



Swansea University
Prifysgol Abertawe



Swansea University E-Theses

The use and abuse of EFL dictionaries: How learners of English as a foreign language read and interpret dictionary entries.

Nesi, Hilary

How to cite:

Nesi, Hilary (1994) *The use and abuse of EFL dictionaries: How learners of English as a foreign language read and interpret dictionary entries..* thesis, Swansea University.

<http://cronfa.swan.ac.uk/Record/cronfa42801>

Use policy:

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence: copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder. Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

Please link to the metadata record in the Swansea University repository, Cronfa (link given in the citation reference above.)

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

The Use and Abuse of EFL Dictionaries:

**how learners of English as a foreign
language read and interpret
dictionary entries.**

Hilary Nesi

A thesis submitted for the degree of Philosophiae Doctor

**University College of Swansea
University of Wales**

1994

ProQuest Number: 10807577

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10807577

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

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



DECLARATION

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

Signed  (candidate)

Date..30/8/94.....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed  (candidate)

Date..30/8/94.....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed  (candidate)

Date..30/8/94.....

ACKNOWLEDGEMENTS

First, I would like to thank all 373 English language students who gave up their time to take part in the experiments recorded here. I would also like to thank those lecturers and course tutors who kindly allowed me access to the students, often at considerable inconvenience to themselves: Dorothy Whittington from Aston University, Philip King from Birmingham University, John McKenny from the Instituto Politécnico de Viseu, Noritah Omar and Arshad Abdul Samad from Universiti Pertanian Malaysia, Eryl Griffiths and Steve Chadwick from Eurocentres Cambridge, Frances Brindley from Coventry Technical College, and Richard Badger from Henley College of Further Education.

I have a debt of gratitude to all my colleagues at the Centre for English Language Teacher Education (CELTE), Warwick University, who were always supportive and ready to listen to ideas and problems: Thelma Henderson, Julia Khan, Shelagh Rixon, Anna Buick and Antonia Pulsford, and especially Peter Brown, Peter Jewel and Judith Kennedy, who helped rate the appropriacy of sentences, and Meriel Bloor, who took a special interest in the progress of the studies, and made many helpful suggestions throughout.

The secretarial staff at CELTE - Anne Beale, Pratibha Chauhan and Louise Tate also deserve particular thanks, for patiently explaining wordprocessing techniques, and putting up with countless interruptions to their work without ever complaining once.

Tom Bloor, from Aston University, Andy Gillett, from Hertfordshire University and Richard Badger, from Henley College of Further Education sacrificed their time too to help me rate the appropriacy of sentences, and I would like to thank them.

Finally I would like to thank my supervisor, Paul Meara, who has helped me so much during the course of this research project. I think it is a very difficult job to be a good supervisor, and achieve the right balance between freedom and control - but Paul has succeeded, and I am very grateful for all that I have learnt from him, both directly and by example.

Table of Contents

List of abbreviations	i
List of Appendices	ii
Abstract	iii
Introduction	1
Chapter One, A review of research into dictionary use and dictionary requirements	5
1.1. Questionnaire-based research	6
1.1.1. Barnhart (1962)	6
1.1.2. Quirk (1975)	9
1.1.3. Jackson (1988)	13
1.1.4. Tomaszczyk (1979)	14
1.1.5. Baxter (1980)	28
1.1.6. Béjoint (1981)	32
1.1.7. Battenburg (1989)	41
1.1.8. Concluding comments on questionnaire -based research into dictionary use and dictionary requirements	53
1.2. Test-based research	56
1.2.1. Bensoussan, Sim and Weiss (1984)	56
1.2.2. Black (1986)	67
1.2.3. Atkins and Knowles (1990)	72
1.2.4. Tono (1988)	91
1.2.5. Tono (1989)	94
1.2.6. Bogaards (1991b)	98
1.2.7. Luppescu and Day (1993)	104

1.2.8. Concluding comments on test-based research into dictionary use and dictionary requirements	110
1.3. Observation-based research	113
1.3.1. Ard (1982)	113
1.3.2. Hatherall (1984)	119
1.3.3. Miller and Gildea (1985)	124
1.3.4. Neubach and Cohen (1988)	133
1.3.5. Ahmed (1989)	144
1.3.6. MacFarquhar and Richards (1983)	151
1.3.7. Bogaards (1990)	157
1.3.8. Concluding comments on observation-based research into dictionary use and dictionary requirements	167
1.4. Conclusion	169
Chapter Two, The effect of dictionary use on performance in a multiple-choice reading comprehension test	174
2.1 Introduction	174
2.2 Pilot study	175
2.3 Study One	177
2.4 Study Two	180
2.5 Discussion of results	184
2.6 Conclusion	195
Chapter Three, The effect of different dictionary defining styles on productive dictionary use	197
3.1. Introduction	197
3.2. Study Three	211
3.2.1. Pilot study	212
3.2.2. Changes to the test program	215
3.2.3. The main study	221

3.3. Conclusions	234
Chapter Four, The effect of language background and culture on productive dictionary use	238
4.1. Introduction	238
4.2. Study Four	240
4.3. Conclusions	257
Chapter Five, The role of examples in productive dictionary use	258
5.1. Introduction	258
5.2. Study Five	264
5.3. Conclusions	275
Chapter Six, Conclusion	276
6.1. A summary of the findings and their implications	276
6.2. A critique of the experimental method employed in Studies Three, Four and Five	281
6.3. Towards a theory of productive dictionary	287
6.4. Questions that remain to be answered	291
Bibliography	295
Appendices	311

List of abbreviations

Abbreviations of dictionary titles:

COD	=	Concise Oxford Dictionary
CULD	=	Chambers Universal Learners' Dictionary
DFLE	=	Dictionnaire du Français Langue Étrangère Larousse
LASD	=	Longman Active Study Dictionary
LDAE	=	Longman Dictionary of American English
LDOCE	=	Longman Dictionary of Contemporary English
OALD	=	Oxford Advanced Learner's Dictionary
OED	=	Oxford English Dictionary
POD	=	Pocket Oxford Dictionary
SOD	=	Shorter Oxford Dictionary
WNWD	=	Webster's New World Dictionary

Other abbreviations:

COBUILD	=	Collins Birmingham University International Language Database
EAP	=	English for Academic Purposes
EFL	=	English as a Foreign Language
ELT	=	English Language Teaching
ESL	=	English as a Second Language
KWIC	=	Key Word in Context

List of Appendices

Appendices are numbered according to the chapter to which they refer.

2.1 The reading comprehension test

2.2 Data from Study Two

3.1 Data from the Pilot Study

3.2 Data from Study Three

3.3 A sample of sentences produced by OALD users

3.4 A sample of sentences produced by LDOCE users

3.5 A sample of sentences produced by COBUILD users

4.1 Data from Study Four

4.2 A sample of sentences produced by the Malaysian group

4.3 A sample of sentences produced by the Portuguese group

5.1 Data from Study Five

5.2 A sample of sentences produced with access to examples

5.3 A sample of sentences produced without access to examples

Abstract

This thesis begins with a review of research into dictionary use. A number of experimental design problems are discussed, in particular the unreliability of questionnaire responses, and the need for detailed accounts of individual dictionary consultations whilst sampling in numbers sufficient to represent specified populations.

The experiments reported in subsequent chapters investigate issues raised in the review. The first two studies find that dictionary use during a reading comprehension test affected completion speed but not test scores. The apparent failure of dictionary use to improve comprehension is attributed to the test itself, the dictionaries, and the users' choice of look-up words.

The ability of users to interpret dictionary entries is investigated in three further studies which use computers to gather data on large numbers of individual consultations. The findings indicate that there is little difference between three major EFL dictionaries in terms of speed of consultation and overall productive success. They also indicate that Malaysian ESL subjects, who have higher vocabulary scores, are slower in their reading and less successful in their interpretation of entries than Portuguese EFL subjects. Finally, the findings suggest that overall productive success is unaffected by the presence or absence of examples.

The experimental findings lead to the conclusion that dictionary consultation is a process in which users match pre-existing beliefs about word meaning and behaviour against segments in the dictionary entry. Such segments are often selected because they are familiar-sounding and conceptually accessible, but may contain only incomplete or non-essential information. Where pre-existing beliefs and dictionary information conflict, dictionary information is sometimes overridden. Thus word knowledge acquired from a single consultation is often insufficient to ensure productive success.

Although it is probably inevitable that word knowledge will be acquired slowly, through multiple encounters, modifications to the dictionary entry and the training of users might help to avoid serious misinterpretation of dictionary information.

Introduction

All literate people own a dictionary. A dictionary is also the first thing most adults buy when they start to learn a new language. Even those who approach language learning via a structural syllabus are aware that lexis carries more meaning than grammar, and that a stock of lexical words are the first thing one needs in order to survive in a foreign language situation. English dictionary use is now a feature of high school syllabuses in most parts of the world where English is taught as a second or a foreign language; the role of dictionaries in language development is now recognised, at least by school syllabus writers and course designers. At tertiary level, English learners' dictionaries have become an essential study aid wherever English-medium learning takes place, through lectures, or, most commonly, through textbooks and journal articles which are unavailable in the learners' mother tongues. Consequently, dictionary skills development is a common component of those English language support courses which operate within universities to help overseas students cope with English-medium study.

Thus we can safely assume that the market for learners' dictionaries will remain strong, whatever the fashion in language teaching; the huge investments that have been made to develop new types of EFL dictionary in recent years reflect this strength. Unfortunately, the time, money and effort that have been channelled into creating new merchandise for the dictionary market are not matched by similar efforts in the field of research into dictionary use. Thanks to corpora studies tied to dictionary projects (the COBUILD database, for example, and the Longman Citation Corpus) we now know a great deal about word frequency across genres, collocational patterns and lexical grammar, but we still know relatively little about the ways in which this information can best be expressed for the benefit of language learners. If English language teachers around the world often ignore the dictionary skills element on the courses they teach (as my MA in ELT students inform me), this is scarcely

surprising, given that so little information is available regarding the lacks and needs of learner dictionary users. Materials for teaching dictionary skills tend to emphasise those design features of the dictionary of which the lexicographers are most proud, rather than those features which learners find most difficult to deal with. Dictionary skills are presented from the lexicographers' point of view, and require learners to mould their needs to the resources of the dictionary - learners' own skills and strategies for negotiating meaning are only rarely investigated.

Dictionary use is not an easy subject to explore. Essentially it is a private affair, easily distorted by intrusive investigative techniques. I am very conscious that no single research method can create a complete picture of the way in which a learner, or group of learners, uses a dictionary in his or her everyday life. The first chapter of this thesis describes and discusses a wide range of possible approaches to the study of dictionary use, in order to summarise key findings and identify areas of contention, and to evaluate the strengths and weaknesses of each approach. The issues raised in Chapter One have guided my choice of research topics in the following chapters, and the review has also helped me to avoid certain methodological pitfalls in my own studies - as perhaps future reviewers will be guided by the pitfalls they identify in my work.

Chapter One is divided into three sections, corresponding to the three basic methods of data collection employed in the studies I review: questionnaires, tests and observation. A broad range of approaches is subsumed under each section; many studies adopt more than one data-gathering tool, and many are also highly original in their methods. The three-part structure of Chapter One is not, therefore, intended to be anything other than a convenient way of grouping and comparing very disparate studies, which often defy categorisation.

Following the initial review chapter, this thesis describes a series of five studies to explore different aspects of receptive and productive dictionary use. In every

experiment the subjects were young adults, who had begun or who were about to begin advanced studies in the medium of English. Their English language needs were primarily academic rather than social or occupational, and this is reflected in the choice of dictionaries for the experiments: the academically-oriented Oxford Advanced Learner's Dictionary, Longman Dictionary of Contemporary English and Collins COBUILD English Language Dictionary.

The first two studies, reported in Chapter Two, are similar in approach, the second being a refinement of research techniques developed in the first. Both of these studies examine the role of dictionaries in reading comprehension, and use testing as a method of acquiring data. Study One reports on the use of a variety of different dictionaries, both monolingual and bilingual, but in Study Two subjects were restricted to the Oxford Advanced Learner's Dictionary.

The remaining three studies are concerned with the written production of dictionary users, and adopt an original research method involving the use of computers to monitor subjects' behaviour and record what they write. The first of these experiments, Study Three, sets out to compare the behaviour of users of three different learners' dictionaries - Oxford Advanced Learner's Dictionary, Longman Dictionary of Contemporary English and COBUILD - but it also examines a number of strategies which users of all three dictionaries seem to hold in common. The conclusions arising from this study provide the research questions which are addressed in Chapters Four and Five; Study Four explores the possibility that language and cultural differences can affect the speed and success of dictionary use, while Study Five examines the role of dictionary examples. Entries from Longman Dictionary of Contemporary English were used in both of these studies.

My concluding chapter summarises the key findings from all five studies and discusses their implications. Some of the findings were unexpected, and some reflect

severe dictionary-using difficulties amongst subjects who were otherwise advanced learners, capable of English medium study. Taken together, they suggest certain patterns of dictionary use and misuse which help to explain why dictionary consultation is not the instant answer to lexical deficit it is sometimes believed to be. My conclusions do not provide easy solutions to the problem of misinterpretation of dictionary information, nor do they point to any one dictionary as the best choice for the EFL dictionary buyer, but it is hoped that they will offer some insight into the interface between dictionary and user, to the benefit of EFL learners, teachers and lexicographers.

Chapter One

A review of research into dictionary use and dictionary requirements

This chapter provides a detailed critical review of prior studies into dictionary-using behaviour. Dictionary use and dictionary requirements can be investigated in a number of different ways. In his paper for the the EURALEX Leeds seminar (1985), Hartmann identified four categories of investigation:

- 1) research into the information categories presented in dictionaries (dictionary typology)
- 2) research into specific dictionary user groups (user typology)
- 3) research into the contexts of dictionary use (needs typology)
- 4) research into dictionary look-up strategies (skills typology).

I am concerned with the needs and strategies of advanced English language learners using monolingual learners' dictionaries. Hartmann's categories three (needs typology) and four (skills typology) are of greatest interest to me. At present, there is little variation in the types of information (Hartmann's category one) available in advanced learners' dictionaries, although dictionaries differ in their defining styles and the extent to which they use non-verbal data, a factor which could be of significance when examining look-up strategies. Similarly my user group (Hartmann's category two) is relatively homogeneous; the major English learners' dictionaries are written primarily for advanced learners of English studying at secondary and tertiary level.

It also seems to me that Hartmann's categories one and two are much more amenable to investigation than his categories three and four. An examination of the dictionaries themselves will reveal what type of information is available to the user, and data collecting methods such as the questionnaire will establish who owns what dictionary, but these methods are less likely to tell us **why** the user acquired it, and **how** she uses it in her daily life.

In this chapter I describe in detail a wide range of prior investigations into the wants and needs of dictionary users, and the skills required to extract needed information from dictionaries. I evaluate the experimental methods used in these studies and the conclusions that have been drawn from them, in order to provide the reader with a background to my own experiments and the methods I have chosen. Most of the studies reviewed in this chapter are concerned with English language learners, but some studies of native-speakers have been included because they introduce techniques and provide findings applicable to the study of the way English language learners use dictionaries. I have grouped the papers according to data collection method: the use of a questionnaire, the use of tests, and observation.

1.1. Questionnaire-based research

In this section I will review seven major questionnaire-based studies into the use of English dictionaries. All seven studies hold many features in common, and ask many of the same questions, but the first three studies under discussion, Barnhart (1962), Quirk (1972) and Jackson (1988) are concerned with native speaker dictionary use, whilst the remaining four, Tomaszczyk (1979), Baxter (1980), Béjoint (1981) and Battenburg (1989), deal with the dictionary use of non-native speakers.

1.1.1. Barnhart (1962)

Barnhart's study is included in this survey because it is, as far as I am aware, the earliest survey of dictionary users' wants and needs. Collected in 1955, his results have been replicated in several more recent investigations, and have, according to Hartmann (1987) proven of considerable interest to lexicographers about to decide on what kind of material should be included in the dictionary and how it should be presented for the benefit of the user.

Barnhart's paper discusses the central issue of commercial dictionary editing: the problem of selecting and balancing material where there are tight space restrictions. Educated Americans use a college dictionary of a size to permit some three million

running words, which is usually taken to allow for between 120,000 and 150,000 entries, in comparison with the 250,000 words estimated to be the working vocabulary of the English language. How can the editor decide which words to include, and how they should be treated?

Barnhart maintained that the selection of material for college dictionaries was largely based on the personal interests of the editor, as opposed to the interests and needs of the buyer. His survey aimed to discover more of the buyer's point of view, in order to inform the selection process (and presumably create a more marketable dictionary). Freshman composition students constituted a large share in the market for college dictionaries, so Barnhart circulated 108 questionnaires to teachers of freshman composition in 99 USA colleges, who reported on dictionary use by about 56,000 students. The questionnaire appears to have consisted of a single request to teachers to rate in order of importance to their students six types of information commonly given in college dictionaries - the six types of information being: meaning; spelling; pronunciation; synonym studies and lists; usage notes; etymologies. The teachers' replies are not reported in any detail in Barnhart's paper, but are only summarized as follows:

their replies indicate that the college freshman uses his dictionary most frequently for meaning and almost as frequently for spelling. Pronunciation is third with synonym studies and lists, usage notes, and etymologies far behind.

(1962:458)

It is unclear precisely what is meant by "lists".

Discussion

Barnhart's survey is flawed because he did not address the users themselves, but rather their teachers. We are given no indication as to how the teachers obtained information about the users' habits; were their replies an account of what they had

observed, drawn from personal memories of college dictionary use, or based on preconceived notions of the way dictionaries are used?

The simplicity of the questionnaire design suggests that Barnhart was not seeking answers to editorial problems, but rather a starting point for discussion of these problems. Results of the questionnaire are briefly stated on the second page of his article, and the remaining seventeen pages are devoted to a detailed discussion of selection procedures. Barnhart does appeal to the questionnaire results when he examines the two top priorities for the college dictionary - the need to provide information on both meaning and spelling. These are described as conflicting priorities, because the spelling out in full of inflected and derived forms, and the granting of headword status to derived forms thrown out of alphabetical order by some change to the root, takes up space that could have been devoted to clarifying the meaning of a more difficult word. The editor's "first problem, then, is to choose whether he will give more importance to information about spelling or about meaning". The questionnaire results justify, according to Barnhart, the decision to prioritise meaning in favour of spelling:

since meaning difficulty outranks all other uses of the dictionary in importance so far as college freshmen are concerned, the editor usually compromises and enters derivatives without meaning difficulty but with simple spelling or pronunciation difficulty as run-ons instead of as main entries.

(1962:460)

However, as the questionnaire appears to consist of a single request to teachers to "rate six types of information according to their importance to the college freshman", teachers' replies cannot really have contributed to the ensuing discussion of dictionary coverage, the choice of pronunciation key, and the choice of usage labels. Moreover, the weighting that Barnhart appears to recommend in his concluding comments on dictionary editing do not entirely fit the priorities suggested by teachers:

the editor has given from 54 to 61 per cent of space to definitions, .7 to 1.3 per cent to illustrative phrases, from 5 to 8.5 per cent to etymologies, from 2.1 to 4.4

per cent to synonym studies and lists, and around 28 per cent to entries, pronunciations, parts of speech, inflected forms, usage notes and other material.

(1962:475)

There is certainly no suggestion that items ranking as least useful according to the questionnaire results might be omitted.

According to Barnhart, the role of the dictionary editor is to balance the interests of publisher, linguist, etymologist and subject specialist "in order to furnish a book which will be acceptable to all the parties concerned, as well as to the general public". He is probably the first dictionary designer to seriously survey the needs of dictionary users, but clearly he does not consider those needs as the only requirements the dictionary should fulfill.

1.1.2. Quirk (1975)

Like Barnhart's, Quirk's paper is concerned with native speaker use of standard English dictionaries; it is relevant to this chapter because it has been influential in the design of studies into learners' dictionary use, breaking as it did new ground by examining the personal needs of dictionary users. According to Hartmann (1987), Quirk's is the first scholarly attempt in Britain to assess the dictionary user's attitudes, expectations and prejudices.

Quirk's study took place in 1972. Unlike Barnhart, Quirk directly addressed his users, issuing a questionnaire to two hundred and twenty students in the middle of the first year of their studies at University College London. The students came from a range of disciplines; approximately half from the humanities and half from the sciences. Male and female subjects were equally represented.

Quirk's questionnaire was much more detailed than Barnhart's, with thirty questions on the following topics:

- * When the subject last used a dictionary
- * Average frequency of dictionary use

- * Concern to consult a particular dictionary
- * The dictionary normally consulted
- * The subject's ownership of a dictionary
- * Knowledge of both British and American dictionaries and basis of preference, if any
- * Ownership and use of a dictionary in the parental home
- * The subject's reason for his or her most recent use of a dictionary
- * The subject's most usual reasons for using a dictionary
- * The subject's failure to find what he or she wanted
- * The subject's suggestions for improving dictionaries
- * Should citations be from named (and well-established) authors?
- * The comprehensibility of definitions
- * The adequacy of definitions in respect of the subjects' own knowledge
- * The use of a dictionary for pronunciation
- * The adequacy and comprehensibility of pronunciation symbols
- * The use of a dictionary for form-class information
- * Should dictionaries be complete, even with well-known words like "throw"?
- * Should dictionaries have encyclopaedic entries?
- * The use of a dictionary for etymology
- * Should dictionaries contain American English words?
- * Should dictionaries contain slang words?
- * The use of a dictionary for synonyms and antonyms
- * The adequacy of a dictionary for finding synonyms and antonyms
- * Should dictionaries contain regional dialect words?
- * Should dictionaries contain phrases and idioms like "take your time"?
- * The subject's further suggestions for improving a dictionary.

It was found that 192 of the 220 subjects owned dictionaries. This majority represented all students equally, irrespective of field. 161 of the dictionaries normally consulted by students came from the Oxford family, eg COD, POD, SOD and OED.

However, there was a two to one majority of "indiscriminate users", who were not particular about which dictionary they used on any occasion. The results also suggested that "the "dictionary habit" appears to go with a tendency to discriminate in the selection of dictionaries".

In the Humanities 52% claimed to use a dictionary weekly, 36% claimed to use a dictionary monthly and 24% claimed to use a dictionary infrequently.

In the Sciences 22% claimed to use a dictionary weekly, 46% claimed to use a dictionary monthly and 40% claimed to use a dictionary infrequently.

Findings on different types of dictionary use are summarised in Table 1.1 below:

Table 1.1: Different types of dictionary use

	Meanings	Spellings	Word-games	Other uses
a) subjects' use	149	5	15	29
b) use in parental home	82	38	57	41

"Other uses" included synonym finding, etymology, usage and pronunciation. Quirk's findings regarding the relative value of different types of information for the user support those of Barnhart. Meanings and spelling were stated to be the most commonly sought types of information; little interest was expressed in etymology, except by students of English, and even less interest was expressed in dictionary information on pronunciation and parts of speech.

Quirk concludes that dictionary makers and dictionary users have different priorities: "some of the dictionary features which seem of particular centrality to lexicographers are decidedly peripheral to the ordinary user" (1975:80). Nevertheless, Quirk maintains that these features should not be dropped from dictionaries, as "in the group

of questions which sought to establish what the general image of a dictionary was in the minds of the subjects, the criterion of absolute completeness was strongly supported".

Discussion

There seem to me to be two basic problems with this study. The first concerns the accuracy of the responses, and the fact that it is unlikely that subjects were able to give correct information for many of the questions. Some questions put subjects' memory to the test; they would probably have had difficulty recalling, for example, when and why they last used a dictionary. Moreover, questions concerning reasons for using a dictionary required the subjects to retrospect, something which many may have lacked the concentration and the inclination to do outside a controlled experimental setting. The language and the frame of reference of Quirk's questions were also very different from those of a first year undergraduate. Quirk records that an overwhelming majority of the university students experienced difficulty in understanding "the metalanguage in which definition is expressed". If this is so, it seems unlikely that all subjects understood the linguistic terminology in the questionnaire, or that, if they did understand it, the unfamiliar register did not affect their response. Altogether, the questions seem to require a capacity to recall, retrospect and comprehend beyond the abilities of the average dictionary user.

The second problem with this study concerns the type of questions asked, which reflect the researcher's preconceived ideas about the type of information a dictionary ought to contain. Quirk's questionnaire asks users to comment on what already exists in their dictionaries, and, despite the invitation to suggest improvements, there is little encouragement to think laterally and suggest departures from the conventional dictionary format.

Both these problems are common to much questionnaire-based research, and have not been entirely resolved in later studies of dictionary use.

1.1.3. Jackson (1988)

This study is a smaller and simpler version of the one conducted by Quirk fourteen years earlier, and yields similar results.

Jackson administered his questionnaire to fifty students beginning a degree course in English language and literature, and to thirty-six students beginning a degree course in Speech and Language Pathology and Therapeutics. The questionnaire consisted of six questions on the following topics:

- * The subject's ownership of a dictionary
- * The average frequency of use
- * The occasions on which a dictionary was used
- * The subject's usual reason for using a dictionary
- * The adequacy of dictionary information
- * The subject's suggestions for improving a dictionary.

Most of the questions could be answered by ticking one or more alternative, although subjects were allowed to specify their own personal response if it differed from the alternatives offered.

Responses to the questionnaire indicated that:

- * 63% of students owned dictionaries from the Oxford family (Shorter, Concise or Pocket) and only two students did not own dictionaries (question one);
- * 50% consulted a dictionary once a week or more, 43% consulted a dictionary once or twice a month and 7% consulted a dictionary less frequently than once a month (question two);
- * 84% used a dictionary while writing essays, 74% used a dictionary while reading, 46% used a dictionary for crosswords, 44% consulted a dictionary for general interest and 40% used a dictionary for word games (question three);
- * 93% used a dictionary for looking up meanings, 92% used a dictionary to check spellings, 72% used a dictionary to check whether a word exists, 10-11% used a dictionary to check pronunciation, 10.5% looked up etymological information, and 4.5% used a dictionary to check a part of speech (question four);
- * 73% thought that their dictionary provided them with all the information that they needed, and 14% thought that their dictionary contained too much information (question five).

Suggestions for improvement included changing the layout and typography, extending coverage, providing clearer and longer definitions, more instances of usage, and more explanation of pronunciation.

Discussion

Both Quirk and Jackson offered a multiple choice format for many of their questions, which means that subjects did not always have to invent their own replies. However, Jackson's questionnaire was easier to complete than Quirk's; the questions were less detailed, and there was less technical jargon. For this reason slightly better accuracy might be expected, although the study was less ambitious in the quantity of information it sought.

Jackson's subjects appeared to use their dictionaries more frequently than Quirk's, but in both surveys dictionaries appear to be used most frequently to check meanings and spellings, and subjects showed little interest in pronunciation, parts of speech and etymology. These findings are very similar to those of Barnhart in the United States.

Jackson's subjects also appeared to be highly satisfied with their dictionaries, and their suggestions for improvements fall well within conventional lines. Interestingly, many of their ideas for improvements match the improvements that have been made to modern learners' dictionaries; learners' dictionaries typically have more elaborate pronunciation guides, more notes on usage, and clearer and longer definitions. Could it be that Jackson's subjects were, wittingly or unwittingly, comparing their own dictionaries with those they had seen for non-native speakers? As the market for learners' dictionaries grows they are given increasingly prominent display space on bookshop shelves, and it may be that Jackson's subjects had already encountered and admired some learner dictionary features.

1.1.4. Tomaszczyk (1979)

Tomaszczyk was the first researcher to investigate the dictionary requirements of non-native speakers of English. He was motivated by the observation that among

foreign language learners there was a widespread "feeling of dissatisfaction with the dictionaries they use", and states his aim to obtain information relevant to the production of better non-native speaker dictionaries:

the study was undertaken in the hope that an examination of the ways in which language learners use dictionaries, and of their attitudes and expectations towards them would provide some information about the extent to which various groups of users depend on dictionaries, help pinpoint those of the current lexicographical solutions that are, as well as those that are not, felt to meet their needs, and give lexicographers some clues they might want to use in their attempts to make better dictionaries.

(1979:103)

Tomaszczyk's survey was more detailed and on a larger scale than the three surveys previously described. He drew his data from 449 questionnaires completed by foreign language learners at tertiary level (group 1: 284 subjects) and foreign language instructors and translators (group 2: 165 subjects).

Group 1 consisted of 55 foreign students at American colleges, 62 foreign students at Polish universities, and 167 Polish students of university foreign language departments. The mean age of the subjects in this group was 21, and the average period of foreign language study was 5 years. The vast majority (97%) had had formal instruction in the language they reported on.

Group 2 consisted 60 language instructors, 25 translators of belles-lettres, and 80 technical translators. The mean age of the subjects in this group was 57, and the average period of foreign language study was 30 years. 82% had had formal instruction in the language they reported on, but they also stressed the importance of self-instruction and foreign stay.

In some places in his study, where there are marked differences between subjects within groups 1 and 2, Tomaszczyk subdivides group 1 into 1a - 117 non-Polish students, and 1b - 167 Polish foreign language students. He likewise subdivides group

2 into 2a - 85 language teachers and literature translators, and 2b - 80 technical translators.

Sixteen different languages were reported on, including English (190 subjects), Russian (84 subjects), Polish (64 subjects), French (47 subjects) and German (46 subjects). The vast majority of subjects (91%) reported reading or speaking at least one other foreign language, which implies experience of dictionaries in more than one language.

Tomaszczyk's questionnaire is not reproduced in his paper, but we are told that it contained "57 items concerning personal and language learning history, current language use, use of dictionaries, and the evaluation of the information contained in them" (1979:104). From answers to these questions Tomaszczyk draws conclusions about the needs of non-native dictionary users in terms of six "language skills" (listening, reading, speaking, writing and translating to and from the L2) and "dictionary information types" (not all of which are specified in his paper).

In his findings, Tomaszczyk distinguishes between four different categories of dictionary: monolingual (M), bilingual (B), restricted (ie dictionaries of slang etc) (R), and technical (T). He further divides bilingual and technical dictionaries into L2 - L1 and L1 - L2. The questionnaire revealed that "almost all subjects, no matter how sophisticated they are, use bilingual dictionaries" (1979:106). Monolingual dictionary use was considerably less than bilingual dictionary use (59.9% as opposed to 94% (for bilingual L2 - L1) and 77.5% (for bilingual L1 - L2)).

Amongst teachers and students of English, Hornby's ALD was reported to be the most frequently used monolingual dictionary, used by 89 of the 138 monolingual English dictionary users. The remaining 49 subjects seem to have consulted dictionaries intended primarily for native speakers, such as COD, Webster's 7th Collegiate and the American Heritage.

The questionnaire asked subjects to list the dictionaries they used and rate them on a four-point scale: "excellent", "adequate", "inadequate" and "definitely bad". Table 1.2 summarizes Tomaszczyk's subjects' replies: The dictionaries named by the subjects are grouped in the categories M (monolingual), B (bilingual), R (restricted) and T (technical). The figure on the left-hand side of each of the four columns represents the total number of dictionaries of a given type each group listed, and the figure on the right represents the average overall rating. These were arrived at by assigning numerical values to each rating, from 4 to 1, where 4 = "excellent".

Table 1.2: Dictionaries used and their overall evaluation

	1a	1b	2a	2b	Total
M	36 3.3	167 3.4	131 3.4	30 3.6	364 3.44
B L2-L1	105 3.0	208 2.9	112 2.8	90 2.8	515 2.92
B L1-L2	48 2.7	185 2.7	89 2.4	72 2.7	394 2.69
R	4 3.0	199 3.1	95 2.9	24 3.1	322 3.04
T L2-L1	12 3.2	-	7 2.9	97 2.8	116 2.84
T L1-L2	-	-	10 2.6	56 2.6	66 2.57

According to these results, monolingual dictionaries were rated more highly than any of the other dictionary types, with a score of 3.44 (between adequate and excellent), compared to 2.92 for bilingual L2 - L1, 2.69 for bilingual L1 - L2, 3.04 for restricted dictionaries, 2.89 for technical dictionaries L2 - L1 and 2.57 for technical dictionaries L1 - L2. Tomaszczyk reports later in his paper that subjects asked to evaluate the usefulness of information categories within dictionaries judged bilingual dictionaries to be as good, or almost as good as monolingual dictionaries as far as spelling, receptive grammar and function words were concerned. In other respects bilingual dictionaries were rated 15-30% lower than monolingual dictionaries.

Tomaszczyk also summarises findings regarding the extent to which subjects looked up word meanings in monolingual and bilingual dictionaries, and the degree to which they were satisfied with the information they found. These findings are presented in Table 1.3 below.

Table 1.3: Information about meaning of L2 items (definitions and equivalents)

Dictionary type	M	L2-L1	L1-L2
No of subjects	257	422	348
Look up meanings	85.4%	95.3%	97.7%
Are satisfied	91.2%	73.9%	65.6%

These findings accord with those in Table 1.2; subjects express significantly greater satisfaction with monolingual dictionary entries. It is to be noted in this table that a slightly higher proportion of the subjects who have L1 - L2 dictionaries claim to use them (in comparison with L2 - L1 ones). Tomaszczyk points out that this "is understandable in view of the fact that they have two types of dictionary to aid them in the comprehension of L2 items (L2-L1 and Monolingual)".

The questionnaire also elicited information about dictionary use for the language skills. The responses revealed that dictionaries were used most frequently for translation, and, in the case of group 1, most frequently when translating from an L2 into the first language. After translation, dictionaries were most frequently used as an aid to writing, followed by reading. Dictionary use accompanied listening and speaking less frequently. Table 1.4 below summarises this information.

Table 1.4: Frequency of dictionary use for the six language skills

Group 1 (per cent of subjects in the group)

	Frequently	From time to time
Listening	15.5	39.4
Reading	61.6	30.6
Speaking	23.6	40.1
Writing	67.3	25.7
Translation into L2	61.6	19.7
Translation from L2	67.6	22.2

Group 2 (per cent of subjects in the group)

Listening	7.3	26.7
Reading	14.5	62.7
Speaking	7.3	30.3
Writing	29.1	41.8
Translation into L2	46.1	27.3
Translation from L2	44.8	37.6

Questions relating to the frequency of dictionary use in the six language skills also provided data on users' preferences for bilingual or monolingual dictionaries, or combinations of the two types. Tomaszczyk analysed the answers of 228 subjects who possessed both types of dictionary, and found that for each of the six language skills the majority chose to consult only their bilingual dictionary. Table 1.5 summarises the information Tomaszczyk obtained. (B = bilingual dictionary and M = monolingual dictionary. In columns 2 and 4 the order of the letters indicates the order of use for a combination of both types of dictionary. The figures in parentheses show the number of subjects who use dictionaries for the given skills, other figures are percentages of

these totals; in some cases the figures do not total 100%, but Tomaszczyk does not comment on this.)

Table 1.5: Use of bi- and monolingual dictionaries when both types are available

Group 1 - 160 subjects

Dictionary type	B	BM	M	MB
Listening (66)	58.6	16.7	19.7	5.4
Reading (112)	42.9	33.0	14.3	9.8
Speaking (71)	57.7	22.5	15.5	4.2
Writing (108)	44.4	23.1	19.4	13.0
Translation into L2 (111)	67.6	24.3	3.6	5.4
Translation from L2 (114)	55.3	31.6	5.3	7.0

Group 2 - 68 subjects

Dictionary type	B	BM	M	MB
Listening (19)	31.6	10.5	52.6	5.3
Reading (38)	31.6	21.0	34.2	13.2
Speaking (20)	40.0	15.0	40.0	5.0
Writing (37)	21.6	27.0	21.6	29.7
Translation into L2 (39)	43.6	23.1	5.1	28.2
Translation from L2 (46)	37.0	32.6	4.3	26.1

Only amongst the experienced speakers of group 2 was the monolingual dictionary preferred, and then only for the receptive skills of listening and reading. A very large proportion of the subjects in both groups preferred bilingual dictionaries for translation both to and from the L1. Next in preference across all the language skills came the combination of bilingual dictionary followed by monolingual dictionary. The opposite combination of monolingual dictionary followed by bilingual dictionary

was generally unpopular, however, and was adopted only by some of the experienced speakers in group 2, for writing and translation.

Tomaszczyk notes that:

it comes as no surprise that the proportion of monolingual dictionary users should rise with their increasing sophistication Indirectly, this result supports the general belief that the extent to which foreign language learners rely on their mother tongue diminishes as their L2 proficiency increases.

(1979:110)

Tomaszczyk stresses that the results given in Table 1.5 are likely to be accurate, as they were extrapolated from answers to questions on more general habits of dictionary use, and the order in which subjects listed their preferences suggests that they were not merely copying from previous answers or "filling in boxes".

The penultimate part of Tomaszczyk's paper reports on the subject's responses to questions concerning "habits and feelings on the main types of information people usually look up in dictionaries". Information from the first set of questions has already been summarised in Table 1.3. We are told that other questions in this part of the questionnaire "concerned information about status, usage and currency of words, foreign words and phrases ie L3 L4 etc items in an L2 text, comparisons (similes), proverbs and sayings etc".

Tables 1.6, 1.7 and 1.8 below are my attempt to summarise those findings mentioned by Tomaszczyk in the course of his discussion. Tomaszczyk does not present these findings in any systematic way and does not provide sufficient data to enable me to complete every section of Table 1.6.

According to Tomaszczyk the low level of satisfaction for group 1a (apparent in Table 1.7) might be due to the fact that most of them used small, often outdated dictionaries.

Table 1.6: Habits and feelings concerning different types of dictionary information (all subjects)

Type of information	% use dictionary	% satisfied
Synonyms	74%	-
Spelling	72%	88%
Idioms	72%	-
Stress/pronunciation	65%	85%
Swear words/obscenities	45%	22%
Word division	36%	-
Etymology	19%	-

Table 1.7: Levels of interest and satisfaction concerning the dictionary information types listed in in Table 1.6

Group	% use dictionary	% satisfied
1a (foreign students)	62%	40%
1b (language students)	69%	56%
2a (language teachers and literature translators)	53%	50%
2b (technical translators)	34%	51%

Table 1.8: The use of dictionaries for grammatical information (group 1a - language students and group 2b - language teachers)

Type of information	% use dictionary
Receptive grammar/ function words	70%
Productive grammar	59%

Tomaszczyk adds that over 35% of language students explicitly stated that they did not consult general dictionaries for productive grammatical information. Of the 69

students of English who possessed a copy of OALD, only 23 consulted it for productive grammar, but of these 23, 83% were satisfied.

