

palgrave
macmillan

Researching Vocabulary

Norbert Schmitt

Researching Vocabulary

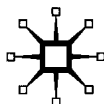
A Vocabulary Research Manual

Norbert Schmitt

University of Nottingham, UK



palgrave
macmillan



© Norbert Schmitt 2010

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6-10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The author has asserted his right to be identified as the author of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published 2010 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN: 978-1-4039-8535-4 hardback

ISBN: 978-1-4039-8536-1 paperback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Schmitt, Norbert.

Researching vocabulary : a vocabulary research manual / Norbert Schmitt.
p. cm.—(Research and practice in applied linguistics)

Includes bibliographical references and index.

ISBN 978-1-4039-8536-1 (pbk. : alk. paper) —

ISBN 978-1-4039-8535-4 (alk. paper)

1. Language and languages—Study and teaching. 2. Vocabulary—Study and teaching. 3. Second language acquisition. I. Title.

P53.9.S365 2010

418.007'2—dc22

2009046796

10 9 8 7 6 5 4 3 2

19 18 17 16 15 14 13 12 11 10

Printed and bound in Great Britain by
CPI Antony Rowe, Chippenham and Eastbourne

Quick Checklist (Principal sections which discuss these issues)

Target lexical items

- Do any lexical characteristics potentially confound your results? (2.1–2.4, 4.5)
- Have you taken frequency into account? (2.5)
- Does L1 influence potentially confound your results? (2.6)
- Is your sampling rate sufficient to make your results meaningful? (4.6)
- Have you considered including formulaic sequences as well as individual words? (3)

Measurement instruments

- Are they valid, reliable, and appropriate for your participants? (5)
- Are they suitable for answering your research questions? (whole book)
- Are you measuring receptive or productive mastery, or both? (2.8)
- Have you considered measuring word knowledge aspects besides meaning and form? (1.1.5, 4.3, 5.3)
- Have you considered measuring depth of lexical knowledge? (5.3)
- Have you considered measuring lexical organization and speed of processing? (2.4, 2.11, 5.4, 5.5)
- If the study is focused on acquisition, is previous lexical knowledge determined or controlled for? (5.1.2)
- If the study is focused on acquisition, are there delayed posttests? (4.4)

Participants

- Are there enough participants to make the study viable? (4.2)

Corpus issues

- Is the corpus you use appropriate for your research questions? (1.1.4, 3.8, 6.2)

Reporting

- Were the units of counting clearly described? (5.2.1)
- Did you discuss the absolute size of any gain/attrition? (4.7)
- Did you report effect sizes? (4.7)
- Are your interpretations and conclusions warranted based on your results? (4.7)

Bottom line

- Is your study interesting?
- Is your study useful to anyone?

General Editors' Preface

Research and Practice in Applied Linguistics is an international book series from Palgrave Macmillan which brings together leading researchers and teachers in Applied Linguistics to provide readers with the knowledge and tools they need to undertake their own practice related research. Books in the series are designed for students and researchers in Applied Linguistics, TESOL, Language Education and related subject areas, and for language professionals keen to extend their research experience.

Every book in this innovative series is designed to be user-friendly, with clear illustrations and accessible style. The quotations and definitions of key concepts that punctuate the main text are intended to ensure that many, often competing, voices are heard. Each book presents a concise historical and conceptual overview of its chosen field, identifying many lines of enquiry and findings, but also gaps and disagreements. It provides readers with an overall framework for further examination of how research and practice inform each other, and how practitioners can develop their own problem-based research.

The focus throughout is on exploring the relationship between research and practice in Applied Linguistics. How far can research provide answers to the questions and issues that arise in practice? Can research questions that arise and are examined in very specific circumstances be informed by, and inform, the global body of research and practice? What different kinds of information can be obtained from different research methodologies? How should we make a selection between the options available, and how far are different methods compatible with each other? How can the results of research be turned into practical action?

The books in this series identify some of the key researchable areas in the field and provide workable examples of research projects, backed up by details of appropriate research tools and resources. Case studies and exemplars of research and practice are drawn on throughout the books. References to key institutions, individual research lists, journals and professional organizations provide starting points for gathering information and embarking on research. The books also include annotated lists of key works in the field for further study.

