaking. Designed to assist interpretations of the construct of productive vocabulary knowledge, Fitzpatrick and Clenton (2017) present a 'capture model' to graphically represent the potentially different knowledge 'captured' by various productive vocabulary tasks. Organized in terms of quantity and quality (of vocabulary knowledge), this specific model has the potential to inform the extent to which elicited aspects of vocabulary knowledge relate to speaking constructs. Chapter 11 in this volume represents a first attempt, to our knowledge, to employ the capture model in terms of how such knowledge might relate to oral ability. Clenton et al. consider the potential relationships between vocabulary knowledge 'captured' by different vocabulary tasks, reporting how task scores might relate to fluency measures (De Jong 2013, 2015; De Jong & Mora, 2017). They find that specific vocabulary task (Lex30; Meara & Fitzpatrick 2000) performance relates to specific fluency aspects,

to the extent that productive vocabulary knowledge, for the pre-intermediate participants in their study, relates to specific aspects of fluency (silent pauses). Uniquely, they also report the extent to which various 'vocabulary skills' relate to the aspects of fluency. Their chapter offers a first attempt to investigate the extent to which skills such as speed or delays in response might relate to spoken output. Such findings might be of potential interest because they represent the beginning of a potential scale of spoken vocabulary knowledge (e.g., pre-intermediate users demonstrate specific vocabulary knowledge as indicated by one vocabulary task, along with specific 'vocabular skills'; other studies allude to higher productive vocabulary task scores being indicative of higher proficiencies and therefore 'more fluent' participants [e.g., De Jong et al., 2013]). Clenton et al., however add an intriguing additional component of 'vocabulary skills'. We wonder whether such findings can be extended to relate to the kinds of abilities one might expect of an scale of spoken vocabulary knowledge. For instance, González-Fernández and Schmitt (2019: 17) suggest a 'reliable implicational scale of *written* vocabulary knowledge', to the extent that they report that 'the written word knowledge components do seem to be ordered in how well they are known' (p. 18), and so we wonder whether such a scale might be possible for spoken vocabulary knowledge. In addition to exploring whether we might be able to devise an implicational scale of spoken vocabulary knowledge, it would be useful to determine the extent to which such a scale might relate to the other three skills (i.e. the extent to which implicational scales of written, spoken, read, or written vocabulary knowledge relate to one another). An additional line of enquiry relates to changes according to speaking task type. Clingwall, Clenton, and Fraser (forthcoming) present a study, in part based on Clenton et al. (Chapter 11 in this volume) in order to examine the extent to which vocabulary knowledge task performance relates to each of the three different IELTS (monologic, quasi-dialogic, and dialogic) speaking tasks. Adding other vocabulary tasks (e.g., the Productive Levels Task [Laufer and Nation, 1999]) to such comparisons would likely guide research towards an implicational scale of spoken vocabulary knowledge.

A further concern related to the use of corpora relates to the use of monolingual corpora to evaluate bilingual performance. Some have gone as far as to suggest that using first language speaker corpora with second language use is 'like comparing apples and oranges' (Slabakova, 2013: 53–4). Monteiro, Crossley, and Kyle (2018) argue, in a paper evaluating L2 written performance, that 'lexical norms from non-native speaker corpora should … represent the language acquired by L2 learners more accurately' (p. 5) not least because L2 users are influenced by their L1 lexicon, and L2 users are more likely to be influenced by formulaic language frequency than L1 users (Monteiro et al., 2018). Monteiro et al. report that their use of L2 corpora 'explained more variance' in their L2 participant group compared to L1 corpora. Such findings, however, are based on written L2 corpora, and we suggest a goal for speaking studies would be to evaluate L2 speaking performance from L2 spoken corpora.

As I began, this brief chapter is not anticipated to provide an exhaustive list of research that relates vocabulary knowledge to that of speaking ability. I hope to have outlined a number of current trends that have potential to be addressed by researchers in the field. The following research questions can, of course, be taken as they currently stand, but can also be developed to suit individual researcher needs and interest.

Potential research questions

1. Does an exploration of multiple aspects of vocabulary knowledge account for the vocabulary used in spoken output?
2. Is it possible to formulate an implicational scale of spoken vocabulary knowledge and vocabulary skills?
3. To what extent does an implication scale of vocabulary knowledge relate to other implicational skills (i.e. reading, writing, listening)?
4. To what extent does analysis of spoken output differ in comparisons between corpora from monolingual and bilingual data?

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, *24(*4), 425–438, https://doi.org/10.1093/applin/24.4.425

Adolphs, S. & Schmitt, N. (2004). Vocabulary coverage according to spoken context. In Bogaards, P., and Laufer, B. (eds) *Vocabulary in a second language* (pp. 39–49). Amsterdam: John Benjamins Publishing.

Clingwall, D., Clenton., J., & Fraser, S. (forthcoming). *Does speaking task type influence the vocabulary L2 speakers use?*

Daller, J. Milton, & J. Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916.

De Jong, N.H., Groenhout, R., Schoonen, R. & Hulstijn, J.H. (2015), Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223–243. doi: 10.1017/S0142716413000210

De Jong, N., & Mora, J. (2017). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, 1–13. doi:10.1017/S0272263117000389

Fitzpatrick, T., and Clenton, J. (2017), Making sense of learner performance on tests of productive vocabulary knowledge. *TESOL Quarterly*, *51,* 844–867. doi:10.1002/tesq.356

González-Fernández, B., & Schmitt, N. (2019). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics.* https://doi.org/10.1093/applin/amy057

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. https://doi.org/10.1177/026553229901600103

Meara, P., & Fitzpatrick, T. (2000). Lex 30: An improved method of assessing productive vocabulary in an L2. *System*, *28*, 19–30.

Milton, J., & Fitzpatrick, T. (2014). *Dimensions of vocabulary knowledge*. Basingstoke: Palgrave Macmillan.

Monteiro, K., Crossley, S., & Kyle, K. (2018). In search of new benchmarks: Using L2 lexical frequency and contextual diversity indices to assess second language writing. *Applied Linguistics.* https://doi.org/10.1093/applin/amy056

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

Nation, I. S. P. (2013). *Learning vocabulary in another language*. 2nd ed. Cambridge: Cambridge University Press.

Saito, K. (2019), To what extent does long-term foreign language education help improve spoken second language lexical proficiency? *TESOL Quarterly*, *53,* 82–107. doi:10.1002/tesq.468

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Slabakova, R. (2013). Adult second language acquisition: A selective overview with a focus on the learner linguistic system. *Linguistic Approaches to Bilingualism*, *3*(1), 48–72. (doi:10.1075/lab.3.1.03sla).

Uchihara, T., & Kazuya, S. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability, *Language Learning Journal*, *47(*1), 64–75, doi: 10.1080/09571736.2016.1191527

Uchihara, T., & Clenton, J. (2018). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*. https://doi.org/10.1177/1362168818799371

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27,* 33–52.

Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*, 63–81.

Part V
Writing

14 Vocabulary and writing

Current research, tools, and practices

Paul Booth

Measures of lexical sophistication

How vocabulary contributes to the sophistication and richness of a written text can be measured in different ways. Meara and Bell (2001: 6–7) coin the terms 'extrinsic measures' of lexical richness versus 'intrinsic measures' of lexical variety, respectively, in order to highlight the difference between external based criteria based on frequency lists and internal criteria based on the text itself. The first part of this overview focuses on the extrinsic measures of lexical sophistication, and the second on intrinsic measures.

Lexical sophistication: frequency profiles

The Lexical Frequency Profile (LFP) analyses the words used in a text. The output is a profile of the percentage of lexis at various frequency bands: the first 1,000 most frequent words (1k), the second (2k), the University Word List, and 'not in lists' (Laufer and Nation, 1995). The LFP was able to highlight the more advanced learners' productive knowledge of rarer words. The results from Laufer & Nation's (1995) suggestion that for advanced learners, that is, post Cambridge First Certificate, their use of lexis is sensitive to differences in topics. The LFP has also been used with measures of passive vocabulary size: the Vocabulary levels Test (Nation, 1990), the productive version of the Vocabulary Levels Test (Laufer & Nation, 1999) for controlled active vocabulary size. Laufer and Paribakht (1998) found that learners' passive vocabulary, controlled active, and free written vocabulary, as measured by the LFP, developed at different rates in which productive vocabulary was slower and less predictable than the development of passive vocabulary. A more advanced lexical measure, the WebVocabprofile at www.lextutor.ca (Cobb, n.d.) also measures lexical frequency but with a wider range of options.

One feature of this type of lexical profile is that words are analysed individually regardless of whether they form part of what Wray (2002: 200) terms a 'holistically learnt string'. For example, the software would analyse 'I don't know' word by word rather than as a complete whole. As such, categorizing the productive vocabulary from a written text into various frequency levels on

a word by word basis may give a false impression if the various lexical items are formulaic chunks of language composed of words at various frequency levels. The other problem (as highlighted by Bogaards, 2000) is that a word is defined in the program as a base form with its inflected and derived forms, with the result being that the program does not pick up incorrect derivatives or inflections.

P-Lex

Although this software shares some similarities with Web Vocabprofile because they both use frequency lists to determine the rarity of a word, the similarity ends there. Meara's *P-Lex* (Meara, 2007a) is not text-length-dependent. Meara (2007b: 1–2) explains that the software calculates lexical richness by analysing 10 word segments of a text and then counts the number of 'difficult' words in the text. It then calculates the number of blocks containing difficult words and the probability of this happening. The 'difficult' words are those which are not found in the list of high frequency words which are listed in the P-Lex Manual. The statistical measure which P-Lex uses is the Poisson distribution which is calculated from the formula below.[1] The Poisson distribution describes the likelihood of rare events occurring. In this context, though, the key factor is the distribution of certain that is, 'difficult' words occurring in a length of text. The program calculates the closest fitting Poisson curve and reports this curve by means of a central parameter, (λ) lambda. The output profile displays the proportion of 10-word segments which contain 0 difficult words, 1 difficult word, 2 difficult words, and so on. Putting aside the differences in mathematical calculations of lexical richness of the VocabProfile and P-Lex, the P-Lex is less 'wasteful' of learners' texts.

N-*grams*

In order to understand the extent to which writers use words that are connected, *N*-grams are a measure of words which normally occur next to each other using corpora as a baseline; for example, bi-grams are *of the* and *in the*. Kim, Crossley, and Kyle (2018) researched the use of *n*-grams in written L2 English. Their study highlighted that writers who are judged to be more lexically proficient used a greater proportion of *n*-grams and advanced (low frequency) content words. The results indicated that higher L2 proficiency was related to bi-grams and tri-grams that are more strongly associated and frequently used (p. 133).

Intrinsic measures: lexical diversity and type token ratio

Another way of measuring lexical sophistication in writing, rather than measuring the text against frequency data, is to measure the text using type token ratio (TTR). One of the most common measures of lexical variability

(or diversity) is traditionally conceptualized as the number of different words (word types) used in a text or transcript, or in terms of the relationship between the number of types and text length. This has been calculated by type-token ratio (TTR) that is, the number of word types divided by the number of word tokens.

Problems with TTR have been well documented (e.g., Jarvis, 2002). One fundamental issue is that the TTR falls as the number of words increases. A person theoretically only has a finite amount of words at their disposal and so as the text increases in tokens, then the likelihood of repetition of tokens of the same type increases. This is why the TTR is high to begin with when there is less repetition but then gradually decreases over a larger sample of words. Malvern et al. (2004) tackle this phenomenon by producing a method of measuring lexical diversity that is a measurement made over a series of points in order to establish the pattern of fall of the curve rather than any particular value on it (Malvern et al., 2004: 59). Parameter D^2 (for diversity) calculates a mean segmental TTR for a random selection of words from the text.

The statistic which is calculated is not any particular point on the curve but it is the pattern of fall of the curve which is calculated. The parameter is a mathematical ideal curve which is the closest fitting curve to the actual TTR curve from real language. The program (*vocd*) 'can read a transcript of the language sample, then plot the TTR verses tokens curve between $n = 35$ and $n = 50$, deriving each point from an average of 100 trials on sub-samples of words of the token size for that point' (Malvern et al., 2004: 55). Skehan (2009: 108) describes the *D* value as 'an index of the extent to which the speaker [or writer] avoids the recycling of the same set of words'.

This measure of lexical diversity has been used in a cross-sectional study (Malvern et al., 2004: 153–76) of nearly one thousand narrative compositions written by English school children of the ages 7, 11, and 14 years. One of the aims was to look at the relationship between lexical diversity and the quality of writing as assessed in accordance with the National Curriculum guidelines. Lexical diversity, as measured by *D*, was sensitive to writing quality and showed continuous development across levels in writing as defined under the National Curriculum.

External and internal measures of lexical sophistication

External measures of lexical richness using frequency data (e.g., P-Lex and Web Vocabprofile, *n*-grams) and internal measures (e.g., Parameter *D*) have their advantages and disadvantages. External measures do not take into account repetition of word tokens and so do not discriminate between learners who repeat rare tokens of the same type and those who use rare tokens of different types. Thus, it would be possible to inflate the lambda score or lexical profile of rare words simply by repeating a small number of rare words. In the case of lexical diversity the same learners would achieve a low score.

The distortion which comes from frequent use of rare words is particularly acute when the sample population uses technical vocabulary to the extent that what is rare in one environment may not be in another. Another issue is that words used in general English may also have a technical meaning (Coxhead & Demecheleer 2018: 86). Measures based on frequency counts from large corpora tend not take into account the frequency of words in any particular environment. Student engineers, for example, may be exposed to technical vocabulary which, for them, is highly frequent and may not be perceived as difficult. However, this type of low frequency lexis may have a limited range and so probably would be considered as rare (i.e., beyond the 2000 frequency level) because technical words tend to not appear outside of a certain environment. Consequently, a written sample of L2 English from students studying engineering may include a disproportionate amount of rare lexis which could give a false impression of the lexical sophistication of the student.

A critical review by Meara and Bell (2001: 6) of an internal measure has pointed out that diversity, in this case *D*, does not take into account the difficulty of the words. Their argument is that diversity measures do not take into consideration the rarity of the word. So, for example, sentences which differ in terms of lexical rarity but have the same TTR would be statistically similar. Malvern et al. (2004: 124) defend their measure of lexical diversity by arguing that the proportion of words that are rare is a function of the number of different words, which is in turn a function of the number of tokens. So overall no measure of the sophistication of a text can take into account all the different levels of lexical richness, so it is important to take into account what factors are to be measured in learners' written text.

Notes

1 $P_N = (\lambda^N * e^{-\lambda})/N!$ (Meara 2007b, p.1)
2 $TTR = D/N * [(1 + 2 * N/D)^{1/2} - 1]$ N = the number of word tokens.

References

Bogaards, P. (2000). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *23*(3), 321–344.