In the final part of his paper Tomaszczyk reports on the answers to a set of yes/no questions, and outlines some of the comments subjects made while completing the questionnaires. In brief, responses show that:

- * 71% of subjects looked up, or expected dictionaries to carry, proper names of people and places
- * 60% thought that pictures and drawings in monolingual dictionaries would make some words easier to understand
- * 85% of group 1 and 65% of group 2 thought that dictionaries should be supplemented with some kind of reference grammar, cross referenced to the main text of the dictionary
- * 58% consulted dictionaries from time to time for no particular reason, just to browse
- * In the appendices, 73% looked up abbreviations, 70% geographical names, 58% irregular grammatical forms, 49% given names, 43% weights and measures, 41% family names
- * When buying a dictionary 65% of the subjects went by the number of entries, 27% by somebody else's advice, 13% by price and 9% by the size of the book.

Subjects' comments tended towards the view that as much information as possible should be included in dictionaries. The majority wanted more extensive treatment of every type of information, and although some suggested that such elements as slang and archaic forms were redundant, the majority favoured their inclusion. This response matches that of the native speakers in Quirk's survey, who desired "absolute completeness" from their dictionaries.

Discussion

While applauding Tomaszczyk's pioneering work, Hartmann (1987) criticises this paper because Tomaszczyk's presentation of numerical evidence is not always clear, the statistical analysis is incomplete, and the questionnaire is not reproduced. These flaws in the presentation make the paper difficult to assimilate, and deny the reader access to some information that the survey was designed to reveal. Moreover some of

Tomaszczyk's findings appear contradictory, raising questions about the accuracy of the questionnaire responses. Before discussing the general implications of Tomaszczyk's work, I will identify some problems with his data with reference to Tables 1.2, 1.4 and 1.8, which summarise findings from different sections of the questionnaire.

TABLE 1.2

The most serious question raised by the findings summarised in Table 1.2 is why users, who rate their monolingual dictionaries so highly, should nevertheless prefer to use their bilingual dictionaries for every kind of language activity. This survey does not provide us with information regarding subjects' reasons for rating a dictionary highly or otherwise, but it appears that users do not equate the "quality" of a dictionary with its "usefulness". In other words, monolingual dictionaries were probably perceived to be "good" for other reasons than their readability, accessibility, and general user-friendliness. Tomaszczyk refers to the school of thought that has condemned the use of the first language in the second language classroom, and has disapproved of bilingual dictionaries. It may be that Tomaszczyk's subjects had been influenced by this attitude, feeling that in some way the monolingual dictionary was superior, despite the difficulties they encountered when consulting it. Tomaszczyk notes that "seven students of English stated explicitly that they find definitions in Hornby's dictionary hard to understand" (1979:110); it is possible that this was a common problem, and that other subjects blamed themselves, rather than the monolingual dictionary, when they could not understand an entry.

The relatively low satisfaction rating of restricted dictionaries is explained by Tomaszczyk as the result of higher expectations on the part of the restricted dictionary user:

an experienced dictionary user knows what he can expect of a general dictionary and will usually not consult it unless he is pretty certain that the information he is interested in is there and even if, on occasion, the book lets him down, quite possibly he makes

allowances for it being "only" a general dictionary. In the case of restricted dictionaries, on the other hand, the standards applied will be much higher, and the disappointment more acute.

(1979:113-4)

This attitude is not directly attested in the data, but is purely inference on Tomaszczyk's part.

TABLE 1.4

Tomaszczyk's distinction between "frequently" and "from time to time" may cause problems here, as it is open to different interpretations. Presumably intended to register degrees of frequency (in contrast to a tick in the "never" box), the terms are themselves contrastive. In the original table (Tomaszczyk 1979 p105) the symbol "S" is used to represent the meaning "from time to time", but the letter suggests "seldom", and if an "S" box appeared on the questionnaire it may well have been ticked by some subjects as an indication of infrequent use. Tomaszczyk admits that group 1's claim to use dictionaries most frequently when translating from an L2 into the L1 is rather surprising:

intuitively one would expect more extensive use of dictionaries for L1 - L2 translation than the other way round.

(1979:108)

Perhaps, as Tomaszczyk suggests, group 1 respondents were influenced in this section of the questionnaire by answers they had previously given regarding frequency of language use. As they had claimed to translate more frequently from the L2 than from the L1, they now claimed to use their dictionaries more frequently when translating from the L2. In contrast group 2, the experienced speakers, claimed to use their dictionaries more frequently for translating into an L2, despite the fact that they were more frequently required to translate into their first language.

Tomaszczyk was surprised that his respondents claimed to use a dictionary at all while speaking, as in ordinary conversation there is no time for dictionary

consultation. 15.5% of the learners recorded that they "frequently" used dictionaries when speaking, but possibly subjects understood "speaking" to mean something different from that which Tomaszczyk intended:

conceivably the subjects meant the occasional use of a dictionary when preparing all kinds of talks, speeches, and oral reports, something language students do quite frequently, but that can hardly be called speaking.

(1979:108)

Alternatively, the subjects may have been unreliable informants, a possibility which Tomaszczyk admits:

another possible interpretation is that when completing a questionnaire some people will check any box there is.

(1979:108)

TABLE 1.8

Tomaszczyk notes with surprise that 70% of language students and teachers claimed to consult dictionaries for receptive grammar and function words. Grammatical information thus appeared to be important for most subjects, and in fact Tomaszczyk also reports that 85% of learners and 65% of speakers "decided that dictionaries should be supplemented with some kind of reference grammar". However, only 59% of subjects reported an interest in productive grammar in the dictionaries, despite the fact that grammatical information is of greater importance for language production than for language reception. Stranger still, Tomaszczyk found that "35% of language students explicitly stated that they do not consult general dictionaries for such information". Tomaszczyk's account is very unclear at this point (p112), as he takes a subgroup of his group 1b - those learning English, and then creates a further subgroup of OALD owners. It is uncertain whether the figure he gives for satisfaction with grammatical information in bilingual dictionaries (56%) is based on the responses of this group of OALD owners, or a larger population. However, it is certainly puzzling that learners claim to consult a dictionary type for which lower levels of satisfaction

have been recorded, despite the fact that they have access to a dictionary which was rated highly for grammatical information by those who used it.

General comments

Regarding the study as a whole, many of the inconsistencies in Tomaszczyk's findings may be attributed to his method of data collection and the large scale of his study, even though he has been praised by Hartmann (1987) for the comprehensiveness of his approach. It is not always possible to interpret Tomaszczyk's data precisely; for example, a large minority of the English language learners were using dictionaries designed for native speakers, but we have no means of distinguishing their responses from those of OALD users. Also, when Tomaszczyk reports on the proportion of subjects who looked up such things as geographical names in the appendix of their dictionary, we do not know what proportion of the remainder actually used dictionaries which contained appendices for geographical names, and thus whether they were in fact exercising a choice not to look up such information in the back of their books.

Results which do not indicate the frequency of the subject's professed behaviour may also mislead. The question "do you look for geographical names in the appendix of your dictionary?" will elicit a yes response from the subject who has done so only once, as well as from the subject who habitually uses the appendix for this purpose.

Contradictory responses from subjects leads Tomaszczyk himself to doubt the validity of some of his data. There is some evidence to suggest that subjects did not know their dictionaries sufficiently well to judge them. For example, Tomaszczyk notes that some group 1 subjects complained that their dictionaries did not list irregular verbs and contracted forms, when in fact the dictionaries did list these things.

Tomaszczyk appears particularly disgruntled with the non-Polish respondents:

although group 1a subjects are not quite representative of beginning language learners in that they all have lived in L2 environments, the erratic nature of some of

their responses, as well as the strange remarks, comments and requests they made suggests that on the whole they are not ready to use reference books profitably.

(1979:117)

This is a strange conclusion from a researcher apparently concerned with dictionary users' needs and wants, and perhaps reflects an attitude discernable throughout all the questionnaire-based surveys, that more adaptation is needed on the part of the user than on the part of the dictionaries, and that, although minor changes may be possible, the basic structure and scope of dictionaries is immutable.

Although flaws in the research design render some of Tomaszczyk's findings suspect, a number of important facts do seem to emerge. Tomaszczyk's subjects seemed to prefer to use bilingual dictionaries rather than monolingual dictionaries. In comparison to the native-speaker respondents of the questionnaires conducted by Quirk and Jackson, Tomaszczyk's subjects attach far greater importance to grammatical information in dictionaries, and register less satisfaction with the dictionaries they use. They appear to share the native-speaker dictionary users' interest in meanings and spellings, and their disregard for etymologies.

1.1.5. Baxter (1980)

This paper, like Tomaszczyk's, concerns non-native speaker dictionary use, and examines in particular the question of the value of monolingual, as opposed to bilingual, dictionaries.

Unlike the four researchers previously discussed, Baxter did not intend the information from his survey to inform future dictionary design, but rather to measure the influence of bilingual dictionary use on a specified group of learners. The first part of Baxter's paper describes a pedagogical problem - the inability of language learners to express their ideas in an acceptable way when the precise lexical item does not come to mind - and suggests that one possible cause of this inability might be the

use of bilingual as opposed to monolingual dictionaries. The second part of Baxter's paper reports on his survey concerning bilingual dictionary use:

having recognized the potential influence of bilingual dictionary use, the next step is to determine the extent to which this influence may actually be operating.

(1980:331)

Baxter's hypothesis that bilingual dictionary use discourages the development of paraphrasing language, while monolingual dictionary use will teach the learner alternative methods of expressing meaning is interesting, but so far remains unproved. Other writers (eg Thompson 1987) have challenged the view that monolingual dictionary definitions do provide a suitable spoken language model for the learner, and Brown (1979) blames the "bilingual reflex" (the belief that one can and should match every word in one's native language with a corresponding term in the language one is learning) on "a rigidly applied grammar translation method" rather than on bilingual dictionary use.

Baxter's study

Baxter's survey was a little smaller than Tomaszczyk's, but still on a large scale. In the summer of 1979 he obtained 342 replies to a questionnaire administered to Japanese students at three national, four-year universities in Japan. The respondents were 62 English majors (18.1%) and 280 non-English majors (81.9%) (1st year 19.9%, 2nd year 57.9%, 3rd year 13.2%, 4th year 7.6% graduate level 0.3% and unspecified 1.2%).

Questions concerned the ownership and use of monolingual English dictionaries, bilingual Japanese-English dictionaries and bilingual English-Japanese dictionaries (these last two are generally not combined in one volume in Japan).

In his paper Baxter reproduces both questionnaire and results. The questionnaire consisted of seven questions:

1. When did you buy your first dictionary?

2. How many bilingual dictionaries have you bought since you started studying English?
3. How many monolingual dictionaries have you bought since you started studying English?
4. Please give the name of the dictionary which you now use most often.
5. How often do you use the following types of dictionary? (bilingual Japanese-English, bilingual English-Japanese, monolingual English)
6. In your studies, what for you has been the most important type of book you have used?

and a seventh question, for which we are not given the exact wording, which asked students to state which type of dictionary, monolingual or bilingual, they preferred, and why.

The answer to every question except 4 and 7 was given in a multiple choice format. Question 6 offered a selection of possible answers but also allowed the respondent to formulate his own reply.

Responses to the questionnaire provide the following picture of dictionary use by Japanese university English language learners:

students begin their studies of English in junior high school, buying their first dictionary, a bilingual English-Japanese one, at that time. Over the next few years, two more bilingual dictionaries are acquired. Only if the university major of a student is English, will a student purchase a monolingual English dictionary. At the university level, as with the lower levels, an English-Japanese dictionary is used most often. Non-English majors rarely if ever refer to a monolingual English dictionary, and while English majors do so more frequently, few of them use a monolingual dictionary daily. By comparison, most of them give their English-Japanese dictionaries daily use. Very significantly, students attribute to the bilingual dictionary, in contrast to other reference books, the greatest degree of importance in their studies of English.

(1980:333)

Thus Baxter's fears were confirmed. As in Tomaszczyk's survey, monolingual dictionary use was reported as being considerably less frequent than bilingual

dictionary use, but monolingual dictionaries were not accorded the high status they were given in the Polish survey. In answer to question seven, Baxter's respondents claimed that their bilingual dictionaries were easier to use, and "many students criticized monolingual dictionaries, complaining that definitions were difficult to understand", whereas in Tomaszczyk's study only seven respondents made this specific complaint. Baxter gives no indication of the type of monolingual dictionaries his respondents were using, although he does say that respondents using native-speaker monolingual dictionaries were the ones to complain most about the need to look up words in the definitions.

Respondents in this study and Tomaszczyk's expressed a preference for the L2-L1 dictionary, as opposed to the L1-L2 dictionary. Reliance on an English-Japanese dictionary suggests dictionary use while translating from English and reading in English; Tomaszczyk's analysis of frequency of dictionary use for the six language skills indicates considerable dictionary use in L2 reading and translation, but writing, which is more likely to require an L1-L2 dictionary, was the skill he found most frequently associated with dictionary use. Baxter did not collect data on the frequency of language skills.

Discussion

Baxter's questionnaire was probably the easiest to answer of all the questionnaires reviewed so far, because the questions did not require the respondent to retrospect about his or her look-up strategies, but rather, in the majority of cases, to provide factual information about dictionary ownership. Questions 6 and 7 require the respondents to evaluate their dictionaries, but only in general terms, without reference to specific activities or look-up events.

The results of this survey contain none of the inconsistencies of Tomaszczyk's results, but also give a less comprehensive and less detailed picture. Baxter's main intention was to gauge the importance of bilingual as opposed to monolingual English

dictionaries for university students in Japan; the only insight afforded into the reasons behind the respondent's choice of dictionary comes from the answers to question 7, of which only a small sample are reproduced in Baxter's paper.

According to Tomaszczyk's findings, bilingual dictionaries in Poland were used more frequently but were less admired than monolingual dictionaries. According to the responses to Baxter's question 7, Japanese learners of English tend to react negatively to monolingual dictionaries, and hold their bilingual dictionaries in high regard.

Factors affecting the difference in status accorded to the two types of dictionary in the two surveys might include the **quality** of monolingual and bilingual dictionaries available in Japan and Poland at the time of the surveys. Baxter's respondents may have been using unsuitable (ie native-speaker) monolingual dictionaries, but superior bilingual works of reference. It also seems likely that Tomaszczyk's respondents had on average reached a higher level of foreign language proficiency. In Baxter's survey, results are not divided according to the respondents' year of study, so there is no record of whether monolingual dictionary use increases with growing language proficiency, as it appears to do in Poland.

1.1.6. Béjoint (1981)

This paper is probably the best known and most frequently cited study of non-native speaker dictionary needs; its stated aim was "to reveal how French students of English used their monolingual general English dictionaries", but its findings have been applied to the dictionary use of learners of English in general.

Béjoint claims that his work was influenced by that of Tomaszczyk in 1979; his findings were published in 1981, so although no date is given for his survey it must have been carried out between these two dates. Béjoint's questionnaire was administered to 122 French students of English at the University of Lyon, of which 63 were in their second year, 43 in their third year and 16 in their fourth year. The

questions were in French, but appear in translation in Béjoint's paper. They are reproduced in full below:

1. Do you own a monolingual English dictionary?
2. Which dictionary (or dictionaries) do you own?
3. Why did you choose the one(s) you bought?
4. When did you buy it?
5. What other monolingual general English dictionaries do you know?
6. If you use several dictionaries, is there one that you prefer? Why?
7. How often do you use a monolingual English dictionary?
8. Which types of information do you look for most often in your dictionary?
Meaning/ syntactic information/ synonyms/ spelling/ pronunciation/ language variety/ etymology
9. For which sort of activity do you most often use your dictionary? Version/ written comprehension/ written composition/ theme/ oral comprehension/ oral composition
10. Do you sometimes browse through your dictionary without looking for anything in particular?
11. How carefully did you study the introductory matter?
12. Do you ever use the information contained in the appendices?
13. Do you use the codes that indicate how a word should be used?
14. Are you satisfied with your monolingual English dictionary? More, or less satisfied, than with your bilingual dictionary?
15. Can you recall occasions when you could not find what you were looking for?
16. Can you mention any words that you were unable to find in your dictionary?
17. What kind of words do you look up most often in the dictionary?
(Often/sometimes/never) Idioms/ encyclopaedic words/ culture-specific words/ abbreviations/ slang words/ common words/ taboo words/ proper names
18. Do you use the following? Examples and quotations/ synonyms/ pictures.
19. Under which headword would you look up the following compounds? Artificial insemination/ boil down to/ false alarm/ magnetic tape/ come down with/ lose sight of/ rid of/ fountain pen
20. Do you think your dictionary is too simplified, or on the contrary too detailed?

The first five questions concern dictionary ownership. Replies revealed that 96% of the students possessed a monolingual dictionary, and of these 45% possessed OALD,

27% LDOCE and 14% COD. 85% of the respondents had chosen their dictionary because their tutor had recommended it, 55% in their first year of study, and 29% in their second year. Few knew of any other monolingual dictionaries.

In answer to question 6, most respondents seemed satisfied with the dictionary or dictionaries they habitually worked with, naming completeness of coverage as the reason for their preference.

Questions 7 to 13 concern the ways in which students use their dictionaries. 40% of students claimed to consult their dictionaries at least once a day, and 52% at least once a week. 87% of respondents "placed meaning among the three most sought-after pieces of information", but only 25% of respondents mentioned spelling and pronunciation. Etymology was the least frequently mentioned information category. Dictionaries appeared to be used more frequently for decoding than encoding, and more frequently for the written medium than for the spoken medium. 55% of respondents said that they sometimes "browsed". 89% of respondents admitted to reading the introductory matter "less than thoroughly", and 55% admitted to not using the codes. Four types of appendix information were consulted: abbreviations (40%), irregular verbs (30%), units of measurement (27%), and proper names (12%).

Questions 14, 15 and 16 concern the degree to which the students were satisfied with the dictionaries they used. 77% claimed to be satisfied with their monolingual dictionaries; 36% preferred them to their bilingual dictionaries, explaining that "the monolingual dictionary is more useful when you need to know the exact meaning of a word, or when you need synonyms". 17% preferred their bilingual dictionaries. Respondents also commented that "the use of the monolingual dictionary called for greater linguistic sophistication on the part of the user".

Many students could not recall occasions when the dictionary did not provide the information they were looking for. Expressed causes of dissatisfaction were:

unsatisfactory definitions 29%; words missing 28%; unsatisfactory syntactic guidance 25%; excessively long entries 16%; incomprehensible coding 10%; pronunciation not indicated or not clear 9%.

Many respondents also could not remember being unable to find particular words, but, for those who could recall instances, slang words, Americanisms, technical words and compounds seemed to be the categories most frequently found lacking.

Questions 17, 18 and 19 concern look-up strategies. 68% of respondents claimed to look up idioms most often, while 66% indicated that they never looked up common words. 70% claimed to use examples and quotations, 68% synonyms and 24% pictures. Respondents were in disagreement over the choice of headword in compound phrases, but on the whole they rejected the notion of separate main entries for compounds, and tended to choose the last word as headword in nominal compounds.

In answer to question 20, 50% of respondents thought that their dictionaries were too simplified, and 45% thought that the level of their dictionaries was "just right".

Complaints made in response to the open final question included: insufficient examples; idioms and Americanisms; no proper names; unclear layout, typography and illustrations; lack of conformity between dictionaries regarding systems of phonetic transcription. To these complaints can be added those expressed in response to question 15: unsatisfactory definitions; words missing; unsatisfactory syntactic guidance; excessively long entries; incomprehensible coding.

Discussion

THE FINDINGS

The survey was limited to an examination of monolingual dictionary use, so no comparison can be made with the findings of Baxter and Tomaszczyk regarding bilingual dictionaries, except to note that, in Béjoint's study, only 17% of informants claimed to prefer bilingual dictionaries to monolingual dictionaries. This finding contrasts very markedly with the findings of Tomaszczyk and Baxter, who both record that subjects preferred to use bilingual dictionaries. However, as far as monolingual dictionary use is concerned, Béjoint's results are not inconsistent with those of Tomaszczyk and Baxter.

Replies to the first six of Béjoint's questions reflect tutors' advice to their students. Third and fourth year students had been advised to buy OALD, therefore almost all of them possessed a copy; second year students had almost all acted on their tutors' advice by buying LDOCE. We are not told (indeed the answers on the questionnaires may not have revealed) whether the 14% who possessed a copy of COD used it as their second or their only English dictionary, but it is not surprising that in question 6 "a preference for EFL dictionaries, as opposed to dictionaries designed for native speakers, appears only dimly", given that so few students possessed native speaker dictionaries, and only the "more advanced" had been recommended to use them. Only 16 out of Béjoint's 122 informants were postgraduates, and we are not told which "more advanced" students had been recommended to use COD.

In their replies to question 7, many of Béjoint's respondents claimed to use their monolingual dictionaries on a daily basis. The respondents to the surveys by Baxter and Tomaszczyk appear to have used their monolingual dictionaries less frequently. Béjoint finds it "difficult to see why" his results differ from those of Baxter and Tomaszczyk, but they may well be a reflection of a) the nature and frequency of the students language assignments, and b) their use of bilingual dictionaries (79% of Baxter's Japanese English majors used an English-Japanese dictionary daily, but we have no record of the French students' bilingual dictionary use).

Alternatively, some of the responses to questions about frequency of use in each of the surveys may not have been a strict reflection of the truth, but may instead have reflected subjects' beliefs about desirable patterns of dictionary use in their place of study.

As in Tomaszczyk's survey, meaning appears to rank highly as information frequently looked up. 87% of Béjoint's subjects placed meaning among "the three most sought-after pieces of information". However this does not necessarily imply, as Béjoint suggests, that students would find some sort of gloss as suited to their purposes as a full dictionary entry. Meaning may be referred to so frequently in questionnaire responses (in native speaker as well as non-native speaker surveys) because all the other types of enquiry entail understanding of word meaning. In order to check the spelling of a word, for example, it is necessary to check that the word located in the dictionary is not just a homophone, or near homophone, of the word required. Again, pronunciation may depend on the meaning as well as on the spelling of the word (the two pronunciations of BOW, for example), and while encoding it is necessary to check meaning together with register labels, as a word may have an informal, archaic, or regional marker on just one of several meanings.

Answers to question 9 suggest that the students made less use of dictionary information for encoding purposes than for decoding purposes. This finding is reinforced by the answers to questions 11 and 13, which indicate that few students read dictionary introductions or use dictionary codes, both of which are designed primarily to aid encoding. The responses to Béjoint's question 9 match those of Tomaszczyk regarding the comparative infrequency of dictionary consultations during oral activities, but do not entirely accord with those of Tomaszczyk as far as encoding and decoding are concerned. Tomaszczyk's results suggest a higher frequency of dictionary use in writing than in reading, although Béjoint obscures this

fact by claiming his results are "very similar to Tomaszczyk's". The reliability of Tomaszczyk's results has already been questioned, but it should also be noted that Tomaszczyk was not only concerned with monitoring monolingual dictionary use, and Tomaszczyk's respondents seem to have made heavy use of bilingual dictionaries in the encoding process.

Béjoint does not draw from his findings any implications of inadequacy on the part of the monolingual dictionaries. Instead, he recommends dictionary-makers not to waste their time refining their encoding information, given that students are not using the existing information as well as they might. I think a more appropriate response to the discovery that students do not make full use of dictionary encoding information might be to suggest ways of altering the dictionaries so that this information becomes more accessible.

In question 12, students are asked about their consultation of information, such as abbreviations, listed in appendices. (The same question appeared in Tomaszczyk's survey). However, of the three dictionaries most frequently consulted by Béjoint's respondents, only OALD lists abbreviations in a separate appendix - both LDOCE and the newer edition of COD list them in the body of the book. Answers elsewhere in the questionnaire reveal that 49% of students look for abbreviations "sometimes" (question 17), and 40% of all students search for abbreviations in an appendix (question 12), yet only 45% of students own OALD (question 2). Do OALD users need to look up abbreviations more often than LDOCE and COD users? A more likely explanation is that some students have misconstrued question 12, which does not seem to take differences in dictionary layout into account.

Again, the 30% of users who look for irregular verbs in the appendix (question 12), and the 24% who use pictorial information (question 18), cannot be consulting COD, which has no verb lists or pictures.

The 77% satisfaction rate Béjoint records in his results for question 14 probably reflects the students lack of familiarity with alternative dictionaries, and their acceptance of their tutors' judgement. It may also be that they lack enthusiasm for bilingual dictionaries because their department has discouraged their use. Students without strong convictions will probably echo departmental policy in their answers, especially when the survey is conducted by that same department, as is the case here.

Many students failed to record problems with their dictionaries in answer to questions 15 and 16. It is possible that they regarded the questionnaire as some sort of test of their own competence; Béjoint comments that "informants are often reluctant to admit to a failure to understand". Apart from those problems that respondents failed to recall, or were reluctant to admit to, there must also have been occasions when the respondent never became aware that the look-up process had been unsuccessful - that, for example, the wrong definition had been selected, or that the definition had been misread. We cannot therefore assume that the answers to questions 15 and 16 accurately reflect the degree of difficulty French students actually experience with their EFL dictionaries.

Neither can we assume that answers to question 17 are completely accurate, given that some of Béjoint's categories are unclear. What constitutes a "culture-specific" word, or a "common" word, is open to interpretation, and indeed it is not surprising that so high a number of respondents denied looking up "common words", as a common word is by definition one that is frequently encountered and well-known. A native speaker's list of common words would be far longer than an advanced learner's, however, whose list in turn would be longer than that of a beginner's!

As Béjoint points out, students' responses to question 19 largely depend on the dictionary they are accustomed to using. LDOCE is the only dictionary commonly

used by the respondents to have separate main entries for compounds, but in LDOCE the compounds can also often be accessed through either of their separate elements. It is not surprising that respondents are not of one mind in any given case, as each of the three commonly used dictionaries has a different policy regarding compounds, and there are also inconsistencies within each dictionary. Whatever their reply, respondents might be successful or unsuccessful in their dictionary search depending on the dictionary they use.

In analysing respondents' evaluative comments on their dictionaries, it would be helpful to have information about the type of dictionary each respondent was referring to. This information may be recoverable from the original data, but is not presented in Béjoint's paper. Béjoint tells us that French students often complain about the incomplete coverage of Americanisms in British dictionaries, and it would be interesting to know whether LDOCE users were more satisfied in this respect, as LDOCE does give better coverage of American words. Many of the problems that students complained of - complicated definitions, not enough examples, unsatisfactory syntactic guidance - are more typical of COD than of OALD and LDOCE, which argues against Béjoint's suggestion that "dictionaries intended for native speakers would seem to be as useful to our students as EFL dictionaries" (1981:220).

THE QUESTIONNAIRE

Béjoint's questionnaire suffers from two major faults: some questions depend too much on the users' powers of analysis, retrospection and recall, while others require background knowledge and awareness of the possibility of alternative solutions that the respondents do not possess.

An illustration of the first problem is the way that respondents are expected to supply from memory a fairly detailed analysis of their look up habits in response to questions

7, 8, 15, 16, 17 and 18. The second problem is illustrated by the way learners are asked to judge their dictionaries in questions 6, 14 and 21, although the answers to question 5 show that the majority of respondents have no experience of monolingual dictionaries other than the one recommended by their tutors. As a result, respondents for the most part appear to echo opinions voiced by their department in answer to questions 6 and 14. Respondents are also expected to understand Béjoint's own (linguist's) terminology, as in question 8, for example, and are asked to express their needs within the framework of an existing dictionary format, rather than in terms of whatever they truly feel that they require. In Béjoint's conclusion he suggests that "students need to be taught how to use the monolingual dictionaries which they already possess so as to get the most out of them". Although dictionary training may be regarded as a valid need, it is not one which the respondents have themselves expressed, and it reflects a desire to mould the learners to the requirements of the dictionary rather than the dictionary to the requirements of the learner.

1.1.7 Battenburg (1989)

This study is very similar to Béjoint's, although fewer subjects were involved, and the questionnaire was somewhat shorter. Of the seven questionnaire-based studies discussed in this section, Battenburg's is the only one to elicit information from learners from a wide variety of first language backgrounds.

Battenburg distributed his questionnaire to 60 non-native speakers studying at Ohio University in 1984. These subjects had been randomly selected from a larger population of non-native speakers studying at three language proficiency levels: 20 had an elementary level of English, and were attending an intensive English programme, 20 were at intermediate level, and were attending either intensive or regular university classes, and 20 were advanced learners, attending regular university classes.

Although the subjects appear to have been evenly distributed across the three language levels, the spread of first languages was rather uneven. The subjects came from seven different language backgrounds; there were 24 Arabic speakers, 19 speakers of Chinese languages, 8 Korean speakers, 4 Urdu speakers, 3 Spanish speakers, one Portuguese speaker, and one Icelandic speaker. However, Battenburg claims that there was no evidence that differences in language background created significant differences in dictionary use.

Subjects were asked to: identify the dictionaries they owned; state the frequency with which they used them; identify the types of information they looked up; give their reasons for deciding to buy a particular dictionary; state the frequency with which they used a dictionary in conjunction with particular language skills; state their degree of satisfaction with their dictionary; suggest improvements to their dictionary.

The results showed that the respondents owned bilingual, monolingual learners' and native speaker English dictionaries. The range of ownership is summarised in Table 1.9 below.

Table 1.9: Ownership of different types of dictionary

	Elementary	Intermediate	Advanced
Bilingual	100%	50%	55%
English MLD	90%	35%	70%
NS English	55%	15%	100%

Bilingual dictionaries were owned by the largest number of subjects, native speaker dictionaries by the smallest number. All the elementary level subjects owned bilingual dictionaries, and all the advanced learners owned native speaker dictionaries.

Curiously, there were much lower levels of dictionary ownership amongst the

intermediate students. Battenburg reports that most of the monolingual dictionary owners possessed either OALD, LDOCE or LDAE (Longman Dictionary of American English).

The questionnaire required subjects to rate frequency of use on an "always/often/sometimes/never" basis. Their responses are summarised in Table 1.10, where the leftmost figure in each column indicates the percentage of subjects who "always" used a dictionary, the rightmost figure the percentage who "never" used one.

Table 1.10: Frequency of dictionary use

	Elementary	Intermediate	Advanced
Bilingual	40/45/15/00	15/25/20/40	15/25/15/45
English MLD	40/45/10/05	25/05/10/60	20/20/25/40
NS English	05/20/25/55	00/00/15/85	05/50/45/00

Battenburg comments that "in general there was a correlation between dictionary usage and ownership". Bilingual dictionary use decreased, and native speaker dictionary use increased with language proficiency. The use of a monolingual learners' dictionary also decreased at more advanced levels of language proficiency.

The intermediate group, which registered the lowest ownership of dictionaries, also used dictionaries least frequently - 40% used a bilingual dictionary "always" or "often", and 30% used a monolingual learners' dictionary "always" or "often". None of this group used a native speaker dictionary frequently. The advanced group used bilingual dictionaries with the same degree of frequency as the intermediate group, and learners' dictionaries slightly more, in addition 60% of the advanced group used a native speaker dictionary "always" or "often".

The questionnaire also required subjects to identify the information types they looked up, once again on a "always/often/sometimes/never" basis. The results are summarised in Table 1.11 below.

Table 1.11: Frequency of consultation of dictionary information types

	Elementary	Intermediate	Advanced
Spelling	25/40/25/10	25/40/25/10	10/45/35/05
Pronunciation	30/20/35/15	25/15/35/25	00/10/55/30
Parts of speech and syntactic patterns	25/30/40/10	10/25/20/45	00/00/65/35
Definitions	55/25/15/05	60/20/15/05	25/30/50/00
Etymology	00/00/35/65	00/05/25/70	00/05/50/45
Illustrations	20/10/35/35	10/10/35/45	05/25/40/30
Derived forms	10/60/20/10	10/20/55/15	00/25/25/50
Synonyms	60/25/5/10	15/35/30/20	00/30/40/35
Cross-references	15/30/15/40	00/20/40/40	05/15/65/25
Usage labels	00/10/15/75	05/30/25/40	05/10/70/15
Pictures+diagrams	25/20/15/30	05/15/40/40	00/20/35/45

In his analysis of these results Battenburg comments on the strong preference for definitions expressed by all subjects, and their lack of interest in etymology. He notes that advanced ESL students look up the meaning and the pronunciation of words less often than less proficient students, but he did not find any great differences in the way students consult information such as spelling, illustrative sentences, and cross references. In general, he found that spelling is more important to subjects than illustrative sentences, and illustrative sentences are more important than cross-references. Elementary students seem to use both synonyms and derived forms more commonly than other dictionary users, and Battenburg suggests that this may be

because such information is particularly useful for building vocabulary. Battenburg notes that usage labels are consulted most by intermediate level students, possibly because dictionary users with a lower language proficiency level may be unable to understand and benefit from such material, while advanced students are proficient enough not to need it.

Subjects were asked to indicate whether they looked up certain types of information given in dictionary introductions and appendices. Their answers are summarised in Table 1.12 below.

Table 1.12: Respondents' use of dictionary introductions and appendices

Introductory Information	Elementary	Intermediate	Advanced
Pronunciation guide	25%	75%	70%
Guide to dictionary	25%	45%	30%
Appendices	Elementary	Intermediate	Advanced
Abbreviations	30%	50%	70%
Nations+money tables	30%	25%	40%
Irregular verb forms	75%	80%	65%
Weights + measures	25%	25%	50%
Family tree	15%	15%	5%
Spelling table	25%	80%	50%
Proper names	35%	40%	55%

Elementary level students appeared to use the front and back matter of the dictionary least. Oddly the intermediate group, who claimed to use their dictionaries least frequently, and to own fewer dictionaries than subjects in the other two groups, appear to use the introduction and appendices to their dictionaries to almost the same extent as the advanced group. The yes/no format of the question is probably

responsible for this result; the wording of this and other questions will be discussed in the discussion section below.

Subjects gave the following reasons for choosing their dictionaries: amount and quality of definitions 30%; advice of teacher 20%; advice of a friend 20%; number of entries 15%; preferred size 10%; appropriate price 5%.

Battenburg comments that there was no significant difference between groups in response to this question. He suspected that "students had no overriding reason to select one reference text in place of another".

Subjects were also asked to rate the frequency with which they used a dictionary when engaged in different activities. The questionnaire responses are summarised in Table 1.13 below. Once again, the "always/often/sometimes/ never" scale is used.

Table 1.13: Frequency of dictionary use according to activity

	Elementary	Intermediate	Advanced
Reading	20/55/25/00	50/40/10/00	05/20/70/05
Writing	10/35/40/15	35/25/35/05	00/45/55/00
Speaking	05/05/20/70	00/05/50/45	00/00/30/70
Listening	00/20/40/40	00/05/50/40	00/00/45/55
Translation L1-Eng	10/15/40/35	10/35/50/05	00/30/40/30
Translation Eng-L1	35/35/30/00	30/35/30/05	00/30/50/20

Predictably, little dictionary consultation is recorded for speaking and listening activities. The greatest dictionary use appears to occur while reading (for elementary and intermediate subjects) and writing (for advanced subjects). Translation from English into the L1 is also recorded as an event where considerable dictionary use

takes place. This is surprising in view of the fact that many of the subjects appear to have been EAP students, who would not have been required to translate into the L1 in the course of their studies in the USA.

Subjects were asked to record their degree of satisfaction with the information and organisation of their dictionary. Table 1.14 is a summary of their responses, recorded once again in terms of "always/often/ sometimes/never".

Table 1.14: Subjects' satisfaction with their dictionaries

Elementary	Intermediate	Advanced
15/30/50/05	30/55/15/00	00/75/25/00

Elementary learners appeared to be least satisfied with their dictionaries. Puzzlingly, in view of the fact that they owned and used fewer dictionaries than those in the other two groups, intermediate students showed an even higher level of satisfaction than the advanced learners.

The following improvements to existing dictionaries were suggested: clearer explanations 20%; more examples 20%; larger and more readable print 15%; better pictures and diagrams 10%; more entry words 10%; more specialized vocabulary 5%. Battenburg reports that, for this question, there was little difference in responses between the various levels.

Discussion

Battenburg's questionnaire is similar to Béjoint's, but there are three major differences between the two studies:

- * Battenburg does not restrict his questions to monolingual dictionary use
- * in Béjoint's questionnaire, an "often/sometimes/never" option is offered only once, and on other occasions where Battenburg asks for an answer along a scale of frequency, Béjoint instead asked his subjects to identify the most frequent category.

For example, Béjoint asked his subjects "For which sort of activity do you most often use your dictionary?", and listed the activities from which they should choose, while Battenburg asks subjects to match dictionary use with activities on an "always/often/sometimes/never" scale. Problems associated with Battenburg's approach are discussed below

* 63 of Béjoint's subjects were in their second year, 43 in their third year and 16 in their fourth year. However, Béjoint rarely differentiates between the responses of the three different groups. For Battenburg, on the other hand, the division into language level groups is a very important one, and his results are reported at every stage in terms of the language proficiency of the respondents.

Battenburg's questionnaire also differs from Béjoint's in a number of minor ways:

- * he does not ask his subjects whether they "browse"
 - * he does not ask his subjects to recall occasions when look up was unsuccessful
 - * he does not ask his subjects to decide on the headwords for compounds
 - * he does not ask his subjects their opinion on the degree of detail given in their dictionaries
 - * he calls his language activities by the names by which they are known in American universities, while Béjoint's language activities are those familiar to French students.
- Thus Battenburg reports on reading, writing, speaking, listening and translation, while Béjoint writes of version, written and oral comprehension, written and oral composition, and theme.

As a result of these changes, Battenburg's questionnaire is slightly shorter than Béjoint's. The omission of Béjoint's questions regarding unsuccessful look up is probably an improvement, as it places less of a burden upon respondents (Béjoint acknowledged that many of his subjects were unable to remember the words and information they could not find in their dictionaries, and his subjects may have been unwilling to admit to their dictionary use problems, anyway). Nevertheless,

Battenburg's questionnaire does not entirely avoid the problems inherent in Béjoint's study, and his questionnaire still requires a certain amount of recall and retrospection, particularly regarding the use of dictionary information types. It is doubtful whether the average dictionary user really thinks about his use of synonyms and illustrative sentences, for example, even at the moment when he is accessing the dictionary. Unless specifically required to do so, most dictionary users are likely to read the dictionary entry solely with a view to solving their specific query, and will not analyse the usefulness of component parts of the entry. Retrospection about the usefulness of dictionary entry components is therefore highly suspect.

Like Béjoint, Battenburg uses terminology which would be rather obscure for the non-linguist. However, it should be borne in mind that Béjoint's subjects were all students of language, while Battenburg's subjects were apparently studying, or were about to study, across a range of disciplines, and were learning English simply as an academic tool. More importantly, Béjoint's survey was conducted in the subjects' first language, whereas Battenburg's questionnaire is in English. It seems unrealistic of Battenburg to require subjects with elementary English to comment on their use of "syntactic patterns", and "derived forms" in the dictionary. The findings reported in Table 1.11 are thus suspect for two reasons: the subjects may not have been able to describe with any accuracy the details of their past dictionary use, and the subjects may not have understood the meaning of some of the categories they were required to comment on. It is also possible that subjects made different interpretations of the adverbs "always", "sometimes" and "often". This possibility will be discussed below.