The overall objective of the series is to illustrate the message that in Applied Linguistics there can be no good professional practice that isn't based on good research, and there can be no good research that isn't informed by practice.

Christopher N. Candlin and David R. Hall
Macquarie University, Sydney

Preface

This is a vocabulary research manual. It aims to give you the background knowledge necessary to design rigorous and effective research studies into the behavior of L1 and L2 vocabulary. It can also help you better understand other people's research and interpret it more accurately. In order to keep the manual to a reasonable length, I assume that you already have an understanding of basic research methodology for language research in general, and also have a basic understanding of statistics. I also assume you have a general understanding of vocabulary issues. The manual will build on this knowledge and discuss the issues which have particular importance for vocabulary research. The exception to these assumptions of previous knowledge is statistical knowledge about corpus linguistics (e.g. *t-score* and *MI*), which is more specific to vocabulary research, and so the calculations behind these statistical procedures are spelled out in Chapter 3. In addition, I have almost always built descriptions of terminology and concepts into the text, but in a few cases have added Concept Boxes to supplement the text.

I did not want this book to be just my personal take on vocabulary research, but rather wished it to be a consensus state-of-the-art research manual. While it inevitably reflects my own interests and biases (and uses many of the studies I have been involved with for illustration), I have been extremely fortunate that many of my friends in the field of vocabulary studies have been willing to read all or parts of the book and provide comments. I often incorporated their insightful critiques more-or-less directly into the text, and the final version of the book is greatly improved by the process. As a result, I feel that the book does reflect a (somewhat personalized) consensus view of good vocabulary research practice. While many of my colleagues might do certain things differently than indicated in this book, it does indicate the major issues which need to be considered to carry out worthwhile vocabulary research, and hopefully will help you to avoid many of the pitfalls that exist.

Although most of the issues discussed in this handbook pertain to vocabulary research in any language, the majority of research to date has been on English, including my own personal research. Almost inevitably, this has led to the majority of examples and citations referring to the English language. There is no value judgement intended in this, and I hope you are able to take the ideas and techniques and apply them to the languages you are researching.

This handbook can't tell you the exact research methodologies to use, as every lexical study is different, entailing unique goals and difficulties. However, I have tried to provide enough background information about the nature of vocabulary and discussion of possible research methodologies to help guide you in thinking about the issues necessary in selecting and developing sound methodologies for the lexical research you wish to do.

I love vocabulary research, and with so many questions still unanswered, I want to encourage as much of it as I can. I hope this book stimulates you to begin researching vocabulary yourself, or to keep researching if you are already at it. It is a fascinating area, and I hope to hear your results at a future conference and/or read them in a future journal.

Norbert Schmitt
Nottingham June 1, 2009

Acknowledgements

I would like to thank the University of Michigan for giving my wife a Morley scholarship to study in Ann Arbor in July 2008. This allowed me to write a large portion of this book in the wonderful environs of the Rackham Building on their campus. All that was missing from the atmosphere was Indiana Jones sliding through the library on his motorcycle.

Colleagues who have graciously commented on the entire manuscript include Paul Nation, Birgit Henriksen, Averil Coxhead, and Ronald Carter. Their many perceptive comments have improved the final version, and helped to make it more complete. I also owe a debt of thanks to numerous colleagues who commented on the parts of the book where their particular specialisms were covered, or who contributed material. Their input has added much to the rigor of the book: Frank Boers, Tom Cobb, Kathy Conklin, Zoltán Dörnyei, Philip Durrant, Catherine Elder, Nick Ellis, Glen Fulcher, Tess Fitzpatrick, Lynne Flowerdew, Gareth Gaskell, Sylviane Granger, Kirsten Haastrup, Marlise Horst, Jan Hulstijn, Kon Kuiper, Batia Laufer, Phoebe Lin, Ron Martinez, Paul Meara, Imma Miralpeix, Anne O'Keeffe, Spiros Papageorgiou, Sima Paribakht, Aneta Pavlenko, Pam Peters, Diana Pulido, Ana Maria Pellicer Sánchez, Paul Rayson, John Read, Ute Römer, Diane Schmitt, Rob Schoonen, Barbara Seidlhofer, Anna Siyanova, Suhad Sonbul, Pavel Trofimovich, Mari Wesche, Cristina Whitecross, and David Wood.