Cobb, T. (n.d.). Web Compleat lexical tutor. Retrieved from www.lextutor.ca/vp/

Coxhead, A., and Demecheleer, M. (2018). Investigating the technical vocabulary of plumbing. *English for Specific Purposes*, *51*, 84–97. doi: 10.1016/j.esp.2018.03.006

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57–84.

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development and writing quality.*Modern Language Journal*, *102*(1) doi: 10.1111/modl.12447

Laufer, B. (1990). Why are some words more difficult than others? Some intralexical factors which affect the learning of words. *IRAL*, *28*, 293–307.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322.

Laufer, B., & Nation, I. S. P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*, 33–51. doi: 10.1191/026553299672614616

Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, *48*(3), 365–391.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. Basingstoke: Palgrave Macmillan.

Meara, P. M. (2007a). *P_Lex. v 2.0.* Swansea: Lognostics.

Meara, P. M. (2007b). P_Lex The Manual [Online] www.lognostics.co.uk/tools/index.htm [Accessed 26 October 2007]

Meara, P. M., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, *16*(3), 5–19.

Nation, I. S. P. (1983). Teaching and testing vocabulary. *Guidelines*, *5*(1), 12–25.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Heinle & Heinle.

Skehan, P. (2009). Lexical performance by native and non-native speakers on language learning tasks. In Richards, B,. Daller, H., Malvern, D., Meara, P., Milton, J., & Treffers-Daller, J. (eds) *Vocabulary studies in first and second language acquisition: the interface between theory and application* (pp. 107–124). Basingstoke: Palgrave Macmillan.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

15 Specialized vocabulary in writing
Looking outside ELT

Averil Coxhead

Introduction

I first came across builders' diaries when I was working on a research project called 'Language in the Trades Education' (LATTE) project with my colleague, Jean Parkinson, and colleagues from WelTec, a polytechnic in Wellington (New Zealand), James Mackay and Emma McLaughlin. Part of the research in this project focused on technical vocabulary in the trades (see Coxhead, 2018: Chapter 8), specifically on identifying this vocabulary in spoken and written texts in two construction trades (plumbing and carpentry) and two engineering trades (automotive technology and fabrication). All students in carpentry courses were required to write a builder's diary and 55 of these diaries were collected as part of the LATTE project. Using vocabulary in writing or speaking is an important aspect of learner knowledge (Nation, 2013), and research in second language learners' vocabulary use in writing shows that it is often not an easy task. It is seen as 'risky' (Laufer, 2003), and can be affected by the first language (L1) of the writer, time on task, the topic of writing, and the expectations of a teacher or an activity (Coxhead, 2011, 2012) in an EAP context. The diaries, then, are an opportunity to find out more about the use of technical vocabulary in writing by students in trades education.

With a long-term interest in the use of vocabulary in writing by second/foreign language users of English, predominantly in EAP, my attention was drawn to the student diaries. They had some features which I thought were particularly important for developing writing skills and opportunities for vocabulary use. These features of the diaries are that they were: mandated, a daily occurrence, intended to mirror professional writing, multi-modal, and an opportunity to use vocabulary productively in writing in a technical document.

In his chapter on 'Helping Learners Write' in *Teaching EFL/ESL reading and writing* (2009), Paul Nation provides a range of principles for the teaching of writing. A core element of these principles is the Four Strands (Nation (2007) of meaning-focused output (learning through writing and speaking), meaning-focused input (learning through reading and listening), language-focused learning, and fluency. I would like to focus on the elements provided by Nation (2009) that relate to meaning-focused output in particular, because they relate closely to the concept of the builders' diaries (see below). By

connecting the diaries and the principles from Nation, I hope to build an argument as to why they might be useful for ELT professionals.

With an eye to understanding how the builders' diaries written by students might be beneficial for the development of writing skills and production of vocabulary in particular, here is an abbreviated set of principles from Nation (2009: 33–4) about meaning-focused output in relation to writing. Note that some of these points are direct quotes and some are summaries. This is because, many times, Nation's key ideas are already succinct and direct.

- 'Learners should do lots of writing and lots of different kinds of writing' (Nation, 2009: 93).
- Writers need to have an audience in mind and aim to communicate a meaning to that audience.
- 'Writing should interest learners and draw on their interests' (p. 93).
- There should be a feeling of success in the writing, as much as possible.
- 'Learners should use writing to increase their language knowledge' (p. 93).

The student diaries in this study are overwhelmingly handwritten and the context for the writing is outside a language-focused course of learning, which means that the principles from Nation (2009) that learners need to develop computer-based writing skills and the role of needs analysis in writing instruction are not so relevant. With our colleague in the US, Susan Conrad, shining light on student and professional writing in civil engineering (see Conrad, 2017, 2018; n.d.), I was also interested in researching the technical lexical elements of the builders' diaries from the LATTE project and to look at ways that these diaries might be adapted for learners, teachers, and courses in English language teaching outside the carpentry context at WelTec. That's why the title of this chapter is 'Specialized vocabulary in writing: Looking outside ELT'.

Using vocabulary in writing

Corson (1985: 110) makes a strong statement about vocabulary knowledge when he writes, 'knowing a word means knowing how to use it'. Use of vocabulary (including aspects such as grammatical function, collocations, and any constraints of use) is one of three types of knowledge in Nation's (2013) framework of form, meaning, and use. Form in writing refers to how a word is written/its spelling and meaning includes the form and meaning connection, any concepts which a word refers to, and the associations it has. This framework suggests that that there is much to learn about using a word in writing (Nation, 2013). This is an important point for teachers. Using words in writing is one of the most difficult skills for language learners, according to Laufer and Goldstein (2004).

It is well known that learners know or can recognize more words than they use in English (see Malmström, Pecorari, and Gustafsson, 2016). We also

know that native speakers need to develop understandings and use of technical vocabulary in disciplines (Coxhead, 2018). For anyone who has written anything in a language that they have had to work hard at learning, using vocabulary in writing in a second language can be anxious-making and difficult (Coxhead, 2011). In my research in an EAP context, I found that learners in an EAP course might actively avoid using words in their academic writing for many reasons. For example, they might not feel able to connect a word to the topic that they are writing about, they may get confused about words that look similar, it might be difficult to gauge the audience or the register for writing, or they might simply lack the background knowledge of a topic and therefore the vocabulary required to write about it (Coxhead, 2011). One consequence of avoiding using particular words in writing is that learners might resort to using high frequency words that they know well rather than taking a risk with lesser known vocabulary. The difficulty with this practice, as Corson (1997) notes, is that the substituted word may be 'less-adequate, or simply wrong' (pp. 698–9). It is difficult to look in a corpus to find words which learners have not used.

In a study on the use of items from Gardner & Davies's (2014) Academic Vocabulary List (AVL) of 3,014 lemmas by student writers in the British Academic Written English (BAWE)[1] corpus, Durrant (2016) found a large number of items were shared between the list and the corpus. He also found a core 427 items that occurred in more than 90% of the student writing across disciplines, indicating that the writers relied on a small number of items in their writing. Writers who had more years of study at university used more AVL items than those writing early on in their academic studies. Durrant (2016) also found variation in the amount of use of AVL items depending on disciplines. This research is an example of examining vocabulary use in writing, drawing on a specialized word list, which is a focus of this chapter also. Let's now turn to the diaries themselves.

The builders' diaries

As part of the carpentry course at WelTec, students have to maintain a diary of their work, as they would do once they had finished their course and were working in the profession. These diaries contain regular accounts of the classwork and building site work in a course where they build a house over of a year (working alongside plumbing students, amongst others). These houses are then sold and moved so the next year's students can use the same building site for their houses. The diaries are assessed as part of the course and are modeled on diaries that builders keep in their everyday professional work. They include pictures and diagrams, as well as short passages of writing. Students work on diaries at Level 3 (the first year of their qualification) and at Level 4 (second year).

Figure 15.1 shows an example of the first page of a diary from a student at WelTec. Note the activity on the left about mathematic formulas for carpentry and the writing, drawing, and picture of people measuring out

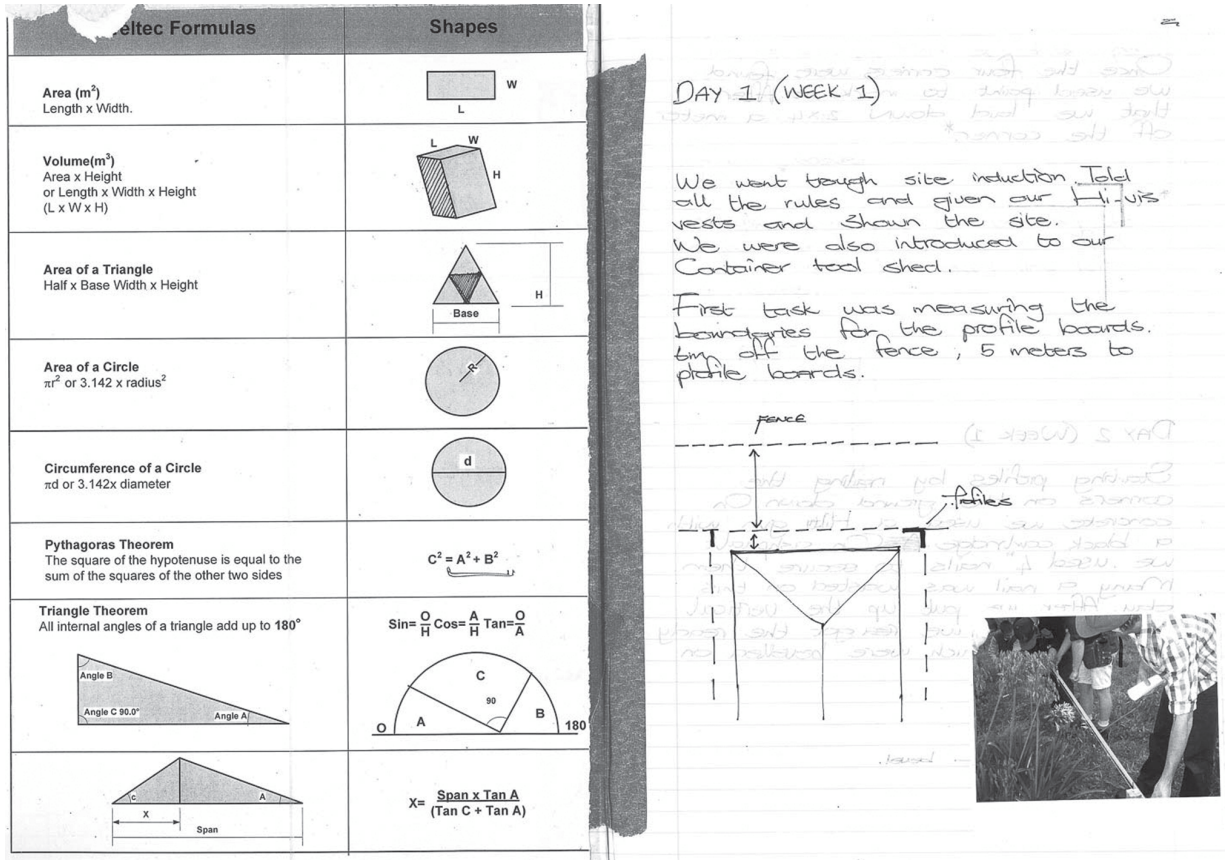


Figure 15.1 An extract from a Builder's Diary from a WelTec carpentry student.

boundaries on the right. It is important to state at the outset that students in these classes could be first, second, or foreign users of English. No distinction is made in this chapter between these writers, largely because of the imbalance between the groups, but also because, in essence, it is the diary writing activity itself which is of primary interest here. Note that the diaries also contain a small proportion of texts which are shared because some students used their diaries at times to take notes from the board in class. The focus of the analysis in this chapter is the writing by the students themselves, rather than the whole texts.

Parkinson et al. (2017b) interviewed carpentry tutors about the diaries and identified six purposes for the builder's diary, four of which are related to students. They are: assessment as part of courses, keeping records of how they carried out a particular piece of work so they could check back when needed, developing a writing habit to carry forward into work after training, and providing evidence of the work that they have carried out to show to potential employers. The other two purposes relate to professional builders: to provide a record for billing customers for the number of hours which have been done, and to provide as evidence in case there is a dispute in court.

An interesting analysis of the diaries so far has involved looking closely at the photos (Parkinson et al., 2017b). The findings suggested that as the course progressed, the photos became less about the people in the pictures and more about the work being recorded. For Parkinson et al. (2017b), this change was evidence for a growing builder identity for the students. A qualitative analysis of 44 student diaries, reported in Parkinson et al. (2017a), showed development in writing by students from a personal account (similar to paragraph one in Figure 15.1) with photographs of classmates on site (see also Figure 15.1) through to a more impersonal diary from a Level 4 student with a concise list of the work that had been undertaken and the date, e.g. *9/5/13 Machine all frame joints/Prime* (Parkinson et al., 2017a: 67).

Readers might have already noticed some familiar words in that diary entry being used in a way that is not familiar (without a background in carpentry, that is): *machine*, *frame,* and *prime* as verbs, and *joints* in a *frame*. These examples point to the technical vocabulary of carpentry (see below). I wondered whether the students who had to write diaries focused on vocabulary in any way, perhaps in making sure that they used the correct technical terms in their writing or if they consciously used the diaries to keep track of technical vocabulary from their courses. So, this aspect of the diaries is part of this chapter, reporting on data from interviews with students and questionnaires.

Specialized vocabulary in carpentry

One of the main focuses of the LATTE project was identifying technical vocabulary that students were exposed to and expected to learn in their trades courses at Weltec. McLaughlin & Parkinson (2018) investigated language-related episodes

in the tutor discourse from the LATTE project and found that vocabulary-related episodes were more frequent in theory-based classes than on the site where the students were building houses. In a way, this finding echoes those from an analysis of technical vocabulary in spoken vs written texts in carpentry in the LATTE project, where we consistently found higher levels of technical lexis in written texts compared to spoken texts (see below).

Typically, technical vocabulary might be expected to occur mostly inside a field or be known by people who have studied or worked in that field (Chung & Nation, 2003). That said, there are also everyday words which can have specialized meanings in a particular context such as *host* or *string* in computer science (see Coxhead, 2018). Chung and Nation (2003) used a semantic scale to identify technical vocabulary in an anatomy textbook and in an applied linguistics textbook, finding that up to one word in three in a sentence could be technical in the anatomy textbook, and one word in four in a sentence in the applied linguistics textbook. Recent research from the LATTE project has found that more than 30% of the vocabulary in a written text in fabrication could be technical (Coxhead et al., 2019). Clearly then, one avenue for research for the builders' diaries from WelTec is to investigate the use of technical vocabulary by students in these texts. To do this research, we need to first look at the Carpentry Word List (Coxhead et al., 2016) which was developed as part of the LATTE project.