Battenburg's decision to divide questionnaire responses according to language level might have provided useful insights into the development of learners' dictionary use unexplored in the studies by Baxter and Béjoint. Unfortunately the results suggest that the three groups do not represent a true language learning continuum. The differences between the elementary and the advanced group sometimes conform to our

expectations, but the intermediate group's responses do not generally bridge the gap between elementary and advanced behaviour. For example, only 35% of the intermediate group claim to possess a monolingual learners' dictionary, as opposed to 90% of the elementary group, and 70% of the advanced group. Likewise only 15% claim to own a native-speaker dictionary, as opposed to 55% of elementary students and 100% of advanced learners. These facts do not provide evidence of gradual change in dictionary using habits, but rather completely different habits among discrete sets of users. It is unlikely that students should cease to own dictionaries they possessed at any earlier language learning stage, unless that stage was many years behind them. It seems difficult to imagine the elementary group, who all claim to own bilingual dictionaries, suddenly disposing of these dictionaries the following year. Surely they are more likely to keep them for occasional use, while adding a more comprehensive monolingual dictionary to their collection?

The intermediate group provide unpredictable data on a number of other occasions. They reported higher frequency levels for dictionary use during activities than the elementary and advanced learners (see Table 1.13), and a higher degree of satisfaction with their dictionaries (see Table 1.14), yet they also reported using their dictionaries less frequently (see Table 1.10). One reason for these strange results might be differences in interpretation of the adverbs "always/often/sometimes/ never". This question type is problematic because there are no clear cut distinctions between three of the categories; only "never" is absolute in any real sense, and "always", although it appears to be meaningful, cannot have been interpreted literally by, for example, the 5% of elementary subjects who claimed to "always" use a dictionary while speaking, or the 50% of intermediate subjects who claimed to "always" use the dictionary while reading. In an English speaking environment, these subjects are bound to have had occasion to read and speak in English without the benefit of a dictionary.

A further complication in the interpretation of results is that, although we are told the percentage of students who owned each of the three dictionary types, we are not given any indication of the degree of overlap. Table 1.9 could indicate 100% ownership of dictionaries amongst the intermediate students, for example, or it could indicate that just 50% of the group owned dictionaries, with many of these students owning more than one. More precise information about dictionary ownership would be recoverable from the data, and should have been presented in Battenburg's findings. I would also have preferred numerical data to be presented in terms of student numbers, rather than percentages, in view of the fact that there were so few subjects in each language level group. (When Battenburg reports on 5% of a group, he really means just one individual).

Like Béjoint and Tomaszczyk, Battenburg generalizes about the kinds of information contained in dictionary introductions and appendices. The information types listed in Table 1.12 seem to reflect the contents of OALD's front and back matter, although the majority of respondents used bilingual and native speaker dictionaries which do not normally offer this quantity of language information. Interpretation of Table 1.12 is further clouded by the yes/no question format used to obtain the data for the table. (This question format is also used by Tomaszczyk when asking about subjects' use of appendices). In order to be included in the calculation, respondents need only have consulted a dictionary information type once in their lives. This may account for the high levels of introduction and appendix use recorded for the intermediate group, who have a lower level of dictionary use according to Table 1.10.

Table 1.12 does not show any significant development in language knowledge or dictionary use across the three levels of proficiency. Battenburg's advanced students appear to make only slightly less use of irregular verb tables, for example, than the elementary and intermediate students. This contrasts with Béjoint's finding that his subjects had ceased to use irregular verb tables by their fourth year of study. The

results may be affected by the yes/no reply system, which does not indicate frequency of use.

Some of Battenburg's findings may be of value, particularly when they support the findings of previous questionnaire-based surveys. Those listed in Table 1.11, although they are the product of retrospection, correspond quite well with the findings of other surveys on native speaker and non-native speaker use: it would appear that dictionaries are consulted primarily for their definitions, and while spelling and examples are of some interest to users, word etymology is considered to be relatively unimportant. The primacy of definitions is to be expected, as it is really necessary to ascertain a word's meaning before moving on to use other types of dictionary information, such as cross-references and derivations.

Battenburg also obtained very similar responses to Tomaszczyk and Béjoint when he asked his subjects to suggest ways in which existing dictionaries could be improved. Like them, he did not encourage respondents to think beyond existing formats, but rather to advise relatively minor changes to the dictionaries already in their possession. Béjoint's and Battenburg's subjects called for clearer layout and typography, and in all three studies, subjects asked for more entries, and more entry information in their dictionaries. Indeed, subjects in all the questionnaire-based surveys seemed to place a high value on comprehensiveness in dictionaries. These findings may be of some interest to lexicographers, but unfortunately in all the studies too many dictionaries were under consideration. In Battenburg's study in particular subjects must have referred to a wide variety of different bilingual dictionaries, and we do not know what features individual titles had in common. The subject who demands clearer printing in a cheap pocket edition may be quite satisfied with the printing standards of a larger volume, and unless we know what standards he bases his comments on, we cannot tell what his ideal dictionary would be like. When

subjects call for larger, clearer, and more comprehensive dictionaries, what are they comparing them to?

1.1.8. Concluding comments on questionnaire-based research into dictionary use and dictionary requirements

This section of chapter one has described and evaluated seven studies which employ questionnaires as their primary means of data collection. Many of these studies are well known, and are often cited in support of lexicographical decisions. It is difficult to see how much of the data recorded in the seven surveys just described could have been collected, if not by means of responses to questionnaires. Questionnaires are excellent for obtaining information on large numbers of subjects, concerning matters that are not readily observable. The surveys by Quirk (120 subjects), Jackson (86 subjects) Tomaszczyk (449 subjects) Baxter (342 subjects) Béjoint (122 subjects) and Battenburg (60 subjects) were on too large a scale for interviewing to be practicable, and testing or protocol analysis alone could not have revealed factual information concerning, for example, the subjects' ownership of dictionaries.

However, certain defects are evident in the studies which lead me to query the usefulness of questionnaires as a primary or sole means of data collection. Below are outlined the problems associated with the use of questionnaires in dictionary research.

Problems associated with the use of questionnaires

The questionnaire as a method of data collection has come in for considerable criticism in studies of dictionary design research. Hartmann (1987), in his critique of research methods, suggests that dictionary user surveys should be moving away from questionnaire-based research:

more and more the suspicion is gaining ground that indirect surveying of population samples needs to be supplemented or replaced by more carefully controlled direct observation.

(1987:15)

Crystal (1986) points out that the retrospective answers requested in many questionnaires place too heavy demands on the respondents' memories. He challenges the veracity of Quirk's findings by asking whether anyone could "confidently write down when they last used a dictionary, why they used it, and how often they consult one".

Moreover, Crystal complains, questionnaires tend to direct subjects towards a certain kind of response, and discourage imaginative and original suggestions for dictionary design:

because we know what "should" be in a dictionary, as good linguists and lexicographers, we ask questions relating only to these notions - questions to do with lexical relationships, form class, etymology and so on. But an ideal lexicographer should always be striving to go beyond this - to discover whether there are other parameters of relevance to the user.

(1986:78)

Hatherall (1984) argues that subjects' preconceptions about desirable dictionary use often interfere with their representation of the facts. He takes examples from responses to questionnaires by Béjoint (1981) and Hartmann (1982) to illustrate his objections to the data collection method:

are subjects here saying what they do, or what they think they do, or what they think they ought to do, or indeed a mixture of all three? Do they all define the categories in the same way as the researcher? When all is said and done, do we not, on this basis, arrive at a consensus on how subjects are likely to behave when faced with a particular questionnaire, rather than authentic data on what they use the dictionary for?

(1984:184)

The same sort of doubts seem to have troubled the researchers themselves. At several places in his paper Tomaszczyk queries the accuracy of responses; one inconsistency between sets of answers to two questions concerning grammar in the dictionary prompts him to comment:

the discrepancy between the two findings may be indicative of the limited usefulness of asking such questions.

(1979:112)

When Béjoint comments on the surprisingly low number of students in his survey who admitted that they could not understand the dictionary codes, he explains that "informants are often reluctant to confess to a failure to understand" (1981:217); if this is the case, such reluctance may invalidate responses across a range of questions.

The main criticisms of questionnaire-based research seem to be that:

1) results are often a measure of the respondents' perceptions, rather than objective fact. The respondents' desire to conform, their (perhaps unconscious) wish to appear in some way better than they really are, or their inability to recall events in detail may distort the data

2) researcher and respondent do not necessarily share the same terms of reference. Linguistic concepts cannot be accurately expressed without metalanguage, but the specialist terms the linguist uses have no meaning, (or a different meaning) for the non-linguist. In large scale surveys where there is no opportunity for researcher and respondent to negotiate meaning there is likely to be considerable misinterpretation of both the questions and answers.

One way of overcoming the defects of questionnaire-based research is to use questionnaires alongside other methods of data collection. The ethnomethodological research approach supports the simultaneous use of a variety of data collecting techniques, the results of which can be cross-checked, or "triangulated". Ideally, valid data should be verifiable across a range of different types of investigation.

Questionnaires might also be modified to improve their reliability. For example, the questionnaire might be designed to elicit only personal information about the respondents, without any element of interpretation or judgement. Respondents might be asked, for example, to state their first language, subject specialism, year of study and the dictionaries they own. However, although data collected by this means would be highly reliable, it would not provide much insight into the causes and effects of

dictionary use. For this we must employ some other method whereby behaviour can be recorded but can remain uncoloured by subjects' own misrecollections and misrepresentations of events.

These considerations led me to reject the questionnaire as a primary means of data collection in my own studies, and to turn my attention to testing as a possible source of reliable and objective information about dictionary use. In the next section the studies I review have all gathered information under test conditions.

1.2. Test-based research

In this section I group a number of studies which aim to discover more about dictionary user's skills and habits by conducting tests on experimental subjects. Seven studies are discussed: Bensoussan, Sim and Weiss (1984), Black (1986), Atkins and Knowles (1990), Tono (1988), Tono (1989), Bogaards (1991), and Luppescu and Day (1993).

The studies are primarily accounts of controlled experiments; two of them (Atkins and Knowles 1990 and Tono 1988) assess subjects' dictionary-using skills, while the remaining four are concerned with discovering how dictionary use affects learners' performance under test conditions. By testing their subjects under controlled conditions the researchers presumably hope to collect more reliable and objective data than that acquired by questionnaire based research. In the first two studies, however, questionnaires were administered to supplement test data, thus providing opportunities for the triangulation of data as discussed at the end of the preceding section.

1.2.1. Bensoussan, Sim and Weiss (1984)

This is the earliest paper to report a test-based investigation into learner dictionary use.

The situation which gave rise to the study was one of conflict between examination administrators and teachers. Examination administrators, the authors say, are against the use of dictionaries in the examination room because they think that dictionaries might help the candidate too much, encourage cheating, and cause candidates to waste time. In contrast, the authors claim that many EFL teachers would like learners to be able to use dictionaries in the examination room, as tests should test the candidate's ability to function in a natural reading situation - outside the examination room learners usually have the opportunity to consult a dictionary. Bensoussan et al also mention, however, that some EFL teachers are against dictionary use in class, preferring their students to rely on contextual guessing techniques.

The authors' survey of current views reflects both sides of the controversy. Some writers approve of (monolingual) dictionary use, and suggest that dictionary skills should be taught to learners. Other sources are cited in criticism of dictionary use.

The idea of developing contextual guessing strategies as an alternative to dictionary use is developed further. Bensoussan et al have strong reservations about the success of guessing techniques. Their own research (Laufer and Bensoussan 1982, Bensoussan and Laufer 1984), records failure in lexical guessing because learners have preconceived notions about word meaning:

instead of using context to help them understand a particular word, students are likely to alter the meaning of the context so that it is in keeping with the word they mistakenly think they know.

(1984:264)

The authors, however, do not adequately distinguish between this sort of lexical guessing strategy and taught contextual guessing techniques which are designed to replace it. In fact, nowhere in this paper is contextual guessing adequately defined, nor are the component strategies listed.

The authors refer to the work of Goodman (1969) to support their view that context can be ineffectual as a means of teaching vocabulary. Goodman points out that native-speaker language knowledge does not guarantee access to a text, because if the reader lacks relevant knowledge, he cannot supply this semantic component and he cannot read. Bensoussan et al do not point out, however, that Goodman's argument should not be taken as an argument in favour of dictionary use as opposed to contextual guessing, because if the reader lacks the appropriate background knowledge the text will probably remain inaccessible to him whether he consults a dictionary or not.

The authors also cite Johns (1980), who believes that contextual guessing can only successfully occur when a sufficient number of words in the text are already known to the reader, thereby enabling the reader to cross a "threshold". John's limit on unknown words in the text - "more than approximately 50 per 1000 words" is rather low, and does not take into account other factors that may affect overall readability of a text, such as background knowledge, conceptual difficulty and syntactic complexity. In their paper the authors do not draw conclusions about the importance of his theory to any comparative study of learner performance with and without dictionaries. A large proportion of unknown words in a text presumably inhibits contextual guessing and dictionary look-up alike; if Johns' calculations are correct, any kind of meaning search will be impracticable when more than 50 per 1000 words are unknown, and a dictionary will be of no benefit to the reader.

Bensoussan et al also do not make the links between contextual guessing and dictionary look-up explicit, although they cite authors who clearly regard some kind of contextual guessing as the first stage in a process which is completed by dictionary consultation. Their summary of Scholfield (1982b) indicates the importance of context to successful dictionary use:

Scholfield argues that, far from being a mechanical process, use of the dictionary involves the prior knowledge of the reader and his ability to make hypotheses about the context while reading.

Gove (1969) is also quoted as saying substantially the same thing:

words do not exist by themselves - they are surrounded by other words and live in a context of association and related ideas from which a consultor (the student) takes to the dictionary some little bit of understanding.

(1969:198)

All this would seem to suggest that learners who are capable of making hypotheses, and who can "recognise the context of association and related ideas" will benefit from access to a dictionary while reading. The experiments of Bensoussan et al described below were designed to test the benefit of dictionary use while answering questions in a reading test.

The experiments

The experiments were designed primarily to determine two things:

- 1) to what extent the use of monolingual and/or bilingual dictionaries affects examination performance (ie test scores)
- 2) to what extent the use of dictionaries affects the amount of time taken to complete the test.

Four experiments were carried out, each with impressively large numbers of subjects.

The first experiment, a pilot study, was the simplest. 900 first year students answered multiple choice questions on ten reading passages. Half used monolingual dictionaries, half used no dictionary at all. There was no significant difference in test score. No further details are given of this experiment, but the authors note that they found the results surprising, and were thus prompted to conduct further, more elaborate, studies.

In study 1, ninety-one students, mostly in their first year, took a three-hour reading comprehension test using either a monolingual dictionary, a bilingual dictionary, or no dictionary at all. In the first twenty minutes of the test students were asked to tick the words they intended to look up, and when answering the questions they were

asked to underline the words they in fact looked up, and note their parts of speech.

Test scores were compared with the type of dictionary consulted, and the type of dictionary consulted was also compared with:

- 1) the number of words students had originally intended to look up
- 2) the number of words they in fact looked up
- 3) the frequency of the parts of speech of the different words indicated.

Dictionary use did not affect test scores, regardless of dictionary type. Students indicated that they wished to look up many more words than they actually looked up when answering the test questions. Bilingual dictionary users looked up more words than monolingual dictionary users. No significant differences were found between high, medium and low student proficiency levels in terms of the numbers of words marked, the number of words looked up, and the different parts of speech looked up.

In Study II 670 first year students took part in a two-hour reading test and could choose to consult a monolingual dictionary, a bilingual dictionary, or none at all.

Study III was a replication of Study II with 740 subjects. In both studies the type of dictionary chosen by each subject was noted, and scores were compared against dictionary type and the number of minutes taken to complete each test.

It was found that 59% of the subjects (Study I) and 58% of the subjects (Study II) chose to use a bilingual dictionary, and about 20% chose to use a monolingual dictionary. The remainder either chose not to use a dictionary, or would have preferred to but forgot to bring one.

Little difference was found between the test scores of different groups of subjects.

There was also little correspondence between dictionary use and the time taken to finish the test, although students who did not use a dictionary tended to finish fastest, and students who used bilingual dictionaries tended to be slowest. There was also a slight tendency for slower students to obtain lower marks, which leads the authors to infer that "students who choose to use bilingual dictionaries tend to be slower and

weaker in taking reading comprehension tests in English and, by extension, in reading English texts". The authors admit, however, that this is speculative, and in the tests as a whole bilingual dictionary use was not found to have any bearing on results. Study I found that students with bilingual dictionaries tended to look up more words than students with monolingual dictionaries, and this factor may account for a slight loss of speed amongst bilingual dictionary users in Studies II and III.

The questionnaire

The questionnaire was administered in order "to clarify the test results and to understand the underlying attitudes and expectations of dictionary users". The respondents were 404 first year students taking an English reading comprehension course, 10 teachers of these students, and 13 third year Psychology students who had a very good command of English. All the respondents were from Haifa University, but although we are told that the first year students were following the same reading course as those who took part in Studies II and III, it is unclear whether these respondents had also acted as experimental subjects in the studies.

Respondents were asked the following questions:

- 1) What is your native language?
- 2) Do you bring a dictionary to class?
- 3) Which dictionary do you bring to class? (English monolingual/ bilingual/ none)
- 4) Do you do your homework with a dictionary?
- 5) Which dictionary do you use to do your homework? (English monolingual/ bilingual/ none)
- 6) When I use the dictionary: I can read faster / I can read slower and more carefully / there is no difference in the speed of my reading.
- 7) When I use a dictionary: I understand sentences better because I understand each word / although I understand each word I don't always understand the meaning of the sentence / both of the above statements may be true.
- 8) When I use the dictionary I look for: every word I am not sure of / only the words I really don't know (I try to guess the words I'm not sure of) / only the most important words (I try to guess the others) / only the longest, most difficult words.
- 9) Do you also use the dictionary for any of the following purposes?

- *spelling (also British and American varieties)
- *punctuation
- *short forms and abbreviations
- *idioms and special phrases
- *examples of usage
- *synonyms and antonyms (words of similar and words of opposite meaning)
- *possible range and register of usage (formal, slang, regional variations etc)
- *part of speech (noun, verb, preposition etc)
- *verb patterns
- *pronunciation

10a) Are you, on the whole, satisfied with your ability to use an English-English dictionary?

10b) If "No", can you briefly explain in what ways you are not satisfied?

11a) On the whole, are you satisfied with the information provided in the English-English dictionary you use?

11b) If "No", can you briefly explain in what ways you are not satisfied?

12) If I use a dictionary during a test, my mark will be higher: yes/no.

13) What was your end of semester grade?

According to findings from this questionnaire, the most proficient students used their dictionaries less and were more critical of them. Almost half of these students did not expect dictionary use to affect their test scores. On the other hand the first year students tended to believe that "they merely needed to look up words in order to understand the text". This same group also complained about the linguistic difficulty of dictionary definitions, their length, and the problem of choosing appropriate meaning in definitions of polysemous words.

The teachers of these students did not think students used dictionaries effectively, but did expect the use of dictionaries to affect test scores.

Respondents generally tended to feel that dictionary use, if it made any difference at all to their speed of reading, would slow them down.

The authors do not reproduce in this paper their findings concerning choice of information type in the dictionary.

Comments on the questionnaire

The questionnaire used in this study was designed as a back-up to the experiments. The first and last questions simply elicit necessary background information, and the multiple choice format makes many of the other questions relatively easy to answer; the problems of metalanguage are to some extent solved in question 9 by the paraphrasing of technical terms. The simplicity of the questionnaire, however, puts it at risk of over-simplifying the facts. For example, dictionaries are divided into two categories only - monolingual and bilingual - so no account can be taken of the style and quality of the dictionaries when analysing responses to other questions. (It is interesting to note that the stated preference for bilingual dictionaries supports the earlier findings of Tomaszczyk, Baxter and Béjoint.) Many of the non-factual questions (eg questions 6, 7 and 12) seem to be designed to check whether respondents share the authors' original hypothesis that dictionary use will raise test scores and increase the time taken to complete tests. There seems little point in sounding out learners' opinion on this issue, given that the experiments have been designed to test the hypothesis objectively. There is in fact very little information elicited in the questionnaire that can have direct bearing on the analysis of the experiment results. One piece of information which does seem to have bearing - the fact that students have difficulty using their dictionaries - is dismissed by the authors in their discussion of the experiment findings (see "Comments on the experiments" below).

In their penultimate section, entitled "A Student's Eye View" Bensoussan et al do have recourse to the questionnaire responses in order to hypothesise about "how most students go about using the dictionary during a test". Other methods of data collection may be more suitable when data is required for the formulation of hypotheses of this sort - the questionnaire data records the problems that learners admit to when using a dictionary outside the exam room, and the problems Bensoussan et al extrapolate

seem on the whole to be more typical of the learner-writer than the learner-reader ("nuances of register" for example, are far more important for the encoder). There is therefore no real evidence that these problems are the main ones facing learners when consulting a dictionary during a reading comprehension test.

I suspect that the questionnaire was designed without a clear sense of purpose, and although it was later used to inform a number of the authors' arguments, the data it provided was not sufficient to provide significant new evidence in any area.

Comments on the experiments

Very large numbers of subjects were involved in all the experiments reported in this paper. Although the size of the survey is impressive, it inevitably led to huge variations in scores. Mean scores for the tests were similar on four separate occasions, but possible significant differences in performance may have been obscured by the range of variation in scores across the entire experimental population.

Moreover, before we can generalize from Bensoussan et al's findings we would need to know more about one variable not discussed in the paper - the types of text subjects had to read, and the types of test question they had to answer. We are told that different tests and texts were used in each study, but we are given little further information about them.

The choice of test may be a very significant factor. Most tests are designed in the knowledge that subjects will not have access to a dictionary, and therefore tend not to be hinged on the understanding of single lexical items; thus it seems likely that they will not favour dictionary use. However, where test administration allows for dictionary use (the Oxford/Arels Preliminary, for example, or the test constructed by Bensoussan herself (1983)) it seems more likely that results would be influenced by access to a dictionary. Tono (1989, discussed later in this section) found that

dictionary users were significantly more successful than non-dictionary users on his specially-constructed reading comprehension test.

Also, within either kind of test the level of difficulty of the text will affect the efficacy of dictionary use. For example, if the text contains a large number of unknown words, the "threshold effect" (described by Johns (1980) and discussed earlier) might obtain, and learners might be unable to work out the meaning of individual vocabulary items by guesswork or by dictionary use. Similarly, a syntactically difficult text may prevent both contextual guessing and successful dictionary use. Bensoussan et al themselves claim that, without knowledge of the syntactic rules, "a working knowledge of the missing or unknown words cannot be guessed, even with the help of a dictionary" (1984:271). On the other hand, if the texts contained few unknown words dictionaries would be unnecessary and it would make little difference whether the students used them or not.

Data from study 1 give us some idea of the difficulty-level of the reading texts used. On first reading the texts students indicated that they wished to look up an average of 61 words per 500 to 700 words. This is far higher than Johns' "threshold" of 50 words per 1000, beyond which, he claims "perception of overall structure may be effectively blocked". However it may be that the test did not require "perception of overall structure". In practice the dictionary-using sector of the test population only looked up an average of nine words each. Bensoussan et al account for this by suggesting that "motivation for looking up words arose from the need for answering the question, not the desire purely to understand the text". Time constraints may have played a part, and the students may have been able to work out meaning from context in some cases, but it would appear that the students could answer at least some test questions without understanding a high proportion of the words in the texts.

The results recorded in the four experiments were probably also influenced by two other variables - the dictionaries used, and the ability of the students to use them.

We are not given much information about the students' choice of dictionary; we are only told that the three most frequently consulted dictionaries were OALD and two bilingual dictionaries: The New Bantam-Megiddo Hebrew and English Dictionary (Levenston and Sivan 1975) and The English-Hebrew and Hebrew-English Dictionary (Ben-Yehuda and Weinstein 1974). There is no indication of the coverage of the two bilingual dictionaries (although we might expect the "Bantam" to be less than comprehensive), and there is no indication of whether the words students chose to look up were actually listed, or whether the information the dictionaries provided was appropriate. We are also not given any idea of the proportion of students who chose other dictionaries to consult.

The authors claim that most of the subjects in their experiments had received some training in monolingual dictionary use. The questionnaires reveal, however, that many first year students still experienced considerable difficulty with their dictionaries. Poor look-up strategies might at least partly account for the failure of dictionary users to improve their test score, but the authors choose not to dwell on this problem:

one explanation could be that students simply do not know how to use the dictionary efficiently during an examination. In this case, by definition, test results would not be affected. This interpretation, however, is both too simplistic and too pessimistic to be useful.

(1984:271)

Certainly it must be a depressing consideration for the authors, as the possibility of simple inadequacy on the part of the user would mean that it is less safe to generalize from their findings.

In these experiments and the ensuing discussion Bensoussan et al do not properly account for variations in test type, text type, dictionary type and coverage, and learners' dictionary skills. However the experiments do have some limited value, viewed within their original context. They suggest that dictionary use will not help first year Israeli university students to answer the sort of reading comprehension tests that were set at the time when the experiments took place.

1.2.2. Black (1986)

In this study, Black was primarily interested in the role of examples in learners' dictionary definitions.

Black cites evidence which suggest that examples added to dictionary definitions might enhance the learning process. Placing a word within an example might make it more imageable, and studies by Paivio (1971) suggest that words which are more imageable are also more memorable. Moreover examples may relate to the dictionary user's personal experience, resulting in additional cognitive processing on the part of the user. According to Craik and Tulving (1975) additional processing improves memory of a word.

Black also cites other researchers, however, who have produced evidence which suggests that definitions with examples might impede memory. The addition of examples might make dictionary definitions more readily accessible, and according to Abbot, Black and Smith (1985) information which is easily linked to existing knowledge tends to be less memorable. Texts which are difficult require more processing for comprehension and so may be better remembered (O'Brien and Myers 1985).

Black's two experiments were designed to contrast subjects' comprehension and subsequent retention of words defined in three different ways: by an abstract

"dictionary-style" definition, by a combination of definition and example, and by examples only. She predicted that words defined by examples only would be most effectively comprehended, whilst those defined by an abstract definition only would be least understood. She does not appear to have made any predictions regarding the memorability of the three defining styles.

Texts and target words for the studies were arrived at by conducting a pilot experiment with 20 subjects. The subjects were shown four texts (two narrative and two expository) on fairly familiar topics, and were asked to underline all words in the texts whose meanings they did not know and could not infer. The subjects were also asked to rate their underlined words on a scale 1-5, to indicate the extent to which they felt they knew the word (5 = completely unknown). As a result of the piloting, one of the four texts was abandoned because it proved to be too easy. Eight target words were selected for each text from those that had been most frequently underlined and had been given high average ratings.

For each of the twenty-four target words three types of definition were prepared: an abstract dictionary-style definition, an abstract dictionary-style definition with one example, and a definition consisting solely of example sentences. All three types of definition for the same word were the same length, and had a matching number of clauses. All senses of polysemous words were given. Each definition was written on a separate index card. Index cards were also prepared for the control condition with the target word only, and no definition. Multiple-choice questions were prepared to test comprehension of each of the twenty-four target words.

16 subjects took part in Experiment One, and 24 took part in Experiment Two. They were all studying in Britain, and were preparing to take the Cambridge Proficiency in English examination. They were judged by their teachers to be at a similar level of language proficiency, although this was not tested formally. There was no statistical

difference between the two groups in terms of the overall length of time the subjects had been learning English (the mean length was six years) or the length of time they had been studying English in Britain (the mean was four months).

In both experiments the subjects were instructed to read the three texts and look up each underlined word in the card index they had been given. Each subject's card index contained some cards with an abstract dictionary style definition for the target word, some cards with a dictionary style definition plus a single example, some cards with a series of examples instead of a definition, and some cards with no defining information at all (the control condition). The definition condition for each word was rotated across blocks of four subjects. The subjects were warned that a comprehension test would follow but they were not told that it would focus on the underlined words.

Having read all three texts, subjects took the comprehension test. In both experiments the subjects had access to the texts while answering the questions, but in Experiment One subjects were also allowed to consult the definitions in the card index, while in Experiment Two this was not allowed.

Results

In both experiments there was a highly significant difference between scores for words which had not been defined, and scores for words which had been defined ($p < 0.001$). However no significant difference was found between the scores for words defined by each of the three methods. Performance in the control condition was 55% correct in Experiment One, and 48% correct in Experiment Two.

Although three words were understood by less than half the subjects there was no common distribution of results for these words to suggest the influence of one particular type of defining information. The problems appeared to be caused by word difficulty, context, and the test questions. There was no significant correlation

between performance in the control condition and the two defining conditions with examples, but in both experiments there was a significant positive correlation between performance in the control condition and the abstract dictionary-style definition condition ($p < 0.05$). Variation in the quality of context information probably explains the correlating behaviour patterns in the two groups without access to examples.

Discussion

It would be dangerous to infer from Black's results that a dictionary can be just as successful if it saves space by omitting example sentences and phrases. Black's definitions in all three styles were of identical length; the examples were substitutes for part or all of the traditional definition, they were not additional features.

The results do seem to suggest that learners' comprehension of dictionary definitions is unaffected by the presence or absence of examples of use. Apparently, any one of the three methods of presenting word information is as effective as any other. The design of Black's experiments is somewhat flawed, however, and this casts doubt on the validity of her findings.

One problem with the experiment is that subjects were required to look up words that they did not necessarily need to look up, and were later tested on all the words, regardless of whether the defining information had played a part in comprehension. The subjects were not asked whether they knew or could guess the word meanings, but their high scores on words in the control condition indicate that they had only benefitted from defining information in about 50% of cases.

Target words were chosen at the pilot stage on the grounds that they were unknown and irretrievable from context. The mismatch between the subjects' anticipated lexical knowledge and their actual lexical knowledge in the experiments might have been due to faulty piloting; it is possible that the subjects who identified target words in the

pre-experiment activity knew fewer words than the experimental subjects, or did not speak cognate languages, or had less developed contextual guessing skills. However, we must assume that Black controlled for these factors across all groups of subjects. A more likely explanation for the mismatch is that the experimental groups were provided with extra context to facilitate guessing at the testing stage, in the form of the multiple choice questions. Some of the questions seem to narrow down the possible range of meaning suggested by the original context, so that it becomes easier to guess meaning correctly. The question which tests the meaning of *sloth*, for example, in the phrase "*People with the motion capacity of a frozen sloth*", asks the subject to decide whether the people are *very slow*, *very cold*, *very rapid* or *very lazy*. Given that the text indicates that they are elderly, obese and infirm, the subject should have little difficulty in picking the correct answer.

While the findings strongly suggest that defining information is a useful aid to word comprehension, it is clear that a great deal of guessing was taking place in the two experiments. Contextual guessing is an important and necessary accompaniment to dictionary use, but the double context of reading passage and test items was so rich in these experiments that essential differences in the quality of the three definition types may have been obscured. In other words the absence of examples in the abstract dictionary-style defining condition may have been offset by the presence of abundant contextual clues. Scores in the example-free defining condition and the control condition showed a significant correlation, which may suggest that subjects were adopting the same strategy of inferencing from context under both conditions, because it was impossible to refer to examples. Given a poorer or less accessible context, subjects would not have found as much information to supplement the example-free defining condition, and so this condition might have yielded less satisfactory results.

Although she does not discuss the possibility that the test itself helped subjects to infer meaning, Black concludes that the texts used in the experiment were probably to blame for the lack of variation between scores under different defining conditions, because they were conceptually easy and dealt with familiar topics.

One way to enable possible significant differences between defining conditions to reveal themselves would be to repeat this experiment, using more difficult texts.

Alternatively, subjects could be set a writing task to demonstrate their word knowledge, rather than a receptive task centred on a reading text. Miller and Gildea (1987) attempted to test the value of defining by examples by this means, but found that subjects tended to model their own writing on the examples they had been given, rendering it difficult to judge the degree of comprehension that had taken place.

Miller and Gildea's work will be discussed in 1.3. below.

Black aimed to measure the memorability of word meaning expressed by different means, as well as comprehensibility. Although scores for the second experiment are slightly lower, they are not greatly so (tests for significant difference were apparently not applied). This similarity of result is probably due to the fact that the test was taken immediately after the texts had been read, before the subjects had had time to forget the original context for the target words. Some variation in memorability might have emerged if the subjects had been tested the following day, and/or after a further lapse of time.

1.2.3. Atkins and Knowles (1990)

The project described in this paper was devised by Atkins in 1984, but received the official sponsorship of EURALEX and the AILA commission on Lexicography and Lexicology in 1986. The project has not yet been completed. Its aims are to discover:

- 1) how foreign learners of English use their dictionaries

- 2) how effective these dictionaries are in helping learners encode, decode and translate
- 3) whether bilingual and monolingual dictionaries are equally effective
- 4) students' attitudes to bilingual and monolingual dictionaries
- 5) how much instruction is being given in dictionary use
- 6) how dictionaries fail students and how dictionaries might be improved.

It was also hoped that the project would "focus theorists' attention on problems where academic research would be most helpful to the lexicographer", and would encourage more detailed research into dictionary use.

In the early stages of the project Atkins, Lewis, Summers and Whitcut designed and ran a pilot study, and compiled questionnaires and tests. These tests were adapted in the light of comments made at the EURALEX 1985 seminar on The Dictionary and The Language User, where initial findings were presented. The project was then implemented with the aid of nine university-based European "agents", each with a local network of teachers.

Data was collected by means of a questionnaire and two tests: a "placement test" and a "dictionary research test". The project was limited to the study of learners from four language groups: French, German, Italian and Spanish. 1600 sets of papers were distributed (400 in each language), but only 723 complete sets were returned. Many more respondents (1140) returned the questionnaire. The language and country distribution among those who completed the questionnaire was as follows:

German 17.72% (Austria 8.36%, W.Germany 6.00%, Switzerland 3.36%)

French 19.36% (Belgium 1.45%, France 17.91%)

Italian 29.45% (all in Italy)

Spanish 33.45% (all in Spain)

50% of these respondents were in full-time secondary education, 34% were studying in higher education, and of the remainder some were following adult education courses or were studying English privately, while a few had completed their studies.

The questionnaire

The questionnaire (known as the Dictionary User Profile Form) was drawn up in French, German, Italian and Spanish. It consisted of 17 questions, summarised below.

- 1) Country of residence and mother tongue.
- 2) Number of years of English study.
- 3) A more detailed breakdown of schooling in English (type of institution, number of years, and frequency of lessons).
- 4) Reasons for studying English. (5 reasons were suggested. Subjects were also invited to state any other reasons they might have for learning English)
 - i) to prepare for an exam (subjects were asked to state which),
 - ii) to study English in higher/further education,
 - iii) to study another subject which requires knowledge of English,
 - iv) to improve job prospects,
 - v) to travel.
- 5) Whether subjects were taught English through the medium of English. (Always / almost always / 50% of the time / rarely / never)
- 6) Whether subjects used textbooks, and if so, which.
- 7) Whether subjects had been taught in class how to use a dictionary to study English, and if so, whether the training was systematic.
- 8) For which areas of study knowledge of English would be useful. (Suggested areas included English/American literature, Science, Engineering, Medicine etc.)
- 9) Dictionaries that subjects owned for use in the study of English (title, publisher and date of publication).
- 10) The number of years that subjects had owned these dictionaries.
- 11) Why subjects bought them. (suggested reasons included recommendation by teacher/parent/friend/bookshop, cost and appearance, and for second, third and fourth dictionaries also the coverage and degree of specialization)
- 12) Frequency of use of a monolingual dictionary. (Never / rarely / weekly / daily)
- 13) Frequency of use of a bilingual dictionary. (Never / rarely / weekly / daily)

14) Subjects' use of dictionaries that did not belong to them, in a library, in class or at home. (Never / rarely / weekly / daily)

15) Titles, publishers and years of publication of these dictionaries.

16) The type of dictionary (ie monolingual or bilingual) subjects normally used for the following activities. (Subjects were also invited to indicate any other activities for which they used a dictionary, and the types of dictionary they used for these activities)

i) to find out the meaning of an English word, eg while reading,

ii) to find the equivalent of an English word in context,

iii) to find out how to use an English word already known, eg while writing.

17) Which dictionaries subjects found most useful.

The Placement Test

The placement test was originally "devised by a British Council-approved language school in London for the purpose of assigning new students to an appropriate class".

It consisted of 100 questions, to be answered within one hour. The questions and rubric were entirely in English. The first sixty questions tested knowledge of English sentence structure, while the remaining forty questions tended to focus on discourse structure and on meaning, with cloze passages and short passages for reading comprehension. Subjects were assigned one of four grades on the basis of their score in this test: A (81-100%), B (66-80%), C (51-65%) or D (0-50%).

The Dictionary Research Test

The Dictionary Research Test consisted of 44 questions to be answered without any time limit. With the exception of the translation passage, which was different for each language group, the questions were the same for all students. Instructions, however, were given in the students' first language. A control group took the test without a dictionary, but all other subjects were allowed to use a dictionary of their choice for certain questions in the test. Subjects were asked to identify the dictionary they had chosen, and after questions which permitted dictionary use they were required to indicate whether they had consulted a dictionary or not.

There were 44 questions in the test, mostly answerable by multiple choice. Atkins and Knowles group the questions into seven categories "according to the linguistic process or aspect of dictionary skills they were designed to test". These were: "knowledge of English grammatical terms, understanding of grammatical metalanguage used in learners' dictionaries, finding of multiword items (set phrase, phrasal verb and compound noun), selection of correct lexical item for several types of context, preposition selection, comprehension of English passage, and translation from English". The test itself, however, is divided into nine sections. I will describe the test section by section below.

Section 1 (six questions) - identifying parts of speech in pairs of sentences where the same word has different grammatical functions (eg 1/1. We walked round the garden 1/2 The world is round, not flat). Dictionary use is not permitted.

Section 2 (four questions) - matching word forms with abbreviated information of the kind that accompanies these forms in many dictionaries (eg 2/9. neg of CAN = cannot). Dictionary use is not permitted.

Section 3 (four questions) - selecting the correct lexical item for a particular context (eg 3/13. It's a bad ...behaviour/practice/custom/habit.... to bite your nails). Dictionary use is permitted.

Section 4 (four questions) - knowing where to find multiword items (or in one case, a derivative) in the dictionary (eg 4/15. requires subjects to state whether lame duck would appear at the entry for LAME, the entry for DUCK or at its own separate entry). Dictionary use is not permitted.