Comments from my editors Chris Candlin and David Hall did much to sharpen both the thinking and presentation of the material. Of course, everyone had slightly different views on the best research methodologies and other content of the book, and so the final distillation of the various points of view is my personal interpretation for which I alone am responsible.

Finally, to my wife Diane, for commenting on the manuscript, but more importantly, for taking me to places like Carcassone, Ann Arbor, Auckland, and Copenhagen where writing various parts of the book was a pleasure. I love you more than ever.

The author and publishers wish to thank Wiley-Blackwell, Elsevier and Lee Osterhout for permission to reproduce copyright material:

Figure 2.1 The Relationship between Historical Origin and Register, G. Hughes *A History of English Words*, 2000, Malden, MA: Blackwell p. 15

Figure 2.12 ERP plots showing N400 and P600 phenomena, Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, and Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means

for exploring the neurocognition of second language processing. *Language Learning 56, Supplement 1*: p. 204.

Figure 2.13 fMRI brain location results; Hauk, O., Johnsrude, I. & Pulvermüller, F. Somatotopic representation of action words in the motor and premotor cortex. *Neuron 41*, 301–307 (2004), Elsevier Science

Contents

<i>Quick Checklist</i>	xii
<i>General Editors' Preface</i>	xiii
<i>Preface</i>	xiv
<i>Acknowledgements</i>	xvi

Part 1 Overview of Vocabulary Issues

1 Vocabulary Use and Acquisition	3
1.1 Ten key issues	3
1.1.1 Vocabulary is an important component of language use	3
1.1.2 A large vocabulary is required for language use	6
1.1.3 Formulaic language is as important as individual words	8
1.1.4 Corpus analysis is an important research tool	12
1.1.5 Vocabulary knowledge is a rich and complex construct	15
1.1.6 Vocabulary learning is incremental in nature	19
1.1.7 Vocabulary attrition and long-term retention	23
1.1.8 Vocabulary form is important	24
1.1.9 Recognizing the importance of the L1 in vocabulary studies	25
1.1.10 Engagement is a critical factor in vocabulary acquisition	26
1.2 Vocabulary and reading	29
1.3 A sample of prominent knowledge gaps in the field of vocabulary studies	35

Part 2 Foundations of Vocabulary Research

2 Issues of Vocabulary Acquisition and Use	47
2.1 Form-meaning relationships	49
2.1.1 Single orthographic words and multi-word items	49

2.1.2	Formal similarity	50
2.1.3	Synonymy and homonymy	52
2.1.4	Learning new form and meaning versus 'relabelling'	52
2.2	Meaning	52
2.2.1	Imageability and concreteness	53
2.2.2	Literal and idiomatic meaning	53
2.2.3	Multiple meaning senses	54
2.2.4	Content versus function words	54
2.3	Intrinsic difficulty	55
2.4	Network connections (associations)	58
2.5	Frequency	63
2.5.1	The importance of frequency in lexical studies	63
2.5.2	Frequency and other word knowledge aspects	64
2.5.3	L1/L2 frequency	66
2.5.4	Subjective and objective estimates of frequency	67
2.5.5	Frequency levels	68
2.5.6	Obtaining frequency information	70
2.6	L1 influence on vocabulary learning	71
2.7	Describing different types of vocabulary	75
2.8	Receptive and productive mastery	79
2.9	Vocabulary learning strategies/self-regulating behavior	89
2.10	Computer simulations of vocabulary	97
2.11	Psycholinguistic/neurolinguistic research	105
3	Formulaic Language	117
3.1	Identification	120
3.2	Strength of association – hypothesis tests	124
3.3	Strength of association – mutual information	130
3.4	A directional measure of collocation	131
3.5	Formulaic language with open slots	132
3.6	Processing formulaic language	134
3.7	Acquisition of formulaic language	136
3.8	The psycholinguistic reality of corpus-extracted formulaic sequences	141
3.9	Nonnative use of formulaic language	142
	Part 3 Researching Vocabulary	
4	Issues in Research Methodology	149