The Carpentry Word List

The Carpentry Word List (Coxhead et al., 2016) was developed using a quantitative analysis of corpora that we gathered for the LATTE project, and a qualitative analysis by carpentry tutors from WelTec. We asked the tutors to help us with decisions about technicality of words, in particular, those which had general and possibly technical meanings. For example, we wanted to know if *wall* was a technical word according to these experts. The Carpentry Word List (Coxhead et al., 2016) contains 1,424 types, and is accompanied by a list of technical abbreviations and a list of technical proper nouns.

This word list is based on types, or single words, because research into technical vocabulary, in particular Chung's work in medical English with Paul Nation (see above), found that while some members of a word family might be technical, other members of the family might not. An example from Coxhead et al. (2016) to illustrate this point is that while *flashing* or *flashings* are technical words in carpentry (flashings are strips of metal which are used to stop water coming into a building where a roof joins another surface), their family member *flash* is not. Table 15.1 contains some examples of high frequency types from the Carpentry Word List, along with some abbreviations and proper nouns. These examples show how high frequency words in English can also be technical words (Nation, 2016), for example, *building* and *wall*. Note also that *fixing* is a technical word in carpentry, that does not mean *repairing*. Instead it refers to items that are used to hold or connect materials in building, such as bolts, screws, and nails.

Table 15.1 Examples of technical words from the Carpentry Word List

12 most frequent types in the Carpentry Word List	*12 technical abbreviations*	*Meanings of the technical abbreviations*	*12 technical proper nouns*
requirement(s)	FFL	Finished Floor Level	Aqualine
figure	FFT	Flexible Flashing Tape	Braceline
building	FGL	Finished Ground Level	Branz
wall	FOPS	Falling Object Protective Structure	Ecoply
timber	FSP	Fibre Saturation Point	Ezybrace
roof	GALV	Galvanised	Flexibrace
concrete	GIB	Gibraltar Board	Fyreline
installation	GLULAM	Glued Laminated Timber	Gantt
construction	GRC	Glass Reinforced Concrete	Goldline
fixing	H1	Building Code Clause H1 (Energy Efficiency) H1.1 H1.2 H3.1 H3.2 H4 H5 = hazard class that	Handibrac
calculation	H1.1	determines LOSP[1] treatment	Hiab
activity	HBG	House Building Guide	James Hardie

1 LOSP = Light Organic Solvent Preservative

The word list was originally based on a written corpus, and when the spoken corpus for the LATTE project became available, we updated it with lexical items which met the selection criteria from the spoken texts.

Coxhead et al. (2016) found that listening was a particularly demanding skill for carpentry students, as they have to follow sometimes quite complex instructions from tutors with multiple steps, and try to remember everything. Chan (2013) points out that much learning in trades education takes place through watching, copying, and practising. Figure 15.2 shows an example of a carpentry tutor <T:> taking a class on installing insulation (in this instance, pink batts) in the walls and ceiling of a house. Note the here and now nature of the instruction overall (e.g. *through here*, *like that*), and how the students <S:> have to follow a number of steps to carry out the task.

We used the Carpentry Word List to find out the proportion of technical types, abbreviations and proper nouns in the spoken (tutor talk, like the example in Figure 15.2) and written (textbooks, teaching materials, and workbooks) texts from the carpentry courses at WelTec. We found that the coverage of the written texts at 38.35% was much higher than the coverage of the spoken texts at 10.69% (Coxhead et al., 2020: 113). The technical abbreviations covered 1.44 of the written texts and the proper nouns covered

<T:> What happens is that normally you just run your knife through it. Where's your knife? No, I mean over here... just through along the pipe... Just run the knife through there.
<S:> <unintelligible>
<T:> The other side, not the fluffy side... and then you ping back through there like that and that comes down over that pipe, so you just cut those along where the pipes are, especially those ones there cos they're up higher.
<S:> Yeah
<T:> So, you just run a knife through the side of it... through there... through there... and just pull those down around it like that... This one needs a cut around there like that, alright? Like... just throw something in there I think...
<S:> We need to cut the bottom of the...
<T:> Oh yeah, so what you do is run your knife alongside the pipe...
<S:> <unintelligible>
<T:> Where <unintelligible> knife?
<S:> Over there.
<S:> Using that one as it is...
<T:> Oh no using the blade.
<S:> Yeah, it won't go in.
<T:> Just run the knife like this, where... Just grab me another stick, yeah... That's good... It's nice and tight. Just cut it... Follow that cut the insulation on the other side, on the side where the pipe is.

Figure 15.2 An example of carpentry instruction on installing insulation in a house.

0.31% (Coxhead et al., 2020: 113). The word list without the abbreviations and proper nouns covered 36.71% of the written corpus and 10.3% of the spoken corpus. This comparison between spoken and written texts is important, because learning through listening and doing is a vital part of trades-based training as we have already seen, and reading is considered problematic in this context (see Coxhead et al., 2020). Furthermore, if we want to analyse the vocabulary use in student writing, we need to take into account the fact that they encounter much more speaking than writing in their studies, and that the diary is the main form of writing in the course. That is, it is the main evidence that we might have of written technical vocabulary in use for these students.

Research questions

1. Do the student writers see the diaries as useful for their technical vocabulary learning?
2. What proportion of the student diaries does the Carpentry Word List cover?
3. Is this coverage similar over high scoring and low scoring diaries?

4. How does the coverage of Carpentry Word List over the diaries compare to pedagogical documents used at WelTec and written by professionals in the field?

Methodology

The builders' diaries

A total of 55 diaries were gathered in the LATTE project, with permission from the diary writers. These texts were scanned or copied, changed into electronic form, and checked carefully in consultation with a carpentry tutor where necessary.The diaries ranged in length from a very short 800 words through to 11,000 words (Parkinson et al., 2017a: 47). The total number of running words in the diaries was just over 227,000. All diaries were kept for the current analysis, even the shortest one.

Parkinson et al. (2017b) worked with tutors from carpentry to identify diaries that they considered to be higher or lower quality, using a four-point scale. This process divided the diaries into two main groups, with the higher diaries (2 and 4 on the scale) totalling nearly 125,000 running words and the lower diaries (scoring 1 and 2 on the scale) containing just over 102,000 running words. It needs to be noted that the students also used their diaries at times to record notes from theory classes, so some diaries on some days may contain a short amount of the same text. Not all tutors treated the diaries in the same way, with some tutors taking a stricter approach than others.

Participants, interview, and questionnaire data

Ten carpentry students responded to the questionnaires and three took part in interviews. The questionnaires were filled in after class on campus, and the responses reported here just pertain to the diary writing exercise. The interviews were also conducted there with a colleague from the LATTE project who was a member of staff who was well known to the students. Interviews were recorded and transcribed. Only responses related to diaries are reported here. See Appendixes 1 and 2 for the interview questions and questionnaire.

Data analysis

The diaries were analysed for Carpentry Word List items using Heatley, Nation, and Coxhead (2002) Range program and Nation's (2012) BNC/COCA frequency and supplementary word lists which had been adapted for the study. Briefly, this process involved identifying all the words in the diaries that belonged to, but were not in, Nation's BNC/COCA lists of the first 1-25,000 words by frequency, marginal words (e.g. swear words, fillers), abbreviations, proper nouns, and compounds and then 'backfilling' the

Nation lists to include all these lexical items. Any leftover words that were clearly related to carpentry but were not in Nation's lists were added to a list of carpentry words, and we developed similar lists for the other trades in the LATTE project (see Coxhead et al., 2016 for more details). This process ensured that all lexical items could be identified and counted by the Range program. It is important to note that the spelling errors made by the students were kept in the texts, but the correct spellings of words were noted in <brackets> because Range ignores words inside these brackets, meaning it does not count them. This means that items such as *weatheboards* and *weatherproffing* are not corrected in the texts.

Results and discussion

In answer to the first research question as to whether the student writers saw the diaries as useful for their technical vocabulary learning, students in interviews reported using the diaries for vocabulary learning purposes. Here is one of the students in an interview, talking about using the diaries to keep track of vocabulary from his course. This quote was originally used in Coxhead et al., 2016: 28) in relation to the importance of memory in the carpentry courses and the sheer number of new words for students to learn. This time, the focus is on the diaries themselves as a learning tool for vocabulary.

> But um a lot of words, I try to remember in my head but I forget them, but I try to remember them as much as I can but sometimes I write them down in my diary just so I er make me remember it when I go back over my diary.
>
> (Coxhead et al., 2016: 28)

The same student suggested that new students should also use their diaries for keeping track of vocabulary, preferably right from the start of the course. He said,

> Definitely write it in their diaries when they are doing their diaries, because the diaries are the most important thing, I wish I had started my diary earlier in the year, like every day because I have lost a lot of words that I could have known … things that help me … I forgot my diary for a couple of weeks and I forgot the words …

Unfortunately, he was not the only student who lost his diary. Another reported losing his diary for a week and panicking because he slipped behind in his studies quickly. The diary was his way of keeping track of both his work and the vocabulary from class. Writing diaries was seen as challenging by the students, mostly because, as the following learner states, falling behind makes the exercise problematic. He said, 'Well, I reckon what is difficult for me is my diary, keeping up to date and stuff.'

In the questionnaire, we asked the students to note words that they thought any new student would need for studying carpentry. We then compared their answers with the Carpentry Word List and found 28 items that were recommended by the students and in the Carpentry Word List (including *bevel, battens, construction, cavity, ceiling, claddingls, dwangls, nog,* and *flashings*). There were words that would have been in the list if they were spelled correctly, including *Ribbenboard/Ribbion board* [ribbon], Faicer/s [fascia], and *hamer* [hammer]. Other suggestions for words new colleagues might need in their studies included *2x4*, *no marks*, *man up*, and *hard up* (meaning making sure one surface is flush against another).

Use of Carpentry Word List vocabulary in the builders' diaries

One of the key questions about the builder's diary activity is whether the students would produce specialized vocabulary in their writing. Because of the connection between the work that the students were doing in their courses (Nation, 2009) and their future professional writing (Parkinson et al., 2018), it could be expected that technical vocabulary would be used and possibly as much as in the professional texts used in the courses reported above. The coverage of the Carpentry Word List abbreviations and proper nouns over the builder's diary corpus was 33.74%, including frequent use of items such as *truss*, *cut*, *stud*, and *steel*. Figure 15.3 shows an example of entries from a builder's diary at the beginning and the end of the carpentry course.

The entry in the first column is more narrative-based than the sample in the second column, with a step-by-step description of the procedures followed by the writer and his classmates as they measured the primary boundary of the dwelling. Note the spelling difficulties with the word *joists* in the second column, which the writer eventually gets right! These examples also demonstrate technical vocabulary from the field of study in use, for example, *boundary, square, fence,* and *measure* in the first diary entry in the first column, and *joists*, *level*, *bearers* and *piles* in the second column which are all in the Carpentry Word List.

Another question about the diaries and technical vocabulary is whether there was any difference between the Carpentry Word List coverage in the higher scoring diary writers and the lower scoring diaries. Here we found that the lower scoring diaries used slightly more items from the Carpentry Word List (33.71%) than the higher scoring diaries (30.66%). The abbreviations and proper noun usages were roughly the same in the two sets of diaries. Both the higher and lower scoring diaries were made up of mostly high frequency words. The first 3,000-word families from Nation's BNC/COCA (2012) lists covered 83% of the diaries on average. This coverage is around 4.6% lower than the professional writing in the trades texts reported in Coxhead et al. (2020). The abbreviations covered 1.62% of the diaries and the proper nouns covered 0.09%. It is important to say here that the professional writing corpus and the builder's diary corpora were not the same size – the professional

An early diary entry by a student writer	*One of the final entries from the same writer*
First thing we did was establish the primary boundary. We measured off 2 meters from the fence line, dividing [Name]'s and [Name]'s classes, and 5 m off the fence opposite our container, and marked out the A, B, C, and D corners with spray paint. We found B, C and D after we had corner A. We measured out with a tape measure our width of 9 … , then we found corner D by measuring 16.090 m our length. Corner C we measured down 16.090 from B and 9 … m across from D. Of course, we tried to get all the length square and parallel from one another, but it isn't possible to get accurate. So, we later used 3-4-5 or Pythagoras Theorem.	Bearers Bearers are set on top of the piles. On top of the bearers sit the joists. Bearers are used in the subflooring and decking to hold or bear and spread the weight of the house. Piles Piles are set in between the ground and bearers. Piles hold the house off the ground away from moisture to lower the chance of rot and insect attack. It also holds the house level from potentially un-level ground. Josists <Joists> Josists <Joists> are set on top of the bearers and on top of the josists <joists> is the flooring (which in this case is 19 mm plywood) Josists <Joists> help spread the weight of the house and people when they walk over the floor. We also use double joists for load bearing walls and perimeter joists around the house.

Figure 15.3 Early and late entries by the same writer in his builder's diary.

writing corpus contains 300,500 running words or tokens (Parkinson et al., 2017a: 40) compared to just over 227,000 running words in the diaries. It could also be the case that the writers of the lower scoring diaries relied more heavily on text written in class by the tutors than the higher scoring diaries. A line by line analysis would help us discover whether this is in fact the case.

How could the builder's diary concept be used in ELT courses?

The diaries have a number of features which make them particularly useful as a writing activity for language learners. First, they are mandated, which means they are a regular and assessed part of the learning curriculum. The tutors collected the diaries every week for assessment. If anything was not clear in a diary entry, the tutors would follow up with the student in question to clarify meaning or check understanding. If we look at the principles from Nation (2009) cited earlier about teaching learners to write, we can perhaps see how the diaries stack up (Table 15.2).

With the longest diary standing at 11,000 words but an average of just over 4,000 words, the diaries provided some students with plenty of opportunity

Table 15.2 Comparing the diaries to Nation's (2009) principles for teaching writing

Principles from Nation (2009)	*Relationship to the builder's diaries*
'Learners should do lots of writing and lots of different kinds of writing' (Nation, 2009: 93)	The students wrote in their diaries every day. The writing involved descriptions of work which they had carried out, photos of the work, and diagrams.
Writers need to have an audience in mind and aim to communicate a meaning to that audience.	There were several audiences for the diaries; the tutors, the students themselves during the course and once they had left the course and were working, and future employers.
'Writing should interest learners and draw on their interests' (p. 93)	The writing is about the learning taking place during the carpentry course. The students learned through watching, doing, and following instructions. They therefore had all the experience they needed to write (Nation, 2009). We did not ask the students whether they were interested in the diary writing, but they valued their diaries highly overall.
There should be a feeling of success in the writing, as much as possible.	Students reported reading back over their diaries and seeing evidence of their progress in their work.
'Learners should use writing to increase their language knowledge' (p. 93)	The diaries provided a clear place to use technical lexis in writing. For example, explaining photos and daily activities on the building site requires the use of technical words. Some of the students used their diaries to keep track of technical vocabulary of their course of study.

for writing. They could bring themselves and their classmates into the writing, using pictures to illustrate activities such as measuring and hammering. The writing was also topic-based, with students in interviews noting that they had spent sometimes up to a week working on, and therefore writing about, framing for walls/the internal structures of a house. This meant that diary entries were focused on particular topics for periods of time. Another feature of the diaries is that they were used to support students in their writing development, because tutors and students knew that writing would be part of future work tasks (Parkinson et al., 2018). This element of the writing speaks directly to the development of professional knowledge and language (Woodward-Kron, 2008), which combine to form a new identity as a builder.