Section 5 (five items) - selecting the correct preposition for a particular context (eg 5/19. I'm surprisedwith/by/for/at.... you). Dictionary use is permitted.

Section 6 (five items) - understanding a short reading passage. Most questions required subjects to choose the correct rephrasing of quotations from the passage (eg 6/24. When the writer says "We must concentrate our efforts on increasing the skill level of drivers", he means we must do all we can to increase drivers skills).

Dictionary use is permitted.

Section 7 (five items) - filling gaps in an English passage by translating from an L1 text. Different texts were used for each language group, and the words and phrases to delete seem to have been chosen on the grounds that they posed particular problems for the L1-English translator (eg 7/29 "*auch in der Bundesrepublik*" translates in this context as "including here in the Federal Republic"). Questions tend to depend on the correct translation of a lexical word (eg "*qu'il n'exploitera pas*" = "which he will not make use of", rather than "which he will not exploit"), but may involve knowledge of the grammatical system of English. French candidates, for example, are asked to translate "*rêve encore d'être explorateur*" in their passage, with a choice of valency patterns for "dreams": "still dreams to be an explorer/ still dreams of being an explorer/ still dreams that he is an explorer/ still dreams of an explorer". Dictionary use is permitted.

Section 8 (six items) - filling gaps in an English passage. This was the only section where answers were not provided in a multiple choice format. Gapped items seem to test collocational knowledge (eg *get* a grip on yourself; to be conversant *with*).

Dictionary use is permitted.

Section 9 (five items) - selecting the correct lexical item for a particular context. Identical to section 3, but with less common vocabulary (eg 9/41. In the height of summer Venice is full/inhabited/running/swarming ... with tourists). In some questions, more than one answer is meaningful, but only one will fit the syntactic

context, eg "They spent many happy hours *reminiscing/ recalling/ reminding/ thinking* their childhood". Dictionary use is permitted.

Questionnaire Results

Figures which only represent the results of the Dictionary User Profile questionnaire come from a dataset of 1140 respondents, whereas those which include information gathered from tests come from the far smaller dataset of 723 respondents. Thus all figures incorporating placement test grades, or involving cross-tabulation of test results with nationality, come from the smaller dataset.

Analysis of questionnaire responses revealed that 19.3% of the respondents had studied English for less than five years, 62.3% for between five and nine years, and 18.4% for ten years or over. Table 1.15 below is taken from the smaller dataset and shows the distribution of placement test grades according to native language.

Table 1.15: The distribution of placement test grades according to native language

	Grade A	Grade B	Grade C	Grade D	TOTAL
French (18.1%)	10	28	54	39	131
German (16.2%)	9	64	37	7	117
Italian (28.4%)	43	46	50	66	205
Spanish (37.3%)	69	78	54	69	270
TOTAL (100%)	131	216	195	181	723
	(18.1%)	(29.9%)	(27.0%)	(25.0%)	

Questionnaire results revealed that the great majority of respondents (69.2%) received more than half their instruction in English through the medium of English, but 60.4% had never been taught how to use a dictionary, and only 12.9% had had "precise and systematic instruction" in dictionary skills. As Atkins and Knowles point out, these results are particularly revealing because all participants in this study had been volunteered by their teachers, who had been "interested enough in dictionary use to devote a considerable amount of class time to this research". Lack of instruction in dictionary use was greatest in France (79.2%), Spain and Austria (70.7%), whereas only 4.5% of Germans claimed to have had no dictionary instruction.

Only 9.2% of the questionnaire respondents said that they did not own a dictionary; 49% claimed to own one, 30.4% two, 9.2% three and 2.2% four or more. The most influential factor in the selection of their first dictionary was "teacher's recommendation". Price and illustrations were the least influential factors.

Answers to questions on bilingual and monolingual dictionary use revealed that 57.9% of respondents used a bilingual dictionary "often, nearly every week", while 30.8% made the same claim for a monolingual dictionary. Only 0.4% claimed never to use a bilingual dictionary, while 27% claimed never to use a monolingual. Cross-tabulation of questionnaire results and placement test results revealed that those who never used monolingual dictionaries tended to belong to the lower ability range; only 4.9% of grade A respondents never used a monolingual dictionary, as opposed to 37.9% of grade D respondents.

Table 1.16 below compares respondents' grades on the placement test with their choice of dictionary type. It shows a steady rise in monolingual dictionary use as ability increases, and a proportionate drop in bilingual dictionary use, although bilingual dictionary use exceeds monolingual dictionary use at every level.

Table 1.16: Choice of dictionary type compared with ability level

L	Grade A	Grade B	Grade C	Grade D	TOTA
Monolingual used "often"	29.0%	28.7%	16.9%	14.9%	22.1%
Bilingual used "often"	45.0%	56.5%	64.1%	68.5%	59.5%

The questionnaire also required respondents to choose between a monolingual dictionary and a bilingual dictionary for three types of task - decoding, translating and encoding. An analysis of replies revealed:

* for understanding an L2 expression 34.7% would choose monolingual, 59.9% would choose bilingual and 5.4% would choose both

* for translating from L1 into L2 9.6% would choose monolingual, 87.7% would choose bilingual and 2.7% would choose both

* for information on usage of a known L2 term 55.0% would choose monolingual, 41.6% would choose bilingual and 3.4% would choose both.

This tendency to choose a bilingual dictionary for decoding and translating, and a monolingual dictionary for encoding words already known, was not reflected in all subgroups of respondents. 70% of post-school students decided that they would use a monolingual dictionary for decoding, while for information on the usage of a known L2 term, only 43% said they would choose a monolingual dictionary, while 51.1% preferred a bilingual. The data on choice of dictionary type has not yet been analysed in terms of further subgroupings, such as nationality.

Test Results

Results from the Dictionary Research Test revealed that approximately 85% of subjects were able to identify parts of speech, and approximately 96% of subjects were able to interpret correctly grammatical metalinguistic labels. However, subjects were less familiar with the conventions of layout in the dictionaries they consulted. In section four of the test, which was designed to discover where dictionary users

expected to find multiword items such as LAME DUCK, the location expected by the subjects differed widely from the actual location in dictionaries. Data for two multiword items are presented in Table 1.17 below.

Table 1.17: Expected and actual locations of multiword items

Possible locations:	<i>lame</i>	<i>duck</i>	<i>lame duck</i>	(don't know)
Location expected by students:	47.0%	43.0%	7.0%	3.0%
Actual location in dictionaries:	50.0%	12.5%	37.5%	----
Possible locations:	<i>do</i>	<i>without</i>	<i>do without</i>	(don't know)
Location expected by students:	64.0%	32.0%	2.0%	2.0%
Actual location in dictionaries:	100%	0%	0%	---

There was little variation of results between different language groupings, and advanced students were no more successful in locating multiword items than beginners were.

Of the respondents who completed the Dictionary Research Test, 75% chose to consult a bilingual dictionary, and 25% a monolingual dictionary. The comparative success rate of monolingual and bilingual dictionaries has not yet been analysed across a range of questions, but figures are available for question 3/13 (It's a bad behaviour/practice/custom/habit/ ... to bite your nails).

Table 1.18 below shows that results from the two sets of users do not differ greatly.

Table 1.18: Percentages of correct answers given by monolingual and bilingual dictionary users to question 3/13 of the Dictionary Research Test

	Grade A	Grade B	Grade C	Grade D	TOTAL
Bilingual users:	97.4%	77.2%	68.4%	46.5%	68.9%
Monolingual users:	94.3%	84.4%	62.2%	40.9%	76.4%

The results of question 3/13 have also been analysed to compare dictionary users with non-dictionary users across the Placement Test grades. Table 1.19 reproduces the figures given by Atkins and Knowles. It will be noted that the figures do not tally - according to the results given, 74.8% is the average grade for dictionary users, while 70.5% is the average grade for non-dictionary users. However, it is possible that some of the other figures in the table have been juxtaposed, as 95.8% is the percentage of correct results for dictionary users rather than non-dictionary users (see Table 1.18 above, where this figure can be arrived at by averaging the grade A scores of monolingual and bilingual dictionary users).

Table 1.19: Percentages of correct answers given by dictionary users and non-dictionary users to question 3/13 of the Dictionary Research Test

	Grade A	Grade B	Grade C	Grade D	TOTAL
Dictionary users:	100%	83.3%	65.5%	50.5%	64.0%
Non-dictionary users:	95.8%	77.6%	67.9%	40.7%	73.1%

Atkins and Knowles also report results for question 5/21 (I haven't much faith ... with/by/on/in... what he says) where dictionary use has been compared to non-dictionary use in a similar way. For this question "approximately 70% of dictionary

users got the answer correct, as opposed to only 60% of those who did the question without a dictionary".

It is unclear whether the non-dictionary users whose results are recorded here were part of the control group who were not allowed to use dictionaries, or simply subjects who chose not to refer to their dictionary for this particular question.

Discussion

Atkins and Knowles admit that "the students tested did not constitute a statistically balanced sample as regards grouping on grounds of nationality, native language, level of English studies and type of academic institution attended". The sample was unbalanced when the 1600 sets of papers were first distributed, but the uneven return rate skewed the groupings still further. It should be noted that, while 1140 questionnaires were returned, the researchers received only 723 complete sets of papers.

As must inevitably be the case when many different questions are addressed in the same piece of research, the available dataset shrinks still further as the specificity of the information required increases. For example, in order to discover more about user preferences it was necessary to allow respondents the choice between using a bilingual dictionary, a monolingual dictionary, or no dictionary at all when answering questions in the Dictionary Research Test. However, as only 25% of the 723 respondents chose to use a monolingual dictionary, and of these some chose not to use their dictionary for some of the questions where dictionary use was allowed, an analysis of answers to any one question from the Dictionary Research Test can be taken from a dataset of up to 542 bilingual dictionary users, but a maximum of only 180 monolingual dictionary users (perhaps far less).

Atkins and Knowles propose further cross-tabulation, by asking, for example, "how many French-speaking Grade A students using a dictionary got the wrong answer compared with those who used no dictionary?", and "how did the dictionary-using students who had had instruction in dictionary use fare compared with those who had had no instruction?". However, although this research is intended to provide the possibility of cross-tabulating in a wide variety of ways, the size of the dataset for each enquiry must be borne in mind. It is possible, for example, that the dataset for French-speaking grade A students using dictionaries for a particular question is too small for any meaningful generalization to be made.

Atkins and Knowles do acknowledge that trends and tendencies indicated by the results of their research will need to be explored in studies with a narrower focus. Indeed, one of the avowed purposes of this project is to spark off ideas for future enquiry.

I will now comment on the design of the Dictionary User Profile Form and the Dictionary Research Test.

The Dictionary User Profile Form

The EURALEX project is the largest piece of research examined in this chapter; because of its size, it would have been costly and time-consuming to obtain the required information by observing or interviewing subjects, thus the questionnaire was an obvious choice as a means of data collection. On the whole, the questions on the Dictionary User Profile Form were more objective and factual than those asked by Quirk, Tomaszczyk, Béjoint and Battenburg; they did not require subjects to admit their own failings, and they did not require understanding of linguistic metalanguage. This emphasis on factual, easy-to-answer questions probably improved the questionnaire's reliability. Furthermore, no questions necessitated detailed recall, although questions 12, 13 and 14 did ask subjects to recall the frequency with which

they used different types of dictionary, and question 16 asked for information about dictionary use for different language activities. The inconsistent answers to question 16, which unexpectedly suggest that experienced users make greater use of bilingual dictionaries for information on the usage of known L2 terms, may reflect the general unreliability of answers to questions which ask respondents to comment on how they would behave in certain imagined situations.

Reliability is of course a major concern in questionnaire based research. For the researchers in the first section of this chapter (Barnhart, Quirk, Jackson, Tomaszczyk, Baxter, Béjoint and Battenburg) reliability was virtually the only concern; all information on their respondents' use of dictionaries was of value, as they had no specific hypothesis about dictionary use to test, and the questionnaire results were not intended to interact with results obtained by other means. In the case of Bensoussan, Sim and Weiss (1984) and the EURALEX project, however, the questionnaires were one part of more complex pieces of research, and therefore the choice of questions to ask was also very important. Bensoussan, Sim and Weiss were testing a hypothesis about the effect of dictionary use in language tests. To be useful, their questions had either to test this hypothesis directly, or provide information which would shed light on results obtained by means of the tests. (As I explained earlier, their questionnaire was not entirely successful in this respect.) In the EURALEX project, data has been amassed with only very general aims in view, but with the intention that the cross-tabulation of results obtained by different means will provide answers to more specific questions. To be useful, questions on the Dictionary User Profile Form should therefore elicit information which is important in itself, and/or information which is useful when cross-tabulated with the results of one or both of the tests. In fact, most of the questions do meet one or both of these requirements, but there are a few questions which do not seem entirely justified.

For example, I consider that question 4 - reasons for studying English - strays beyond the scope of the project as a whole. It is unclear whether the researchers wish to discover respondents' perceptions about the purpose of learning English, or whether they intend to conduct a miniature needs analysis. In either case, given that the respondents were chosen as representative of language learners in different types of school and at different levels of study throughout Europe, it is unlikely that any clear picture will emerge, and it is difficult to see how results could be usefully cross-tabulated with results from the tests. Such a question would be more useful if this survey focussed on a specific group, whose training might be modified in the light of findings (as was the case in the surveys conducted by Béjoint and Baxter).

The justification for the inclusion of question 6, which requires learners to name the textbooks they use, is also unclear. It may be interesting in itself to know which sorts of textbook are used among which sorts of learners, and in this case more complete information could be obtained from other sources such as education departments and publishers. However, such information is only of value in a survey of dictionary use if it gives insight into learners' dictionary needs and wants, perhaps after cross-tabulation with scores on the placement test, and/or results from the Dictionary Research Test. Textbooks will vary from one type of school to another, and one country to another, and it may be difficult to establish the role textbook use has to play in the emergence of trends in the data. It would be interesting, however, to use the data to identify patterns of dictionary use amongst those who used textbooks with glossaries. It might be possible to explore the hypothesis that the use of textbook glossaries inhibits the development of dictionary skills.

The purpose of question 5 - whether subjects were taught English through the medium of English - is also not entirely clear. The question looks as if it were designed to test the hypothesis that subjects taught through the medium of English would be more successful language learners, but as the project was primarily designed as a

hypothesis-forming rather than hypothesis-testing tool, the variables have not been sufficiently controlled for it to be possible to establish via cross-tabulation clear causative links between teaching method and learner success. In any case, the question seems once again to lie slightly beyond the scope of a survey into dictionary use.

Other questions relate directly to dictionary use and provide useful background information for the interpretation of test results. The data must be interpreted carefully, however, to avoid jumping to unfounded conclusions. For example, answers to question 11 suggest that price is not a factor when choosing a dictionary, but Atkins and Knowles point out that this is probably due to the fact that the majority of the respondents were schoolchildren, who did not yet buy books for themselves. Price may be an influential factor, but for the parents and teachers of respondents rather than for the respondents themselves.

Of all the questions on the questionnaire, question 16, discussed earlier, makes the most radical attempt to explore the reality of dictionary use, but yields inconsistent results for the following reasons:

- 1) for the sake of clarity, the three language activities are only briefly delineated, which means that different respondents may have quite different situations in mind (for example, an enquiry concerning usage in the third situation might be an enquiry about grammar, collocation, pragmatics or simply spelling)
- 2) dictionaries are only divided into two types - monolingual and bilingual, with no further distinction into L1-L2 and L2-L1 (and no opportunity to name a specific dictionary for the job)
- 3) the question simplifies the choices a learner really has to make when completing a language task. Dictionary choice presumably depends not only on the nature of the task, but also on the level of difficulty of the word and the text as a whole. It may also depend on further factors such as the time available for look-up.

In order to establish users' choice of dictionary for particular tasks, it is probably necessary to present them with these tasks, and monitor their behaviour. This is what occurs in the Dictionary Research Test.

The Dictionary Research Test

A number of administrative problems spring to mind when examining the rubric of the dictionary research test. Testees are asked to use the dictionary they normally use; this suggests that they were not supplied with dictionaries on arrival in the examination room. In studies by Bensoussan, Sim and Weiss (1984) some of the subjects who would have liked to use dictionaries had to be classed as non-users because they forgot to bring their own copies; in my own experience it is common for students to refuse to bring a dictionary to lessons because it is "too heavy", but if students arrived without dictionaries for this test, there would be no point in testing them. The answer paper I have been given as a sample has been filled in by a candidate who never once consulted a dictionary, although a bilingual dictionary is named on page one. Clearly we need data on candidates who actually used dictionaries, as opposed to those who merely named them.

There is a further problem for the test administrators in the restrictions on dictionary use during the test. Some questions allow dictionary use, others not, but it would be difficult for invigilators to enforce this rule.

In sections one and two of the test candidates are not allowed to use their dictionaries. These sections do not, therefore, provide data on dictionary using habits, but only on basic skills needed to interpret dictionary information. Scores were high for both sections, particularly for section 2, where 96% of respondents seemed able to interpret four abbreviations found in dictionaries. It is worth noting, however, that not only were there very few questions in section 2, but also the abbreviations that were tested

are not those that needed to be understood in order to answer later questions in the test. Grammatical questions in sections 5 to 9 are concerned with the use of prepositions and valency patterns, and the grammatical coding systems which give information on these features are often very difficult for users to interpret. It would of course be impossible in this test to assess users' understanding of this type of information directly; dictionaries vary too much in their coding systems, and in any case the more elaborate codes are supposed to be supported by explanations in the front or the back of the book. Nevertheless, the fact that candidates understand four simple and transparent dictionary abbreviations is no guarantee that they can handle more complex dictionary codes successfully.

Subsequent sections, with the exception of section 4, permit dictionary use and require knowledge of word meaning or word behaviour in context. We are not told whether items were chosen intuitively, or from a corpus of learner errors, but some of the multiple choice distractors do look like typical non-native speaker mistakes. I assume that the answers to these questions will be most interesting in cases where candidates answered wrongly; their choice whether or not to use a dictionary will be an important consideration in the analysis of these answers. Right answers will be less interesting; correct answers without dictionary consultation will be of little value, those who answer correctly with a dictionary may or may not have used a dictionary entry to inform their choice. If Atkins and Knowles intended to compare testees' dictionary use with the success of their answers, it is a pity they did not request more information regarding the testees' strategies. For some questions, there are a number of words that the testee might wish to look up, but the test only requires an indication of whether a dictionary was used or not. Testees do not have the opportunity to list the words they looked up, or how satisfied they were with the information in their dictionary.

It seems to me that the real interest in question 4 is not whether testees can anticipate the organisational patterns of dictionaries, but rather whether certain patterns of

expectation emerge, which might inform the organisational policy of dictionaries in the future. The results so far analysed suggest that dictionary users often do not know exactly where a particular word or expression will be listed, but these results do not appear to have been analysed according to the dictionary each respondent habitually uses - users could be much worse at locating entries than it appears, if those who chose the location favoured by most dictionaries in fact used a dictionary with a different pattern, and alternatively they could be better, if every user who chose a location which few dictionaries favoured was in fact a user of one of those few dictionaries!

It is a pity that users do not get the option to select more than one location for each question in this section, as many dictionaries list multiple word entries more than once.

One further general comment about the Dictionary Research Test is that the dictionary-using situations supplied in the test do not reflect ordinary, unprompted dictionary use. This is particularly true of encoding questions, where the testee is supplied with a choice of possible words; normally the encoder must draw on his own vocabulary store. Testees are also only allowed to access one dictionary, although this may not reflect their ordinary behaviour: results from the questionnaire indicate that the majority of dictionary users tend to use bilingual dictionaries for decoding and translation, and monolingual dictionaries for encoding when the L2 word is already partially known.

The criticism that the EURALEX Dictionary Research Test does not examine natural processes of dictionary look-up is taken up by Tono (1988), when he compares the EURALEX test with one conducted by the Okayama Prefecture in Japan in 1985. The description of the Okayama testing project is the first of two papers by Tono which investigate Japanese students' dictionary-using skills, and it is the next paper I will discuss.

1.2.4. Tono (1988)

In this paper Tono describes studies involving a test devised by Okayama Prefecture Senior High School Circle of English Education. The aim of the studies was to examine how effectively High School students could use English-Japanese dictionaries for recognition purposes.

Tono compares the Okayama Dictionary Using Skills Test to the EURALEX test devised by Atkins et al. and finds it has a wider coverage; his summary of the skills and language elements tested in the two tests is given in Table 1.20 below:

Table 1.20: A summary of skills and language elements tested by the EURALEX and Okayama tests

	EURALEX	Okayama
Alphabetisation		x
Finding the word on the page		x
Reference speed		x
Sound system		x
Stress system		x
Parts of speech	x	x
Grammatical terms	x	x
Polysemy	x	x
Vocabulary selection	x	x
Word forms, inflections		x
Count v non count nouns		x
Derivative forms		x
Synonyms and antonyms		x
L2-L1 translation		x
L1-L2 translation		x
Compounds, phrasal verbs etc	x	

The Okayama test was divided into seven sections:

- 1) the sound system
- 2) alphabetisation
- 3) a) parts of speech, b) and c) analogy of meanings
- 4) reference speed (subjects were asked to look up as many as 12 words to choose the right definitions within three minutes. None of them looked up all the words within the time limit)
- 5) the stress system
- 6) a) parts of speech labels b) inflections c) count/non-count nouns d) derivatives e) synonyms/antonyms.
- 7) a) usage b) social/cultural background.

Instructions for the test were spoken and there was a strict time limit for each item, the entire test lasting exactly 50 minutes. In the case of the first study the instructions had been recorded on audio tape, but for the second study they were read aloud. Tono tells us that "questions 1 to 3 had to be answered without a dictionary, and the rest (4 to 7) with a dictionary". No dictionaries are specified, but Tono notes that "the Okayama test was made in consideration of several popular bilingual learner's dictionaries".

Tono reports that the test was first administered by the Okayama group in 1984. 1,055 1st year senior high school students took part, and English language proficiency varied considerably in the twelve different schools where the test was administered. About half the subjects went to high schools where standards of English were quite high, and the rest went to commercial, technical and vocational schools where the standard was low.

Because the Okayama group conducted this study, rather than Tono himself, he only obtained access to the mean scores rather than the raw scores for each individual. He therefore administered the tests again to 76 third grade students in Tokyo Metropolitan College of Aeronautical Engineering. These students were older than

the high school subjects, but according to Tono their English language proficiency was "mediocre",

Discussion

Tono reports that the original Okoyama study was set up as an alternative to questionnaires regarding dictionary use because:

even though in the questionnaires the students answered yes to the question of whether they could use dictionaries well, most of the high school teachers felt that the students had difficulty finding words and appropriate meanings in dictionaries.

(1988:109)

However the test results can neither prove or disprove the teachers' belief about the students' abilities. The percentage of correct answers for each item in the Dictionary Using Skills Test is reported, but the scores alone are relatively meaningless, as we have no means of knowing the percentage-level at which efficient dictionary skills begin to be indicated.

In a later paper (Tono 1989) Tono reports on subsequent research which found a positive correlation between the test results reported here and results in English proficiency tests. As Tono rightly points out, "this does not indicate automatically that a positive relationship exists between dictionary reference skills and reading comprehension itself" (1989:193). A correlation between the two scores tells us that good readers also tend to be skilful dictionary users, but it does not tell us whether ability in one of these two areas aids progress in the other. The correlation might be merely indicative of overall intelligence and motivation, and similar correlations might be obtainable between scores in unrelated areas, such as English proficiency and Mathematics.

Analysis of the raw scores of the test administered in Tokyo reveals that there is low correlation between each test item in the Dictionary Using Skills Test. Tono feels that

54

this is counter-intuitive; he had expected the ability to guess a word meaning or its part of speech to positively correlate with other dictionary-using skills. As Tono points out, the results could be interpreted as indicating that dictionary skills can be acquired independently of each other, or alternatively the unexpected results might be put down to faults in the test design.

Tono identifies a number of areas where the test itself might be at fault; he finds the time limitation unnatural, and complains that variations in the conventions of different bilingual dictionaries were not taken into consideration, but most of all he criticises both the Okayama test and the AILA/EURALEX test for failing to assess the "whole process of dictionary look-up". Certain skills such as locating the words you need to look up, and choosing from a number of definitions the one most appropriate to a given context, are difficult to assess in a discrete point test. Tono concludes by arguing in favour of an essay-type test for the assessment of reference skills, although he does not give details of the format he envisages for such a test. He also suggests that future tests should distinguish more clearly between dictionary use for comprehension and dictionary use for productive purposes, as the two purposes actually require different skills.

1.2.5. Tono (1989)

Tono's 1988 paper reported on the results of a paper devised by the Okayama Prefecture Senior High School Circle. In his 1989 paper the test he describes is of his own devising, and specifically concerned with receptive dictionary use. This study had three objectives:

- 1) to determine if there is any significant difference in the subject's performance between reading with the help of a dictionary and that without its help
- 2) to identify what kinds of reference skills are most relevant to better performance in reading comprehension tasks
- 3) to identify possible measures of dictionary reference skills.

32 subjects took part in the study; 17 in their first year and 15 in their second year of junior high school. All the subjects had received special training in dictionary skills. Tono's experiment required the subjects to take two series of tests, the first to assess their ability to use a dictionary, the second to assess their reading comprehension.

A Dictionary Reference Skills Test Battery (DRSTB) was developed specially to test the subjects' dictionary skills, and consisted of four subtests in six categories:

- 1) Placing words in alphabetical lists: i) one-word entries; ii) two-word entries (Subjects were to locate words in alphabetically ordered word lists.)
- 2) Finding meanings: iii) one-word entries; iv) two-word entries (Subjects were to look up words in a dictionary and find meanings.)
- 3) Finding the most appropriate meanings: v) polysemes (Subjects were to identify the correct meaning among those under the same entry, on the basis of contextual information.)
- 4) Finding idioms: vi) idiomatic phrases (Translation into Japanese of English sentences containing idioms.)

The test battery contained two levels of tests, the more difficult tests being administered to the second year group, who were also allotted slightly more time to complete them (32 minutes as opposed to 26 minutes). Samples of the DRSTB questions are included in an appendix to Tono's paper.

The reading comprehension tests consisted of two passages, each followed by 10 multiple choice questions. Subjects were allowed 20 minutes to read and answer questions on each passage, but had no access to a dictionary in the first test (RC1), whereas they could use their dictionaries in the second test (RC2). All dictionaries used were bilingual English-Japanese; first year subjects used Eiwa Kihongo Sho-Jiten (Basic English-Japanese Dictionary), Kodansha, 1981, while second year subjects used Global English-Japanese Dictionary, Sanseido, 1983.

Results

The results of these tests showed that subjects performed better in reading comprehension with dictionaries than without. They made "a significantly higher

proportion of errors when they did not use dictionaries than when they did use them". Tono suggests that the "primary reason" for the difference between these results and those of Bensoussan et al is that his subjects had received special training in dictionary use. (According to Bensoussan et al their subjects were supposed to have received some dictionary training too, although their training does not appear to have been specific to one particular dictionary, and questionnaire responses suggested that many subjects still experienced difficulty using their dictionaries.)

As far as the relationship between dictionary skills and reading comprehension was concerned, results were less conclusive; amongst the first year subjects there was a good correlation between DRSTB scores and achievement in RC2 (where dictionaries were used), but amongst second year subjects DRSTB scores correlated better with RC1 (where dictionaries were not used). Tono admits that this is "puzzling", but argues that continuous dictionary use may actually be the reason why subjects obtained higher reading comprehension scores when they did not have access to a dictionary:

at the beginning level, those who are good at handling dictionaries perform better in reading comprehension with dictionaries than those who are not, while neither group was any better at reading comprehension without a dictionary. On the other hand, as they use dictionaries more and more, those who can use them effectively can perform better in reading comprehension without dictionaries.

(1989:197)

Discussion

Tono admits that his dictionary skills test battery needs to be "re-examined for validity and reliability", but he does not identify any specific problems in the tests. I suspect that marking the DRSTB might have been problematic, as the questions are somewhat open-ended. Subjects are told that they must find meanings "with a dictionary", but there is no indication that they should quote directly from the definitions in their answers, and subjects who choose to express meaning in their own

words may lose marks not because they have not understood a definition, but because they cannot express their ideas clearly. In section D subjects are required to translate sentences containing English idioms; clearly this is a test of something more than merely dictionary skills. It is fairly easy to guess the meanings of the idioms in the sample questions from context (eg "her shrieking voice gets on my nerves"), so a correct answer would not necessarily be proof of efficient dictionary use.

Tono does not mention any control over whether subjects actually used their dictionaries in either DRSTB or RC2. In RC2 subjects were permitted to use dictionaries, but, as the study by Bensoussan et al shows, subjects do not necessarily use dictionaries when they are made available to them.

Without evidence to the contrary, it remains a possibility that subjects gained higher scores on RC2 because they found it an easier test, rather than because they used their dictionaries. This possibility would have been eliminated if Tono had allowed half his subjects to use dictionaries with RC1 rather than with RC2. Instead, it was always the same second test that was taken with access to dictionaries, and there was no control for variation between the two tests, and for variation as a result of taking the test second, having had 20 minutes to "warm up".

Tono's suggestion that improved dictionary skills enable subjects to read more efficiently when they are not using their dictionaries seems rather implausible. One researcher on the Okayama dictionary-using skills test project (cf Tono 1988) found a correlation between dictionary skills and language proficiency, but no causal relationship was established. In this study too we might expect some correlation between dictionary skills scores and reading comprehension scores, whether or not a dictionary was used, but success for efficient dictionary users in RC1 rather than in RC2 is most likely due to an inappropriate test level. Second year subjects may have found the reading comprehension passages so easy that they did not necessitate

dictionary use at all. In this case, easier questions in RC2 would account for the fact that second years achieved a higher score in the second test. (The first year subjects also achieved better scores in the second test, but may have found both passages more difficult and may have used their dictionaries in RC2, which would account for the better correlation between DRSTB and RC2 scores amongst the first year subjects.)

Tono's RC2 was designed to be taken with the aid of a dictionary, and since both reading comprehension tests were written specifically for this experiment it seems likely that they included questions answerable with the aid of the dictionaries the subjects were going to be allotted. Tono does not give examples of his reading comprehension test questions, neither does he discuss whether the questions hinged on understanding of individual words. He also does not appear to have monitored the words his subjects looked up, or their coverage in the two English-Japanese dictionaries. For this reason we lack the information we need to fully account for Tono's findings.

1.2.6. Bogaards (1991b)

Bogaards begins by summarising current opinion for and against the use of monolingual dictionaries by language learners, while drawing attention to the fact that extremely little experimental research has been carried out to substantiate rival claims. The pro-monolingual lobby argues that monolingual dictionaries give more complete information about word formation and grammar, and are not constrained by the near-impossible task of accurately translating single words. Opposing writers point to the following weaknesses in monolingual dictionaries:

- 1) they are not written with a particular first language in mind, and therefore do not address the problem of "false friends"
- 2) they do not compare and contrast words and concepts in the foreign language and the mother tongue
- 3) they have to provide lengthy and often clumsy definitions where a translation equivalent would be simpler and more direct

4) learners cannot look up words in a monolingual dictionary that they do not already know.

In view of the claims outlined above, Bogaards raises the following research question:

Le dictionnaire monolingue, et plus spécifiquement le monolingue destiné aux apprenants d'une langue étrangère, est-il plus efficace que le bilingue?

1991:94

He sets out to investigate the usefulness of a bilingual dictionary and two types of monolingual dictionary for the language student, in terms of both task performance and vocabulary retention.

The subjects used in this study were Dutch-speaking first-year university students of French. 44 took part in the first stage of the experiment, and 55 in the second stage.

In the first stage of the experiment, the subjects were given 45 minutes to translate a 150-word Dutch text into French. The text was grammatically simple, but contained 17 words which were judged to be difficult because they were not in the productive vocabulary of the subjects.

The subjects were divided into four groups to work under the following conditions:

group one (12) a bilingual dictionary (not named),
group two (10) Dictionnaire du Français Langue Étrangère Larousse (DFLE).
group three (12) Petit Robert.
group four(10) no dictionary.

There was no significant difference in terms of language proficiency between the four groups, although group two had a slightly higher grade average.

Groups 1 to 3 were asked to underline those words in the Dutch text which they had looked up in the dictionary.

The second stage of the experiment took place fifteen days later, when, without advance warning, subjects were asked to translate into French the 17 difficult words from the translation passage. On this occasion some of the original subjects were

missing. Also a fifth group was formed, consisting of 14 students who had not undertaken the original translation task.

The subjects' translations and test papers were analysed to establish:

- 1) the percentage of words looked up (groups 1 to 3)
- 2) the percentage of words correctly translated after dictionary consultation (groups 1 to 3) and without dictionary consultation (groups 1 to 4)
- 3) the percentage of words correctly translated 15 days later (groups 1 to 5).

Bogaards decided not to subject his data to statistical analysis to test for significant differences between groups, judging the number of subjects to be too small. He therefore recommends caution in interpreting the following findings.

1) Users of the Petit Robert chose to look up only 34.8% of the words (71 out of a possible 204 look-ups), users of DFLE chose to look up slightly more words - 44.7% (76 out of a possible 170 look-ups), and users of bilingual dictionaries chose to look up considerably more - 70.6% (144 out of a possible 204 look-ups)

2) Users of the Petit Robert also made the fewest correct translations after dictionary consultation - 14.2%. DFLE users made slightly more correct translations - 21.2%, and bilingual dictionary users made considerably more - 60.3%. Overall, the most successful translations were produced by bilingual dictionary users, and the least successful by subjects in group 4, who did not have access to any kind of dictionary.

3) In the vocabulary translation test 15 days later the success of DFLE and the bilingual dictionary was reversed. DFLE users made the most correct translations - 51.6%, and the bilingual dictionary users came second - 48.5%, followed by the users of Petit Robert - 44.7%. Group 4 subjects, who had not had access to dictionaries for

the translation, fared less well - 41.8%, and group 5 subjects, who had not carried out the translation task, fared worst - 29.4%.

Discussion

As Bogaards suggests, these results indicate that the use of any kind of dictionary leads to better results in translation and vocabulary learning. This is consistent with the findings of Luppescu and Day (1993), who recorded vocabulary test scores for (bilingual) dictionary users 50% higher, on average, than scores for subjects who had not used dictionaries. It is also consistent with the findings of Black (1986), who found that, in a multiple choice comprehension test, there was a highly significant difference between scores for words which had been defined and scores for words which had not been defined. Bogaards' findings are particularly valuable because, whereas Luppescu and Day (1993) and Black tested their subjects' vocabulary knowledge immediately after dictionary consultation, Bogaards waited fifteen days to establish whether subjects had really learnt the words they had looked up.

Bogaards' findings also tally with those of Tono (1989), who found that subjects performed better in reading comprehension with (bilingual) dictionaries than without, but they do not accord with the findings of Bensoussan, Sim and Weiss (1984), who found that dictionary use did not effect reading comprehension test scores, regardless of dictionary type. These last two experiments were not, of course, strictly comparable to Bogaards' because they tested reading comprehension rather than the ability to translate and remember lexical items.

Bogaards' findings suggest that bilingual dictionary use is more effective than monolingual dictionary use for translation tasks, but does not result in the most effective vocabulary learning. Bogaards discusses the possibility that bilingual dictionary look-up requires less thought and concentration than monolingual dictionary look-up, and results in lower retention because the subjects do not need to make the same effort to interpret entries:



*les étudiants qui travaillent avec le bilingue ont
tendance à y vérifier même les mots qu'ils connaissent.
Ils ne semblent guère prendre le temps de réfléchir....
Les étudiants qui ont utilisé un dictionnaire bilingue
semblent avoir noté les bonnes traductions sans
s'intéresser aux problèmes posés.*

(1991:100)

DFLE users actually scored a higher percentage of correct answers in the second test than they did in the first. The 17 test words were not taught in class, so Bogaards hypothesises that these subjects had found the correct translations by themselves, because the puzzling monolingual dictionary entries motivated them to continue searching for the correct answers.

This explanation seems plausible, but it does not explain why users of the Petit Robert, who presumably had the hardest task interpreting dictionary entries, retained fewer words than the unreflecting bilingual dictionary users. The Petit Robert group retained most of the words that they had translated correctly in the first test, but do not appear to have discovered and learnt new translations in the period between the two tests. Bogaards does not discuss the differences between the styles of the two monolingual dictionaries in any detail, but one explanation for the disparity in results between groups 2 and 3 could be because DFLE is written in a livelier and more thought-provoking style.

In Bogaards' experiment, as in Bensoussan, Sim and Weiss' studies (1984), bilingual dictionary users looked up many more words than monolingual dictionary users. Some words were looked up with almost equal frequency in both types of dictionary, but other words were often looked up in bilingual dictionaries and only rarely looked up in monolingual dictionaries. Given that all the words under consideration had entries in all the dictionaries used, and the subjects in all three groups had similar proficiency levels and language backgrounds, Bogaards seeks an explanation for this behaviour in the words themselves. He argues that words belonging to a known lexical field, such as *jek* (*blouson*), *verband* (*bandages*), and *waakhond* (*chien de garde*), were often looked up in monolingual dictionaries because it was easy for

subjects to find an entry point for their search - for example a search for the translation equivalent for *jek* could begin with *vêtement* or *manteau*, while the common words *malade*, *médecin* or *blessure* might be a starting point on the way to the word *bandages*. No such starting point would present itself for abstract expressions such as *kennelijk* (*manifestement*), and *achterover* (*à la renverse*), and this seems to explain why such words were less frequently looked up in monolingual dictionaries than in bilingual dictionaries.

This theory of dictionary-using behaviour clearly relates to the criticism of monolingual dictionaries summarised earlier; learners cannot look up a word in a conventional monolingual dictionary unless they know of it already. The data does not provide any means of proving the theory, however, and it must be borne in mind that the numbers of look-ups for each word in each condition were too small to permit tests for significance. It is also possible that subjects in group 1 were working with their own bilingual dictionaries, and were therefore slightly advantaged with respect to the other groups (presumably subjects are more likely to make use of a familiar dictionary than an unfamiliar one). No titles are given for the bilingual dictionaries, which suggests that subjects were using a variety of different titles brought in by themselves, rather than a class set.

In the first test (translation from Dutch into French), Bogaards records a greater number of unsuccessful look-ups with the Petit Robert (42 out of 71) than with DFLE (40 out of 76), and he suggests that look up may have been more successful in cases where the unknown word formed part of a lexical set which was listed in the entry for a commoner word within the same field. For example, four out of five subjects successfully found *cuisse*, and four out of seven successfully found *mollet*, in DFLE, where the names for the component parts of the leg are listed systematically at the entry for *jambe*. On the other hand, only two out of five found *cuisse* in the Petit Robert, and only one out of six found *mollet*, because under *jambe* the word *cuisse* is given but not explained, and the word *mollet* does not appear.