4.1	Qualitative research	149
4.2	Participants	150
4.3	The need for multiple measures of vocabulary	152
4.4	The need for longitudinal studies and delayed posttests	155
4.5	Selection of target lexical items	158
4.6	Sample size of lexical items	164
4.7	Interpreting and reporting results	166
5	Measuring Vocabulary	173
5.1	Global measurement issues	173
5.1.1	Issues in writing vocabulary items	174
5.1.2	Determining pre-existing vocabulary knowledge	179
5.1.3	Validity and reliability of lexical measurement	181
5.1.4	Placing cut-points in study	187
5.2	Measuring vocabulary size	187
5.2.1	Units of counting vocabulary	188
5.2.2	Sampling from dictionaries or other references	193
5.2.3	Recognition/receptive vocabulary size measures	196
5.2.4	Recall/productive vocabulary size measures	203
5.3	Measuring the quality (depth) of vocabulary knowledge	216
5.3.1	Developmental approach	217
5.3.2	Dimensions (components) approach	224
5.4	Measuring automaticity/speed of processing	242
5.5	Measuring organization	247
5.6	Measuring attrition and degrees of residual lexical retention	256
6	Example Research Projects	260
Part 4 Resources		
7	Vocabulary resources	279
7.1	Instruments	279
7.1.1	Vocabulary levels test	279
7.1.2	Vocabulary size test	293

7.1.3	Meara's_lognostics measurement instruments	306
7.2	Corpora	307
7.2.1	Corpora representing general English (mainly written)	309
7.2.2	Corpora representing spoken English	320
7.2.3	Corpora representing national varieties of English	323
7.2.4	Corpora representing academic/business English	324
7.2.5	Corpora representing young native English	325
7.2.6	Corpora representing learner English	325
7.2.7	Corpora representing languages other than English	326
	7.2.7.1 Parallel corpora	326
	7.2.7.2 Monolingual corpora	327
7.2.8	Corpus compilations	331
7.2.9	Web-based sources of corpora	333
7.2.10	Bibliographies concerning corpora	335
7.3	Concordancers/tools	335
7.4	Vocabulary lists	345
7.5	Websites	347
7.6	Bibliographies	351
7.7	Important personalities in the field of vocabulary studies	352
	<i>Notes</i>	359
	<i>References</i>	362
	<i>Index</i>	385

Part 1

Overview of Vocabulary Issues

1

Vocabulary Use and Acquisition

This is a vocabulary research manual whose primary goal is to provide readers with a solid foundation of vocabulary research methodology, both in terms of good research practice, and in terms of the common pitfalls to avoid. But in doing research, we must always make methodology serve the research issues we are interested in exploring. The issues which attract the most attention (and thus research) in the field of vocabulary concern the nature of lexis, its employment in language use, and the best ways of facilitating its acquisition. In order to design good vocabulary research on these issues, one must be on good terms with what the field already knows about these issues. There are a number of good overviews/collections which should be reviewed to gain a general understanding of vocabulary and its behavior (e.g. Bogaards and Laufer, 2004; Carter, 1998; Coady and Huckin, 1997; Daller, Milton, and Treffers-Daller, 2007; Hunt and Beglar, 1998, 2005; McCarthy, 1990; Meara, 2009; Nation, 1990, 2001; Read, 2000, 2004; Schmitt, 2000, 2008; Schmitt and McCarthy, 1997; Singleton, 1999). This chapter of the book will follow up on the information in these publications and highlight ten key issues which must be taken into account when designing vocabulary research. They are outlined below and have direct implications for the discussion of methodology in the following chapters of the book. I will then identify a number of important vocabulary issues about which we do not yet have much knowledge, and how these gaps affect lexical research.¹

1.1 Ten key issues

1.1.1 Vocabulary is an important component of language use

Quote 1.1 Wilkins on the importance of vocabulary for communication

Without grammar very little can be conveyed, without vocabulary nothing can be conveyed.

(1972: 111)