The idea of using these diaries has been taken up by plumbing tutors at the polytechnic because they foster writing skills, allow students to demonstrate their knowledge, and provide a powerful learning tool for tracking and learning vocabulary. That said, in interviews, students noted that some

tutors had different expectations about the diaries, and instead put more of a focus on building the house during the course. Other limitations of this study, clearly, are the fairly small number of diaries, texts of different lengths, and small numbers of interview and questionnaire responses. Judicious editing of the diaries would help to identify shared sections of texts in the diaries to drill down further to the texts which were generated by the student diary writers themselves.

There is value in following up this study with a more in-depth analysis of the development of a productive knowledge of vocabulary and tracking the development of use, for example, from early diaries through to later diaries. We might find, for example, that students increase their technical vocabulary use as they move through the course. That said, we might also find that the diary writers became more succinct as they write.

Conclusions

In this chapter, we have looked into writing in a carpentry course at a polytechnic in New Zealand. Builders' diary writers in this study were first and second language speakers of English, in class cohorts with theoretical and practical elements. The diaries were used to keep track of vocabulary and were seen by some students as particularly central to their learning, so much so that loss of a diary could incite panic. The Carpentry Word List (Coxhead et al., 2016) covered nearly 34% of the diaries, with the lower scoring diaries containing roughly 3% more technical words from the Carpentry Word List than the higher scoring diaries. The diaries used roughly the same proportion of technical abbreviations as the professional text writers, but fewer proper nouns overall.

One of the purposes of this chapter was to encourage ELT teachers to look outside our field to find out more about the writing activities of other domains, in this case, trades education. It is important to find writing activities that reflect Nation's principles and connect our students' writing to their future needs. We could look into technical writing to find out whether vocabulary from word lists such as the Carpentry Word List used here occurs in student writing, or whether our writing tasks might be analysed or reworked by taking technical vocabulary use into account. We could then find out more about technical vocabulary use in writing, including barriers and facilitating factors. After all, we cannot complain that our students do not use technical or specialized vocabulary in writing if we have not done our best to understand why that might be the case or done what we can develop the knowledge and understanding to do so.

Note

1 BAWE was developed at the Universities of Warwick, Reading, and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre

for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (Department of Applied Linguistics, Reading), and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the Economic and Social Research Council (ESRC) (RES-000-23-0800). More details can be found at the corpus website: www.coventry.ac.uk/bawe/.

References

Chan, S. (2013). *Learning a trade: Becoming a trades person through apprenticeship*. Wellington: Ako Aotearoa.

Chung, T., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, *15*(2), 103–116.

Conrad, S. (2017). A comparison of practitioner and student writing in civil engineering. *Journal of Engineering Education, 106*(2), 191−217.

Conrad, S. (2018). The use of passives and impersonal style in civil engineering writing. *Journal of Business and Technical Communication*, *32*(1), 38−76.

Conrad, S. (n.d.). *The civil engineering writing project*. Retrieved 15 February 2019 from www.cewriting.org/.

Corson, D. (1985). *The lexical bar*. Oxford: Pergamon Press.

Corson, D. (1997). The learning and use of academic words. *Language Learning*, *47*(4), 671–718.

Coxhead, A. (2011). What is the exactly word in English? Investigating second language vocabulary use in writing. *English Australia*, *27*(1), 3–17.

Coxhead, A. (2012). Academic vocabulary, writing and English for Academic Purposes: Perspectives from second language learners. *RELC Journal*, *43*, 137–145.

Coxhead, A. (2016). Acquiring academic and discipline specific vocabulary. In Hyland, K., & Shaw, P. (eds) *Routledge handbook of English for academic purposes* (pp. 177–190). London: Routledge.

Coxhead, A. (2018). *Vocabulary and English for specific purposes research: Quantitative and qualitative perspectives*. London: Routledge.

Coxhead, A., Demecheleer, M., & McLaughlin, E. (2016). The technical vocabulary of Carpentry: Loads, lists and bearings. *TESOLANZ Journal*, *24*, 38–71.

Coxhead, A., McLaughlin, E., & Reid, A. (2019). The development and application of a specialised word list: The case of Fabrication. *Journal of Vocational Education and Training*, *71*(2), 175–200.

Coxhead, A., Parkinson, J., Mackay, J., & McLaughlin, E. (2020). *English for vocational purposes: Language use in trades education*. London: Routledge.

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, *43*, 49–61.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327.

Heatley, A., Nation, P., and Coxhead, A. (2002). Range. Available from www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*(4), 567–587.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*, 469–523.

McLaughlin, E., & Parkinson, J. (2018). 'We learn as we go': Acquiring a technical vocabulary during apprenticeship. *English for Specific Purposes*, *50*, 14−27.

Malmström, H., Pecorari, D., & Gustafsson, M. (2016). Coverage and development of academic vocabulary in assessment texts in English Medium Instruction. In Göpferich, S., & Neumann, I. (eds) *Assessing and developing academic and professional writing skills* (pp. 45–69). New York: Peter Lang.

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching*, *1*(1), 2–13.

Nation, I. S. P. (2009). *Teaching EFL/ESL reading and writing*. New York and London: Routledge.

Nation, I. S. P. (2012). *The BNC/COCA word family lists*. Retrieved from: http://www.victoria.ac.nz/lals/about/staff/paul-natio

Nation, I. S. P. (2013). *Learning vocabulary in another language.* 2nd ed. Cambridge: Cambridge University Press.

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing* (pp. 145–151). Amsterdam: John Benjamins.

Parkinson, J., Coxhead, A., Demecheleer, M., Mackay, J., Matautia, L, McLaughlin, E., & Tu'amoheloa, F. (2017a). *The language in the trades education project*. Wellington: Ako Aotearoa.

Parkinson, J., Demecheleer, M., & Mackay, J. (2017b). Writing like a builder: Acquiring a professional genre in a pedagogical setting. *Journal of English for Specific Purposes*, *46*, 29–44. doi:10.1016/j. esp.2016.12.003.

Parkinson, J., Mackay, J., & Demecheleer, M. (2018). Situated learning in acquisition of a workplace genre. *Vocations and Learning*, *11*, 289–315.

Woodward-Kron, R. (2008). More than just jargon: The nature and role of specialist language in learning disciplinary knowledge. *Journal of English for Academic Purposes, 7*(4), 234–249.

APPENDIX 1

Interview questions for Carpentry students to help build a picture of how they learn the vocabulary and what we can learn from that (Parkinson et al., 2017a, p. 126)

Notes

Use these as a guide, but follow-up with any other relevant things they say about learning and using carpentry specific words.

Questions to guide the interview

1. What carpentry words did you know before you started this course?
2. What is different between the words you used when you started the words you use now?
3. What do you do when you hear a new word?
4. What do you do when you see a new word in an assessment or workbook?

5. How do you learn that word?
6. What does your tutor do to help you learn new words?
7. When do you use them and where do you use them?
8. How difficult is it to learn the new words:

1	2	3	4	5
Always easy	Often easy	Okay: sometimes easy, sometimes difficult	Often difficult	Always difficult

   a. Why? What are the biggest challenges?

9. Do you use glossaries?
   a. Why/why not?
   b. When?

10. What helps you learn and use carpentry words?

N.B. Find out if the learner speaks any other languages, background in English learning

APPENDIX 2

Questionnaire to students on writing: Language in the Trades (adapted from Coxhead, Demecheleer & McLaughlin, 2016, pp. 55–58)

1. What writing do students need to do in courses that you are taking?

	Daily	*Weekly*	*Monthly*	*Never*
Report on what you do in the workshop				
Report on work done on site				
Summaries				
Short answers to questions in workbooks				
Reports written in teams/groups				
Notes on work complete e.g. builder's diaries/ Record of work				
Short answers to questions in assessments				
Other (please specify)				

Are you assessed on any of these? ■ Yes ■ No
If yes, which ones are you assessed on?

2. How are you assessed on them?
3. What other language tasks do students have to do in courses you are taking?
4. What kind of words do you need to know to learn Carpentry?

5. What's the most difficult thing for you about learning new vocabulary in your trade and why?
6. What do you do when you hear a word or phrase that is new?
7. How does your tutor support your learning new words or terms related to your trade?
8. If you were advising a friend about taking this trade course next year, what advice would you give him or her about how to learn the vocabulary that they need?

What reading, writing, speaking and listening do you think you will have to do in your job?

16 Lexical development paths in relation to academic writing

Paul Booth

Introduction

A 'snapshot' of learners' lexical profiles does not give an indication of how L2 lexis develops over time. In particular, it does not give an indication how the learning style construct, memory-analysis, may relate to any development in lexical profiles over time. Although earlier studies on L2 lexical development (e.g., Schmitt, 1998, and Laufer, 1995) exist, research in this area seems to be sparse and learning style has not previously been considered in terms of lexical development.

Beginning learners have only a very limited amount of lexis to draw upon. As learners become more proficient then they have a larger store of lexical items to draw upon. Vocabulary size is manifest in responses to receptive tests such as EVST (Meara and Jones, 1990), *X-Lex* (Meara and Milton, 2003a), *Y-Lex* (Meara and Miralpeix, 2006), and productive tasks such as Lex30 (Meara and Fitzpatrick (2000). So we can expect that more advanced learners may show more variation in their productive lexis than less advanced learners. In fact, Laufer and Nation (1995) suggest that this is indeed the case at the 1,000 word frequency level. One of the ways of understanding variation could be to understand how learning style may contribute to learners' lexical production.

There has been some disagreement over how lexical frequency profiles (LFP) can distinguish learners with low levels of lexical knowledge. Meara (2005) focuses on different frequency levels to see whether different sized vocabularies ought to produce large differences in the profiles and if similar sized vocabularies would not. Meara's computer simulations show that there are not significant differences between groups when the source vocabularies are very different. However, Laufer's (2005) counter-argument is that there are differences between vocabulary *use* by actual learners and what Meara concludes as what should happen given the simulated vocabulary size and the 'texts' generated which were also simulations. Laufer (2005) questions whether vocabulary size and lexical use run parallel and uses this as a foundation to query Meara's simulation. Despite the fact that they seem to describe a different phenomenon: simulation and vocabulary use, the process of vocabulary development is the motivation for the current study in this chapter. Given

the complex nature between vocabulary size and vocabulary use the focus of this study is on productive lexical development rather than discriminating learners of different proficiency levels.

Lexical development

One way of measuring development is to look at lexical sophistication, which is the knowledge and use of infrequent words. Daller, Turlik, and Weir (2013) found that production of 'advanced types', lexis beyond the 2,000 frequency level, over a two-year period showed a Loess curve (locally weighted scatterplot smoothing) line in which there is no increase at the beginning of the period, but then there is an increase in the middle and then a flattening out towards the end (p. 206). Such a study indicates that learners' lexical development is typically non-linear and so an understanding of why productive lexis beyond the 2,000 frequency level is variable is worth investigating. Lexical development was also studied by Laufer (1994) and is important because she used the Lexical Frequency Profile (LFP), a more recent version of this tool is used in the current chapter, Web VocabProfile (Cobb, n.d.), with more recent word lists which reference the British National Corpora.

The Lexical Frequency Profile (LFP) classifies the vocabulary from essays that are typed into the software to calculate the percentage of words at the first 1,000 most frequent, the second 1,000, the University Word List (Xue & Nation, 1984) and not in any list. Laufer (1994) argues that one way of measuring lexical richness can be to capture words beyond the 2,000 frequency as this is where development occurs in that words 'beyond 2000' frequency tend to reflect a larger lexicon. Laufer's (1994) study presented the Lexical Frequency Profile (LFP) in a revised form which is productive of lexis beyond the basic 2,000 frequency band in order to capture any changes in lexical quality.

The current study also measures vocabulary free production beyond the 2,000 frequency level as a mark of vocabulary richness. The rationale for examining profiles beyond the 2,000 measure is to give a single measure of lexical richness in free production. Laufer's (1994) study aimed to determine whether there would be an increase in the productive lexicon of advanced learners of English. This was a longitudinal study designed to highlight lexical development in academic writing. In Laufer's (1994) study participants wrote two different compositions at two different points in time, in order to compare the lexical frequency to measure any gains in lexical richness.

The LFP (1994) was used in its condensed form, which was the percentage of lexis beyond the 2,000 frequency level. The participants were first-year university students from an Israeli university and whose first language was either Hebrew or Arabic. Their L2 English proficiency level was similar to the Cambridge First Certificate of English intermediate level of proficiency. The compositions were written at three different points in time: Time 1 was the entrance exam which comprised of a choice of three statements from which

the students wrote for or against type essays (all learners, $N = 48$); Time 2 was at the end of the first semester that is, the same essay again (group 1, $n = 23$); Time 3 was at the end of the second semester that is, the same entrance exam (group 2, $n = 25$).

Laufer's results show that the mean percentage of beyond the 2,000 frequency level words for group 1 grew from 9.96% to 13.17% (3.21% increase). Group 2 grew from 8.48% to 10.04% (1.56% increase). T-tests showed significance at: $p = .01$ and .03 respectively. In Laufer's study, the post-test of group 1 after one semester showed a greater increase than the post-test of group 2 after two semesters. Logically, the post-test after two semesters should show greater gains. This may be one of the problems with related sample testing. Group 1 may have simply been more effective or more motivated learners of English than group 2. The results from the relatively small sample of both groups, $n = 23$ and $n = 25$, may have been influenced by individual differences in language learning. It is important to identify these because they may be related to their approaches to language learning.

At least one other additional potential influence is that it is not clear what effect the composition topics had on the learners' lexical profiles. The first group had the same statement for the post-test as the one they had for the entrance exam. The second group was 'given the composition of the entrance exam' (Laufer, 1994: 25). It is not clear whether they could choose a different statement or not. Although the LFP is stable across two compositions written by the same learners, that is, non-significant differences in percentages of words from the second thousand frequency level, UWL, and 'not in any list' words (Laufer & Nation, 1995), it is still unclear the extent to which the differences in question topics might have influenced the frequency profiles.

It can be controversial to link rarer lexis to academic development. Rarer, academic lexis is examined further on the basis that Morris and Cobb (2004) suggest that knowledge of rarer lexis may be analogous with academic performance. Morris and Cobb's study is somewhat controversial because it infers a style over content issue; however, it is one worth exploring because it has parallels to lexical knowledge, indicating a certain type of academic aptitude.