Bogaards does not examine other relevant dictionary entries in this paper, but his findings support the view that DFLE is a more useful dictionary than the Petit Robert for foreign learners of French. However, the difference in success rate between the two dictionaries in either test is not great, and the data does not suggest that DFLE is vastly preferable to the Petit Robert. Bogaards does not discuss the fact that, although users of the Petit Robert made a greater number of mistakes after dictionary consultation, as an overall percentage more unsuccessful look-ups occurred with DFLE (23.5% as opposed to 20.6%). Moreover, although *cuisse* and *mollet*, and certain other words, seem to have been more easily traceable via DFLE, there are some words which are dealt with more successfully by the Petit Robert. For example, only one out of the four DFLE users translated *bandages* correctly as opposed to four out of six users of Petit Robert, and all four DFLE users got *faire demi tour* wrong while two out of four Petit Robert users got it right.

In conclusion, although the number of subjects was quite small, Bogaards' study provides strong evidence that dictionary use helps the language learner in translation tasks, and considerable evidence that bilingual dictionaries are more helpful than monolingual dictionaries for such tasks. The study also provides a certain amount of evidence that monolingual dictionary users remember the words they have looked up better than bilingual dictionary users. The evidence that monolingual learners' dictionaries are more useful for language learners than monolingual native-speaker dictionaries is less conclusive, because only two monolingual dictionaries were examined, and the differences between the results for the two user groups was not great.

1.2.7. Luppescu and Day (1993)

Like Bensoussan et al (1984) and Tono (1988), the authors of this paper set out to compare the test scores of subjects who had been given access to dictionaries with the

test scores of a control group. However, although the subjects were required to read a short passage before they were tested, the test was designed to assess their vocabulary knowledge rather than their overall reading comprehension.

Two hypotheses were tested in the experiment:

- 1) that there would be no significant difference in the measure of vocabulary learnt by bilingual dictionary users and the measure of vocabulary learnt by those who did not use a dictionary
- 2) that dictionary users would take significantly longer than non-users to read a text.

The subjects were 293 first and second year Japanese University students studying English.

In the first phase of the experiment the subjects were required to read a 1,853-word story. This story was judged to be at an appropriate level, but it contained at least seventeen words previously identified as being unknown or difficult for college-level Japanese EFL students. 145 subjects were randomly assigned to the experimental group and were allowed to use their bilingual dictionaries while reading. The remaining 148 were assigned to the control group and were not allowed to use their dictionaries.

In the second phase of the experiment all subjects were tested on their knowledge of vocabulary occurring in the story. The original test had 27 items, but only the responses for 17 items identified as "target words" are discussed in this paper, and from these 17 items a further two were ultimately discounted when item analysis indicated that subjects' responses were not representative of their overall performance. The test had a multiple choice format with three distractors and a "don't know" option. Access to dictionaries was not permitted during the test.

Test scores were arrived at by assigning two marks for each correct answer, one mark for each "don't know" answer, and no marks for incorrect answers. The number of "don't know" responses proved to be small, although significantly smaller for the experimental group than for the control group, and in the reporting of results only correct and incorrect answers were considered.

The mean score of subjects in the group which had access to dictionaries was found to be significantly higher than that of the control group [$p = <.001$]. This result appears to disprove the authors' first hypothesis that there would be no significant difference in scores between the two groups.

However, results were not uniform across the full range of test items; while certain words (SOB, CHANT, STARE and FAINT) proved more difficult for non-dictionary users, others (APPEAR, SCARE, HAPPEN, TERRIBLE, STRANGE) were more often misunderstood or not known by subjects who had been given access to dictionaries.

It was found that the experimental group took on average almost twice as long to read the passage. However there was almost zero correlation between time taken to read the passage and performance in the test.

Discussion

There are two defects in this experiment which may affect the validity of the results; the failure to monitor dictionary use, and the faulty design of the vocabulary test.

Lupescu and Day did not monitor either the amount of use made of dictionaries by the experimental group, or which words they looked up. They explain that their study "was not able to consider how, or even *if* dictionaries were actually used by the students". The whole experiment rests on the assumption that members of the

experimental group did use their dictionaries, and that they looked up the words that were to be featured in the subsequent test. The story which was used in this experiment had, however, been edited "to provide opportunities for the [difficult] words to occur with ample frequencies and in sufficient contexts to allow the subjects to make reasonable guesses about their meanings." It therefore seems likely that some of the experimental group would have chosen to guess unknown words from context. Given the likelihood that some subjects in the experimental group were not actually dictionary users, we probably need data for three subgroups of subjects rather than just two: dictionary users, those who were permitted to use a dictionary and chose not to, and the control group.

Of course even amongst dictionary users there may be great differences in the degree of dictionary use. The researchers assume that members of the experimental group looked up an identifiable set of items:

Given the work which established that a comparable pilot group of participants in general had difficulty recognizing or did not know the target words, we also must infer that when the students in the dictionary condition used their bilingual dictionaries, they were most likely looking up the meanings of the target words.

1993:269

This may seem a reasonable assumption, but it remains a possibility that dictionary users either already knew, or guessed from context the identified target words, and looked up other words instead.

The experimental results indicate that certain words actually proved more difficult for subjects in the dictionary-user group than for subjects in the control group. Luppescu and Day hypothesise that dictionary users were confused by the multiple entries for polysemous words:

perhaps the use of a dictionary in some cases may be misleading or confusing, that is, if a student is not able

to find the appropriate meaning in the dictionary from among all the possible meanings listed.

1993:273

However we have no means of knowing whether the subjects actually looked up these multiple-entry words. They are amongst the commonest words in the test; in the Thorndike and Lorge frequency lists APPEAR, HAPPEN, and STRANGE occur 100 or more times (only two other test words, CLEAR and FIRE, occur as frequently as this), TERRIBLE occurs more than 50 times, and SCARE occurs 37 times per million words. We are not given details of the subjects' language level, but these words certainly fall within the minimum basic vocabulary that a university student with "six years of junior high school and high school English study" might be expected to know. The wide range of use of these words is reflected in the number of entries assigned to them in a typical English-Japanese dictionary, but it should be noted that they are not the only words in the test to have multiple meanings; the adjective CLEAR, for example, has ten entries in the Longman Dictionary of Contemporary English, whilst APPEAR has only five entries, and HAPPEN, STRANGE, TERRIBLE and SCARE only three. It therefore seems wisest not to place any interpretation on the fact that the experimental group had significantly greater difficulty than the control group in coping with these words.

The test was supposedly designed to test knowledge of words that were previously unknown to the subjects, that appeared in the reading passage, and were looked up by subjects in the experimental group. However Luppescu and Day do not establish whether these words were really previously unknown, neither do we know whether they were looked up. An equally important criticism of the experimental design is that the test itself was written in such a way that subjects with full understanding of the target words could still give incorrect answers. Although instructions for the test were given in Japanese, all multiple choice words were in English. Thus the subjects were in fact tested on their understanding of all of the words presented, not simply the seventeen test items. In cases where the target word was extremely common, the

answer and the distractors were sometimes much harder words; the choices of synonym for HAPPEN, for example, were TO PEEL, TO LAY, TO DEVELOP, and TO OCCUR. Naturally those in the experimental group did not look up these words during the first phase of the experiment, as the words were not present in the reading passage.

Two of the test items were discounted by the researchers because they yielded unrepresentative results, but I found two further items problematic: CLEAR, where both DIRTY and DULL would be suitable antonyms, and TERRIBLE, where both ATTRACTIVE and DELIGHTFUL would be suitable antonyms. This further suggests that the test may not have been an accurate tool to measure the number of target words that subjects actually knew.

Despite defects in the test and in the experimental procedure, which prevent us from gaining a very precise picture of the extent and value of dictionary use, it seems reasonable to reject Luppescu and Day's first hypothesis and accept their conclusion that dictionary use must have been responsible for the highly significant difference in scores between the two groups. However, we should question the validity of the scores as indicators of vocabulary learning. It is doubtful whether words can truly be said to have been learnt if there is no evidence of their long-term retention. The test appears to have been administered immediately after the reading of the passage, and the researchers themselves acknowledge that immediate effects of dictionary use on vocabulary knowledge may not remain.

The second hypothesis - that dictionary users would take longer than non-users to read a text - was confirmed by the experiment but never really open to question. It seems obvious that readers who set themselves to read not only a short story but also a number of dictionary definitions will take longer than readers who limit themselves to the story. The amount of extra time required is a matter of interest, but only if we

can establish exactly what extra material was read in that time. It would be interesting to know how long each look up process took, and whether the story-reading flow was seriously interrupted as a result of looking words up. Unfortunately, this experiment does not provide that sort of data.

1.2.8. Concluding comments on test-based research into dictionary use and dictionary requirements

In the studies reviewed in this section testing was chosen as the means of data collection. Test-based research can enable the researcher to prove or disprove hypotheses in a more conclusive fashion than is possible with questionnaire-based research, which usually generates rather than tests hypotheses. The researchers may also have chosen this type of research because they felt that the information it provided would be more reliable and objective than information obtained by simply questioning subjects.

Certainly, many of the findings reported in this section are counter-intuitive, and would probably surprise the majority of language learners who, according to the questionnaire findings reported in 1.1, regularly depend on dictionaries as learning aids.

Particularly surprising are the cases where dictionary use does not seem to improve reading test scores (Bensoussan, Sim and Weiss 1984), and where dictionary examples do not seem to help learners understand word meanings (Black 1986), but even where the major findings of the test-based studies match expectations regarding the value of dictionaries, some of the details of the findings are unexpected. For example although Bogaards (1991), Tono (1989) and Luppescu and Day (1993) all report the advantages of dictionary use, Luppescu and Day found that their subjects learnt some words more successfully when they did **not** have access to a dictionary, and Tono only observed a correlation between dictionary reference skills and test performance in the test where dictionaries were **not** used.

It would appear that test-based studies are particularly useful in cases where there is a discrepancy between observable behaviour and widely-held belief. Whereas subjects in questionnaire-based research report what they believe, subjects in test-based research can only do what is possible. However, the validity of test findings very much depends on correct test administration, and appropriate test design. Problems with both of these are evident in the studies reviewed in this section.

Problems associated with test-based research

In several of the tests reported in this section, dictionaries were allocated in a rather haphazard fashion; in some cases the dictionary-using group was self-selecting, and therefore different in kind from the control group, and in some cases researchers failed to take into account the type of dictionary used, and the subjects' familiarity, or lack of familiarity, with the dictionary.

In some studies, the mere fact that the dictionary had been allocated to a subject was taken as proof that the subject had been a dictionary user in the test. Some subjects may not have used their dictionaries, or may have used them only rarely. Clearly, if dictionary use is going to affect test results at all, quantity of use is an important factor.

In order to generalise from the findings of performance tests, the ways that dictionaries are used during the test should correspond in at least some respects to the way dictionaries are used in natural, non-test conditions. However, it is extremely difficult to create a reading test which actually poses the same questions that successful readers would need to ask themselves when reading "normally". If researchers adopt an existing test, as did Bensoussan, Sim and Weiss, the test may assess general language skills and strategies, rather than the comprehension of a particular text. If, on the other hand, researchers design their own reading test with a

study of dictionary use in mind, as did Tono (1989), there is a danger that the test items will be biased towards the type of information available in dictionary entries.

In those studies which set out to assess the effect of dictionary use on vocabulary learning (Black 1986, Bogaards 1991, and Luppescu and Day 1993) there may have been slightly fewer test design problems. However the multiple-choice format (used by Black, and Luppescu and Day) may have affected the validity of results by providing a further context for each word, and therefore facilitating contextual guessing as an alternative to dictionary use.

Finally, the marking of tests and the analysis of data seemed to pose problems in one or two of the studies discussed in this section. Tono's Dictionary Reference Skills Test Battery looks, from Tono's description, as if it would be difficult to mark objectively. On a much larger scale the EURALEX project (Atkins and Knowles 1990) has still to be analysed in full. The vastness and the diversity of the data, and the fact that so many research questions have been addressed simultaneously, have rendered the project unmanageable in its final stages.

It can be seen from this review that research which has used testing as its primary means of data collection has tended to present a slightly less positive picture of the effectiveness of learners' dictionaries than that reflected in the findings of questionnaire-based research. Test-based research, however, does little to explain the causes of any unsatisfactory results it records, because it tends to focus on the end-product of dictionary use, rather than the process by which results are achieved. It is possible that these causes may be investigated more effectively by means of observation-based research, because this type of research aims to report not only the results of the dictionary-using task, but also the attitudes and strategies of dictionary users. In the final section of this chapter I therefore set out to evaluate the contribution of observation as a method of collecting data concerning dictionary use.

1.3. Observation-based research

Seven studies are discussed in this section: Ard (1982), Hatherall (1984), Miller and Gildea (1985), Neubach and Cohen (1988), Ahmed (1989), MacFarquhar and Richards (1983), and Bogaards (1990). Essentially the authors of these papers have attempted to explore in greater depth the attitudes and behaviour patterns elicited retrospectively by many of the questionnaires discussed in 1.1. and 1.2.

As in questionnaire-based research, the studies are concerned with generating hypotheses, rather than testing them. However, whereas the data regarding user behaviour obtained by questionnaire may be suspect, because subjects misunderstand questions, fail to recall, or falsely claim to behave in ways that they perceive to be desirable, observation-based research avoids these problems by setting subjects observable tasks, and collecting data either during the task itself, or immediately following its completion.

Some, but not all of the papers in this section involve interviews and protocol analysis. The remainder rely for their data on the products of written tasks, but remain distinct from test-based research because the skills or language competence of the subjects is not at issue; their dictionary-using behaviour is of interest to the researchers, because it can provide insights to inform dictionary choice and dictionary design.

Because observation-based research focusses on the dictionary-using behaviour of the subjects, rather than their test performance, this type of research is also able to avoid some of the problems of test-based research discussed in 1.2.8. Using this research method it is possible to monitor more closely the type of dictionary subjects use for a given task, the degree to which they use it, and the time they take to look up words.

1.3.1. Ard (1982)

This paper has a four-part structure, and four related issues are discussed in it:

1. the attitudes of ESL students, teachers and methodologists towards bilingual dictionaries
2. the form and content of bilingual dictionaries
3. students' use of bilingual dictionaries
4. recommendations: a role for bilingual dictionaries in ESL writing.

The first part of the paper draws on personal experience and published sources in the field of English language teaching methodology to support the view that students generally have a positive attitude towards bilingual dictionaries, while teachers and methodologists dislike them. The second part of the paper considers the strengths and weaknesses of bilingual dictionaries. Ard proposes that a classification according to user needs should distinguish between dictionaries designed for speakers of the defined language and those designed for speakers of the defining language, and between dictionaries for productive and receptive use. He identifies the following shortcomings in bilingual dictionaries: their failure to explain the differences in meaning between words listed as equivalents, their failure to indicate the frequency, collocations and connotations of words, and their failure to include in their entries a choice of words with opposing meanings.

The details of Ard's study of bilingual dictionary use are given in the third part of the paper. The study aimed "to determine how and how successfully students actually use bilingual dictionaries", and Ard took as his subjects the students in a "high-intermediate" ESL writing class at the University of Michigan. Data was collected in a variety of ways, the sources being retrospections by students, in-class compositions in which bilingual dictionaries were consulted and students' oral protocols while writing.

There is little discussion in the paper of students' retrospections. Ard notes that he discovered that students use bilingual dictionaries more frequently at home than in class, because of the time restrictions placed on them during lessons.

Data from students' compositions was collected by asking students who were in the process of writing to circle words they looked up in bilingual dictionaries and subsequently used. Ard cites excerpts from compositions written by Japanese and Spanish-speaking students, although he does not tell us the number of students involved in the study, or the number whose compositions are cited. The main finding seems to be that, despite the lower overall writing ability of the Spanish-speaking students, they used their bilingual dictionaries more successfully than the Japanese students. He ascribes this to the fact that Spanish and English are much more closely related to each other in terms of lexical typology than are English and Japanese: "there is more likely to be a one-to-one translation between a Spanish word and an English word than between a Japanese word and an English word". The Spanish speakers' use of bilingual dictionaries was not entirely error-free; their strategy of choosing "English words morph-orthographically close to a Spanish word suitable in the context whenever such a choice is available" is usually a successful one, but can cause problems by directing users away from a more appropriate non-cognate word. As an example of this Ard cites a Spanish student's rejection of the best choice - HOPELESS, to express the meaning of *desesperado*, and his inappropriate choice of a word that looked similar to the Spanish - DESPAIRING.

Protocols were made of the writing processes of two ESL students. We are told that one was a Japanese female who habitually used a bilingual dictionary at home and in class, while the other was an Arabic-speaking male who never used a bilingual dictionary. However, we are not told how these two subjects were chosen, nor indeed whether the Japanese subject was one of those whose compositions had been discussed earlier in the study.

Ard reports that L1-influence was observable in the products of both subjects in cases where bilingual dictionaries were not consulted. The Arabic speaker made explicit

reference to Arabic when thinking-aloud; he appeared to translate directly from Arabic when writing **FINALLY** and **EVEN THOUGH**, and to reflect upon **IN ORDER TO**, which was a direct translation from Arabic, before deciding to write **TO KEEP ... FROM** instead. The Japanese speaker made no reference to Japanese in her protocol, but chose without consulting her dictionary the calque **SALARY MAN** rather than **WHITE COLLAR WORKER**. Ard cites this behaviour as evidence that "prohibiting bilingual dictionaries does not eliminate L1-influence".

Ard also cites the protocol of the Japanese subject to argue that dictionary use did not greatly increase the overall time spent on the composition process. According to Ard "the use of a bilingual dictionary involves a considerable expenditure of time", but so too do other kinds of problem solving within the writing process; it took 52 seconds for the subject to decide on **TEETH TREATED** with the aid of a dictionary, but there were twelve pauses of between 5 and 43 seconds in her protocol which did not involve dictionary use. Unfortunately, although the protocol transcripts are attached in an appendix to Ard's paper, no time-scheme is recorded, and we cannot recover from the protocol what proportion of the entire time available to the Japanese student was spent on dictionary look-up.

Ard also cites the Japanese speaker's protocol to reinforce the point he made when analysing the effect of dictionary use on in-class compositions. As was the case with these compositions, the composition written in the think-aloud experiment contains errors resulting directly from bilingual dictionary consultation. The Japanese speaker used her dictionary three times during the experiment, to produce **MISTAKES OF TEETH TREATED** (an error Ard describes as a "paronym", ie the choice of a word morphologically related to an appropriate form), **STEP OF LIVING** (a collocational error) and **LIVING ...COMFORTABLY** (acceptable in context). However, both of the experimental subjects also made similar lexical errors on occasions when they did not consult a dictionary. In his concluding section Ard points out that "it has not been

proven that the use of a bilingual dictionary leads to errors where no errors would otherwise occur", especially in view of the fact that the learner turns to the dictionary when in ignorance about the correct word to use, and "it is unlikely that the desired concept could be expressed in English *without* the use of a bilingual dictionary, either".

In the final part of his paper, which reflects on the role of bilingual dictionaries in the ESL writing process, Ard suggests that the teaching profession's condemnation of bilingual dictionaries goes hand in hand with an undue emphasis on error avoidance. He argues that bilingual dictionaries have a part to play in improving learners' expressive abilities, and teachers must accept that when learners struggle to express new concepts they are bound to make more mistakes than when they merely repeat what they already know.

Discussion

This paper is part reflection on current teaching styles and attitudes, part experimental report. However, Ard's experiments play a largely supporting role; they are not reported in full, and we are not given sufficient details of his method to permit accurate replication. This approach to experimental reporting is potentially dangerous, because the facts are not allowed to speak for themselves; we are only permitted to examine those findings which Ard considers relevant to his argument, and are denied access to other details which might possibly be less conclusive.

Thus, in his account of dictionary use during in-class composition writing, Ard does not summarise the complete range of look-up strategies recorded for all class members, but only cites a few lines from the compositions of some Spanish speaking and some Japanese speaking students. Ard's claim that bilingual dictionary use is more successful for speakers of languages with a similar lexical typography rests,

therefore, on a tiny amount of evidence, and further data potentially available within the same class is ignored.

Protocol analysis is of course highly time-consuming, and it is understandable if Ard selected only two subjects for this treatment. However we are given no indication of why the protocols of these particular two were chosen, and whether the Japanese speaker was one of those whose compositions were previously cited. Given that the earlier data was used to compare Japanese and Spanish-speaking dictionary users, Ard's line of enquiry might have been more effectively followed through had he chosen a Spanish-speaking second subject. Given that dictionary users make many lexical choices which do not involve dictionary use, it would still have been possible to compare language produced with and without the aid of a dictionary had a dictionary user been chosen as a second subject. More relevant data could certainly have been obtained regarding the role of the dictionary in writing if two dictionary users rather than one had been selected. The reader is left with the impression that subject selection was entirely haphazard; an alternative possibility is that Ard in fact collected data from a larger number of subjects, and selected for publication only those two protocols which served to illustrate his arguments regarding the role of dictionaries.

The think-aloud technique works comparatively well in the case of the Arabic speaker, who explains the reasons for his choice of certain lexical items, but is less successful in the case of the Japanese dictionary user. Her oral report does not provide any insight into her reasons for using a dictionary, or the method she employs to select an appropriate word. In effect, her protocol is analysed in the same way as the in-class compositions of the earlier experiment - with reference to the finished product rather than the writing process. The think-aloud technique is intended to shed light on subjects' thought-processes, but some subjects seem far better able to express their thoughts than others, and training in the technique might have been advisable in

this case. A video recorder was used for data collection, and the recording could have provided us with some data on the amount of time taken at various stages of the writing process, and also perhaps data on the subject's physical handling of her dictionary; however the information we are given about timing is very incomplete, and no reference to search style is made in the paper at all.

While recommending further and more adventurous ways of developing learners' active vocabulary, Ard says that bilingual dictionary use:

is one among many methods, including the use of a thesaurus. Unfortunately, these alternatives are so rarely mentioned in ESL classes that they are not really alternatives.

(1982:18)

I am surprised that he does not acknowledge **monolingual** dictionary use, which is in much greater favour amongst teachers and EFL methodologists. It seems quite possible that some of Ard's subjects were monolingual dictionary users; this is particularly likely in the case of the linguistically more sophisticated Arabic speaker, who claimed not to know of the existence of a comprehensive English-Arabic dictionary. Ard never considers this possibility, and if he has denied his subjects the opportunity to access monolingual dictionaries in the way they might otherwise have done, his data does not provide, as I presume it intends to, a record of normal dictionary-using behaviour.

1.3.2. Hatherall (1984)

Hatherall's objections to questionnaire-based research are summarised in 1.1.8. He argues that questionnaires often yield very misleading information, yet they are very useful for investigating dictionary use amongst large samples of the population. Direct observation is a more reliable method of data collection, but this is usually too time-consuming to permit the study of a sample large enough to be representative. Moreover it is difficult for subjects to behave normally while being observed. If

subjects alter their behaviour during a direct observation experiment, the data collected will be as unreliable as that obtained via questionnaires, where subjects may distort the reality of their dictionary use as they recall it.

Hatherall's paper is not so much a research report, more a proposal for an alternative data collection method, which he claims will combine the advantages of both the indirect and direct approaches. Using Hatherall's technique, large groups of subjects can be monitored as in questionnaire-based research, but there is less danger of user behaviour being falsified, as subjects are not required to recall and interpret actions remote in time and place.

In the pilot study reported on in this paper, subjects were given one hour to translate into German part of an article from the business section of *The Sunday Times*. We are given no details about the subjects, except that they volunteered for the experiment, but they would appear to be native English speakers studying German at college-level. The text, which is reproduced in full in Hatherall's paper, was intentionally difficult, to ensure that an adequate amount of data on dictionary use could be collected in a short space of time.

While they translated, the subjects were allowed free access to dictionaries and "other reference works", and were requested to note down every occasion when they looked up a word, giving details of the dictionary they used, the search item, and the extent to which they considered the dictionary information to be useful.

Hatherall admits that the double task of translating and simultaneously recording dictionary use is an unnatural one, but his instructions to the subjects exhort them to approach the task as far as possible in "a natural way".

After one hour, subjects submitted their translations and report forms and were asked to complete a short questionnaire concerning the strategies they had adopted to complete the task, and their views regarding the text, their translation of it, and the experiment generally.

Hatherall does not provide any numerical data relating to the experiment; we are not told how many subjects took part, or what proportion behaved in any given way.

However, the following behavioural tendencies are presented as findings:

- * most students do not read the whole text through in advance of translating
- * more advanced students use the dictionary more often than less advanced students
- * most students use only the English-German section of their bilingual dictionary while translating from English into German. Advanced students are only slightly more inclined to consult the German-English section
- * students do not look up closed-set items such as prepositions, or common words
- * in verb-noun expressions such as "conduct an in-depth survey" and "take a strain", subjects look up the noun first. If they check the verb entry, it is only after they have found the noun entry information unsatisfactory
- * students tend to translate word-for-word, and the dictionaries they use do not discourage this behaviour.

Hatherall recommends that in future studies adopting this method of data collection, subjects should not be given free access to a variety of types of reference book, but should instead be allotted one type of dictionary only, or possibly two for the purposes of comparison. He also recommends that subjects should form a homogeneous group, at the same level of language proficiency. He considers it important to set a time limit for the experiment, because the ease and speed with which information can be accessed from a dictionary is an important consideration, but in his pilot study he found that the time he had allowed for the translation of the *Sunday Times* text was unrealistically short.

Discussion

The procedure Hatherall recommends seems feasible; if this method is adopted subjects do not need to be trained in think-aloud techniques, and data from a large number of subjects can be collected at the same time, without the researcher needing any special skills or equipment. Usually the most time-consuming stage in qualitative research, however, is not the data collection itself, but the analysis of that data. The responses to a well-designed questionnaire with a multiple choice format can easily be summarised and expressed in tabular form, but it is much more difficult to categorize look-up strategies that vary with each individual subject, and it takes much longer to sift through such data in search of emerging trends.

This may be the reason why the results of Hatherall's pilot study are incomplete. Certain trends are identified, but the original data must have been considerably richer; perhaps the information regarding the subjects' choice of dictionary and search items, and their satisfaction with the entries for these items, did not lend itself to categorisation. Much of the information that is reported as findings from the pilot study is easily quantifiable, and could have been obtained by less elaborate means: scrutiny of the finished translations would reveal that subjects tend to translate word-by-word, and a simple request to subjects to underline search words in the text could be used to discover which words they had looked up.

Hatherall's finding that subjects did not look up common words is consistent with the findings of Béjoint's questionnaire-based survey discussed in Chapter 2 section 1. 66% of Béjoint's subjects said that they never looked up common words. Hatherall's findings that subjects do not look up closed-set items, and look up nouns rather than verbs in noun-verb expressions, are consistent with Bogaard's findings, discussed later in this section. It would be interesting to know how advanced in German Hatherall's subjects were, and the extent of the grammatical information provided by their

dictionaries. To find the correct choice of preposition, for example, they may have chosen to look up grammatical collocations at the verb entry in their bilingual dictionaries. We do not know whether instances of this type of behaviour were recorded.

The finding that advanced students used their dictionaries more often than less advanced students differs from the findings of a number of questionnaires.

Bensoussan, Sim and Weiss (1984) found that 3rd year university students used dictionaries less frequently for decoding than did 1st year students, Battenburg (1990) also reported that advanced students used dictionaries less frequently than elementary students for reading, writing and translation, and Tomaszczyk (1987) claimed that EFL beginners used dictionaries too often, while advanced students did not use them enough. Of course, as we are given no numerical or statistical information in Hatherall's paper, we cannot judge the significance of his findings, which could be a summary of strong or weak tendencies across a large or small sample.

Hatherall's final observations regarding dictionary user research point to the value of computer-based studies:

if the dictionary user is himself looking up data in a computer rather than in a book, his behaviour can be monitored with ease, at least in terms of what and when (how often). Wholly reliable information in these two areas should prove invaluable in also explaining how and why.

(1984:189)

Hatherall's main concern is to find an accurate way of collecting large amounts of data on every stage in the dictionary look-up process. The method described in this paper records dictionary use as it occurs, but it is still a relatively crude method because it depends for its success on the notes written by the subjects themselves, which will inevitably vary in quality. Moreover, even if the subjects prove to be

honest and conscientious, some distortion will occur because of the intrusive nature of the monitoring system.

1.3.3. Miller and Gildea (1988)

For this study data was collected by a simple method which Miller and Gildea call the LUCAS task (Look Up, Compose A Sentence). The method required subjects to look up given words in a dictionary, and then write sentences using those words.

The aims of the first stage of the study were to discover the kinds of mistakes children make when looking up words, and which kinds of mistakes occur most frequently.

The subjects were 5th and 6th grade children (10-11 year olds). We are not told how many were involved in the study. It is implied that the children were native speakers of English attending schools in the United States, but we are given no details of their language background or proficiency.

When this paper was written the project was not yet complete. Thousands of sentences had already been collected by means of the LUCAS task, but only 457 had been analysed. 249 of these sentences contained 12 relatively common target words taken from 4th grade basal readers, and the remaining 208 sentences contained 10 relatively rare target words, taken from 12th grade basal readers. No details are given of the dictionary (or dictionaries) that the subjects consulted.

Each of the sentences composed by the children was checked for acceptability, and the errors in unacceptable sentences were described. Descriptions of errors were then roughly classified. The results of this analysis showed that 21% of the sentences using common words, and 63% of the sentences incorporating rarer words, were "sufficiently odd or unacceptable to indicate that the author did not have a good grasp of the meaning and use of the word". Sentences fell into one or more of the following categories:

No mistake	273
Kidrule example	68
Wrong part of speech	45
Wrong preposition	28
Inappropriate topic	24
Used rhyming words	14
Inappropriate object	14
Wrong entry	13
Word not used	10
Object missing	8
Two senses confounded	7
No response	4
Not a word	3
Unacceptable idiom	3
Not a complete sentence	3

As can be seen from the above list, the most frequent of the unacceptable sentences were those written in accordance with what Miller and Gildea call the "Kidrule" strategy. They argue that it is not fair to write of "errors" in this category, because the children who wrote them were employing a consistent strategy to solve their dictionary reading problems. This strategy appeared to entail searching within the definition for a familiar word or phrase, composing a sentence containing this segment, and then substituting the target word in place of the segment. For example, in the dictionary used by subjects in Miller and Gildea's study, TRANSITORY was defined as "passing soon; lasting only a short time", so according to kidrule a subject might produce the sentence "I bought a battery that was transitory", in which "transitory" is a substitute for the familiar segment "lasting a short time".

Miller and Gildea describe three phenomena which they regard as evidence for the existence of the Kidrule strategy:

- 1) occasionally subjects forgot to make the final substitution, and composed sentences in which a segment of the definition appeared instead of the target word
- 2) subjects also occasionally wrote a segment of the definition in the page margin; presumably this was intended to aid them during the process of substituting the target word
- 3) in two cases the dictionary only provided a one-word definition of the target word; TANTAMOUNT was defined as "equivalent", and SUCCULENT was defined as "juicy". If the target words were unknown, the Kidrule strategy would be the only logical way to complete the LUCAS task for these words, and in fact Kidrule errors were particularly frequent in sentences containing these two words.

To explore the Kidrule strategy still further, and to test whether employing the strategy as they had defined it would result in sentences similar to those the children had composed, the authors devised a 5-stage "Kidrule simulation":

- 1) find the target word in the dictionary
- 2) read the definition
- 3) select some short, familiar segment of the definition
- 4) compose a sentence containing the segment that has been selected from the definition
- 5) substitute the target word for the selected segment in the sentence, and write it down.

The simulation was run as a computer program, using the definition for PLUMMET, for which all "short, familiar segments" had been identified. The program searched through the Brown University corpus for sentences using the same common words, and then substituted PLUMMET for those words. A few of the sentences generated by the computer simulation were acceptable; many were very much like those produced by children performing the LUCAS task.

From this Miller and Gildea concluded that the Kidrule strategy might be even more widespread than their categorisation suggested. When employing the Kidrule strategy, children might randomly generate some acceptable sentences in the same way as the computer had done, yet although these sentences would be categorised as error-free, the children would not have a clear understanding of why they were appropriate, and might produce further, inappropriate, sentences by following the same procedure.

In a second stage of the study, Miller and Gildea aimed to investigate whether children could learn words more easily from illustrative sentences than from dictionary definitions. We are told that three classes of 6th grade children were used for the experiment, but we are not given any further details about them.

The authors selected 10 relatively rare (12th grade) words, and prepared three different kinds of instructional material to accompany them: definitions taken from a dictionary, illustrative sentences taken from a dictionary, and illustrative sentences taken from the *New York Times*.

The LUCAS task was used once again as a means of data collection, but instead of consulting their dictionaries the children read the definitions or illustrative sentences that had been prepared for them. Although the exact procedure for the experiment is not specified, it seems likely that each of the three classes received one of the three types of instructional material. As in the account of the first experiment, it is not clear whether each schoolchild produced sentences for each of the words, or whether each child produced only one sentence.

All the sentences were rated by two judges as either acceptable, marginal, or unacceptable, with the following results:

	Definition only	Dictionary sentence	NY Times sentence
Acceptable	36%	55%	52%
Marginal	33%	22%	19%
Unacceptable	31%	23%	26%

These results suggest that illustrative sentences are more helpful to children than definitions.

Amongst those children who had access to illustrative sentences rather than definitions, examples reflecting the Kidrule strategy still occurred. Miller and Gildea suggest that this may be because some illustrative sentences did not provide sufficient information regarding word behaviour. The dictionary illustrative sentence for USURP, for example, was *The king's brother tried to usurp the throne*, which suggests that USURP takes a concrete rather than an abstract object. Children produced sentences such as *The blue chair was usurped from the room* - categorised as a Kidrule error. Similar Kidrule examples occurred among the children who had read the dictionary definition for USURP. Here suitable objects for USURP were given within parentheses: "(power, position or authority)", but the schoolchildren did not recognise the function of the parenthetical construction, and wrote sentences such as *During the wrestle, he had usurped his opponent's hair*.

However, Miller and Gildea felt that the children consulting illustrative sentences were using a more complex version of the Kidrule strategy than those who consulted the definitions; "they must first abstract a familiar concept from the unfamiliar word's context of use, and only then apply Kidrule". This leads Miller and Gildea to hypothesize that "perhaps the Kidrule strategy is simply the second half of the more general strategy that children use to pick up new words by hearing them used" (1985:24).

The studies suggest that:

- 1) kidrule is a strategy commonly employed by children when accessing dictionaries, and it may also play a part in the acquisition of vocabulary by less formal means
- 2) children do not learn words very successfully from conventional dictionaries
- 3) children learn words more effectively by encountering them in context than they do by reading their dictionary definitions.

On the basis of their findings, Miller and Gildea propose three computer-based means of developing children's vocabulary. The first proposal is simply that dictionary users should also be given access to the database of instances which informed the lexicographer when composing the dictionary. In this way they would be able to learn from not just one instance of use for any given word, but many. The second proposal is that computers can function as automated vocabulary tutors, giving feedback on learner errors. Children could type in sentences generated by a LUCAS task, and "the machine would be programmed to recognize the kind of error that had occurred and to give immediate feedback". If this proved technically impossible, limiting the children's choice of words in the sentences they composed would make the computer's tutoring task less complex. The third proposal is for the presentation of lexical databases on videodisc. Instead of looking up words in a dictionary, children could call up pictures, graphics and a voice-over giving example contexts for the search word.

Discussion

The LUCAS method of data collection seems to have an important advantage over other language-generating methods described in this section. Large amounts of data can be collected in a short time, with or without a researcher present. However, it would be necessary to check that the children did in fact look up the target words, especially if common words which children might think they already knew were involved. The LUCAS task is quite demanding, and if young subjects are required to

process long lists of target words it is also possible that loss of concentration might begin to affect the data. It is unclear from this paper, however, whether each individual subject was required to compose many sentences or just a few.

The LUCAS task is not a "natural" activity in the way that a translation task (as in Hatherall's paper), or a composition writing task (as in Ard's paper) might be. It therefore provides no data on which words learners might choose to look up while reading or composition-writing. In ordinary life, if children turn to the dictionary they presumably do so because they are strongly motivated to know more about a word that they have already encountered in a meaningful context, but in this experiment they looked words up simply because they were instructed to do so. The extra context of personal experience, which a dictionary user ordinarily refers to when interpreting a dictionary entry, is not considered in this experiment.

Of course, some of the children may well have had some familiarity with some of the target words before they looked them up. Ideally, Miller and Gildea should have controlled for this, perhaps by instructing the children to look up only those words they did not know. A lot of potentially interesting data would have been lost if this method had been employed, because inevitably some children would have guessed at the meaning of some words they did not really know, but the results may have been a better reflection of the value of the dictionary entries to the children, because there would be less likelihood of prior knowledge overriding dictionary evidence. I presume that far fewer acceptable sentences would have been recorded if the children had only looked up the words they did not know. I also presume that the higher number of acceptable sentences formed with common words is a reflection of subjects' prior knowledge of those words, rather than the readability of common-word definitions.

The LUCAS approach is thus on the whole a good way of collecting data concerning dictionary use. Like all qualitative data, however, the type of data generated by the LUCAS task is somewhat resistant to categorisation. The method meant that Miller and Gildea did not have to sift through the huge quantities of the disparate information typically available from oral protocols, but on the other hand they could not inform their decision-making by triangulating the LUCAS data with evidence collected during or immediately after the sentence-writing process. At times the evidence of the sentences alone may not be sufficient to justify the choice of one category over another. Miller and Gildea cite a sentence with the target word TENET as an example of how difficult it is to categorise correctly: they decided that "John is always so TENET to me" was a Kidrule example, and ruled out the possibility of a phonological/orthographical confusion with TENDER, on the grounds that the 10 year old boy writer was unlikely to wish to express this kind of meaning. Although Miller and Gildea were doubtless right in this particular case, they do not appear to have any data on their subjects other than age and sex, and in most cases this information would not be adequate to act as a deciding factor between one category and another. Certain other Kidrule examples which they quote but do not query seem to me to be suspect. For example the sentence "The water was very SUCCULENT" is described as resulting from the substitution of SUCCULENT for the one-word definition JUICY, yet the sentence "The water was very juicy" makes no better sense.

In the account of the preliminary categorisation of sentences there is no mention of tests for inter- and intra-rater reliability. In the later analysis of sentences generated with the aid of the three different types of instructional material, we are told that two judges rated the sentences as "acceptable", "unacceptable" and "marginal". However it seems likely that this is a coy reference to the authors themselves; what is really needed is an independent panel who have no prior expectations of the outcome of the analysis.

A further problem of analysis regards the sentences children wrote after consulting instructional material. Miller and Gildea found that in fact about 10% of the acceptable sentences were closely modelled on sentences shown to the children as illustrations of word use. A copied sentence did not necessarily indicate that the child has failed to understand the word meaning, however, as the children were not told to avoid imitating the examples. We are told that "judgements of closeness of modelling are subtle and subjective", presumably because there is no easily-recognisable cut-off point between sentences which parallel the original and those which do not. Miller and Gildea make light of the problem in this paper, suggesting only that a different study, possibly using multiple choice questions so that subjects could not reproduce source material, might be devised to confirm that subjects really do understand illustrative sentences better than definitions. In a later paper, however, they express greater reservations about the outcome of the study:

A preliminary study indicated that children can write better sentences when they are given a model sentence employing the word than when they are given a definition of the word. Since many of the sentences they wrote were patterned on the models, this result could not be interpreted to mean that the children learnt more about the meaning of a word from illustrative sentences than they learnt from definitions.

(1987:90)

A further study recorded in Miller and Gildea (1987) also casts doubt on the value of showing the subject several example sentences rather than one. The authors found that "the acceptability ratings of sentences written after seeing one model sentence were the same as the rating of sentences written on the basis of three examples". This is a curious finding, as children naturally acquire vocabulary through multiple encounters with the same words. Classroom concordancing is growing in popularity precisely because it provides the opportunity to present the same lexical item in many contexts.

The lack of improvement when several contexts were provided in Miller and Gildea's study may be the result of a flaw in the LUCAS task; perhaps most of the acceptable

155

sentences were really the result of imitating a single example sentence, and the children were not employing any strategies to extract the target word meaning from context. It should be pointed out, however, that Black (1986, reviewed in the previous section) also failed to discover improved word comprehension under defining conditions where one and several examples were available. In Black's study, word comprehension was tested by multiple choice questions, so subjects could not mindlessly copy defining information.

Miller and Gildea's experimental method cannot be reviewed in detail because their paper does not provide sufficient information regarding their subjects, materials, and procedures of data collection and analysis. However, the information that is given suggests that, despite certain problems at the analysis stage, the LUCAS task might be a useful way of obtaining data on non-native speaker dictionary use. The paper provides strong evidence for the existence of a Kidrule strategy; it would be interesting to explore whether Kidrule is only for kids, or whether older English language learners employ a similar technique when reading dictionary entries. Given that learners' dictionaries are distinguishable from native-speaker dictionaries by their heavier dependence on examples (among other things), we also need to explore how successful these examples are in conveying the meanings of words to EFL learners.

1.3.4. Neubach and Cohen (1988)

This study aimed to find answers to seven major questions concerning language learners' dictionary use:

- 1) What strategies and outcomes characterize the use of monolingual and bilingual dictionaries by EFL learners?
- 2) How does the proficiency level of the student relate to the strategies used and the outcomes?
- 3) Does the search for a given word provoke certain strategies and outcomes?
- 4) Does the dictionary aid in performing dictionary comprehension tasks?
- 5) What type of dictionary do students at different levels of proficiency prefer?

6) Are there strategies and outcomes specific to the use of a monolingual or a bilingual dictionary?

7) In what ways might the dictionaries themselves be problematic?

The subjects were six students at the Hebrew University of Jerusalem, enrolled on an EAP reading course. Two were characterised as being at a high level of English language proficiency, two at intermediate level and two at a low level, although the terms "high", "intermediate" and "low" are not related to performance in a standardised test, and we are not told the proficiency range of the group.

Three different dictionaries were used: Longman Active Study Dictionary (for the low level subjects), Collins English Learners' Dictionary (for the intermediate and advanced subjects), and a bilingual English-Hebrew Dictionary, the Megido Modern.

There were two tasks, both followed by structured interviews. In the first task subjects were given ten sentences, each one with an underlined polysemic word, such as *moored*, *rack*, *bearing*, and *dash*, and they were required to look these words up, whether they knew them already or not, first in a monolingual dictionary and then in a bilingual dictionary. The ten sentences had presumably been written by Neubach and Cohen themselves; they explain that they chose polysemic words for subjects to look up because they were interested in the problems students had in choosing the correct entry from a number of alternatives. While they were looking the words up, subjects had to give an oral report of their search process. These protocols were tape recorded and subsequently analysed. Finally, subjects were required to translate the underlined words into Hebrew, and explain why they had selected particular meanings from the range they had looked up in the dictionaries.

The interview which followed gathered data on the subjects' attitudes to dictionaries, and their preferences.

In the second task subjects were asked to read a 150-word passage by Helen Keller, in which ten uncommon words had been underlined. Subjects were free to use any or none of the three dictionaries while reading, and were required to make an oral report of their progress. The protocol was once again tape-recorded for analysis. Finally the subjects were required to summarize in Hebrew what they had read.

The second interview was similar to the first, but subjects were asked their reasons for choosing a monolingual or a bilingual dictionary for the task.

From comments made in the results section of the paper, it would appear that Neubach and Cohen also noted the time subjects took to look up underlined words, although the collection of this information is not recorded in the account of the experimental method.

No details are given of the methods by which Neubach and Cohen analysed and categorised their data. Their findings appear to summarise data collected from the think-aloud tasks, the interviews, and an analysis of the subjects' Hebrew translations, but it is really only in response to their first research question, regarding strategies and outcomes, that they appear to have categorised their data in a systematic way. For this research question they identify fifteen categories of behaviour, but as no numerical information is given to indicate the frequency with which particular strategies and outcomes occurred, it is impossible to judge whether any of the fifteen categories are typical of the group as a whole, or simply characterize one or two of the subjects.

Neubach and Cohen's findings are presented in seven sections, corresponding to the seven research questions that the study aimed to investigate.

In response to question one - "What strategies and outcomes characterize the use of monolingual and bilingual dictionaries by EFL learners?" - 15 strategies and outcomes are listed, divided into two sections: "before the search" and "during or at the end of the search". Subjects' behaviour before looking a word up appears to have been characterized by attempts, successful or unsuccessful, to gather information about the word from context. Neubach and Cohen do not record any instance of subjects accessing the dictionary directly, without regard for context.

The strategies or outcomes listed as occurring once the search had begun suggest that it was common to experience difficulties while looking words up. Only the last outcome on Neubach and Cohen's list is a positive one:

- 1) Reading only the first definition in the monolingual dictionary
- 2) Encountering a problem with vocabulary in a definition in the monolingual dictionary
- 3) Encountering a problem with terminology in the monolingual dictionary
- 4) Encountering an alphabet order problem
- 5) Encountering a problem with the monolingual or bilingual dictionary entry itself
- 6) Encountering a problem with the format for presentation of the definition in a given monolingual or bilingual dictionary
- 7) Experiencing frustration during the search
- 8) Arriving at the word meaning but uncertain about it, whether with a bilingual or a monolingual dictionary
- 9) Arriving at the correct conclusion on the basis of the dictionary entry.

In response to question two - "How does the proficiency level of the student relate to the strategies used and the outcomes?" - not surprisingly the two most advanced students appear to have had fewest problems, and the two weakest students the most problems. There seemed to be a correspondence between language proficiency and the ability to adopt the following successful strategies:

- 1) determining the correct part of speech before a search

- 2) forming correct expectations at word and sentence level
- 3) leaving the context to make the search
- 4) understanding the words in the definition
- 5) understanding the symbols and abbreviations in the definition
- 6) choosing the correct definition.

Presumably not all protocols provided data on all these strategies. Intermediate students appear to have had the same sort of problems as weak students. One of the most advanced students complained that the dictionary was time-consuming, but Neubach and Cohen note that the weak students took longer than the other subjects to look words up (this comment is presumably based on their own observation, rather than an analysis of protocols).

In response to the third research question - "Does the search for a given word provoke certain strategies and outcomes?" - Neubach and Cohen identify the context and the dictionary treatment of a word as the two major factors in determining search success. Subjects looked up MOORED, FARE, and PAWN most easily, according to Neubach and Cohen, because the context for these words was particularly easy to understand. They found that the words subjects had greatest problems with were either not listed in one or more of the dictionaries used for the experiment, or were defined in language that the subjects found difficult to read.

In response to the fourth research question - "Does the dictionary aid in performing reading comprehension tasks?" - Neubach and Cohen found that only the more advanced students benefited from dictionary use. These students formed semantic field expectations before they looked words up; they already understood the main message of the passage, but used the dictionary to deepen this understanding. Weak students were hampered by lack of world knowledge, and failed to recognise implicit and subtle aspects of the text.

In response to the fifth and sixth research questions - "What type of dictionary do students at different levels of proficiency prefer?", and "Are there any strategies and outcomes specific to the use of a monolingual or bilingual dictionary?" - the strong students were found to prefer the monolingual dictionary, while the intermediate and weak students preferred the bilingual dictionary. For the second task, the strong students always used a monolingual dictionary, and the weak students always used a bilingual dictionary, but the intermediate students used a mixture of both, turning to the monolingual only if they were not satisfied with the information they found in the bilingual. According to Neubach and Cohen the advanced students' preference for a monolingual dictionary "stems from a combination of language proficiency, past experience, a certain perfectionism in search style, and intellectual curiosity which makes the search more enjoyable". We are not told whether these strong students preferred the Collins English Learner's Dictionary or the Longman Active Study Dictionary. It was intended that the advanced students should use the Collins dictionary in the sentence translation task, but all subjects had a free choice of dictionaries for the second task, so either of the two monolingual dictionaries might have been used.

Neubach and Cohen make a number of criticisms with respect to the seventh research question - "In what ways might the dictionaries themselves be problematic?". They complain of small print and crowding, and the lack of space between dictionary definitions and examples. They also note problems with the "high register" and incomplete coverage of the bilingual dictionary used, and the fact that sometimes an alternate word given in parenthesis was preferable to the main meaning given. As far as the monolingual dictionaries were concerned, the main problem seemed to be with definitions and examples containing words which subjects did not know.

Neubach and Cohen conclude their paper with some comments on subjects' mistakes, and the dictionary-using skills they need to acquire. The subjects went wrong by

155

looking up words in the wrong places, giving up the search before they found the right meaning, or continuing the search without realising that the correct meaning had already been found. All the subjects had problems with grammatical terms, abbreviations and phonetic script.

Neubach and Cohen identify the following skills as necessary for students to acquire:

- * the ability to check definitions against the original context, especially in the case of polysemous words where there may be four or five definitions for the same word form
- * the ability to extract information from contextual clues before starting to look up a word
- * the ability to recognise inflected forms and reconstruct uninflected forms, in order to know which form to look for in the dictionary
- * the ability to cope with the "mechanics of dictionary use" - alphabetical order, symbols and abbreviations.

Discussion

This study has a number of defects. The sample size is small, the tasks posed problems that Neubach and Cohen do not appear to have foreseen, and the resultant data is not presented systematically. As in Ard's study (1982), the total number of subjects was extremely small, and where dictionary use at different proficiency levels was compared in questions 2 and 5, there were only two representatives for each level. We should be wary of making generalisations based on the dictionary-using behaviour of these three pairs of subjects, not only because the smallness of the sample reduces its representativeness, but also because the three groups appeared to have differed from each other in other respects than just English language proficiency.

We are not told the age or language-learning experience of the subjects, but as they were all native speakers of Hebrew enrolled on the same pre-university course, it seems likely that they shared similar academic backgrounds. In this case their difference in level might not have been a result of their having studied English for

differing lengths of time, but may have reflected differing levels of motivation, or study skills, or intelligence. The two subjects in the lowest group were severely disadvantaged in the second task by the fact that they did not know of Helen Keller; although this lack of knowledge adversely affected their dictionary use, it is not directly related to a low level of English language proficiency.

Neubach and Cohen write of the advanced students' "language proficiency, past experience perfectionism in search style, and intellectual curiosity", which distinguished them from the other two groups. Again this suggests that the groups of students differed from each other in a variety of respects.

The purpose and design of tasks one and two in the experiment are not entirely clear. We are not told of any criteria for the selection of sentences for task one, and we are not told how the text in task two was chosen, although the choice of texts and underlined words was crucial to the outcome of the study.

Presumably Neubach and Cohen wrote the sentences for task one themselves, and presumably they wished to minimize variation in contextual richness and difficulty, yet they account for the relative ease with which some underlined words were looked up by explaining that these words appeared in contexts which were easier to understand. They do not state on what grounds they decided that certain sentences were easier, but the ten sentences do not obviously vary in difficulty level and contextual clues, and it would appear that the researchers did not notice the variation until after the task had been completed. The possibility that some subjects might have already known the meanings of some of the underlined words is not discussed, and does not appear to have been checked before the experiment, although prior knowledge of word meanings would clearly affect the data considerably.

It is curious that Neubach and Cohen should choose to use a passage about Helen Keller when a third of their subjects did not know who she was. The authors do not appear to have anticipated this discrepancy in subjects' background knowledge, although it may have made the difference between dictionary-using success and failure in some searches. It is now widely recognised that measurements of reading comprehension are significantly affected by readers' prior knowledge of the content matter (cf Perkins and Brutton 1988), and a certain level of text comprehension is necessary for successful receptive dictionary use. It is conceivable that the results of question two could be reversed by setting a reading passage on a topic unknown to the two advanced students, but familiar to the lower proficiency subjects!

Just as Neubach and Cohen encountered unforeseen difficulties with the tasks they set, so too did they encounter problems with the bilingual dictionary they had chosen. It might have been better to check coverage in all three of the dictionaries before finally deciding on which polysemous words to feature in task one. When analysing the data from task one, Neubach and Cohen discovered that five out of the ten bilingual definitions for the words they had underlined were faulty. Although this finding is certainly worth reporting, dictionary coverage can be easily assessed without recourse to an observation-based experiment. Prior knowledge of the strengths and defects of the dictionaries could be used to inform the design of the task, enabling researchers to compare user strategies in cases where adequate information was and was not provided.

We are not told how the three dictionaries used in the experiment were chosen, or whether they were familiar to the subjects. The *Megido Modern* dictionary in particular comes in for criticism, and this was the dictionary chosen by lower and intermediate level subjects in task two. It is possible that some of the weak subjects' frustrations and failures may have been caused by the dictionary's defects, rather than their own lack of skill. Perhaps it is outside the scope of the experiment to isolate reasons for the strategies and outcomes reported, but I feel that the choice of the

Megido Modern should be justified, given the problems caused by its "high tone" and poor coverage.

We are also given no explanation for the choice of Longman Active Study and Collins English Learner's Dictionary, and the two dictionaries, although rather different in style, are rarely differentiated in the findings. Subjects who habitually used these dictionaries would be at an advantage in the experiment, but we are not given any information about the subjects' previous dictionary-using experience.

With a systematic analysis of oral protocols, identified trends can be placed in perspective. Not only the type, but also the frequency of different categories of behaviour can be indicated in the findings. Neubach and Cohen did not carry their analysis to this point; their findings tend to be summaries of comments made in individual protocols and interviews, and there is no indication of the proportion of subjects who completed each search successfully or unsuccessfully. This approach to data analysis may distort the real picture of dictionary use, because the think-aloud method of data collection invites emphasis on problem areas of the task; even with competent dictionary users as subjects, more data will be collected for long and frustrating searches than for quick and successful searches - indeed when dictionary look-up goes entirely according to plan there may be little left for the subject to say. Difficulty and failure feature in most of the strategies and outcomes listed in response to research question one. Strategy 15, "formulating the correct conclusion based on the dictionary", might, however, have been relatively common but little talked about.

A further problem with summarised protocols is that summaries may reflect the authors' own interpretation of events. In response to research question 2, for example, there is extensive citing of the weaker subjects' problems, while advanced subjects' problems are downplayed. One of the two advanced students admitted that she did not use a dictionary much because she found it too time-consuming, and the other claimed to have difficulty with dictionary symbols and abbreviations, but these

outcomes are not presented as serious problems, presumably because Neubach and Cohen were influenced by the status of these subjects as advanced and successful students. Any problems that weak students admit to, however, are seized upon as evidence of their ineffective dictionary-using strategies. It is almost as if the authors had decided on their data categories - those characterising high, intermediate and low students - before examining the data itself.

Before we can form valid generalisations from data gathered by the think-aloud technique, we need to consider all contextual variables, make a systematic analysis of the data, and possibly triangulate the data with information gathered by other means. Neubach and Cohen's findings are not fully generalisable; instead they provide a vignette of the thoughts and behaviour of six different dictionary-using individuals.

The report is not without interest, however. It is probably the most detailed study to date of procedures adopted by EFL learners when using dictionaries. The findings support conclusions reached in earlier studies, in particular the study by Mitchell (1983), which is cited by Neubach and Cohen and possibly served as a model for their experiment. Mitchell's research into dictionary use focussed on native-speaker primary schoolchildren, but she reports difficulties similar to those identified by Neubach and Cohen, for example alphabet order problems, and the failure to read beyond the first definition in a long entry. It is interesting to note that Tono (unpublished B Ed dissertation, cited in Battenburg (1991)) found that Japanese university students learning English also tended to focus on material appearing at the beginning of dictionary entries and were often unwilling to read the entire entry.

The similarities between some of the findings in these studies suggest that native speaker children and non-native speaker adults may have further, as yet unrecognised, behaviour patterns in common when using their dictionaries. Mitchell's study is much longer and more detailed than Neubach and Cohen's, and identifies other types of

dictionary user behaviour. It would be interesting to discover whether these also occur in EFL situations.

Neubach and Cohen's study also supports the findings of Tomaszczyk (1979), Bensoussan, Sim and Weiss (1984), and Atkins and Knowles (1990) that monolingual dictionary use increases, and bilingual dictionary use decreases, with increasing linguistic sophistication.

1.3.5. Ahmed (1991)

The study reported in this paper was not primarily concerned with dictionary use. Instead, the two objectives of the study relate to more general issues of vocabulary acquisition:

- 1) to identify the types of microstrategies used by a group of Sudanese learners of English when learning vocabulary, and to assess how frequently these strategies were used
- 2) to discover whether there were any differences in the vocabulary learning strategies used by good and poor learners.

The subjects were 300 Sudanese learners of English, attending universities, Government and private secondary schools, and Government intermediate schools. Prior to the collection of data school officials had categorised these subjects as either "good learners" or "underachieving learners" on the basis of subjective assessment and scholastic records.

Data was collected by recording and observing subjects' performance during a think-aloud activity, and by subsequently interviewing the subjects.

In the think-aloud activity, subjects were first trained in the think-aloud technique, and then given a list of fourteen unknown words to learn, with no time-limit on the

learning period. Thoughts verbalised by the subjects during the learning period were recorded on audio tape, and any observable strategies which were not verbalised were noted at the time. A monolingual and a bilingual dictionary were available for the learners, and so the observers noted dictionary look-up strategies.

The interview was questionnaire-based, and covered four areas:

- 1) what information sources subjects used to find out about difficult words, and what questions they asked of these sources
- 2) what dictionaries the subjects used, and what information they looked for in a dictionary
- 3) what note-taking microstrategies the subjects used
- 4) what techniques the subjects used to practise and memorize words.

The full questionnaire upon which the individual interviews were based is not provided in this paper.

Fifty-two different microstrategies were identified in the data, thirty-eight of which occurred with sufficient frequency, and with sufficient variation between groups to justify inclusion at further stages of analysis. Subjects were grouped according to which of these microstrategies they did and did not use. The frequency with which individuals used a strategy was not taken into account. Using the cluster analysis program CLUSTAN 1B five clusters of subjects were identified - clusters K1 to K5. Subjects within each cluster were believed to display the following characteristics:

K1 - the majority of subjects in this cluster were good secondary school students, good intermediate students forming the second largest component. Typically these subjects helped each other with new words and they also guessed at meaning and used dictionaries. Subjects in this group tested themselves and asked to be tested. They took notes about new words using both the L1 and the L2.

K2 - the majority of subjects were underachieving university students, but a few underachieving secondary school students also formed part of the cluster. Typically these subjects relied heavily on their L1; they asked for the L1 equivalent of new words, made notes of new words in terms of the L1, and memorized them by writing and repeating them with their L1 equivalents. These subjects also used bilingual dictionaries more than was average for the whole group.

K3 - almost all the subjects in this cluster were good university students and private secondary school students. Typically these subjects worked directly in the L2; they asked for an L2 paraphrase of new words, used L2 synonyms when noting the meaning of new words, and memorized them by writing and repeating them with their L2 synonyms. These subjects made more than average use of monolingual dictionaries. K3 subjects also tended to employ a wider than average variety of other strategies for learning, such as checking the meaning of words in context, testing themselves, using a vocabulary book and organizing new words by meaning.

K4 - Good intermediate students made up the majority of subjects in this cluster, but there were also a few good secondary students. Like subjects in the K1 cluster they tended to guess word meanings and ask classmates for help, but they made greater use of the teacher as an information source. In other respects they were also similar to the K1 cluster, but they tended not to practise vocabulary by testing themselves, and not to use a dictionary.

K5 - This cluster was made up of underachieving secondary and intermediate students. They tended to overlook new words or ask classmates for help. Typically they were not dictionary users, and repeating words aloud was their only learning strategy. All practice strategies, and all microstrategies involving the use of the L2, were employed by this cluster less frequently than was average for the group as a whole.

Discussion

Ahmed does not state whether the two categories of learner, "good" and "underachieving", represent the extremes of a continuum from unsuccessful to successful learners - in which case large numbers of potential subjects in the middle achievement range must have been excluded from the experiment - or whether randomly selected subjects were placed in either one category or the other - in which case the lowest of the high achievers would have performed little better than the highest of the low achievers.

Moreover, if standards of achievement varied from one school to another, it might have been possible for subjects judged (by subjective assessment) to be successful at one school to be judged as average or even as unsuccessful by the standards of another. This would mean that subjects categorized as "good" would not all be equally successful learners, and subjects categorized as "underachievers" would not all be unsuccessful to the same degree.

If subjects of average achievement were included in the experiment this would clearly weaken the significance of the findings, but in the absence of evidence to the contrary, I assume that Ahmed used only the best and weakest students in his experiment.

The most revealing data on vocabulary learning strategies would be a record of learners' spontaneous behaviour, at home, in the classroom and elsewhere. This might show us what words learners choose to learn and what they do to learn them under normal conditions, when they are not conscious of being observed. Unfortunately this kind of data is almost impossible to obtain, and Ahmed chose the easier course of recording learners' behaviour in a contrived vocabulary-learning situation, and supplementing this data with the subjects' retrospective accounts of the strategies they employed under normal conditions. Perhaps the defects of both methods of data collection could have been lessened by "triangulating" the data, and discarding any

contradictory evidence (for example that of any subjects who claimed not to write and repeat words aloud when interviewed, but were observed to do so during the think-aloud task). However, there is no evidence in this paper that the two sets of data were compared for validity.

One objection to the think-aloud and observation data might be that the learning task was rather artificial. Subjects were given a list of decontextualized L2 words, without any sort of gloss. They were then required to learn them for a test, although they were not told what type of test they would be given. In real-life learning situations, most learners encounter unknown words in context. They make their own choice as to which words are worth actively learning, and which words to overlook. In the course of reading an extended text they may encounter the same unknown word several times over, and gradually refine their ideas about its range of use and meaning. Recording the way these learners memorize a list of decontextualized words will not provide much insight into the way they learn words normally.

It is possible that some of the subjects in the experiment did habitually learn new vocabulary from lists provided by the teacher, but even so it is unlikely that the words on their school lists were presented in the same way as in this experiment. On the school list the words would almost certainly be made more accessible to the learners, for example by glossing.

In the think-aloud experiment, subjects appear to have had no means of discovering word meaning except through dictionary use. We are told that they were not provided with any information about the words they had to learn, and as they were observed individually there were no opportunities for conferring with fellow-subjects. Under these circumstances it would be natural for them to look up the words in a dictionary, even if they employed other methods to discover word meaning normally. Such dictionary use during the think-aloud task was presumably recorded as a

microstrategy and included in the final analysis as evidence of the typical behaviour of certain types of learner, whether it was really their typical behaviour or not. It is remarkable that some of the subjects appear **not** to have used a dictionary during the think-aloud task. Perhaps these subjects simply made different assumptions about the nature of the forthcoming test, or perhaps they thought they already knew what the words meant (We are not told how the lists of words were compiled, or how Ahmed could be sure that they were indeed unknown to the subjects.)

Data from think-aloud experiments is always to some extent distorted; it is generally accepted that the effort of verbalising behaviour alters the nature of that behaviour. In this experiment there is a further distorting factor because not only the requirement to verbalize but also the task itself requires subjects to behave in a way they would not behave in real life.

The proportion of data obtained by each of the two collection methods is not stated, but it would appear from the list of microstrategies that more information was collected by interview than through think-aloud. The questionnaire covered each of the macrostrategy areas: information sources, dictionary use, note-taking, practice and memorization. Strictly speaking, only strategies concerning dictionary use and memorization would be identifiable from the think-aloud data, and even the dictionary use strategies would be suspect because subjects were not provided with the sort of alternative information sources that they might find in school or at home.

This means that, despite the immense investment in time that the think-aloud experiment required, the findings presented in the paper are primarily based on the subjects' own accounts of their behaviour, rather than direct observation. Although information obtained by questionnaire may help to outline the general approach taken by a group of learners - their degree of reliance on the L1, for example - I doubt that it provides reliable details of learners' microstrategies. I think it is possible that both the

more advanced subjects and the underachieving subjects revealed in their interviews different attitudes towards teachers, classmates, the target language and their own capacity to learn. I think it is also possible that these subjects were influenced by their own attitudes when they claimed to employ certain microstrategies. For example, I find it hard to believe that subjects in K1, K3 and K4 did not overlook unknown words. The most successful language learners tended to belong to these clusters, and those in K3 (university and private school students) were actually taught in the medium of English. Learners working communicatively with advanced-level authentic texts are unlikely to stop and investigate every unknown word they encounter; like native-speaker students, they are likely to learn a lot of vocabulary by continuous exposure in context. Overlooking words characterised those in the least successful cluster, K5, who as secondary and intermediate students were in fact more likely to be given simplified texts with every new word explained. I expect K5 subjects did overlook new words, but their admission of "guilt" probably reflected the fact that they knew themselves to be underachievers. Eager beavers in the successful clusters may simply have given what they perceived to be the right answer.

Ahmed's summary of his findings is puzzling, as many details do not tally with the information given in his tables. We are told that K1 subjects "typically use a bilingual dictionary" (p8), yet according to Table 3 in his paper this is not a positive diagnostic for K1. We are also told that K1 subjects "tend to be more aware than other groups of the sorts of information that dictionaries can provide", but in Table 3 both K2 and K3 are listed as using more different types of dictionary information than K1. Ahmed notes that members of the K3 cluster "often overlook words they do not know" (p9), but Table 3 shows overlooking words as a negative diagnostic for K3. K4 is distinguished from the other groups as being "predominantly bilingual", yet it does not seem to employ the L2 more frequently than K1, and is considerably less L1-dependent than K2. Finally Ahmed claims that "co-operation appears to be a characteristic of poor learners"(p11), but I cannot find evidence for this in Table 3.

151

Although using classmates as an information source is positively diagnosed for K5, an underachieving cluster, it is also positively diagnosed for two good clusters, K1 and K4, and negatively diagnosed for the underachieving K2. Groupwork as an information source is positively diagnosed for K2 but negatively diagnosed for K5 - for neither cluster does it appear to be a preferred information source. It might have been more accurate to suggest that co-operation was a characteristic of school children, both successful and unsuccessful.

Ahmed acknowledges that cluster analysis is not an entirely objective method of analysing data, but my criticisms primarily concern his methods of data-collection, and also to a lesser degree the list of microstrategies used to enable clustering to take place. As Ahmed rightly says, not all the diagnostics are unrelated variables, and some microstrategies should perhaps have been collapsed into one. However I have not noted any obvious contradictions in the data as given in Table 3.

Despite all the problems I have identified within this experiment, the findings do seem to confirm commonly-held beliefs about the behaviour of good learners and poor learners. We expect good learners to use more strategies, be more actively engaged in the learning process, and operate to a greater extent in the L2. But do good learners become good learners because they behave in this way? I think it more likely that they behave in this way because they are good learners, with the power to use strategies effectively. So it is no use advising under-achievers to go and do likewise - they may not be able to operate in the L2, they may not be able to make sense of a dictionary entry, and at school they are unlikely to request tests that they will probably fail.

1.3.6. MacFarquhar and Richards (1983)

This paper involves the "acceptability testing" of dictionary definitions by a group of English language learners - a similar technique was later proposed by Crystal (1986)

as a means of improving dictionary design. In their study MacFarquhar and Richards set out to:

- a) survey the uses and users of dictionaries
- b) consider the lexicographical conventions dictionaries make use of
- c) report on a study in which the comprehensibility of definitions in several well-known dictionaries was evaluated by second language learners.

The two opening sections are primarily a review of the literature concerning dictionary use, with references to the questionnaire-based studies of Barnhart, Quirk, Tomaszczyk, Baxter, and Béjoint. The findings of Tomaszczyk and Baxter regarding users' stated preferences for certain dictionary types are summarised in the claim that "it was found that users of bilingual dictionaries generally found them less helpful than English/English dictionaries". This seems to be a misrepresentation of the results as reported by Tomaszczyk and Baxter, as in Tomaszczyk's study users were found to make greater use of their bilingual dictionaries, despite a high evaluation of their monolingual dictionaries, while Baxter's subjects were roundly critical of their monolingual dictionaries, showing strong preference for bilingual dictionary use.

MacFarquhar and Richards' purpose in reviewing the literature, however, is to foreground the role of the learners' monolingual dictionary, and the challenge faced by these dictionaries to define words accurately in language the learner can understand. They point out that whereas bilingual dictionaries tend to use the "definition by synonym" technique, monolingual dictionaries tend to have longer entries combining a variety of defining styles. A gap in the research exists, according to MacFarquhar and Richards, because, although defining practice has been discussed from a theoretical standpoint, "there is little empirical data on the practical consequences of different methods of definition".

MacFarquhar and Richards identify three main approaches to defining in monolingual dictionaries.

- 1) The use of restricted defining vocabularies of 1500 to 2000 words which are used to define all the words in the dictionary (as in LDOCE).
- 2) Particular effort to write clear and unambiguous definitions bearing in mind the needs of non-native speakers of English, while not restricting definitions to a defining vocabulary (as in OALD).
- 3) The use of virtually unlimited defining vocabulary, depending on the terms being defined (as in WNWD - Webster's New World Dictionary).

Section three of this paper is a preliminary discussion of the relative merits of these three practices, with particular attention paid to two problems MacFarquhar and Richards consider to be inherent in restricted vocabulary definitions. According to MacFarquhar and Richards, users of learners' dictionaries cannot be guaranteed to know even a basic set of defining terms. They also argue that the use of conceptually simpler words in a limited defining vocabulary may actually create longer, clumsier and more inaccurate definitions. Stylistic judgements are of course subjective, but MacFarquhar and Richards cite a number of scholars and lexicographers who have expressed doubts about the precision of restricted vocabulary definitions. West, for example, in his 1935 monograph on definition vocabulary, acknowledged that restricted vocabularies may be "better able to define the concrete than the abstract", and may also result in long and awkward definitions:

in defining with an unlimited vocabulary, we can select one or two apt words which match the idea. In defining with a small vocabulary, we are compelled to explain at length. The less the user knows, the more carefully we have to explain, and the more difficult it is to explain.

(West 1935:13)

MacFarquhar and Richards also refer to other linguists who believe that defining vocabularies achieve simplicity at the expense of accuracy, contrasting this view with

that of Bauer (1980), who argues in favour of restricted vocabulary definitions while acknowledging their occasional clumsiness.

OALD definitions are designed on the principle that "common words should be explained by means of other common words", and "less common words should be defined by the use of a wider vocabulary" (Hornby 1948), while introductions to various editions of Webster's New World Dictionary emphasize the comprehensiveness and precision of their definitions, with no reference to comprehensibility.

The subjects chosen for this study were 180 intermediate and advanced students from Asia and the Pacific enrolled in courses at the English Language Institute at the University of Hawaii.

In preparation for the test, 60 words "representative of the kinds of items learners are reported to consult dictionaries for" were selected, and were printed on individual cards together with their definitions from LDOCE, OALD and WNWD. The order in which the definitions from each dictionary appeared on the card was varied so that each dictionary had an equal proportion of definitions appearing first, second and third. The cards were then randomly grouped, and copied so that each of the 180 participants would receive a random set of ten different cards, and each of the sixty words would be evaluated by thirty different participants.

Participants each received an envelope and were given oral instructions to read each of the dictionary entries and tick the entry they found easiest to understand. Each subject also filled in a questionnaire giving background personal information.

Analysis of the subjects' responses revealed the following intelligibility preferences: LDOCE 51.5%, OALD 28.5%, WNWD 20%. Data from the questionnaire was used

to group subjects into higher and lower proficiency levels, and into bands according to their TOEFL score, but "there was no indication that the general preference pattern leveled out as proficiency increased".

Discussion

I cannot see any serious problems with MacFarquhar and Richard's choice of test words, although observation of student dictionary consultations might have been a better method of compiling the list. Instead the authors chose words of the types that Yorkey (1974), Béjoint (1981) and Cowie (1981) thought learners looked up in dictionaries; these were "rarer lexical items, idioms, culturally specific words, slang terms, phrasal verbs and compounds". Two categories of words that Béjoint's users claimed to consult their dictionary for - encyclopaedic words and abbreviations - are not included in MacFarquhar and Richard's list, possibly because coverage of such items varied too greatly from dictionary to dictionary.

MacFarquhar and Richards were careful to ensure that arbitrary factors such as word order, order of occurrence or fatigue did not affect the subjects' judgement. No sources were given for the definitions, and so subjects could not manifest bias for a particular dictionary unless they recognised the format. Unfortunately we are not told whether the subjects were familiar with any of the dictionaries used in this survey, and whether they had received any training in dictionary use. Such information about dictionary use or ownership could usefully have been elicited in the questionnaire. The validity of the findings of the study might be affected if, for example, the subjects were already regular users of LDOCE.

The finding that such a high proportion of LDOCE entries were judged to be "clearest" is an important one, and as far as I know this is the first study to test the acceptability of different defining styles. When subjects were asked to tick the definition they found easiest to understand, however, there were a number of factors

which might have influenced their choice. Presumably unfamiliar words would make a definition less easy, but so too might other factors unrelated to defining practice, such as numerical codes (Hornby's verb patterns), odd-looking etymological notes (WNWD), and the tilde (more frequently used in OALD and WNWD than in LDOCE). What MacFarquhar and Richards were really monitoring was the users' overall impression of the look of the entry. In order to monitor the actual readability of the entries, subjects would have had to complete some task.

Moreover, MacFarquhar and Richard's findings only go halfway to providing "data on the practical consequences of different methods of definition". The experiment goes some way to proving that learners recognise the words in a restricted vocabulary definition better than those in a non-restricted definition (although we cannot be sure that it is just the limited vocabulary that makes LDOCE entries easier to understand). However the experiment does not confront the second criticism levelled at restricted vocabulary definitions - the criticism that they are inaccurate and difficult to process - because the subjects were not required to prove their understanding of the entries in any way.

MacFarquhar and Richards are aware of the defects of their study:

it should be noted that it was only learners' perceptions which were measured, and not how helpful the dictionaries actually are. A follow-up study could investigate the relationship between perceived intelligibility of definition and the actual learning that takes place.

(1983:122)

Although he does not refer to MacFarquhar and Richards' work, Bogaards (1991, reviewed in 1.2.7.) aimed to measure the helpfulness of different types of dictionary definitions as MacFarquhar and Richards suggest. Bogaards used Dutch-speaking

subjects and French bilingual and monolingual dictionaries. I am not aware of any similar studies with English language learners.

1.3.7. Bogaards (1990)

This is the second paper by Bogaards to be reviewed in this chapter, but whereas Bogaards 1991 compared the value of various dictionary types, Bogaards 1990 is concerned with dictionary users' search strategies. Two questions were addressed in this study:

- 1) Does the dictionary user behave in a systematic way when searching for expressions in the dictionary?
- 2) If yes, how can this behaviour be explained?

Study One

The materials were 32 noun-adjective combinations, 16 French, and 16 Dutch. Both language sets were equally made up of four categories of adjective-noun combination: frequent adjective plus frequent noun (eg *un trou perdu*), less frequent adjective plus less frequent noun (eg *un mandat tacite*), frequent adjective plus less frequent noun (eg *une vieille taupe*) and frequent noun plus less frequent adjective (eg *un vin capiteux*).

The subjects were 28 Dutch-speaking 3rd year university students of French. They were required to underline in each of the 32 noun-adjective combinations the word that they would look up if they were required to check the meaning in a dictionary.

Bogaards' first research question was answered by the finding that the subjects did choose look-up words in a systematic way. In cases where the adjective in the combination was a frequent one, they overwhelmingly chose to look up the noun. Most also chose to look up the noun in cases where both the adjective and the noun

were less frequent. Only in those expressions where a less frequent adjective was paired with a frequent noun did the majority elect to look up the adjective.

The results of the study also went some way towards explaining the reasons for the subjects' look-up choices. Seven possible criteria for look-up word selection are examined.

1) Grammatical and lexical words. The subjects never underlined grammatical words such as prepositions. These words were sometimes provided in parenthesis (eg (*ne pas avoir*) *froid aux yeux*), and the method of presentation may have affected subjects' choice. Bogaards wondered what subjects would do in the case of expressions made up largely of grammatical words, such as *ne faire ni une ni deux*.

2) Word order. In both languages, subjects tended to choose to look up the second element more frequently than the first. This challenges the widely held view that dictionary users look up the first element in fixed expressions.

3) Word frequency. The most important factor in the choice of look-up word was frequency. In 93% of the French expressions, and 76% of the Dutch expressions, subjects underlined the less frequent of the two words. The difference between results for the two languages can be explained by the fact that the subjects were native Dutch speakers, who knew all frequent Dutch words but not all frequent French words.

4) Word class. Subjects preferred to look up nouns in 62% of the French expressions and 69% of the Dutch expressions. Where there was no difference in frequency between the noun and the adjective in the expression the preference was even stronger - 67% for the French expressions and 70% for the Dutch expressions. Word class overrode frequency as a criterion for selection in certain cases: there were more instances of subjects underlining a frequent noun combined with a less frequent

adjective than there were instances of subjects underlining a frequent adjective combined with a less frequent noun.

5) Syntactic hierarchy. Subjects generally chose to look up nouns, perhaps because the noun usually plays a central role in the syntax of noun-adjective combinations. However, in *rouge comme une pivoine*, where the noun is dependent on the adjective, 27 out of the 28 subjects still underlined the noun.