Learning style: memory and analysis

One way of understanding lexical development is to look at not only lexical profiles but also individual differences such as learning style. Laufer's study highlighted how more time in an academic environment in which English is used does not necessarily equate to greater gains of rarer lexis. A plateau effect in lexical development could be analysed in light of learners' learning style. Therefore the study presented in this chapter also looks at learning style from a language aptitude framework. Das (1988: 102) defines learning style as a general, habitual mode of processing information when applied to practical, educational, or training applications. Some research has been carried

out into learning or cognitive style dimensions, but as yet little has been applied to second language acquisition, see Ehrman and Leaver (2003) for a review. Although much of the research in learning/cognitive style has been fragmented, the model of learning style used in this study is examined in relation to language learning. In this case learning style is not one particular approach to learning but a dual mode system in which one relies on processing language as chunk-based and idiom (memory) and the other relies on structure and rule (analysis). This theory of learning comes from a variety of fields: cognitive science (Carr and Curren, 1994) and empirical studies using aptitude profiles (Skehan, 1986; Wesche, 1981).

In cognitive science, experiments have been conducted that relate to how people learn sequentially structured sequences. In these experiments participants learn false grammars or letter strings. The question remains, though, of how learning is mentally represented. The issue is whether structured sequences are represented as generalizations across stored examples (i.e., exemplars) or as a set of abstract rules (Carr & Curren, 1994: 210). This dichotomy in language use has been illustrated by Sinclair's (1991) idiom versus open choice principle. In spoken and written texts, the user has available a number of pre-constructed multi-word combinations adhering to the idiom principle, versus word-for-word combinations (i.e., grammatical creation) making use of the open choice principle.

In second language learning, Skehan (1998: 88–9) also argues in favour of a dual mode system. At one extreme, language is coded and represented as exemplars that require minimal computational demands on the learner, the cost being that the system may not be so easily adapted for the expression of complex meanings. This is because such lexical elements are stored as units longer than a word and not broken down into constituent parts. At the other extreme, learning is rule-based in which language is analysed into parts and produced from rules. The operation of this system is more costly in terms of processing burden, but the benefit is the language system is more open to complexification. I do not suggest that these operations occur separately, rather the learner switches between the two.

The memory-analysis framework is understood to be representative of the complexity which underlies a memory based and rule based system and is not intended to capture all of the complexity clustered within these domains. This dual mode of learning is categorized as memory and analysis and is the basis on which learning style is conceptualized and tested.

Learners who are predisposed to analyse language may use lexis in a qualitatively different manner. Skehan (1998) argues that analysis-orientated learners' lexicon may be more parsimoniously organized, consist of a single-representation lexical system, and would engage in regular restructuring and complexification of their interlanguage system. High memory learners in contrast would have more redundancy and multiple representations of lexical elements and considerable redundancy in their memory systems. What this might mean in terms of productive lexis is that memory-orientated

learners may be less systematic in their use of lexis. There may be more variation in their lexis because of a greater emphasis on exemplar-based language. Analysis-orientated learners may be more systematic in their use of lexis, that is, rely less on exemplars and more on the underlying rules of language. So the aim of the study in this chapter is to explore whether learners' L2 lexis in the production of academic essays develops over one semester. A secondary aim is to examine learners' strengths and weaknesses in memory and analysis in relation to lexical development. A profile score of lexis beyond the 2,000 measure was adopted because more advanced lexis is thought to occur at this level and because a single score is more amenable to statistical analysis, (see Laufer, 1995).

One would expect most learners to develop beyond the 2,000 frequency level or, at the very least, to remain relatively static. Research has yet to reliably establish which type of learner (memory- or analysis-orientated) would be more consistent in their development. Learners with good memories are more likely to accumulate lexis whereas learners with good analysis are more likely to restructure their language which might not show so much quantitative gain but are likely to be more consistent from Time 1 to Time 2. Learners who score poorly in memory and analysis are most likely to remain static in terms of lexical development.

The study

In order to explore whether L2 lexical development, in terms of the use of infrequent or rare words, relates to how learners approach their learning, in terms of memory and analysis, the following outlines the design of the longitudinal study employed in this chapter. To elicit productive lexis, participants wrote discursive compositions the topics of which related to the faculty in which the participants belonged and these were combined into three different composition topic groups. Such grouping might have the effect of producing more varied profiles as learners were encouraged to express their ideas according to different themes. As the focus is on the process of lexical development rather than a single test, the participants were not divided into different proficiency levels but categorized based on their learning styles.

Research questions

In the light of the previous section, the research questions are as follows:

1. Is there any lexical development beyond 2,000 over the period of one semester?
2. Is any development related to strengths and weaknesses in memory and analysis?
3. Are any correlations in Time 1 and 2 lexical profiles related to memory and analysis?

Method

Participants

The group comprised of L2 language learners of English whose writing was sampled at two different times: Time 1 at the beginning of a university semester and Time 2 at the end of the semester. The semester was twelve weeks long. There were 23 males and 10 females. There was a wide variety of first language backgrounds. Most of the language learners had a first language which is not cognate with English, for example, Korean (n = 4), Arabic (n = 3), and the diversity of first language (L1) backgrounds was wide in that there were twenty-five different L1 backgrounds.

The students were mainly in their first year of study from the Faculty of Engineering. Year one students were contacted because they tend to be less jaded by university questionnaires and so are more likely to give up their time for research projects than students in the more advanced stages of their studies. Other faculties included Computing, Information Systems and Mathematics; Science; Art Design and Architecture; Arts and Social Sciences; and Business and Law.

In order to understand the learners' language level in English, a test of receptive vocabulary knowledge was used, *X-Lex*, and Meara and Milton's (2003b) notes accompanying the *X-Lex* because it highlights the comparability of international language test scores. Based on these comparisons, high proficiency learners (n = 25) were classified as those with IELTS scores above 5.5, TOEFL scores of above 520, CBTOEFL scores above 190, Cambridge Advanced English, A-Level English, and GCSE English grade D and above. Where no data were available on their English language backgrounds (3 participants), learners with *X-Lex* scores of above 3745 were classified as high proficiency. Engineering, science, and computing students tended to be high proficiency whereas art students tended to be low proficiency.

The written compositions

The composition questions for the engineering and computing students Time 1 and Time 2 were as follows:

> How has science and technology changed life since you were a child?
> How important is science and technology to the modern world?

The questions were designed to encourage the students to write as freely as possible by using their background knowledge of science and technology but neither question demanded the use of specialist knowledge or vocabulary.

The composition questions for the students from the Faculty of Art, Design, and Architecture and Faculty of Arts and Social Sciences were based on topics which these students had read and written about during their

pre-sessional English course. So although they seem more abstract than the engineering and computing questions, they were in fact familiar topics to this group of learners. Time 1 and 2 were as follows:

> What is the relationship between culture and community?
> What is the relationship between culture and communication?

For the business and law students the questions were:

> Is a good manager born and not made?
> Which qualities would you expect a good manager to have?

Data processing

All participants were tested for their receptive vocabulary knowledge using Meara and Milton's (2003) *X-Lex* (v2.05) to determine proficiency levels if no background data were available. All students were tested for their learning style by using Meara, Milton, and Lorenzo-Duz's (2001) Memory LAT B (visual memory for paired associates) and Analysis LAT C (grammatical sensitivity). All students were shown how to use these tests through a demonstration of each test projected onto a large screen. Written instructions were also provided.

After the aptitude tests, the learners were asked to write a discursive composition of 250 words on one of the questions outlined in the previous section. Although a time limit of 40 minutes was given for the writing section, the slower writers were given more time in order to reach a word count of 250 tokens. Exactly the same procedure for the productive free writing was carried out at Time 2. The texts were then inputted into Cobb's Web VocabProfile/BNC-20 (v3.0), which calculates word frequency by using the British National Corpus. This version of the VocabProfile calculates the percentage of coverage of families, types, and tokens at the various frequency levels, from the one thousand level (1k) to the twenty thousand level (20k) as well as Off-List which is lexis not within the 20k frequency. The Off-List for the VocabProfile is different to the Lexical Frequency Profile in Laufer's study (1994) in which 'not in any list' is beyond the 2,000 frequency. Hardly any of the learners' texts contained any tokens beyond 10k.

All spelling errors were corrected unless they were unrecognizable. The following sample of a participant's text contains spelling errors which make the intended words 'well' (?) 'willing' (?) difficult to deduce:

> When I was a boy we had no computers in school and even the government had few of them, our teachers and as ware* so waling* to learn computer if we could get one, but we never had a chance.
> * word discarded

Errors in the incorrect derivative form were ignored because the software for the VocabProfile counts all the derivatives (i.e., the word family) at the

same frequency level. Semantic lexical errors were few and far between. The criterion for a semantic error was if a word made no communicative sense. However, no errors fell into this category. Only 250 tokens were analysed from each participant from each session. In the case of longer texts, only the first 250 tokens were used for computer analysis because the effects of different text lengths have not been fully investigated.

Memory and analysis

The participants were grouped into high and low memory and analysis dimensions. These benchmarks were calculated from Meara et al.'s (2001) scores from the aptitude tests for the percentage of people who fell into bottom, middle, top scores. Memory was tested via LAT B which tests people's ability to remember pairs of words in a language they will not know next to an English translation. The words are out of context and scroll across the screen. The scores of 43%–73% are classified as the middle (with a score of 58% as the median). Therefore ≤58% were classified as low, 59% ≥ as high. Analysis was tested via LAT C which is a test of inferring grammatical rules from examples in a made-up language and their English translations. Participants are then tested on a set of English phrases of which one of two translation examples is correct. Middle range scores from Meara et al. are 60%–69% so ≤ 64% were classified as low, 65% ≥ as high. The mean score for memory in this study (m = 58.75; SD = 22.275) seems to be in line with Meara et al.'s data for the middle score. The mean score for analysis (m = 58.94; SD = 17.534), on the other hand, is just below the middle band.

Results

Lexical development beyond 2,000 over the period of one semester

Overall, there is only minimal development in lexical profiles beyond the two thousand frequency level. The beyond 2k lexical frequency Time 1: mean = 4.50% (SD = 2.00) and the beyond 2k Time 2: mean = 4.65% (SD = 2.25).

Lexical development in relation to strengths and weaknesses in memory and analysis

With regards to whether lexical development beyond the 2,000 frequency is related to strengths and weaknesses in memory and analysis, the mean scores in Figure 16.1 show that learners who are strong in one of the dimensions but weak in the other tend to develop over the period of one semester. In contrast, learners with low memory and analysis scores show a decrease in lexis beyond the 2,000 frequency level over one semester. High memory and analysis learners show practically no development although they do display higher mean beyond 2,000 profiles scores at Time 1 and 2 than the other sub-groups.

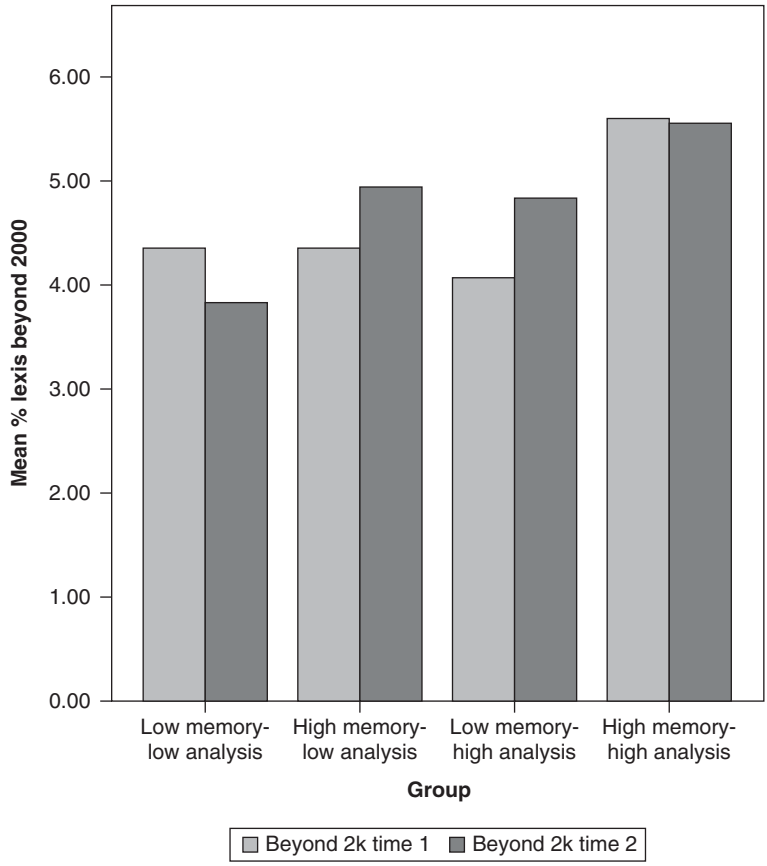


Figure 16.1 Lexical development over one semester.

The mean scores, however, do not show the wide variation between the individuals in each of the four groups. Therefore the next set of data highlights these variations in the sub-groups.

The box-and-whisker plots in Figure 16.2 show that there are differences between these groups for T1 and T2. The bar in the box shows the median, the top of the box shows the 75th percentile and the bottom the 25th percentile. In terms of lexical rarity beyond the 2,000 frequency level, at T1 learners low on the memory-analysis dimensions show the greatest variation. In contrast, learners who score high in memory and analysis show the least variation in their lexical profiles beyond the 2,000 frequency level. However, participant 30 is an extreme case which is difficult to explain.

These lexical profiles show variation across the different sub-groups of learners. The boxplots clustered for Time 2 show little in relation to Time 1. There is greater variation in the Time 2 scores, and the greatest variation is

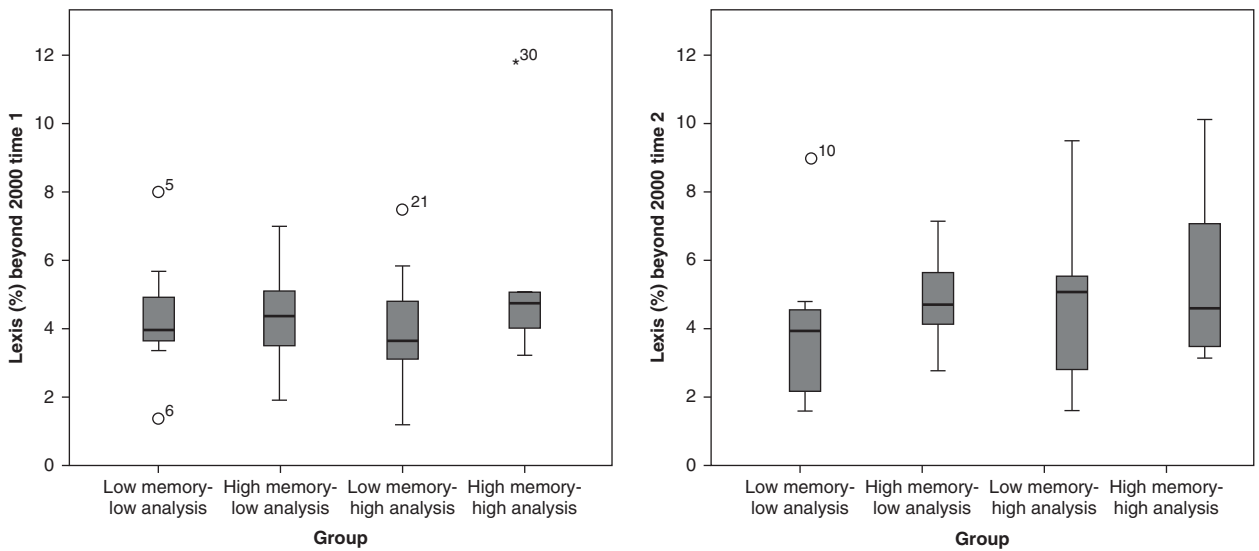


Figure 16.2 Time 1 (left) learner differences in lexis (%) beyond 2,000 and Time 2 (right) learner differences in lexis (%).

from the low memory-high analysis sub-group and a similar but less extreme pattern for the high memory-high analysis group.