6) Semantic value. Bogaards has reservations about the theory that look-up choice is influenced by the relative meaningfulness of the words in the expression. He argues that "semantic load" is not a useful criterion for the selection of likely look-up words, because there is no means of accurately calculating the semantic weight of a word. He provisionally recommends the criterion of frequency as a more successful means of predicting subjects' look-up choice.

7) Figurative use. Bogaards acknowledges that the expressions were not controlled for degree of idiomaticity, and varied in this respect. In some cases subjects behaved atypically in their choice of look-up words within more idiomatic expressions. For example about two thirds of the subjects chose to look up the adjective *froid* in the very idiomatic (*Ne pas avoir*) *froid aux yeux*, although in other, less figurative, expressions containing frequent nouns and adjectives the noun was commonly selected. Bogaards comments on the problem of defining figurative use, and the difficulty of creating sets of expressions equivalent in terms of syntax and word frequency, and comparable for idiomaticity.

Bogaards acknowledges that his first study had a number of weaknesses: the sample was small both in terms of subjects and items, foreign and mother-tongue items were mixed together in the same study, and subjects were given no indication of the type of dictionary under consideration. Study Two was designed to avoid these weaknesses.

Study Two

In this study 52 French expressions and 52 Dutch expressions were used, but two parallel versions of the task were created for each language, so that each subject only had to judge 26 expressions. The same frequency categories were retained, but there was greater syntactic variety in the expressions; many were not adjective-noun combinations, but included verbs, adverbs, grammatical words or two noun elements. The expressions were analysed and coded in terms of word class, complementation, co-ordination and subordination.

This time a monolingual dictionary was specified for the task:

Si vous aviez à vérifier le sens exact des expressions suivantes dans un dictionnaire monolingue, sous quel mot le chercheriez-vous en premier lieu?

(1990:84)

For each expression responses were allocated to one of four categories: the first element, the second element, another element, or "empty" (no response, or more than one word underlined).

615 task sheets were analysed. For the French expressions, the respondents were 287 francophones at high school and tertiary level, and 51 foreign students studying at the University of Poitiers. For the Dutch expressions, the respondents were 244 native speakers of Dutch at high school and tertiary level, and 33 foreign students studying at the University of Leiden. In both cases, the foreign students came from a variety of language backgrounds. The results of the study are reported in terms of three categories of subject for each language: university students, high school pupils and foreigners.

In Study 2 the choice of look-up word did not appear to follow such a clear set of rules. Defining systematic behaviour as that occurring when 75% or more of subjects made the same choice, Bogaards found that behaviour was systematic in about half

the cases. He acknowledges that this might not be enough to convincingly answer his first research question, but he identifies three patterns in the results which seem to provide evidence of "systematic traits" in the way subjects chose which words to look up.

1) In only a quarter of the expressions was there less than 60% conformity in the choice of one element. On the other hand there was between 90% and 100% conformity in subjects' choice for 20% of the expressions.

2) Subjects were equally divided in their choice of element to underline in the case of the two French expressions with the construction *adj.fr.<n.* (frequent adjective subordinated to less frequent noun), and the two Dutch expressions with the construction *n.fr.cod.<v.fr.* (frequent noun as a direct object subordinated to a frequent verb). As results for both expressions with the same coding were identical, subjects would appear to be conforming to some sort of system.

3) There was no statistically significant difference between the behaviour of native speaker students and school children, but there was a difference at $p < .001$ between native and non-native speakers. Were all subjects choosing look-up words at random, there would be no significant differences between groups.

When looking for answers to the second research question, Bogaards limits his discussion to the native speaker data, on the grounds that the non-native speaker group was too small and too heterogeneous to analyse. The results of this study are considered in terms of five main criteria for look-up choice; idiomaticity is not taken into account as a possible influential factor, and semantic load is only considered as a minor influence.

- 1) Grammatical and Lexical words. As in the first study, subjects chose to look up lexical rather than grammatical words. In expressions made up entirely of grammatical elements, subjects chose the element with the greatest semantic load, or the verb.
- 2) Word order. This seemed to play a more important role than Study 1 had led Bogaards to believe. A slight preference for the first element was revealed, particularly in Dutch.
- 3) Word frequency. Subjects generally chose less frequent words, particularly in the case of verbs. Frequency was a more important criterion for the French subjects than for the Dutch subjects.
- 4) Word class. For the Dutch speakers, word class appeared to be a stronger criterion than frequency. They tended to choose nouns, and had a less marked preference for adjectives. Verbs were their least preferred element. For the French speakers this order of preference was reversed, but frequency overrode word class as a criterion. In the case of expressions composed of a less frequent noun and a more frequent adjective or verb both groups concurred in choosing the noun.
- 5) Syntactic hierarchy. In co-ordinated expressions subjects from both language groups tended to opt for the first element. In expressions containing a subordinated element, there was a tendency to choose the hierarchically superior or independent element. This was particularly marked in French; in Dutch the preference for the noun was overriding.

Discussion

As in MacFarquhar and Richard's study, Bogaards' research questions are relatively narrow, being limited to one aspect of dictionary look-up procedure. A small amount

of easy-to-analyse data is obtained from each subject, and thus it is possible to work with a sample sufficiently large to be representative of the population.

The task Bogaards sets is virtually identical to question 19 in Béjoint's questionnaire (1981) and two of the test items described in Atkins and Knowles (1990), but whereas the earlier studies also addressed a wide range of other research questions and only touched on the subject of look up strategies for multi-word expressions, Bogaards created a 52 item test solely to investigate this one aspect of dictionary use. The reported preference for the first element and the lack of interest in looking up grammatical words is compatible with Atkins and Knowles' findings. (Béjoint notes a preference for the final element in multi-word expressions). Interestingly, Bogaards and Atkins and Knowles regard the task from slightly different viewpoints; Atkins and Knowles seem to see it as a test of subjects' ability to anticipate lexicographical practice, while Bogaards is simply interested in knowing how the subjects behave, without any reference to their dictionary-using skills, or the organisation of existing dictionaries.

Bogaards' task is less holistic and natural than many observation-based research tasks, such as those set by Ard, Hatherall, and Neubach and Cohen. As in the studies conducted by Miller and Gildea and by MacFarquhar and Richards, the lexical items that form the basis of the task were chosen by the researcher, not his subjects, and we have no means of knowing which (if any) of these items the subjects would be inclined to look up in real life. A context for the task is provided, but unlike most of the observation-based researchers featured in this chapter Bogaards does not aim to elicit even one stage in the process of normal dictionary-using behaviour. His subjects are primarily representatives of populations of French and Dutch speakers, who can act as specialist informants, rather than representatives of the population of dictionary users who look up definitions for multi-word expressions. For this reason the subjects' background knowledge is not taken into consideration; the assumption appears to be

that all of them (except the Dutch students working on French expressions in Study 1, and the "foreigners" in Study 2) will already know all the expressions given on the task sheet.

Although Bogaards does not discuss the possible practical value of his findings in this paper, an obvious use would be to inform lexicographers' choices regarding the placing of definitions for multi-word expressions. (This topic is raised in a later study, Bogaards 1992). It seems likely that subjects have greatest need to consult a dictionary definition of a multi-word expression when a) the expression is idiomatic, and means something different from the sum of its individual parts, and/or b) the dictionary user is still learning the language, as a native-speaker child, or a non-native speaker. Neither the special demands of figurative language nor the special needs of language learners are fully investigated in this research; Bogaards encounters problems when he explores both these areas, and in his second study he decides not to differentiate between figurative and non-figurative use, and to ignore most non-native speaker data. These decisions may not invalidate Bogaards' findings, but perhaps lessen the studies' practical value. Future research might build on Bogaards' work by retaining his experimental technique, while adding the parameters of idiomaticity and subjects' mother-tongue.

Idiomaticity proved to be an important factor in Study 1, where Bogaards found that figurative meaning on occasion overrode frequency and word class as a criterion for look-up choice. Despite this, Bogaards did not control for idiomaticity in either of his studies, apparently because he lacked a method of identifying figurative use, and because he could not make up lists of expressions in both French and Dutch which were identical in terms of idiomaticity as well as word frequency and syntax. He writes of the role of figurative expressions:

Jusqu'ici, l'importance de ce facteur est très peu claire, ce qui est sans doute dû en partie au caractère vague de la notion d'idiomaticité

mais en partie aussi à la difficulté de trouver et de comparer des expressions équivalentes à tous les points de vue, sauf en ce qui concerne le critère qui nous occupe ici.

(1990:83-84)

Because of this lack of control over idiomaticity, it is just possible that some of the differences that Bogaards noted between the French and the Dutch respondents may have been due to figurative use overriding frequency or word class, rather than the influence of the mother tongue.

The problem of equivalence is overcome in Bogaards 1992 by using expressions from one language only (French), and comparing the responses of subjects with different mother tongues. The problem of recognising figurative language use is not addressed in Bogaards' more recent studies, however. It could probably be dealt with most effectively by appealing to a panel of native speaker judges.

It is interesting to note that LDOCE1 (1978) explicitly states a policy of allowing idiomaticity to override frequency in the placement of definitions of figurative expressions:

An **IDIOM** is usually found under the word that has the most **IDIOMATIC** meaning. Thus **a bone of contention** is under **bone** because *bone* is used in a more **IDIOMATIC** way than *contention*. If all the words are **IDIOMATIC** then it will be included under the most unusual word. Thus **a pig in a poke** is under **poke**.

(LDOCE1 *Guide to the Dictionary* p xxvii)

In LDOCE2 (1987) this policy is abandoned in favour of placing the definition under the first important word in the expression - a policy also adopted by OALD. Research into the user-friendliness of these alternative policies would clearly be of value to the dictionary makers.

Although differences between the behaviour of native and non-native speakers, and between the behaviour of mother tongue French speakers and mother tongue Dutch

speakers, constitute some of the major findings of Bogaards' study, Bogaards' original research questions did not address these differences. The comparison of data from the two language groups appears to have been an afterthought, insufficiently integrated into the experimental design. In Study 1, differences between the data from the two lists of expressions can be put down to any one or more of three possible causes: the fact that one list was in French and the other in Dutch, the fact that the respondents in one case were native speakers and in the other case non-native speakers, and the fact that the two lists of expressions were not controlled for idiomaticity. In Study 2, non-native speakers of both languages made up a very small proportion of the sample. From the finding that their choices were significantly different from those of the native speaker school children and students, Bogaards draws the conclusion that:

il existe des comportements typiquement français ou néerlandais, différents de ceux que manifestent les étrangers. Il est donc permis de croire que les choix que font les sujets dépendent dans une large mesure de leur langue maternelle.

(1990:94)

However it seems just as likely that non-native speakers made different choices because they did not know the language sufficiently well to do as the native speakers did, and exercise the criteria of frequency, syntactic hierarchy, idiomaticity and word class. Presumably there was not sufficient equivalence between the French and Dutch expressions, in terms of the combination of frequency, word class and syntax, to allow Bogaards to correlate native speaker responses across the two languages. Maybe it is impossible to compile lists of expressions in different languages which are exactly equivalent in this way, but failing exact equivalence across languages, native speaker look-up strategies cannot really be compared.

Although Bogaards' evidence for the existence of "typically French" and "typically Dutch" look-up behaviour is open to question in this paper, a follow-up study (Bogaards 1992) provides stronger supporting evidence. Bogaards gave Dutch

learners of French a list of French expressions and found that even the most advanced, near bilingual, learners retained look-up behaviour patterns previously identified as typically Dutch.

Another way to test Bogaards' conclusion might be to set a task for French and Dutch subjects in a third language, such as English.

All in all, Bogaards' method of data collection seems an excellent way of acquiring a large amount of information that is relatively easy to categorize and quantify. The identification of the five criteria for look-up choice, and the discussion of the way these criteria interact, is of great potential interest to lexicographers.

1.3.8. Concluding comments on observation-based research into dictionary use and dictionary requirements

The first five papers in this section can be loosely termed as ethnomethodological and holistic, because they set out to observe "natural" dictionary use, rather than contrived behaviour taking place in a controlled experimental setting. In these studies, subjects were set a variety of language tasks: translating (Ard), composition writing (Hatherall), sentence forming (Miller and Gildea), reading (Neubach and Cohen) and vocabulary learning (Ahmed). In all but Miller and Gildea's study, where children were directed to use a dictionary, subjects were free to consult dictionaries as and when the need occurred.

These studies involved the collection of data which could not be quantified in the same way as test scores or multiple choice questionnaire responses. At the analysis stage, the researchers had to make their own choices about the categorisation and interpretation of the data. Second language research textbooks, such as Seliger and Shohamy (1989) advocate the use of a variety of different data gathering procedures to increase the reliability of the data in this kind of qualitative research. A subject's

oral report of an event, for example, can be compared with a video recording of that same event, with observer's notes, with questionnaire responses and so on, and by triangulating these sources contradictory evidence can be eliminated, and information that is confirmed by all the sources can be given greater credence. Most of the studies discussed in this section make use of at least two different data collection methods, although triangulation is not rigorously carried out in every case.

Seliger and Shohamy also recommend that researchers control the process by which they categorise their data, both by re-assessing samples that they have already categorised (to check intra-rater reliability), and by "second-marking" samples (to check inter-rater reliability). None of the papers discussed in this section make explicit reference to this kind of control, but some provide a much more detailed account of the process of data analysis, and the reasoning behind their categorisation system, than others do.

Qualitative research lays itself open to the charge of subjectivity and is easily reduced to the level of anecdote if the processes of data collection and analysis are not recorded in detail. Yet even when this kind of research is conducted with a high degree of rigour, with checks on inter-rater and intra-rater reliability, and the triangulation of data wherever possible, the question of transferability remains a problem. The richness of the data collected from interviews, task observation and oral reports necessitates that the researcher works with a much smaller sample of subjects than is normal in quantitative research. In the studies conducted by Ard and by Hatherall the precise number of subjects remains unclear, but was probably no greater than six or seven in each case. Neubach and Cohen worked with six subjects. Ahmed, also using the think-aloud method of data collection, managed to obtain protocols for 300 subjects - an enormously time-consuming task - but may have been overwhelmed by the quantity of varied data he obtained; many of the findings he reports are extrapolated from questionnaires rather than the think-aloud reports. Miller and

Gildea obtained more manageable data by setting their subjects to compose single sentences rather than discourse. The relative simplicity of the data made it possible for them to analyse many more responses than Ard, Hatherall, and Neubach and Cohen.

Two of the studies in this section (MacFarquhar and Richards, and Bogaards) simply required subjects to make selections from series of definitions of lexical items, under experimental conditions which differed from their normal working environment. In these studies the data was limited to information regarding user preference, and could be easily quantified and reported. Thus it was possible for the researchers to work with large numbers of subjects: 180 in MacFarquhar and Richards' study, and a total of 643 in Bogaards'.

In observation-based research, it is clearly necessary to make a compromise between size of sample, on the one hand, and level of investigation, on the other. Broadly speaking, the greater the number of research questions, the smaller the sample that can be observed. The studies in this section have made this compromise with varying degrees of success; the best planned studies recognise the necessity of compromise and adjust their research questions and data collection task accordingly, but in some studies there are clearly defects in the experimental design, which threaten to at least partially invalidate the findings.

1.4. Concluding comments on prior research into dictionary use and dictionary requirements

The studies described in this chapter do not provide much directly valuable information for designers of future learners' dictionaries. Some of the studies are concerned with the behaviour of native-speaker dictionary users rather than learners of English as a foreign language; they are relevant primarily because they suggest experimental approaches that might be applied in the future to the study of EFL

dictionary use. The findings of many of the other studies are ultimately inconclusive, either because they report on the beliefs and perceptions of dictionary users, rather than on the observed consequences of dictionary use, or because different studies of similar phenomena have resulted in contradictory findings. However, despite the fact that many of the studies described in this chapter suffered from design faults, and many others did not provide a complete account of experimental procedure, all raised important questions, and revealed new problems for further research to investigate.

From the literature there seemed to emerge three main areas where research was particularly needed.

1) In the first instance, more data was necessary to resolve the question of whether dictionary use improves language learning and language task performance. Three studies summarised in this chapter take a positive view. Tono (1989) reported significantly improved reading comprehension for subjects who used dictionaries, and two studies of vocabulary learning, Black (1986) and Luppescu and Day (1993), found that the retention of new lexical items was significantly higher in cases where learners had access to word definitions. However there were also three pieces of research in the literature which cast doubt on the usefulness of dictionaries. The large-scale studies of Bensoussan, Sim and Weiss (1984) unexpectedly found no appreciable difference in performance in reading comprehension tests between those with access to dictionaries and those without. Neubach and Cohen (1988) appeared to reach a similar conclusion when they reported that only the most advanced students in their study benefited from using dictionaries during reading comprehension tasks; likewise the errors produced by native speaker school children in Miller and Gildea's study (1985) seemed to suggest that they had acquired little or no productive word knowledge through dictionary use.

Implicit in the design of the questionnaires discussed in this chapter was the assumption that a dictionary was a useful tool for students. The questionnaire responses supported this assumption, with almost all respondents reporting regular use of some kind of dictionary, be it monolingual native speaker, monolingual non-native speaker, or bilingual. It remained to be proved, however, whether dictionary use brought these students actual benefit, or was merely the result of habit, supported by their teachers, and the publishers who promote dictionary purchase.

2) Further data was also needed in a second and related area, concerning the type of dictionary information most useful for EFL decoding and encoding. The questionnaire results discussed in this chapter provide strong evidence that monolingual dictionary use increases with proficiency. Ahmed's study (1989) also confirms this trend; his highest-achieving group made the greatest use of monolingual dictionaries, while low achievers used bilingual dictionaries or no dictionaries at all.

In the early stages of English language learning, monolingual dictionaries seem too difficult for learners to use properly; Baxter (1980) reported that most of his subjects disliked monolingual dictionaries and complained that they could not understand the entries, and Neubach and Cohen quote a number of comments by low-level students which reflect disappointment and frustration with the *Longman Active Study*

Dictionary:

I didn't go on after the first definition. I thought all the rest were just examples;

I don't understand this definition. What should I do - look up meanings of words in the definition? Where does it stop?

All these signs and abbreviations frighten me!

Actually the dictionary hardly ever helps me. I don't understand the definition and I feel that it hinders me more than it helps me.

(1988:8-9)

Other studies reflected subjects' esteem for monolingual dictionaries, but also some dissatisfaction. In Tomaszczyk's study (1979) monolingual dictionaries were rated more highly than any other dictionary types, yet most users inexplicably preferred to use bilingual dictionaries for every kind of language activity. And although 83% of the subjects in Béjoint's study (1981) claimed to prefer monolingual dictionaries, subjects also complained that they contained unsatisfactory definitions, insufficient examples and syntactic guidance, excessively long entries and incomprehensible coding.

Most of Béjoint's students used dictionaries designed for non-native speakers, such as OALD and LDOCE, but Béjoint did not see the value of the extra features these dictionaries contained when students did not receive training in how to use them:

Given [the students'] lack of sophistication, dictionaries intended for native speakers would unfortunately seem to be as useful for our students as EFL dictionaries.

(1981:220)

MacFarquhar and Richards' study (1983) suggests that students may benefit from using EFL dictionaries as opposed to native-speaker dictionaries, even without training. The study also suggests that there may be benefits in choosing an EFL dictionary which uses a limited defining language. Subjects in their study considered LDOCE definitions to be more comprehensible than OALD definitions, and considered OALD definitions to be more comprehensible than WNWD definitions. The claimed improved comprehensibility of definitions written with non-native speakers in mind is supported by Bogaards' finding (1991) that Dutch learners of French using a French learners' dictionary completed a translation task more successfully than their fellow-students using a dictionary intended for native speakers.

EFL dictionary entries may vary in the style and range of the defining language, and also in the quantity and type of examples they provide. All three of the major EFL

dictionaries include more examples of use than are found in native speaker dictionaries, yet surprisingly the literature does not provide evidence that dictionary examples are useful. Miller and Gildea (1987) had doubts about the value of examples as an aid to creative language production, and found that native-speaker school children exposed to larger numbers of examples did not make better use of new words. In Black's study (1986) little difference was found between the scores for words learnt with examples in the definitions, and the scores for words learnt without examples. I needed to investigate further to discover whether different defining styles, and the type and quantity of examples, affect the success of practical language tasks involving dictionary look-up.

3) More data was also clearly needed in a third area, to investigate possible variation in the behaviour of different types of EFL dictionary user. There are a number of findings in the literature which suggest that students from different language backgrounds may react differently to the same dictionary information, and may have different dictionary needs. For example many of the variations in the questionnaire findings of Tomaszczyk (1979), Baxter (1980) and Béjoint (1981) can best be explained as reflecting differences in the attitudes of dictionary users in Poland, Japan and France. Ard (1982) concluded that Spanish students are more likely than Japanese students to use their bilingual dictionaries successfully, and Bogaards (1990 and 1992) suggests that typically French and typically Dutch dictionary look-up behaviour might exist. However, although there exist different monolingual learners' dictionaries to cater for different levels of language proficiency, there is as yet little monolingual dictionary provision for differences between learners from different language backgrounds. Are these differences great enough to justify the creation of different styles of dictionary?

The studies described in the following chapters attempt to investigate more fully the three research areas identified here. All have been touched on in a variety of ways in

the literature, but have not yet been fully researched. The questions I seek to answer have direct relevance not only to lexicographers and English language teachers, but also to any learner of English as a foreign language. They are in fact more elaborate versions of two questions teachers are frequently asked in the EFL classroom: "Do I need a dictionary?" and "Which one should I buy?"

Chapter Two

Studies to investigate the effect of dictionary use on performance in a multiple-choice reading comprehension test

2.1. Introduction

In the light of prior research findings, I considered that my first research task was to investigate the extent to which dictionary use affects language task performance. Do learners perform better when they have access to a dictionary, or does dictionary use simply slow them down, without producing superior results? In particular, some of the issues raised by Bensoussan, Sim and Weiss (1984) in their investigation of the effect of dictionary use on the outcome of reading comprehension tests required clarification, because their results ran counter to their own expectations, and to teachers' and learners' intuitions.

Bensoussan, Sim and Weiss had anticipated that learners would be both helped and hindered by using a dictionary in the tests. On the positive side they expected that "the permitted use of monolingual and/or bilingual dictionaries would significantly raise examination scores", and on the other hand they foresaw that "the use of dictionaries would significantly increase the time taken to complete a test" (1984:270). When they conducted a number of experiments to test their hypotheses, these assumptions were seriously called into question. In all four studies no significant difference was found between the test scores of dictionary users and those who did not use dictionaries. There was also little correspondence between test scores and the time taken to finish the test, although students who did not use a dictionary tended to finish fastest, students who used bilingual dictionaries tended to be slowest, and there was a slight tendency for slower students to obtain lower marks.

Bensoussan, Sim and Weiss's research has been presented as providing evidence that dictionary use makes little difference in reading tests: students were just as fast but no more proficient when they consulted a dictionary in the examination room. These results did not confirm the fears of the examination administrators who had wanted to ban dictionaries from the examination room, but neither were they particularly helpful to the English teachers who had wanted to encourage dictionary use. There does not seem much point in advising students to use dictionaries if using dictionaries does nothing to improve test scores. Moreover these puzzled teachers would have every reason to enquire why the test scores did not improve, if, as Bensoussan Sim and Weiss maintained, dictionary use is normally beneficial to readers working with the right level of text.

The data collected by Bensoussan, Sim and Weiss provided convincing evidence against their two original hypotheses. However, perhaps because of the immense scale of the project, the researchers did not focus on those details necessary to answer the questions raised by the rather disturbing findings. Dictionaries are designed to help readers read more efficiently, and reading tests are designed to test reading efficiency; in order to account for the surprising failure of dictionaries to improve reading test scores in this experiment, we need to know more about the dictionaries used, the skills the candidates employed, and the reading comprehension test itself. None of these variables are discussed in any detail in Bensoussan, Sim and Weiss's paper, and for this reason I decided to recreate the conditions of Bensoussan, Sim and Weiss's project, working on a smaller scale but paying particular attention to the interface between candidates, questions and dictionaries.

2.2. The pilot Study

The subjects for this first experiment were 20 overseas students on their first day of an eight-week EAP course at Aston University, Birmingham. 18 had previously taken the

British Council IELTS test, with scores ranging from 4.5 to 8.0 (mean score 5.5). All 20 students intended to pursue postgraduate studies at British universities.

Each student took the same test, which consisted of two texts (812 words in all) and 15 multiple choice questions. The texts were both taken, unadapted, from *The New Scientist*, and according to the Fry readability formula had a readability level of 9th grade (text 1) and 11th grade (text 2). The test was a pilot version for a series being developed at the Centre for English Language Teacher Education, Warwick University, and it would normally have been accompanied by a 90 minute "English Usage" module. 9 or more correct answers out of 15 on this part of the test would be considered an acceptable score for meeting minimum University English language entrance requirements. The test is reproduced in full in Appendix 2.1.

Students were allotted a total of sixty minutes for the test, which consisted of 15 multiple-choice questions. 10 students were given a copy of OALD to use if they wished to do so, but of these ten subjects, only four chose to refer to OALD during the test. Subjects were not allowed access to any other reference books during the test.

In order to investigate the effect of dictionary use on test performance, data for the 4 dictionary users and 16 non-users were compared according to:

- i) the number of minutes they took to complete the test
- ii) their test scores.

Results

The average completion time for those who did not use a dictionary was 33.6 minutes (SD 6.5). The average completion time for dictionary users was 40.7 minutes (SD 4.8).

In this experiment, therefore, dictionary users took substantially longer to complete the test than non-dictionary users. All four dictionary users took 35 minutes or longer,

whereas all but five of the sixteen subjects who did not use a dictionary completed the test in under 35 minutes. It should be noted, however, that the subject who took longest to complete (49 minutes) did not use a dictionary.

Dictionary use had little effect on test scores. The mean score for those who did not use a dictionary was 9 out of a possible total of 15 (SD 1.5) (of these, the six students who had access to a dictionary but chose not to use it scored a mean of 8.8). The mean score for dictionary users was 8.7 out of a possible total of 15 (SD 0.7).

The results of this study tallied with Bensoussan, Sim and Weiss's finding that dictionary use did not effect test scores, but did not confirm their finding regarding speed of test completion. As in Bensoussan, Sim and Weiss's sample, a number of subjects opted not to use the dictionary that was made available to them. In my very small scale study, this meant that I was left with only a tiny number of dictionary users - too few to form a representative sample, or conduct statistical analyses. I therefore decided to repeat the experiment on a larger scale, permitting the dictionary-using group to consult any dictionary of their choice.

2.3. Study One

The subjects for this study were 83 overseas students at Warwick University. They took the reading test on the final day of a four-week preessional EAP course. The majority of the participants were postgraduates, and all of them intended to continue their studies at British universities.

Each student received the same test, which consisted of two texts (812 words in all) and fifteen multiple-choice questions. The texts were identical to those used in Study I. 40 students took the test without access to dictionaries. The remaining 43 students were allowed to use their own monolingual dictionaries (OALD, LDOCE or LASD) or bilingual dictionaries (Japanese, French, Turkish, Chinese, Polish, Korean and Thai).

All subjects took the test in the same examination room, and were allotted a maximum of 60 minutes. Those students who had been given permission to use their dictionaries were asked to draw a circle round any words on the test paper which they looked up. To minimize errors in the data, each subject was asked to write on the question paper whether dictionary use had been permitted, and what type of dictionary (if any) had been used. Subjects were then asked to confirm this information as they handed in the test.

As in the studies by Bensoussan, Sim and Weiss, monolingual and bilingual dictionary use was related to test score and the amount of time the subject took to complete the test. I also compared test score with the quantity of dictionary use (i.e. the number of words looked up), and noted which words subjects had chosen to look up.

Results

Although half the subjects were permitted to use dictionaries, as in Study One not all of this group actually used them. This means that the subject population can be divided into four groups:

DICmo = monolingual dictionary users	n =19
DICbi = bilingual dictionary users	n = 9
DICno = dictionary not used	n =15
NOTav = dictionary use not permitted	n =40

Data from these groups is summarised in Table 2.1

Table 2.1: mean correct comprehension items (max = 15)

Group	:	DICmo	DICbi	DICno	NOTav
Mean Score	:	10.7	10.8	10.8	10.9
s.d.	:	1.5	1.9	1.7	2.1

As in the studies by Bensoussan, Sim and Weiss, the difference in scores between those who used dictionaries and those who did not was non-significant.

Bensoussan, Sim and Weiss reported that there was no difference between high scorers and low scorers in the number of words they looked up. I therefore divided the dictionary users into high scorers (13-15 n=8), medium scorers (10-12 n=12) and low scorers (6-9 n=8). Table 2.2 shows the mean number of words looked up by each group.

Table 2.2: subjects' scores related to mean number of words looked up

Group	:	High	Medium	Low
Mean no of look ups	:	1.6	6.3	2.3
s.d.	:	0.8	6.0	1.7

The data suggests that there might be a tendency for high scorers and for low scorers to use their dictionaries less than intermediate scorers. There was, however, considerable variation between subjects, and generalisations seem unreliable: among the medium scorers, for instance, look-ups ranged from 1 to 23 words.

As in Bensoussan, Sim and Weiss's Study One, bilingual dictionary users seemed to have used their dictionaries slightly more than average; they looked up a mean of 6 words (sd=7.7, range=23).

Bensoussan, Sim and Weiss also reported a correlation between speed of completion and score achieved, with faster candidates gaining a higher average score than their slower companions. I therefore divided my subjects into three groups according to the time span within which they submitted the test: fast (submitted within 40 minutes, n=36), medium (submitted within 60 minutes, n=33) and slow (submitted only when required to do so at the end of the test, n=14). Table 2.3 shows the relationship between completion speed and score.

Table 2.3: Completion speed related to mean number of correct responses (out of a possible total of 15)

Group	:	fast	medium	slow
Mean no.of correct responses	:	11.5	10.4	9.8
s.d.	:	2.1	1.3	1.5

There is a significant difference here: [F(2,82) 5.43, $p < .01$]. Further analysis showed that this effect is due to a significant difference between the fast group and the slow group.

The medium group is not significantly different from the other two.

Whereas in the studies by Bensoussan, Sim and Weiss there was little correspondence between dictionary use and the time taken to finish the test, in this study there appeared to be a rather closer correspondence. In the fast group, only 19% of the subjects used a dictionary; in the middle group 36% of the subjects used a dictionary, while in the slowest group 64% of the subjects used a dictionary. This discrepancy clearly deserved closer investigation.

In summary, scores for the subjects in Study One were unaffected by dictionary use. The results indicated, however, that the fastest subjects also tended to achieve higher scores, and were the least likely to consult any kind of dictionary. There appeared to be a link between speed of completion and extent of dictionary use which had not been recognised in the studies of Bensoussan Sim and Weiss.

2.4. Study Two

In order to resolve this apparent discrepancy between my findings and those of Bensoussan, Sim and Weiss, I decided to repeat the experiment, gathering more precise information regarding subjects' speed of completion. I was also concerned that

Bensoussan, Sim and Weiss's random allocation of subjects to the experimental groups may have introduced some uncontrollable variables into the design of the experiment. In Bensoussan, Sim and Weiss's first study the three conditions - without a dictionary, with a monolingual dictionary and with a bilingual dictionary - were selected "randomly", and in the later studies there was a certain amount of free choice in dictionary use: "Of those students not using dictionaries at all, some students decided that they did not need a dictionary because it was too time-consuming, while others simply forgot to bring them and would really have preferred to use a dictionary" (p267). In my first main study, following Bensoussan, Sim and Weiss, the group to be allowed access to dictionaries was also selected randomly, and within that group those who opted to use bilingual dictionaries were those who happened to have their own bilingual dictionaries with them. It is likely that these students made a habit of using their bilingual dictionaries, carrying them with them wherever they went, and these subjects may have been less confident and weaker than those who did not carry dictionaries. I found that there was also within the dictionary-access group a number of students who considered that they should not have been allotted dictionaries, because they felt themselves to be too advanced. They very ostentatiously pushed the dictionaries aside and indicated that they wanted to join the group who had not been allotted dictionaries. In my second study I therefore decided:

- a) to match the two groups according to language ability;
- b) to conduct the test in two separate rooms, so that both groups were unaware that the other group was taking the test under different conditions;
- c) to allow access to one kind of dictionary only (OALD), rather than a range of monolingual and bilingual dictionaries; and
- d) to record accurately each subject's completion time.

Test administration

65 overseas students at Warwick University participated in this study. On the basis of scores on the preessional course entry test they were divided into two groups of matching ability. The two sets of subjects took the test in separate examination rooms. One group (31 subjects) was not given access to dictionaries during the test. The other group (34 subjects) was given access to OALD during the test.

Prior to the test all 65 subjects were asked to underline on a wordlist those words which they were not familiar with. The wordlist contained all lexical words in the text and question paper, with the exception of common words (those in Bands 1 and 2 of Hindmarsh's Lexicon (Hindmarsh 1980)).

Subjects were then given a maximum of sixty minutes to complete the test. They were required to indicate completion time on the test paper.

On completion of the test subjects in the group with access to dictionaries were required to indicate on the wordlist those words which they had in fact looked up.

Results

As in the studies conducted by Bensoussan, Sim and Weiss, test scores were analysed according to whether the subject had access to a dictionary, and the amount of time the subject took to complete the test. An analysis was also made of the words subjects indicated that they were not familiar with, and which words they actually looked up.

Subjects in both groups were evenly matched on the basis of the preessional course entry-test scores. Results from the pretest wordlist activity confirmed that there was little difference in the make-up of the two groups. The first group, who were to be given access to dictionaries, indicated that a mean of 16 words were unfamiliar to them (s.d. 7.6), the second group, who were not to be given access to dictionaries, were unfamiliar with a mean of 17.9 words (s.d. 8.2).

The subject population can be divided into three groups:

DICu = dictionary users	n =29
DICno = dictionary not used	n = 5
NOTav = dictionary use not permitted	n =31

Data from these groups is summarised in Table 2.4.

Table 2.4: mean correct comprehension items (max = 15)

Group	:	DICu	DICno	NOTav
Mean Score	:	11.0	12.6	10.7
s.d.	:	2.3	2.2	2.4

As in my first study and those conducted by Bensoussan, Sim and Weiss, there was no significant difference in comprehension scores between those who had access to a dictionary and those who did not. However, my results regarding dictionary use and speed of completion, which are summarised in Table 2.5 below, contradicted the findings of Bensoussan, Sim and Weiss.

Table 2.5: Time taken to complete the reading task (mins)

Group	:	DICu	DICno	NOTav
Completion time	:	37.3	38.4	25.8
s.d.	:	7.8	8.0	4.3

Whereas Bensoussan, Sim and Weiss reported that dictionary use made no significant difference to the speed of test completion, my subjects took significantly longer to finish the test if a dictionary was available, irrespective of whether they used it or not [$F(2,62)=24.4, p<.001$]. Some possible explanations for this puzzling result are put forward in the next section.

2.5 Discussion of the results of Studies One and Two

Before starting to investigate the relationship between EFL test performance and dictionary use, Bensoussan and her colleagues had made certain assumptions about the effects of dictionary use. They believed, among other things, that:

1. All students would prefer to avail themselves of the opportunity to use a dictionary during a test when permitted.
2. The permitted use of monolingual and/or bilingual dictionaries would significantly raise examination scores.
3. The use of dictionaries would significantly increase the time taken to complete a test.

(Bensoussan, Sim and Weiss 1984:270)

However, the results of the studies conducted in Israel suggested that these assumptions were false:

According to the three studies, the use of a dictionary has no significant effect on reading comprehension test scores based on multiple-choice questions. Neither does its use affect the time students need to complete the test. Moreover, even when permitted to use a dictionary, many students (mostly those with relatively high English proficiency) did not wish to do so.

(Bensoussan, Sim and Weiss 1984:270)

The studies at Aston University and Warwick University confirmed two of these findings. Test scores were not significantly affected by dictionary use, and dictionaries were not popular with all subjects; 26 of the 87 subjects who were allotted dictionaries in the three studies chose not to use them.

However, Bensoussan, Sim and Weiss's other finding was not confirmed in these studies. In all three tests, dictionary users took considerably longer to complete the test. I can

only speculate about the reasons for this difference. It is possible that Bensoussan, Sim and Weiss's subjects were more efficient dictionary users than my subjects, and therefore wasted less time when using their dictionaries, although there is no evidence for this hypothesis in either set of research findings. It is also possible that Bensoussan, Sim and Weiss's subjects were under more pressure to work at speed, and dictionary users intermingled with non-users may have been sensitive to the pace of the examination room. In my Study Two, where subjects with access to a dictionary worked in a different room from those without access, it was found that the whole group worked more slowly, including those subjects who did not in fact look up words. One interpretation of this behaviour is that, as the majority of subjects were dictionary users and thus worked at a slower pace, the non-dictionary users were not spurred to work more quickly by the sight of their colleagues submitting their completed papers.

Another possible explanation for the difference is that my subjects were making greater use of their dictionaries than their Israeli counterparts. However this does not appear to be the case as far as Bensoussan, Sim and Weiss's Study One is concerned, as their monolingual dictionary users are reported to have looked up a mean of 5 words, and their bilingual dictionary users are reported to have looked up a mean of 13 words in a three hour test, a figure that does not differ greatly from my results. My subjects, taking a one-hour test, looked up an average of 4 words in Study One, and 3.2 words in Study Two. (In the other three studies by Bensoussan, Sim and Weiss, the number of words subjects looked up was not reported.)

Bensoussan, Sim and Weiss were surprised that dictionary use did not help their subjects in their test, and they speculated that the students' lack of dictionary skills, their lack of knowledge of syntactic rules, or the difficulty of the test itself may have caused dictionary users to fare no better than those without access to a dictionary. The results

certainly suggest a breakdown at some point in the process. If reading tests are designed to measure the learner's ability to comprehend text, and if dictionaries are designed to aid reading comprehension, it is not unreasonable to assume that the test, the dictionary or the user is failing in its purpose when dictionary use cannot improve reading test scores. Analysis of my data suggests that the responsibility lies with all three agents: the test, the dictionary and the user.

The test

Bensoussan, Sim and Weiss suggest that some of their testees failed to cope with the text, and could not use their dictionaries effectively, because the text contained too high a proportion of unknown words. This explanation can be ruled out in my studies.