Because of the small sample size, a non-parametric t-test was carried out to determine the differences in mean scores at Time 1 and Time 2. A Wilcoxon matched-pairs, signed ranks test was used for the difference between the beyond 2,000 lexis from Time 1 and Time 2. No significant differences were found for low memory-low analysis, high memory-low analysis, and high memory-high analysis. The largest difference was for the low memory-high analysis sub group, the Time 2 level was higher (*Mdn* = 5.11) than for Time 1 (*Mdn* 3.63) z = -1.71, p = .086 (2-tailed), r = .405 (Cohen's effect size). What this means is that, although the low memory-high analysis group showed the greatest increase in productive lexis beyond 2,000, there is an 8% likelihood that this could be due to chance alone.

There is actually no reason why the beyond 2,000 scores should be homogeneous as various factors could influence their lexical profiles. Learners grouped according to LAT scores should show variation in their lexical profiles if learning style or aptitude is not related to proficiency in a particular language. It is hoped that learning style will shed light upon the pattern of development for learners with different learning profiles. It is the relationship between the lexical profile scores which is likely to be more informative because it will show the trajectory of development. Therefore, the next set of results are set out to answer the third question of whether there is a correlation in profile scores taken at Time 1 and 2 in relation to memory and analysis strengths and weaknesses. Pearson correlations are shown in Table 16.1.

Correlations between Time 1 and 2 lexical profiles in relation to memory and analysis

The data suggest that when learners are low in the LAT B and C scores then there is little or no correlation from Time 1 to Time 2. When memory is high

Table 16.1 Memory and analysis beyond 2K correlation time 1 and 2

Low memory – low analysis	Time 2
Time 1	.144
n	11
High memory – low analysis	Time 2
Time 1	.041
n	7
Low memory – high analysis	Time 2
Time 1	.876(**)
n	9
High memory – high analysis	Time 2
Time 1	.773
n	6

** Correlation is significant: p = .002 (2-tailed).

Table 16.2 Aptitude group and proficiency

Aptitude group	*Proficiency*	
	Low	*High*
Low memory – low analysis	3	8
High memory – low analysis	0	7
Low memory – high analysis	3	6
High memory – high analysis	2	4

then profile scores of lexis is highly variable over the same time period. The data show that when analysis is high, then profile scores of lexis beyond 2,000 are relatively stable across two sets of writing taken over the period of one semester. Learners who are high in both memory and analysis show a weaker correlation but this is non-significant. Interestingly when the data are displayed in the scatter charts and we can see the individual learner profiles from T1 to T2 then it is the learners who score high in analysis but low in memory who show the strongest correlation in lexis beyond 2,000 over time. Recall that when the beyond 2,000 scores are analysed by mean gains from T1 to T2, the mean scores for the sub-groups mask this relationship and we simply see the sub-group gains in lexical development but not the stability of the high analysis group and the variability of the high memory group.

Discussion

The overall aim was to determine whether learners would show lexical development beyond the 2,000 frequency. In particular, the study was to determine whether any development was be related to the memory and analysis learning style framework. Time 1 and 2 lexical profiles were also analysed for any correlations in light of memory and analysis scores to understand patterns of variability. This discussion will first consider to what extent the research questions can be answered. After that I will consider methodological and theoretical issues that arose during the experiments.

Lexical development patterns T1 and T2 means

The first question asked whether there was any lexical development beyond the 2,000 measure over the period of one semester. The group mean showed an increase, however, the gain was relatively small. The second research question asked whether any lexical development over a one semester period is related to strengths and weaknesses in memory and analysis. The lexical profile scores beyond 2,000 showed that lexical development beyond the 2,000 frequency is possibly related to strengths and weaknesses in memory and analysis. Learners who have a predisposition towards memory or analysis show the greatest gains in lexical development beyond 2,000. That is, learners who

have strengths and weaknesses in memory and analysis tend to show development. Learners low in these dimensions, as one would expect, tend to show little if any development. Strangely, high memory and analysis learners do not change very much over time. One probability might be that these learners are simply learners with a greater proficiency in L2 English and so their lexical development may be more static (see Meara & Milton, 1995). In other words, what remains unclear is whether the LAT B and LAT C are linked to language proficiency, as seen by a greater proportion of rarer lexis.

Table 16.2 shows the number of high and low proficiency learners in each sub-group. Recall that proficiency was determined by external examinations of English unless none had been taken, in which case proficiency was determined by *X-Lex* scores which are only a rough indication of language proficiency based on vocabulary recognition scores. For three of the groups there is roughly double the number of high to low proficiency learners. The exception is the high memory-low analysis group which contains all high proficiency learners. Learners who are high on both dimensions tend to be high level learners; but low proficiency level learners are not necessarily low in LAT B and C scores. In this study, lexical development seems to be closer related to LAT C scores than L2 proficiency.

Lexical development patterns T1 and T2 correlations

The third research question asked whether there is a correlation in profile scores taken at Time 1 and Time 2 and whether any correlations are related to memory and analysis. High analysis and low memory (and to a lesser extent high analysis and memory) are related to a stable use of rarer lexis over Times 1 and 2, which is an interesting finding albeit with very small numbers in the sub-groups, and tends to suggest that analysis of language, rather than associative memory, is linked to stable profiles over one semester. The strong correlation between T1 and T2 could indicate the strength of an attractor state, that is, settle in a specific state which is defined by analysis of language. It could indicate that analysis of language bootstraps lexical development (Booth, 2011). That is, learners who are adept at analysing language use this knowledge to help acquire new lexis. Although there is not a linear relationship between rarer lexis and analysis, it appears that with high analysis scores there is stability in beyond 2,000 lexis and that learners who achieve high scores tend to be more consistent in their lexical development beyond 2,000.

What appears transparent is that the learners with good memories (i.e., associative) can and do also make gains in lexical rarity beyond 2,000 but they can also just as easily show a decrease in lexical rarity which is masked when the scores are simply analysed by grouping together the mean for Time 1 and 2 and comparing the differences. This interpretation has important implications for how the data are analysed. Although group means may show quantitative developmental gains in lexical frequency beyond 2,000, individual profiles may not correspond to the group pattern. Both high memory and high analysis sub-groups showed mean lexical gains, but it is only when

the data are broken down across time, that is, correlations between Time 1 and 2, that we can see the differences between the two groups.

Developmental patterns, or a lack of them, were again apparent with learners who are low in both dimensions showing erratic scores over the period of one semester. In fact, with the low scoring learners, there appears to be little or no relationship between the two scores. When lexical profiles beyond the 2,000 measure are erratic in that there is no relationship between T1 and T2 there is no apparent development.

One potential interim conclusion is that memory-orientated learners tend to show uneven gains whereas analysis-orientated learners show more consistent gains. When we simply look at the net gains we miss important developmental patterns. In fact most learners will make net gains but learners appear to take different paths of development. What seems to be more interesting is the variation in trajectory of lexical development rather than the product of development.

Lexical variation was also examined in a study by Bell (2002) whereby written texts were collected from a single subject over 18 months. They were subsequently analysed using a P-Lex (Meara 2001). This is a similar measure of lexical richness in that both the Web VocabProfile and P-Lex make central use of frequency lists. The main difference with P-Lex is that it is based on the observation that certain words occur more rarely than others and that this differential distribution is best described by a Poisson curve and reports this curve by means of a lambda (Bell, 2002: 79–80). The results in Bell's experiment suggested that 'students with low levels of lexical proficiency are more likely to produce consistent scores from one piece of writing to another, and that this effect fades as proficiency rises' (p. 164). The implication from this study is that variation in lexical richness (i.e., use of rare words) may be related to the memory-analysis learning style construct. Learners who obtain high analysis and low memory tend to show a consistency in their free production of lexis beyond 2,000 over the period of one semester. In fact, learners who are high on the analysis dimension but low on the memory dimension showed the greatest lexical development. The other sub-groups showed less lexical development. Learners who are low in both dimensions show erratic lexical profile scores beyond the 2,000 measure.

This result may indicate the effects of grammaticization which could lead to greater stability in the percentage of function words at around 50% in the one thousand frequency band. More developed lexical systems may in fact show greater signs of stability in respect of function words. What this also suggests is that it is not the mean percentage which is more revealing regarding low and high proficiency learners but the standard deviations of the mean scores. Low proficiency and low analysis seem to suggest greater variability in terms of function words and lexical profiles. Development, that is, quantitative gains in rarer lexis, may take two paths: either a memory-based approach which is more erratic and more likely to fluctuate or an analysis based approach which is more consistent and less likely to fluctuate. The next section explores how depth of processing may be related to the process of lexical development.

Depth of processing and lexical development

In this study, learners who show signs of lexical development and are consistent in their use of rare words over one semester are those who show, on average, greater grammatical sensitivity (analysis). What analysis could indicate in the context of this aptitude test is the ability to process language to understand the grammatical patterns. The key word here is 'process' which in Craik and Lockhart's terms is depth of processing. Although information may be held in what they call primary memory, such information is lost at a rate which depends essentially on the level of analysis (Craik & Lockhart, 1972: 677). Consequently, deeper analysis leads to a more persistent memory trace. Kendel's work has also shed light on how we shift from short-term memory to long-term. 'For a memory to persist, the incoming information must be thoroughly and deeply processed. This is accomplished by attending to the information and associating it meaningfully and systematically with knowledge already well established in memory' (Kendel, 2006: 210). Although grammatical sensitivity is seen as a separate ability from memory, it is the ability to recognize grammatical patterns and so process language on a deeper level which seems to be a prerequisite for the storage of information in long-term memory. In other words, lexis which is analysed in terms of its grammar for example may have a better chance of storage in long-term memory because it may be more systematically established in terms of how it is used with other lexis.

Learners who are oriented towards memory (i.e., associative), but not analysis show less consistency in their production of rare lexis over a period of several weeks, although in this study the number in this group was particularly small compared to the others. It may be that associative memory, in this study, is related to short-term or explicit memory but not long-term memory. For these learners then, their store of rare words may not be so permanently available in the long-term memory as those learners who are more able to process lexis more deeply. Learners with above average analysis scores may be better able to commit rare lexis to long-term memory. Future research could test for a correlation between the analysis scores and a long-term memory test of lexis.

The results in this study suggest that learners who analyse and so process language on a deeper level are those who consistently produce lexis beyond the 2,000 measure. Because rarer lexis, on the whole, has a lower surrender value, it may require deeper analysis in order for it to grade into long-term memory. Learners who do not process language so deeply may only have control over rarer lexis which has occurred in their input fairly frequently and so is in their long-term memory not through conscious processing but through repetition. This may explain why their profile scores are more erratic. Moreover, their lack of analysis of language may go hand in hand with a lack of analysis of the writing topic. Learners who analyse the essay topic on a superficial level may also be those learners who analyse language in a similar manner.

Lexical development

Overall, the difference between the percentages of beyond 2,000 words at Time 1 and Time 2 is minimal for the learners in this study. This result may have been compounded by the way the software analyses learners' texts. The VocabProfile software does not distinguish between different word types at each frequency level or whether a word is repeated or not. This means a learner who repeatedly uses the same word (e.g., 'technology') is not differentiated from one who uses different word types or families at this frequency level (2k). The lexical frequency profile is a calculation of the percentage of word coverage at each frequency level. Therefore, lexical development may be attributed to learners who simply repeat words as well as to learners who show a wide use of different words beyond the 2,000 measure.

This is the problem when the software categorizes words according to its frequency level, but does not recognize the fact that some words may be repeated. Learner profiles that show a greater variety (i.e., contain a greater number of word families) are not differentiated from profiles that show repetition (i.e., a lower number of word families but the same number of tokens).

Another possibility of why there is so little development is that the learners have reached a level of proficiency which is adequate for their studies. There may not be the motivation to increase their knowledge of rarer lexis and so they may have reached a plateau in their use of rarer lexis. Learners who are accepted on year one undergraduate engineering or computing information systems and mathematics (CISM) courses only need an IELTS level of 6.0, which is equivalent to an upper intermediate range. For other students in this study, for example, L2 English students for business and law normally need IELTS 6.5 whilst for undergraduate art and design the IELTS score can be lower e.g. 5.5. What is more, the IELTS score is an aggregation of scores for different language skills (speaking, listening, reading, and writing) and so their writing score could, in theory, be lower. In practice though the CISM and engineering students did not have IELTS scores and so we must examine other factors. The writing demands placed on them in their first year of study may not require them to use a large percentage of lexis which is consistently beyond the 2,000 measure. Students from the Faculty of Arts and Social Sciences and the Faculty of Business and Law do have assignments to write, however. Moreover the writing topics used in this study could be answered with high-frequency lexis. No specialist vocabulary was needed but background knowledge of the topic was.

Potential pedagogical implications

There is a danger, as with all learning style tests, that the memory and analysis tests can lead to a self-fulfilling prophecy. Learners who are tested can have their own fears confirmed through low scores on either of the tests. However,

that is not the point of these tests. They need to be used to create an awareness of an individual's particular strengths and weaknesses. A learner who always seems to get a lot of corrections on his or her work should understand that it may be that they cannot, literally, see the patterns of the second language. Alternatively, a learner who scores relatively high on gap-fill exercises but cannot hold a basic conversation in L2 may think that it is due to not being very good at learning a second language. Without greater awareness that we all learn in different ways, it can be demoralizing to the learner.

In a study conducted in France students were asked to compare two teaching approaches they had experienced: a focus on dictionary work and a focus on data-driven learning (DDL) which encourages learners to examine corpora to discover language rather than being explicitly taught. Boulton (2010) found that from open responses to a questionnaire students preferred dictionaries for new or unknown words and meaning or definitions (p. 553), whereas corpora were preferred for the contexts and concrete examples that highlight usage and grammar and to represent practical English frequent usage, and the language of today (p. 553). This could mean that for new lexis dictionaries are preferred whereas for extending existing knowledge DDL may be better, as Boulton highlights.

Learners with strong memories but little analytic strength may depend more on the context to make sense of and recall lexis. However, as these learners may be predisposed to use formulaic chunks of language to convey meaning, morphosyntax may be compromised. Skehan and Foster (2001: 187) make the point that 'context can often substitute for syntax'. Learners who are particularly prone to bypass syntax and rely on the context to a certain extent need to produce language to encourage them to notice and reveal their hypotheses about the L2 (Swain, 1995). By downloading sentences from the British National Corpus which contain lexis learners may find useful (e.g. technical vocabulary), the authentic language can be manipulated in various ways to scaffold the learner to produce language which is more target like. There are various ways in which this can be done. For example, tasks which encourage learners to construct full sentences from sentences which have all the function words removed force them to notice and grammaticize the language. Errors of parallel structure can be inserted so that learners need to notice how to reconstruct the text. Sentences can be fused together so that the learner has to reconstruct the text into coherent sentences. These types of tasks have been used in an English for Academic Purposes (EAP) context. The wider problem in certain contexts is that learners may not have the specialized vocabulary they need in order to communicate with any degree of expertise.

Conclusions

The overall aim was to understand any lexical gains over one semester in relation to strengths and weaknesses in memory and analysis. The learners in this

study showed modest gains in lexical development over one semester. Those with higher analysis scores tended to progress (i.e., produce lexis beyond the 2,000 frequency level) more than learners with higher memory scores. Although both types of learners progressed, the analysis-orientated learners' development was greater. Analysis-orientated learners also displayed a strong correlation between their beyond 2,000 profile taken at the start and end of a semester. Memory-orientated learners did not show a correlation between Time 1 and Time 2. It is argued that the net gains are less revealing about lexical development than the relationship between the two points in time. Correlational analysis highlighted the consistency that the analysis-orientated learners displayed. It is argued that language analysis may help learners to establish stronger memory traces which could lead to greater retention of lexis. Some memory-orientated learners also showed mean gains while others did not, highlighting the inconsistencies in their profile scores. A lack of language analysis is thought to be the reason why. Closer inspection of the data revealed that some learners repeated rare tokens and so inflated their beyond 2,000 percentage of words in comparison to those learners who produced rare tokens of different word families beyond 2,000.

By solely analysing the start and end point, we may miss important developmental patterns in the trajectory of lexical development. Most L2 learners tend to develop in terms of lexical rarity, but some learners may take different developmental paths. This chapter has not only focused on two points in time for lexical development but also the variation which highlights interesting differences between different sub-sets of learners grouped by language aptitude tests.

References

Bell, H. (2002). *Using frequency lists to assess L2 text*. (Unpublished PhD thesis.) University of Wales, Swansea, Wales.

Booth, P. (2011) *The interplay between lexis and learning: a study of second language lexical profiles and learning styles*. (Unpublished PhD thesis.) Kingston University, London.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534–572. doi: 10.1111/j.1467-9922.2010.00566.x

British National Corpus (n.d.). Available from http://corpus.byu.edu/bnc

Carr, T. H., & Curran, T. (1994). Cognitive factors in learning about structured sequences. *Studies in Second Language Acquisition, 16,* 205–230.

Cobb, T. (n.d.). *Web VocabProfile/BNC-20*. Available from www.lextutor.ca/vp/ [Accessed February and May 2007].

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, *11*, 671–684.

Daller, M., Turlik, J., & Weir, I. (2013). Vocabulary acquisition and the learning curve. In Jarvis, S., & Daller, M. (eds) *Vocabulary knowledge: Human ratings and automated measures* (pp. 185–217). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Das, J. P. (1988). Simultaneous-successive processing and planning. In Schmeck, R. (ed.), *Learning strategies and learning styles* (pp. 101–129). New York: Plenum Press.

Ehrman, M., & Leaver, B. L. (2003). Cognitive styles in the service of language learning. *System*, *31*(3), 393–415. doi: 10.1016/S0346-251X(03)00050-2

Kendel, E. R. (2006). *In search of memory: The emergence of a new science of the mind*. New York: Norton.

Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal*, *25*(2), 21–33.

Laufer, B. (1995). Beyond 2000: A measure of productive lexicon in a second language. In Eubank, I., Selinker, L., & Sharwood Smith, M. (eds) *The current state of interlanguage* (pp. 265–272). Philadelphia: John Benjamins Publishing Company.

Laufer, B. (2005). Lexical frequency profiles: From Monte Carlo to the real world (a response to Meara 2005). *Applied Linguistics*, *26*(4), 582–588.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322.

Meara, P. M. (2001). P-Lex: a test of productive lexical resources. v1.1. Swansea: Lognostics.

Meara, P. M. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics, 26*(1), 32–47.

Meara, P. M., & Fitzpatrick, T. (2000). Lex-30: an improved method of assessing productive vocabulary in an L2. *System*, *28*, 19–30.

Meara, P. M., & Jones, G. (1990) *Eurocentres Vocabulary Size Test. 10KA*. Zurich: Eurocentres.

Meara, P. M., & Milton, J. L. (2003a) *X-Lex: The Swansea Vocabulary Levels Test.* v2.05. Swansea: Lognostics.

Meara, P. M., & Milton, J. L. (2003b) *X-Lex: The Swansea Vocabulary Levels Test.* Swansea: Lognostics.

Meara, P. M., & Miralpeix, I. (2006) *Y-Lex: The Swansea Advanced Vocabulary Levels Test*. Swansea: Lognostics.

Meara, P. M., Milton, J. L., & Lorenzo-Duz, N. (2001). *Language aptitude tests*. Newbury: Express.

Milton, J. L., & Meara, P. M. (1995). How periods abroad affect vocabulary growth in a foreign language. *Review of Applied Linguistics* (*Institut voor Togepaste Linguistik*), *107*/108, 17–34.

Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System, 32*, 75–87: doi.org/10.1016/j.system.2003.05.001

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, *48*(2), 281–317.

Sinclair, J. McH. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Skehan, P. (1986). Cluster analysis and the identification of learner types. In Cook, V. (ed.), *Experimental approaches to second language acquisition* (pp. 81–94). Oxford: Pergamon,

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In Robinson, P. (ed.), *Cognition and second language acquisition* (pp. 183–205). Cambridge: Cambridge University Press.

Swain, M. (1995). Three functions of output in second language learning. In Cook, G., & Seidlhofer, B. (eds) *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.

Wesche, M. B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In Diller, K. C. (ed.) *Individual differences and universals in foreign language aptitude* (pp. 119–154). Rowley, MA: Newbury House.

Xue, G., & Nation, I. S. P. (1984) A university word list. *Language, Learning and Communication*, *3*(2), 215–229.

17 Vocabulary and writing

Future research, tools, and practices

Averil Coxhead

Exploring vocabulary use in writing

There have been several key avenues of research into vocabulary use in writing in recent times. One such avenue focuses on what vocabulary is used in student writing using a corpus-based approach. Hyland & Tse (2007), for example, used student and professional writing to explore the concept of general academic vocabulary using Coxhead's (2000) Academic Word List. More recently, this research is extending to the use of multiword units in student writing. Shin, Cortes, and Yoo (2018), for example, examine Korean college learner writing and identify omission errors with definite article use as a problem in the use of lexical bundles in writing. Vo (2019) looked into the use of both single words and lexical bundles in L2 writing for English placement tests, while Staples and Reppen's (2016) researched lexico-grammatical patterns in writing by first year university students with English, Arabic, or Chinese as their first languages. See also Durrant (2019) for an excellent discussion on English for Academic Purposes (EAP) and formulaic language. There is much more work to be done in this area, particularly in technical vocabulary and multiword units (see Chapter 15, this volume).

Another avenue for research is evaluating the effect of pedagogical approaches and their possible effects on vocabulary use in writing. These studies investigate teacher-oriented questions such as why learners do not seem to use vocabulary that has been the focus of instruction in class in their writing, what might be the best approaches to fostering vocabulary use in writing, or whether mandating the use of new words in sentences in writing is useful for learning vocabulary. An early study on classroom-based approaches to instruction on formulaic sequences in an EAP course in the UK by Jones and Haywood (2004) highlighted difficulties such as identifying the sequences that are worth teaching, time constraints in EAP courses, and how to encourage actual use of sequences in L2 academic writing (see Coxhead, 2018a, for a call for replication of this study). Shi (2004) looked at textual borrowing by L2 writers in summary and opinion writing, finding that summary writing resulted in more shared used of language than opinion writing tasks. Coxhead (2012), in a small-scale study of vocabulary use in

EAP, found that L2 writers used different strategies for incorporating highlighted academic lexis in source texts into their own writing, such as quotation, paraphrase, and summarization. Recent research finds support for highlighting lexical items in texts to encourage noticing (see Boers et al., 2017) in that enhanced items are more likely to be recognized by readers after reading. Again, the research finds noticing and recognition of vocabulary are important for learners, but the impact appears to be limited. More help is needed to help learners and teachers with moving vocabulary from recognition to production. Barcroft (2006) investigates pushed output, finding negative effects for memory of vocabulary if L2 writers are tasked with using new vocabulary in writing, and argues that the type of processing required by learners affects their capacity to pay attention to form and meaning. That is, paying attention to form requires mental effort from the learner at the expense of paying attention to meaning, explained in the Type of Processing Resource Allocation (TOPRA) model (see Barcroft, 2015). More research is needed in pedagogical approaches to vocabulary use in writing to support efforts by learners and teachers.

The final avenue of research looks almost behind the writing to find out more about the intents, beliefs, and practices of the L2 writers themselves. This is important work, because while corpora can tell us about the vocabulary that L2 writers use, it is difficult to investigate lexis that is not used by writers. After all, if a learner does not use a word or a lexical bundle in writing, it might not be because they lack knowledge of that lexis. Therefore, we need to know more about what these writers choose to do and why. There is some research into what L2 writers might be thinking or doing when it comes to vocabulary use in writing. For example, L2 writers might not use vocabulary in their writing because they are not motivated to do so or the vocabulary itself is not 'motivated' by a writing activity (Corson, 1985; Laufer, 2003; Nation, 2001). That is, writers can complete an activity without needing to use particular lexical items. Laufer (2003) and Coxhead (2007, 2012) identify risk and confidence as factors that affect vocabulary use. Coxhead (2011, 2012) analysed vocabulary use based on an EAP reading-writing task and interviewed the L2 writers about the task. The participants in this study reported that using vocabulary in writing was not taken lightly and was a considerable source of concern. The writers considered the audience for their writing when choosing which words to use (Coxhead, 2012), and other factors that affect vocabulary use included the knowledge of vocabulary in relation to writing topics (Coxhead, 2007). Individual responses to writing tasks resulted in quite different profiles of vocabulary use, from reproduction of the source text almost in its entirety through to little reference to the source text by a writer with substantial existing knowledge of the topic (thereby not really needing to call on the source text for writing). Finding out more about what our learners choose to do with vocabulary in writing, or what factors are affecting those choices, we might be able to help shift some of the lexical items that learners recognize to being words that learners actually use.

Tools for research

In terms of the tools which could be used to examine vocabulary use in writing, Anthony (2019) and Booth (Chapter 16, this volume) both report on various corpus-based tools and approaches for text analysis. If a more qualitative approach is favoured, Coxhead (2018b) outlines several possible approaches to these kinds of data.

Potential research questions

1. How and why do L2 writers incorporate technical single and multiword units in their writing? What inhibits and encourages this use?
2. What features of pedagogical writing tasks might foster vocabulary use in writing by L2 writers?
3. If a course adopts Nation's (2013) Four Strands approach, what effect might there be on productive vocabulary use in writing by L2 writers?
4. Are the factors affecting vocabulary use in writing in EAP also found in general English L2 writing? What other factors might also play a part and how important are they?
5. How do L2 writers' beliefs and practices in relation to vocabulary use in writing change over time? How do their beliefs and practices vary according to language background? What evidence is there of a change in learners' beliefs and practices over a course of language studies, university studies, or in workplaces?

Readers will find more suggestions for research directions in Webb's (2020) volume on vocabulary studies.

References

Anthony, L. (2019). Tools and strategies for data-driven learning (DDL) in the EAP writing classroom. In Hyland, K., &Wong, K. (eds) *Specialised English: New directions in ESP and EAP research* (pp. 179–194). London: Routledge.

Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research*, *22*(4), 487–497.

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. Amsterdam: John Benjamins.

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, *27*, 448–469.

Corson, D. (1985). *The lexical bar*. Oxford: Pergamon Press.

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

Coxhead, A. (2007). Factors and aspects of knowledge affecting L2 word use in writing. In Davidson, P., Coombe, C., Lloyd, D., & Palfreyman, D. (eds) *Teaching and learning vocabulary in another language* (pp. 331–342). Dubai: TESOL Arabia.

Coxhead, A. (2011). What is the exactly word in English? Investigating second language vocabulary use in writing. *English Australia*, *27*(1), 3–17.

Coxhead, A. (2012). Academic vocabulary, writing and English for Academic Purposes: Perspectives from second language learners. *RELC Journal*, *43*, 137–145.

Coxhead, A. (2018a). Replication research in pedagogical approaches to formulaic sequences: Jones & Haywood (2004) and Alali & Schmitt (2012). *Language Teaching*, *51*(1), 113–123.

Coxhead, A. (2018b). *Vocabulary and English for Specific Purposes research: Quantitative and qualitative perspectives*. London: Routledge.

Durrant, P. (2019). Formulaic language in English for Academic Purposes. In Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (eds) *Understanding formulaic language: A second language acquisition perspective* (pp. 211–227). Abingdon: Routledge.

Hyland, K., & Tse, P. (2007). Is there an 'academic vocabulary'? *TESOL Quarterly*, *41*(2), 235–253.

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In Schmitt, N. (ed.) *Formulaic sequences* (pp. 269–291). Amsterdam: John Benjamins.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*(4), 567–587.

Miralpeix, I., & Meara, P. (2014). Knowledge of the written word. In Milton, James, & Fitzpatrick, Tess (eds) *Dimensions of vocabulary knowledge* (pp. 30–44). Basingstoke: Palgrave Macmillan.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. 1st ed. Cambridge: Cambridge University Press.

Nation, I. S. P. (2013). *Learning vocabulary in another language*. 2nd ed. Cambridge: Cambridge University Press.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, *52*, 261–274.

Shi, L. (2004). Textual borrowing in second language writing. *Written Communication*, *21*(2), 171–200.

Shin, Y. K., Cortes, V., & Yoo, I. W. H. (2018). Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing*, *39*, 29–41.

Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing*, *32*, 17–35.

Vo, S. (2019). Use of lexical features in non-native academic writing. *Journal of Second Language Writing*, *44*, 1–12.

Webb, S. (2020) (ed.). *Routledge handbook of vocabulary studies.* Abingdon, Oxon: Routledge.

Part VI
Conclusion

18 Conclusion

Vocabulary and the four skills – pedagogy and practice

Paul Booth and Jon Clenton

The chapters in this volume help to identify how lexis relates to each of the four skills. In each of the four skills there is the processing, knowledge, and use of lexis in both orthographic and aural form. This is why understanding how the multifaceted nature of lexis relates to skills is a complex undertaking which these chapters have given us insights into. We can start to understand how different skills call upon different areas of the mental lexicon under different conditions for language knowledge and use. A simple binary distinction between vocabulary knowledge and use does not do justice to the complexity in which lexis calls upon different aspects in breadth, depth, and fluency. The following sections bring together the conclusions from the chapters in order to understand how each study relates to a particular skill area.

Listening

The section on vocabulary and listening highlights how under-researched this skill area is. The overview chapter from Suzanne Graham and Penchong Zhang surveys research ideas using multi-media such as television and video and even music which are becoming ever more pervasive due to people having more access to digital media.

Zhang and Graham's main chapter focused on how three different types of aural focus-on-form explanations provided by L2 only, codeswitched (CS), and contrasted focus on form (CFoF) to help learners retain vocabulary knowledge at different concreteness levels of different word types. The chapter highlighted the effects of contrastive focus on form as a way in which to balance the uptake of concrete versus abstract words. Although abstract words might be universal (e.g., *love* and *hate*) their results highlight just how important a first language is to scaffold learners to interpret lexical items which may not directly map onto the first language (Pavlenko, 2009). This additional cross-linguistic information on how a lexical item functions may provide the learner with a richer base from which to understand and remember lexical items. The authors state that contrastive focus on form is effective at different concrete levels and code-switching for abstract words in particular. The authors state that CFoF is in tune with dual coding theory

(Paivio, 1986), in which abstract words are brought up to concrete lexical item level of understanding through verbal-based and imagery-based information via the extra contextual information by contrasting L1 and L2. Zhang and Graham's chapter also focuses on Chinese learners of English and how different teaching approaches have specialized effects on the understanding of different word types.

James Milton and Ahmed Masrai's chapter on the lexicon and listening looked at how the phonological form of a word may be less developed in an L2 than the orthographical for the same word. Their main chapter argues that, for vocabulary size, written skills may more strongly relate to written comprehension than listening skills to aural comprehension. For example, while it is well known that vocabulary size relates to reading comprehension (e.g., Laufer 1992), what remains unclear is how phonological form influences vocabulary knowledge and, in particular, how phonological form relates to listening comprehension. Milton and Masrai show how L1 Arabic speakers have a more balanced lexicon in terms of orthographic and phonological forms of lexical knowledge, and that this is probably due to Arabic orthography having no relationship to English orthography.

Milton and Masrai's forward-looking chapter highlighted the need for research to help uncover how spoken language and storage is operationalized in the mental lexicon and how spoken words are processed. The underlying reason why we lack this knowledge is that, as Milton and Masrai argue, we lack the tools to investigate aural lexis.

Overall what these chapters in this listening section allude to is the need to develop learners' listening skills through informal listening opportunities. Although the uptake of lexis via listening lags behind that of through reading (Milton et al., 2010), what remains important is that learners do not become anxious when it comes to listening in another language, as oral communication is vital in a world of digital technology that makes speaking across countries so much more available. Likewise there needs to be more opportunities to test learners on their listening skills so that more research can develop our understanding of the phonological mental lexicon.

Reading

We turn our attention now to the chapters that discussed vocabulary and reading. Jeanine Treffers-Daller's chapter on current research shows that there is a difficulty order for L2 learners as they move from recall (active then passive) to recognition (active then passive), with recall being more demanding than recognition. Most tests, Treffers-Daller argues, focus on the less demanding vocabulary recall which is likely give an overestimation of learners' vocabulary size.

Irena Elgort's main chapter highlighted the reciprocal relationship between vocabulary and reading. In this chapter, she focused on the variance between age, vocabulary, and reading between higher and lower levels with higher

proficiency early bilinguals and less skilled lower proficiency learners who tend to be older. For the higher proficiency learners what seems to be a more effective way of acquiring lexis through reading is unassisted contextual word learning. What this appears to imply is that because such learners have existing lexical semantic networks in their mental lexicon, their word learning seems to be associated with rate of occurrences of unknown lexis in a text, quality and diversity of encounters, and quality of reading. With Elgort's less proficient learners, their word learning through reading needs to be scaffolded with form-focused and meaning-focused activities. This way of learning Elgort argues, stems from the fact that such learners have fewer and weaker L2 semantic connections in their lexicon, with the implication being that there needs to be mapping of form and meaning in both directions. Several pedagogical approaches are discussed but what benefit both Chinese and Dutch learners are word writing activities that activate both aspects of meaning and form. However, the identified research gap is how to develop non-declarative knowledge of lexis which is understood to be needed in processing language in real time whilst reading.

How vocabulary and reading relate is analysed in relation to reading tests and vocabulary measures. Jeanine Treffers-Daller and Jingyi Huang's main chapter examined the validity of the Test for English Majors Band 4 (TEM-4) in comparison with the York Assessment of Reading Comprehension Secondary (YARC) and vocabulary measures: Vocabulary Size Test and the Vocabulary Knowledge Scale. The vocabulary measures focused on size and depth of vocabulary knowledge respectively. There was no correlation between the two reading tests, which was surprising, but when the TEM-4 was examined the authors found that only literal meanings of the texts were tested. In contrast, the YARC reading comprehension test had inferential questions which went beyond literal meaning of the text. The differences between the TEM-4 and YARC was highlighted by the differences in relationships between the two reading tests and the vocabulary measures. Regression analysis showed that vocabulary depth (VKS) was a significant predictor of reading comprehension from the YARC test, whereas vocabulary size (VST) and depth (VKS) explain variance in reading from the TEM-4. The authors highlight several weaknesses in the TEM-4 that relate back to the need for reading tests to be based on a model of reading which needs to include both decoding and linguistic comprehension of which inferencing is an important part. Contextual vocabulary learning is argued to be augmented by fine tuning of vocabulary knowledge in orthography and semantic representations to facilitate efficient reading.

Elgort's overview of future research, tools, and practices outlines what is needed in pedagogy and our understanding of developing learners' L2 reading. Such research projects could look at augmenting contextual learning through reading with handwriting, inferencing, and dictionary work; pushing learners to use contextually learnt words; repeated reading with repeated instances of multiword expressions.

What these chapters in this reading section indicate is that reading and vocabulary are reciprocal to the extent that one develops the other and vice versa. It is not enough to simply know the meaning of lexical items but that depth of knowledge is vital for both fluency and reading comprehension. Less skilled and lower proficiency learners will need more scaffolding to develop lexical semantic connections to encourage deeper understanding and greater fluency in reading. How we measure reading needs to go beyond literal understanding as texts are embedded with meaning which is co-constructed by the text and reader.

Speaking

Takumi Uchihara's current research chapter highlighted how the knowledge of vocabulary can be examined by assessing the samples of speech (dependent) or by eliciting lexical knowledge and speaking data separately (independent). Current research, Uchihara argues, is mostly based on the independent method in which various interrelated aspects of vocabulary and speaking; for example, fluency and vocabulary knowledge are related. Other research focuses on lexical richness (frequency and diversity) in speech in which indications of lexical size and use are associated with richer language in speech but indications of larger vocabulary size do not necessarily relate to richer speech but are more associated with rater perceptions of accuracy. This chapter also shows the multifaceted nature of human speech in L2 and that the tools to evaluate speech can be human ratings or more objective measures of errors.

Jon Clenton, Nivja de Jong, Dion Clingwall, and Simon Fraser's main chapter focused on a sub-set of L2 speech in which they investigated vocabulary knowledge and skills and how they relate to fluency in speech. Their overall conclusion is that vocabulary knowledge and fluency are probably proficiency dependent, with another conclusion suggesting that productive vocabulary use from a productive vocabulary task (Lex30) overlaps with the speaking fluency output indicated by the Academic Spoken Word List (ASWL): the higher their participants scored in the Lex30 task the more ASWL words were used. This chapter shows that vocabulary knowledge and speaking are not simply binary distinctions but within the knowledge and skill area there is a complex system that can be examined from different perspectives. If we take vocabulary knowledge as the predictor variable it can measured in different ways, which has an impact on which sub-set area of speaking it is predicted to be related to. For example, in this chapter there was no significant correlation between receptive vocabulary knowledge and speaking fluency; however, there was a moderately significant correlation between higher productive vocabulary from the Lex30 and fewer silent pauses. As the authors highlight, speaking is a productive skill but it may not mirror writing and so the relationships found in vocabulary and speaking are probably not found in vocabulary and writing. This is understandable as in speaking (oral language) has to be processed in real time

whereas in writing (orthographic) the time factor is not so acute. Moreover, as these chapters on speaking indicate, phonological form may be at different levels of development to orthographical form in the mental lexicon. What this chapter has tried to do is untangle some of the relationships between knowledge and use of vocabulary with the measurable aspects of spoken language.

The relationship between productive vocabulary and second language oral ability was the focus of the main chapter by Takumi Uchihara, Kazuya Saito, and Jon Clenton. Along with the previous chapter on speaking skills and vocabulary, this chapter also used Lex30 as a measure of productive vocabulary and one of the conclusions is that L2 speakers with larger productive vocabulary are rated as more comprehensible. Moreover, the measure of fluency is also related to ratings of comprehensibility. So, when we consider productive vocabulary there are relationships to speaking skills in terms of fluency and comprehensibility. When the lexis in the speech is examined there are relationships with lexical diversity and lexical sophistication, i.e. lexis beyond the 2,000 frequency level. However, the appropriateness of the lexis is not related to lexical production so the authors recommend that there needs to be a reappraisal in using frequency-based measures because although some learners may use rare words these lexical items may not always be used in a correct manner. Appropriate use of lexis in speech is an important consideration as it can be more important than grammar in understanding what the speaker is trying to communicate.

Jon Clenton's future-facing chapter emphasized that vocabulary knowledge and speaking are both multifaceted and that measures of both vocabulary and speaking need to take this into account. If only one measure of vocabulary is used then it may not reflect the complex nature of lexis and so will be limited in how this knowledge is related to the several aspects of speaking.

The chapters in this speaking section highlight that the relationship between vocabulary and oral ability needs to be examined in a holistic manner, taking multiple levels of vocabulary knowledge and speech production into consideration. Objective measures can uncover certain relationships but we also need qualitative, subjective measures otherwise we do not get the full picture of how vocabulary relates to speech production. The tests themselves also need to be considered carefully as the authors highlight that Lex30 is a measure of vocabulary recall and not about vocabulary use which is why it may not correlate with lexical appropriateness. Likewise, when examining spoken data, L2 corpora should be taken into consideration as L1 corpora do not consider the nature of the L2 lexicon.

Vocabulary and writing

Paul Booth's chapter on current research divided the current tools of written vocabulary knowledge into extrinsic and intrinsic measures. The former compare lexis from written samples with frequency lists and the latter with internal measures of type-token ratio but with more recent tools considering the

number of words produced. Both sets of measures have their strengths and weaknesses so deciding which measure to use with written data needs to consider what aspect of vocabulary knowledge the researcher needs to obtain from L2 writers as, for example, the task topic will influence the words produced.

The context in which specialized vocabulary is used was the focus of Avril Coxhead's main chapter on how technical vocabulary is used in building students' diaries from a trade education carpentry course. One of the main themes running though this chapter is how diaries are central to their learning; for example, how the writing of words is important to the remembering of words at a technical level. Later entries in the diaries showed how there were more words from the Carpentry Word List than at the beginning of the course. This has pedagogical implications in that ELT courses could make diaries obligatory so that learners are writing meaning-focused texts in that diaries can help learners to reflect on their experiences. Interestingly, the diaries which were marked lower by the tutors relied more on technical words from the Carpentry Word List. This could suggest that these learners (not necessarily L2 or EFL) relied more heavily on texts in class written by their tutors. Overall both the high and low scoring diaries had over 80% of the words from the first 3,000 frequency bands.

Paul Booth's main chapter on lexical developmental paths in relation to academic writing also looked at word frequency but his focus was on lexis in relation to memory and analysis learner differences. Overall learners with higher analytical scores tended to develop more lexis beyond the 2,000 frequency band over a period of one semester. This development could be linked to depth of processing (Craik and Lockhart, 1972) as the analysis test focused on finding patterns within texts which could be the result of depth of processing. Learners strong in analysis (c/f memory) also showed greater correlation in the beyond 2k scores over two time periods which may mean a greater consistency in their use of rarer words. Data-driven learning could be one way of facilitating learners' semantic knowledge of lexical items as words are shown in authentic texts.

Other important factors to consider are outlined in Coxhead's future-facing chapter which are learners' intents, beliefs, and practices when writing in an L2. This looks like an important area, as so much research has examined the lexis learners produce in their writing but what also seems important is why learners may avoid certain lexis. This could give us insights into the motivations behind the learner especially in relation to the writing task itself which could have important pedagogical implications.

The chapters in this section have shown that writing can be used as a pedagogical tool for learners to process more deeply technical vocabulary. Lexis in written texts is also shaped by the learner in that more analytical learners may process lexis more deeply. The actual purpose for which learners are writing also shapes the words which are used. Moreover, there needs to be more research which uncovers the writers' beliefs as learners can bypass certain types of lexis, but we are still unsure of why this could be the case.

Final thoughts

Perhaps the strongest thread running through this book is that, although we have labels for vocabulary knowledge and skills, what underlie these labels are areas of complexity which have been explored in this volume. The very nature of aural vocabulary makes it distinct from written. How learners understand and use lexical items draw upon different lexical knowledge. Listening and reading, although both receptive skills, can be seen as distinct in that the processing of each will call upon different areas of the mental lexicon. Likewise speaking and writing are categorized as productive but each call upon different aspects of lexical knowledge. Therefore a learner who is comprehensible in speaking may not be as comprehensible in writing. What these chapters have started to show is that skills draw upon different areas of vocabulary knowledge and that the context in which the skill is used makes a difference to the vocabulary knowledge needed.

While it is not possible to give 'tips' for teaching, as there is no simple cause or effect, these chapters have shown that when we consider vocabulary and each of the four skills in a single volume there are patterns that can be drawn that might inform future research and pedagogy. We need, therefore, to understand that aural lexis is different from orthographic form and that how learners process and use lexis is also shaped by the specific skill area.

References

Brown, D. (2008). Using a modified version of the vocabulary knowledge scale to aid vocabulary development. *Language Teacher*, *32*(12). No page numbers.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, *11*, 671–684.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In Arnaund, P. J. L., & Béjoint, H. (eds) *Vocabulary and applied linguistics* (pp. 129–132). London: Macmillan.

Milton J., Wade, J., & Hopkins, N. (2010) Aural word recognition and oral competence in a foreign language. In Chacón-Beltrán, R., Abello-Contesse, C., & Torreblanca-López, M. (eds) *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Pavlenko, A. (2009). *The bilingual mental lexicon: Interdisciplinary approaches*. Bristol-Buffalo-Toronto: Multilingual Matters.

Index