Bensoussan, Sim and Weiss based their suggestion on Johns' claim that "when more than approximately 50 per 1000 words are unknown, perception of overall structure may be effectively blocked, which in turn means that there is not enough in the way of context to allow successful guessing" (1980:9). Although Johns' "threshold effect" may have operated in the experiments recorded by Bensoussan, Sim and Weiss, where some subjects identified as many as 68 unknown words per 500-700 word text, my dictionary-using subjects identified a mean of only 6.7 unknown words in Study One, and a mean of 16 words in Study Two: the texts totalled 812 words. This suggests that the subjects recognised enough vocabulary to permit the successful application of guessing techniques - techniques which also help in the identification of meaning during dictionary consultation.

A more likely explanation for the apparent uselessness of the dictionaries as an aid to reading comprehension lies in the nature of reading comprehension tests. Most communicative reading tests are primarily concerned with testing reading skills rather than language knowledge. On the whole the student is tested on his understanding of the meaning of the text rather than on what he knows about individual words. Questions

which require the reader to recognise the function of a grammatical structure may be included, but questions depending on the understanding of individual lexical items are generally avoided because they do not enable the tester to generalise about the learner's overall reading ability. The specific lexical item may be one of a tiny number that one learner knows, yet it might be excluded from another learner's vast mental lexicon. If testees do not have access to a dictionary this approach to test design seems only right and fair. Nobody should pretend, however, that such tests reflect a real-life reading situation, where word meanings are not always recoverable from context, and where the information that the reader requires may often depend on precise understanding of a particular word.

Bensoussan, Sim and Weiss's paper does not give details of the types of texts subjects had to read, or the types of test question they had to answer. Analysis of my own test suggested, however, that the choice of test might be a very significant factor in determining whether dictionary users derived benefit from their dictionaries.

Only five of the fifteen questions depended to any degree on the comprehension of individual lexical items, and the test designers tell me that it was their intention that even these questions should be answerable from context alone. The other questions in the test required the reader to analyse language functions (eg "the main point of the first paragraph is to suggest that"), to process anaphora (eg "the car" (line 76) refers to"), or to extract factual information (eg "why are some aid agencies reluctant to buy the car?"). For such questions comprehension of overall context seemed to be more important than the ability to define a key word or expression.

The five questions which did seem to require detailed understanding of specific lexical items are reproduced here in full:

4) Lines 18-20 state that "laminated plywood components are slotted together like a giant jigsaw puzzle". This is so that:

- a) it can be manufactured anywhere
- b) any damaged part can be replaced (CORRECT ANSWER)
- c) people can choose different designs
- d) the car is corrosion and dent proof

6) The word "configurations" in line 25 is most likely to mean:

- a) strengths and weaknesses
- b) components
- c) shapes and sizes (CORRECT ANSWER)
- d) colours and forms

7) The word "just" in line 46 could be replaced by:

- a) alone
- b) right
- c) only (CORRECT ANSWER)
- d) already

13) The expression "in conjunction with" in lines 84-85 means:

- a) in co-operation with (CORRECT ANSWER)
- b) in competition with
- c) in co-ordination with
- d) in cahoots with

14) The word "assessed" (line 89) is most likely to mean:

- a) marked
- b) criticised
- c) measured
- d) judged (CORRECT ANSWER)

It should be noted that some of the words in these questions are more crucial than others.

Question 4 is difficult to answer correctly without some understanding of SLOTTED, and a knowledge of the word JIGSAW helps to contribute to that understanding, but DENT-PROOF can only help in the elimination of alternative d), and LAMINATED is not useful at all (although a subject who did not know the meaning of the word would not be able to guess that it was not important). In question 13 knowing the meaning of IN

CONJUNCTION WITH is not enough, the subject is required to make fine distinctions between CO-OPERATION, CO-ORDINATION and CAHOOTS.

It should also be noted that all five of the questions are to some extent dependent on context, and the reader cannot know which answer is correct without referring to the text. This should not mean that the dictionaries are unable to aid the reader, as few words can be defined independently of context, and it is for this reason that the dictionaries provide examples. It does mean, however, that the dictionary user's task is a harder one, as she must match context with context to find a meaning appropriate to the text she is reading.

Surprisingly, there was not a great deal of correspondence between the words that dictionary users indicated that they had looked up and the words needed to answer these five questions correctly. Table 2.6 lists "keywords" (ie words which needed to be understood before the question could be answered correctly) and shows the number of subjects who indicated that they had looked up these words. (None of the subjects looked up more than one keyword per question.) Table 2.6 also indicates that in most cases, subjects who looked a word up answered the question correctly.

Table 2.6: Look-up rates and correct answers for "keywords"

keyword	looked up	correct
jigsaw (puzzle)	12	8
slot together	5	5
corrosion proof	1	0
dent proof	3	2
laminated	3	2
configurations	9	8
in conjunction with	6	4
in cahoots with	9	4
in coordination with	1	1
assessed	8	3
Total	57	37

In summary, the texts used in the test were not sufficiently difficult to block perception of overall structure and thereby prevent successful dictionary use. The accompanying test items, however, tended to test general reading strategies rather than knowledge of specific words, and for this reason dictionary information was not often directly useful to the testees. Test writers have a policy of avoiding purely lexical questions, but readers functioning under non-test conditions may well need to focus on the meaning of individual words. In this case dictionary information might aid comprehension, whereas it did not appear to do so in Studies One and Two. Most testees who looked up keywords for the few test questions which depended on the comprehension of individual lexical items answered the questions correctly. This suggests, but does not prove, that they benefited from the dictionary information available.

The dictionaries

The data from Studies One and Two did not provide information concerning the comprehensibility of dictionary definitions. An analysis of the monolingual dictionary entries and the test items indicated, however, that the dictionaries did not always supply the information necessary to answer the test items correctly. I have no data on the entries in the bilingual dictionaries, but the monolingual dictionary definitions were not always helpful. This was particularly the case with OALD, where the examples were limited in number and not always applicable to the technological topics of the *New Scientist* texts.

The extent of the dictionaries' failure to supply information that the test candidates needed can be seen from the examples below. The following list gives definitions from the three dictionaries for five of the most important keywords that subjects indicated that they had looked up. (In OALD definitions, the tilde has been replaced by the full form of the word.)

SLOTTED

OALD v. *Provide with slots: make a slot or slots in*. SLOT n. is defined as a narrow opening, slit, groove or channel, or (figuratively) a right or suitable place. Examples give a context of vending machines, and (figuratively) of broadcasting and job-finding.

LDOCE v. *to put or be put into a slot*. SLOT n. is defined as a long, straight, narrow opening, or a place on a list etc. An illustration shows a slot in a vending machine. Two of the four examples for SLOT v. are: You buy this bookcase in sections and slot them together. "This box has a removable lid which slots back in like this", he said, slotting it into the box.

LASD v. *to cut a slot, to put into a slot, to find a place for*. SLOT n. is defined as in LDOCE. Examples illustrate the vending machine sense and the sense "find a place for".

CONFIGURATION

OALD n. *shape or outline; method of arrangement*: the configuration of the earth's surface.

LDOCE n. *the arrangement of the parts of something; shape*: the configuration of the moon's surface

(LASD has no entry.)

IN CONJUNCTION WITH

OALD *together with* (examples for CONJUNCTION, but not for IN CONJUNCTION WITH).

LDOCE *(a) combination of qualities, groups or events*: the army is acting in conjunction with (= in combination with) the police in the hunt for terrorists

LASD *in combination with; together with; along with*. (example as in LDOCE)

IN CAHOOTS WITH

OALD *be planning sth (esp sth disreputable), be in league*

LDOCE *in partnership (with), usu. for a dishonest purpose*: The bank robbers and the police were in cahoots./ The bank robbers were in cahoots with the police

(LASD has no entry)

ASSESS
OALD 1 <i>decide or fix the amount of (eg a tax or a fine)</i> : Damages were assessed at \$100 2 <i>appraise: fix or decide the value of (eg property), the amount of (eg income), for purposes of taxation; (fig) test the value of</i> : Assess a speech at its true worth
LDOCE 1 <i>to calculate or decide the value or amount of</i> : To assess the damage caused by a storm/ They assessed the value of the house at \$60,000. 2 <i>to judge the quality, importance or worth of</i> : EVALUATE : He's so lazy that it's difficult to assess his ability./ It's too early to assess the effects of the new legislation.
(LASD almost identical to LDOCE.)

It can be seen that OALD ignores the (non-figurative) sense of SLOT which is relevant to question 4 (whereas the LDOCE examples at SLOT manage to express the idea that it is easy to remove something that has only been slotted into place). OALD also provides no examples for IN CONJUNCTION WITH, and only provides examples with a financial theme for ASSESS. (LDOCE again guides the reader to the correct answer for question 14 by including the synonym "judge" in definition 2). LASD too lacks necessary information; there are no entries for two of the five keywords that subjects indicated that they had looked up. (Both LDOCE and OALD, however, provide guidance for answering question 6, by mentioning "shape" in their definitions of CONFIGURATION. By their provisos "disreputable" and "dishonest" they also steer the reader away from the IN CAHOOTS WITH alternative in question 13.)

These findings help to explain why dictionary users in the reading comprehension test did not always improve their score, even when they looked up keywords for questions which depended on the understanding of a particular lexical item. Many dictionary entries did not provide the testees with the information they needed to answer the test question correctly. Whether the correct answers to the test questions were themselves accurate definitions of the words in the text remains open to debate, because the answers reflect the communicative value of the word in context, rather than its decontextualised

signification, and value must always depend to some extent on the perception of the reader.

The users

Table 2.6 indicated that not all instances of dictionary use resulted in correct answers to the questions. My data does not provide a means of determining whether lack of dictionary skills led students to make mistakes, but I can identify two other possible reasons for this result: the inadequacy of the dictionary definitions themselves (discussed above), and the users' failure to look up other important keywords. Of the twelve subjects who looked up JIGSAW PUZZLE, for example, four answered the question incorrectly, and this may well have been because the question depended more heavily on an understanding of the word SLOTTED - which none of the twelve went on to look up.

Some dictionary users did not look up keywords at all. Although it is not usually taught as a dictionary skill, the ability to identify relevant words in a text is just as important as the ability to find their meanings in the dictionary. Subjects in this experiment seemed to lack this skill; not one, for example, accessed the dictionary for information about the word JUST (question 7), although the question depends entirely on the correct understanding of that word.

Subjects chose to look up a variety of words apart from keywords. Some of these words were relevant to the test questions, others apparently only of interest to the subject. All the non-key words that subjects indicated that they had looked up in Study Two are listed below in order of frequency. Their position in the texts is also indicated.

RUTTED (6) Text 1 paragraph 1

LOATH (5) Text 1 paragraph 6

POTHOLED (5) Text 1 paragraph 1

MASTERMINDED (5) Text 1 paragraph 1

KIT (4) Text 1 paragraph 6
 STAGNANT (3) Text 1 paragraph 7
 EPOXY/EPOXY RESIN (3) Text 1 paragraph 3
 DRAG (3) Text 2 paragraph 3
 GALVANISED (2) Text 1 paragraph 3
 TILTING (2) Text 2 paragraph 4
 ENTREPRENEURS (2) Text 1 paragraph 6
 PROPULSION (2) Text 2 paragraph 4
 FEASIBILITY (1) Text 1 paragraph 6
 CLEARANCE (1) Text 1 paragraph 4
 TRULY (1) Text 2 paragraph 4
 PLYWOOD-REINFORCED PLASTIC (1) Text 1 paragraph 4
 STABLE (1) Text 2 paragraphs 1 and 4
 CONTENT (1) Text 1 paragraph 2
 SEALED (1) Text 1 paragraph 3
 SPRUNG (1) Text 1 paragraph 4
 VALVES (1) Text 1 paragraph 4
 RELIEF (1) Text 1 paragraph 6
 TILT (1) Text 2 paragraph 4
 NON-TOXIC (1) Text 1 paragraph 3
 ADHESION (1) Text 1 paragraph 7

The data does not provide us with information regarding subjects' motives for selecting words to look up, but it would appear that some were taking the opportunity to use the dictionary to learn new vocabulary, or looked up words to answer their **own** questions about the meaning of the text, despite the fact that they were working under test conditions. According to Bensoussan, Sim and Weiss, the subjects in their experiments looked up words to answer test questions, and were not motivated by "the desire purely to

understand the text". In my study the subjects were mature and self-motivated students, used to taking responsibility for their own learning. Perhaps they used the dictionary in the test situation in the same way as they would use it in "real life", without too much regard for the demands of the test.

The majority of the words looked up occurred in text 1 (16 out of 21), and most of these occur in the opening paragraphs. This suggests that subjects were more enthusiastic about looking up words at the beginning of the test, and lost interest later on. We can only speculate as to whether subjects became bored, disillusioned, or perhaps more confident as they worked through the test.

2.6. Conclusion

When analysing their data, Bensoussan, Sim and Weiss did not relate the words their subjects looked up to the demands of the test questions, they did not check to see whether the dictionaries their subjects used dealt adequately with the words their subjects looked up, and they did not check whether look up of keywords in the test resulted in correct answering. All these considerations are important if we want to find out why dictionary use did not affect reading comprehension test scores. In my studies, it would appear that dictionary use did not affect test scores primarily because the test itself was made up of items which were not likely to be affected by the availability of a dictionary. However, in some cases where dictionary use might have aided the subjects, either the dictionaries themselves did not provide the necessary information, or the users failed to identify the words in the text which were most crucial for correct answering of the test questions.

One further possibility, that the subjects failed to absorb potentially useful dictionary information, could not be investigated in these studies, but remained an important topic for investigation. Studies One and Two provided no means of knowing how well subjects understood the dictionary entries, yet it is obvious that if subjects failed to comprehend

the definitions of the words they looked up, their reading test scores could never improve as a result of dictionary use, no matter how well they identified keywords in the text.

Bensoussan, Sim and Weiss touch on the dictionary skills variable when drawing conclusions from their experimental findings:

One explanation [for the lack of difference in test scores] could be that students simply do not know how to use the dictionary efficiently during an examination. In this case, by definition, test results would not be affected.

1984:271

Although they chose not to investigate this possibility in greater detail, their summary of results from the questionnaire used in conjunction with their tests suggests that their Israeli subjects did indeed experience difficulty finding and interpreting dictionary entries. Problems associated with questionnaire-based studies of dictionary-using habits have been discussed in Chapter One, and we may query the ability of the respondents in Bensoussan, Sim and Weiss's study to remember and report back; the results at best provide an overview of their students' impressions of dictionary use, rather than objective facts about their behaviour. Nevertheless the issue of dictionary comprehensibility remains a vital one; learners' definition-reading skills would clearly have to be investigated, using a methodology which could provide more reliable data by permitting unobtrusive observation of dictionary consultation as it occurred.

It was therefore to the issue of comprehension and interpretation that I turned in my next series of studies, to investigate the extent to which, once learners had identified the word they needed, and had located the necessary information in their dictionary, they were able to make sense of this information and put it to practical use.

Chapter Three

Study Three: the effect of different dictionary defining styles on productive dictionary use

3.1. Introduction

The studies reported in Chapter Two were designed to investigate the effect of dictionary use on reading comprehension test results. No significant difference was found between the scores of subjects who used a dictionary during the test and the scores of subjects who did not, and three reasons were suggested for the fact that dictionary use did not improve test scores: most test items did not require thorough understanding of difficult words in the texts, the dictionaries did not always define words in the sense in which they were used in the texts, and the users did not always look up the words which were most likely to help them answer the test items.

The results suggested that dictionary use would be a worthwhile reading strategy only if all the following conditions were met:

- i) the reading purpose (whether imposed on the reader by others, or decided by the reader herself) necessitated the comprehension of unknown words in the text
- ii) the reader was sufficiently skilled to recognise which unknown words in the text it was necessary to understand in order to achieve the reading purpose
- iii) the dictionary provided meanings for the unknown words which matched the senses intended in the text
- iv) the reader was able to interpret the dictionary definitions correctly.

Although I was able to examine the first three of these conditions in my first two studies, little data was obtainable by this means regarding the fourth condition, which concerns the readability of dictionary definitions, and users' dictionary-reading skills. In the majority of cases examined in my second study, subjects who looked up a keyword went

on to answer the relevant test item correctly, but a causal link between the two events remained unproven. Likewise in the cases where relevant look-up did not result in the correct answering of the test item, I had no means of knowing if the dictionary definition had been misread, or whether some other problem concerning the text, or the test item, played a part.

I therefore decided to investigate the readability of different types of dictionary definition by collecting and analysing written data produced after dictionary consultation. I chose a writing task rather than answers to multiple-choice questions as a means of data collection because I felt that any questions I asked about word meaning might influence the subjects' interpretation of the definitions (as perhaps happened in Black's study (1986)). I also felt that tests of vocabulary knowledge which do not require the subject to make active use of words merely test a subject's potential to recognise words, rather than their present understanding (see Corson 1983). A free writing activity might reveal the subjects' own interpretation of the definitions they had read, and their semantic grasp of the target words.

The productive effectiveness of dictionary definitions has been investigated in the past in two main ways. In one approach a corpus of spontaneously produced errors is compiled, and dictionary definitions are examined to see whether the errors might have been corrected, had the writers consulted a dictionary. In the other approach, examples of writing produced after dictionary consultation are collected, in order to determine what word knowledge the writers have acquired through dictionary use.

The focus of the first type of research has largely been on EFL learning situations; researchers such as Huang (1985), Nesi (1987) and Meara and English (1988) have gathered examples of errors produced by non-native speaker writers, and then investigated the relevant entries in major EFL learners' dictionaries. Huang concentrated

on Chinese learners' errors and their treatment in the *Oxford Advanced Learner's Dictionary* (OALD), the *Longman Dictionary of Contemporary English* (LDOCE) and the *Chambers Universal Learners' Dictionary* (CULD). He was primarily concerned with the grammatical information the dictionaries gave, such as valency, number agreement, and the use of the article, whilst in a similar study (Nesi 1987), I considered the treatment in the same three dictionaries of semantic (lexical) errors produced by EAP students in Britain. My study was taken by Meara and English (1988) as the starting point for a larger-scale investigation of learners' errors and their treatment in the *Longman Active Study Dictionary*. All three studies have usefully identified areas where dictionary definitions ignore or even appear to condone the mistakes writers make, and their findings have led to some changes in learners' dictionary entries.

The second approach to studying the productive use of dictionary definitions is complementary to the first, but so far only one or two small studies of EFL dictionary use have adopted it, notably Jain (1981) and Black (unpublished, but reported in Maingay and Rundell 1987). In contrast, there has been a considerable amount of research of this type using native speaker data; Mitchell (1983) and Miller and Gildea (1985, 1987), for example, used target words and sentences produced by native English-speaking children in large-scale studies of school dictionary use.

A certain degree of overlap is to be expected in the findings of research adopting the first and the second method, especially where weak definitions condone typical misconceptions about the meaning and grammar of words. However, whilst the second approach may well miss many common errors that dictionaries fail to correct, it may also bring to light errors actually induced by dictionary use. Both types of research are therefore necessary if we wish to obtain a rounded picture of the productive effectiveness of dictionary definitions.

I decided to adopt the second research method, which has been so little used in non-native speaker studies, while comparing subjects' performance across the most recent editions of three learners' dictionaries, each noted for a different defining style: OALD, LDOCE and COBUILD. In addition to the studies discussed above, my experimental design owed something to the work of MacFarquhar and Richards (1983) who also compared the usefulness of three dictionaries noted for their different defining styles.

A brief account is given below of the main features of the three dictionaries I decided to use in the study reported in this chapter - OALD, LDOCE and COBUILD.

The **Oxford Advanced Learner's Dictionary (OALD)** appeared in its fourth edition in 1989 - earlier editions were in 1948, 1963 and 1974. It has the advantage of being the best known learners' dictionary, and it is often recommended by teachers on the grounds that they themselves used it when they were learning English. The authority of the Oxford name must also surely add to its popularity. Earlier editions of OALD featured Hornby's verb patterns, where verbs were given one or more coded numbers, from one to twenty-five, with finer distinctions indicated by the addition of a letter. (Verb pattern 18C, for example, was the pattern taken by HAVE in the sense of "wish", "experience" or "cause" - as in *What would you have me do?*). OALD4 adopts a more transparent coding scheme, with 33 different combinations of letters to represent various valency patterns. The preface to OALD4 claims that these codes can be "easily learnt", because the letters are for the most part abbreviations for grammatical categories; "Tn", for example, stands for a transitive sentence, while "Tn.pr" indicates a transitive pattern with an adjunct prepositional phrase - *the visiting speaker thanked the chairman for his kind remarks*.

OALD's policy is to define words in simple language, but without any overall restriction on defining vocabulary. OALD examples have, for the most part, been invented. They

have the virtue of being self-contained, but it is sometimes difficult to imagine how they might fit into naturally occurring discourse.

Longman Dictionary of Contemporary English (LDOCE) was first published in 1978, and appeared in its second edition in 1987. Its chief contribution to the learners' dictionary field has been its use of a controlled 2000 word defining vocabulary with the intention that its definitions should be easier to understand than those of OALD. Some reviewers have queried the wisdom of this. For one thing, the defining vocabulary, which was based on West's General Service List of English Words (1953), claimed to be "the only frequency list to take into account the frequency of **meanings** rather than the frequency of word forms" (LDOCE1 introduction). This meant that some of the words in the defining vocabulary were not particularly high-frequency words, and might not be known to dictionary users; BACTERIA, ASHAMED, COWARDLY, INFECTIOUS and WORM, for example, are all words in the LDOCE controlled vocabulary. It was also argued that some meanings were difficult to express when vocabulary was restricted, and that this led to rather clumsy paraphrases which were actually longer and more difficult to understand than those in OALD. In response to this criticism the 1987 edition of LDOCE occasionally uses non-restricted words in its definitions. However, some oddities remain; STEAK, for example, is defined as a *piece of meat from cattle* because BEEF is not one of the controlled defining words (see Hanks 1987). Also in LDOCE2 examples no longer remain within the restricted vocabulary range. Instead most of them are attested instances of use taken from the Longman Citation Corpus.

LDOCE1 also added to learner lexicography by extending grammatical coding to include adjectives, nouns and adverbs. This experiment is generally agreed to have been even more impracticable from the learners' point of view than the verb patterns originally given in OALD. West (1987) found that in-service teachers rejected both the old-edition OALD and the first edition LDOCE grammar codes: "Teachers felt that both systems

presupposed a linguistic sophistication which most students do not possess, and no one claimed that they had successfully trained their classes to use either system, most stating that they did not even try" (1987:62).

Like OALD4, LDOCE2 has greatly modified its grammar coding system so that meanings are more transparent; "A", for example, indicates an attributive adjective, and [the+P] indicates a plural noun with obligatory definite article.

The appearance of **Collins COBUILD English Language Dictionary** in 1987 heralded a number of interesting new departures for British lexicography. Three features in particular set COBUILD apart from its forerunners: its dependence on a computerised corpus, then running at over twenty million words (the Collins Birmingham University International Language Database), its use of an "Extra Column" to set grammatical information apart from meaning-related material, and its "folk" defining style.

COBUILD concordancing revealed new facts about word patterning which the lexicographer could not arrive at by intuition alone (see Sinclair 1987), and probably no large-scale dictionary project will ever again proceed without a corpus and concordancer. However the COBUILD corpus, at least at the time of the development of the COBUILD dictionary, seems to have been more of a haphazard collection of the texts then available in machine-readable form than a carefully balanced representation of English in the late eighties. The corpus is a little top-heavy with modern literary dialogue, and under-resourced as far as language relating to recent technical developments is concerned (and oral data, which is, of course, always the most difficult to obtain). The literary bias of the corpus can be discerned in many of the examples which include references to fictional characters (although minor changes have sometimes been made to the original citation in the hopes of removing "unnecessary distracting information"). For instance, we find as an example of BEYOND: *"They had no money beyond Sir Arthur's*

salary", and as an example of MATTERS: "*the family, the cottage and Twickenham were all that mattered to me*".

Such decontextualised references may well puzzle the user, and detract from the intended purpose of the example, which is to clarify meaning. Rare and unusual words, which also sometimes appear in COBUILD example phrases, serve a useful purpose according to Fox (1987), who cites the example at the entry for CIVILLY, clearly uninfluenced by the principles of a controlled defining vocabulary: "*I made my farewells as civilly as I could under such provocation*". Fox argues:

many teachers would dismiss this as a bad example because it is difficult. We would argue that its very difficulty makes it a good example for the word it is exemplifying, because it is typical of how the word is actually used.

(1987:146)

Opinions remain divided as to the merits of using straight citations to exemplify meaning, on the other hand the Extra Column feature in COBUILD has met with almost unqualified approval. According to the editor:

one important reason for creating the Extra Column is to keep the main dictionary text simple and accessible. We did not want to keep interrupting the flow of entries with abbreviations and technical terms. Nor did we want to have hundreds of grammar codes, as some dictionaries have, which can only be understood by looking up another part of the book.

(Introduction to COBUILD)

It is good that the extra column provides grammar as an optional feature which the reader can turn to if and when she needs it, but it would be wrong to assume that, just because it has been placed to one side of the entry, the grammar information itself has become more readable and easier to understand. Unfortunately nobody has yet found an economical alternative to grammar coding, and the system employed by COBUILD is only transparent to the extent that the systems of LDOCE2 or OALD4 are transparent (PRONPOSS stands for possessive pronoun, for example, and N PART for partitive

noun). Each COBUILD grammar code is listed and explained at its correct place in the alphabetical entry list of the dictionary, but it is doubtful whether the user finds it any easier to find and consult in this way than if it had been placed separately at the front of the dictionary as in LDOCE, or at the back as in OALD.

COBUILD's folk defining style was adopted because it was considered that it would be easier for the learner to read and internalize. The style makes little use of the abbreviations and typographical conventions which are typical of most dictionaries, whether intended for the native or the non-native speaker, and COBUILD lexicographers contrast it favourably with the ordinary "lexicographic definition style":

In my own experience lexicographic definitions, however elegant and logically constructed (indeed particularly when elegant and logically constructed) can be unhelpful as an aid to learning new meanings. Lexicographic definitions have a curious tendency not to stick in the mind, whereas the immediacy, the accessibility and the vividness of folk definitions often makes them more memorable and consequently more likely to be of help in both decoding and encoding.

(Stock 1988:86-87)

Many teachers and reviewers also respond warmly to the folk definition; Tadros, for instance, in her review of COBUILD proclaims:

anyone who consults the dictionary will realize that items are not, as in other dictionaries, explained in a detached way, but will feel that there is a real concern for them as users.

(1987:20)

It must be borne in mind, however, that COBUILD users who become accustomed to the folk definition style will not get such a good preparation in dictionary reading as a skill, because the style is not to be found in reference works from other publishing houses.

The difference in defining style between the three dictionaries can best be seen by comparing OALD4, LDOCE2 and COBUILD definitions for the same word. Listed below are their definitions of ABASE:

OALD4 - lower oneself/sb in dignity; degrade oneself/sb

LDOCE2 - to make (esp. oneself) lose self-respect; make HUMBLE

COBUILD - If you abase yourself, you behave in a way which shows that you accept that something or someone else is much more important than you are.

Of the three, OALD makes the greatest use of symbols and abbreviations, and COBUILD the least. It can be seen that the OALD and LDOCE definitions are brief and fragmented, while the COBUILD definitions are written in continuous prose. In a study of the readability and productive effectiveness of the three defining styles, it might be expected that the abbreviated and fragmented OALD definitions would cause the greatest problems for learners, while the flowing COBUILD definitions would offer the most help to learners completing an encoding task. After all, the COBUILD team had precisely stated their aim:

to create a dictionary that would not merely help readers in decoding texts, but that would hold up models that would be of assistance to learners in encoding English.

(Hanks 1987: 117)

On the other hand, the length of the COBUILD entries and their sometimes obscure examples might tip the balance in favour of LDOCE, which perhaps represents a "happy medium" in terms of fluency and style.

So far, despite numerous reviews and articles concerning the three dictionaries, no full-scale study has been undertaken to compare their usefulness, and studies which have investigated the causes of productive failure after dictionary consultation have not pointed to any one dictionary style that might help to prevent misreading.

In fact, prior studies involving productive dictionary use offer a rather inconsistent picture of the causes of error after look-up. Jain (1981) provides interesting but somewhat anecdotal evidence to support his view that the dictionaries' neglect of certain aspects of semantic structure is the major cause of lexical error. According to Jain, some errors are caused by confusing one-word synonyms in the dictionary definitions; for example, *DRESS* was given as a synonym for *CLOTHES* in *OALD* and *LDOCE*, so a learner produced the sentence *Are these new dresses for her son?* Jain believes that other errors are caused because dictionaries do not provide an explanation of basic differences between often-confused words, such as *KEEP* and *REMAIN*, and *CONVEYANCE*, *TRANSPORT* and *VEHICLE*, while a third kind of error is caused because the dictionaries fail to give explicit information regarding selection restrictions; for example, although *LDOCE* informs the learner that *COLLIDE* is intransitive and can be used with or without *WITH*, it does not mention further selection restrictions, so one of Jain's students happily produced *I saw his car collide with my own eyes.*

Jain was using early editions of *OALD*, *LDOCE* and *CULD*, and some of the defects he drew attention to have since been corrected. In the second of the two studies examining the written products of EFL dictionary use, Black (unpublished) did not complain of the inadequacy of dictionary information. She claimed that the Longman dictionary entries her subjects consulted would have enabled them to avoid all the errors they made, if only they had been properly understood. In Black's study, conducted under the auspices of Longman ELT Dictionaries and Reference Department, each given word tended to produce a different type of error, and each error type had the same underlying cause, for example semantic failure in the case of *DEBRIS*:

He always gives the debris of the meal to the dog

syntactic failure in the case of *REMINISCE*:

We stayed all morning reminiscing our childhood

and failure to recognize negative connotations in the case of NEW-FANGLED:

The new-fangled inventions of the computer are amazing.

Maingay and Rundell (1987) cite this experiment in support of their view that the best way to help dictionary users avoid errors is to emphasise and repeat information in the dictionary entry:

once a specific aspect of a word's meaning or use has been identified as a primary source of error, the dictionary writer can target the problem and deal with it by "overkill".

(1987:134)

The upshot of both the learner dictionary studies seems to be that errors in productive dictionary use can largely be avoided, if dictionary entries offer more information, expressed in a more accessible way. However this conclusion does not seem to be one shared by similar native speaker studies. The findings of Mitchell, and of Miller and Gildea, suggest that it may not be so easy to prevent lexical errors after dictionary use, because users go about dictionary look-up in very unexpected ways, and often ignore or misread those elements of the definition which are most informative.

Mitchell's study (1983) was part of a larger research project assessing reading strategies in secondary schools in Scotland. Although ultimately designed to analyse the use children made of dictionaries whilst reading, the study required the children to complete a number of production tasks, one of which involved looking up target words in a dictionary and completing sentences in which the target word appeared.

According to Mitchell, the children made mistakes in this task when they focused only on those parts of the definition that they could easily understand, or when they misread an unfamiliar word as a similar-looking familiar one. For example the unfamiliar word **POPLAR** in the definition:

aspen *noun*
a kind of poplar whose leaves quiver even in a light breeze

caused some children to write:

An aspen is a kind of leave

or

An aspen is a kind of quiver

and the unfamiliar word CELL in the definition:

lignin *noun*

an organic substance which, with cellulose, forms the main part of wood and is usually present in cell walls

was transformed by one child into *ceiling*:

Lignin is the main part of ceiling wall or wall.

Mitchell also found that inappropriate responses often contained a word or phrase extracted from the wrong part of the definition, where the meaning of the word was elaborated rather than defined. For example one child read:

vestment *noun*

a ceremonial robe, especially one worn by the clergy during religious services

and wrote:

A vestment is a kind of religious service

while another read:

nozzle *noun*

a projecting spout or end through which something is poured or discharged, such as a fitment on the end of a pipe or hose

and wrote:

A nozzle is a kind of pipe or hose.

Miller and Gildea's experiments have been described in 1.3.3. They were designed independently of Mitchell's study, yet their findings are in many ways remarkably similar. Like Mitchell, Miller and Gildea required 10-11 year old native speakers to look up given words in a dictionary and then write sentences illustrating their use. 457 sentences were analysed, 249 containing 12 relatively common words, and 208

containing relatively rare words, and it was judged that 21% of the sentences using common words, and 63% of the sentences using rarer words were "sufficiently odd or unacceptable to indicate that the author did not have a good grasp of the meaning and use of the word".

Miller and Gildea decided that some of the errors were due to ignorance of the part of speech or lexical grammar of words, but they claim that by far the most common single cause of error was the children's application of a look-up technique they name "kidrule", whereby a short familiar segment of a definition is used to form a part of a sentence, and is then replaced by a target word (see 1.3.3.)

The sentences Miller and Gildea believed to have been produced by this strategy were often weird and wonderful. They cite such examples as:

I was meticulous about falling off the cliff

where the subject appears to have substituted METICULOUS for the segment *very careful* in the definition:

meticulous very careful or too particular about small details

and:

That news is very tenet

where the child appears to have substituted TENET for *true*, the last word in the entry:

tenet opinion, belief, principle, or doctrine held as true.

Miller and Gildea present several pieces of evidence to support their theory that children follow the kidrule process when dealing with dictionary entries. They found, for example, that occasionally some of their subjects composed sentences in which a segment of the definition appeared instead of the target word; these subjects appeared to have reached the penultimate stage in the kidrule process, but had forgotten to complete

the final stage of replacing the familiar segment with the unfamiliar one. They also occasionally found segments of the definition written in the page margins; presumably these were jotted down when the subjects identified them as familiar, and were used as an aid to composition later in the process. Furthermore a computer program, created to test the theory of kidrule by following the stages hypothesized by Miller and Gildea, generated very similar sentences to those marked as kidrule errors in Miller and Gildea's data. (The computer also generated some sentences which made perfect sense, despite the random process - a fact which suggests that kidrule may be more widespread than appears from studies which assume that appropriate sentences are only produced by applying appropriate strategies).

Miller and Gildea do not specify which dictionary they worked with, and they do not discuss ways in which childrens' dictionaries might be improved. Defining style does appear to have had some effect on the appropriacy of the sentences their subjects produced, however. Difficult defining language may have forced the children to adopt the kidrule strategy on occasion. Evidence for this is that when Miller and Gildea's dictionary used fairly difficult words, as in the definition for TENET: "*opinion, belief, principle, or doctrine*", the sentence produced, and quoted as an example of kidrule, was similar in kind to those produced in Mitchell's study by children who did not know the word *poplar* in the definition of ASPEN - in both cases, the children used a familiar word from the final part of the definition to form their inappropriate sentences. Kidrule sentences also occurred when the dictionary provided only a one-word definition of the target word, and indeed some of the kidrule sentences Miller and Gildea cite are similar to those collected by Jain as evidence of the inadequacy of one-word definitions in learners' dictionaries.

However, whilst Jain, and Maingay and Rundell, advocate lengthening and emphasising the entries in learners' dictionaries to reduce production errors, the evidence from the studies by Mitchell and Miller and Gildea shows that native-speaker child users can

misread even long and explicit entries. Indeed, longer entries may create their own particular problems; it is possible that only part of a longer entry will be attended to, and this part may not even be the kernel definition, but may be an example phrase which simply provides context.

EFL teachers, and designers of learners' dictionaries, must be uncertain how to react to these conflicting views. Some of the developments in dictionary design proposed by Jain have already been put into practice, while those proposed by Maingay and Rundell may influence the design of future editions of LDOCE. Yet adult non-native speakers may behave similarly to child native speakers when it comes to dictionary use, and in this case different developments in defining style will have to take place. The studies discussed above seemed to be united in only one respect - the belief that the success of a dictionary consultation can depend on the style of the dictionary entry. The best way to compose a learners' dictionary entry remained unclear, as did the causes of errors produced after learners' dictionary consultation. Answers to these questions were therefore sought in Study Three.

3.2. Study Three

Variations in defining style are likely to affect the speed with which the dictionary entry is read; it seems likely that dictionary entries which learners find easier to read will be read more quickly. Similarly, variations may also affect the frequency with which the dictionary is consulted; learners will probably lose their enthusiasm for dictionary consultation if they are continually disappointed in their searches. The most important effect of differences in defining style, however, will be the variation in success rate. It was therefore decided to investigate these three effects in Study Three. In particular, three specific research questions were addressed:

1. does the choice of learners' dictionary affect the number of words looked up?

2. does the choice of learners' dictionary affect the time taken to consult dictionary entries?
3. does the choice of learners' dictionary affect the number of acceptable sentences produced after dictionary consultation?

Study Three also attempted to investigate errors in productive dictionary use by analysing incorrect sentences with the following research questions in mind:

1. What are the causes of the errors made by adult non-native speakers when they use dictionaries?
3. Can the kidrule strategy account for some of these errors?

This study departed from all previous studies into learner dictionary use in its employment of a computer program (written by Paul Meara at University College Swansea), both to record and time instances of definition look-up, and to record the subjects' own language production.

3.2.1. The pilot study

Materials

For the pilot study, 30 words were chosen from Nation's "University Word List" (Teaching and Learning Vocabulary 1990). In this list words are grouped according to frequency and range. Words in group 1 are those Nation considers to occur frequently in Academic English and in the widest variety of contexts (eg ALTERNATIVE, ANALYSE), while words in the highest group, 11, are those he considers to occur infrequently in Academic English and in a very limited range of contexts (eg CYLINDER, DIGEST).

In this study, 30 words from groups 5 and 6 were used. They are not the commonest words in Academic English, and are therefore less likely to form part of the academic

vocabulary of overseas students. At the same time, they are not limited in their academic use to one area of study, and are therefore more likely to be listed in a General English dictionary. However some of the words in groups 5 and 6 occur fairly frequently in non-academic contexts, and were likely to be known to intermediate/ advanced learners with a background of General English. In order to reduce the list to be used in this experiment to those items which subjects were least likely to know in advance, all words occurring in Hindmarsh's Cambridge English Lexicon (1980), and/or occurring more frequently than fifteen times per million words of General English according to The Teacher's Word Book of 30,000 Words (Thorndike and Lorge 1944) were eliminated.

Those remaining items with definitions longer than one computer screen page in any of the three learners' dictionaries, OALD, LDOCE or COBUILD, were omitted because it was felt that the definition access times would be less accurate if subjects had to scroll pages of definitions. In most cases the items which were dealt with at length were polysemous, and were given more than one entry in all three dictionaries. In seven cases, however, a single COBUILD entry was longer than one screen.

The names of academic subjects (BIOLOGY and SOCIOLOGY) were also rejected, together with technical terms (ELECTRON, MORPHOLOGY, RADIUS, HEMISPHERE, METABOLISM, SCALAR), on the grounds that the testing device chosen for this experiment would not adequately test knowledge of the meaning of these words.

The thirty items chosen for the pilot study are listed in table 3.1 below. They had the positive attributes of being relatively infrequent, and of being defined within a single entry in each of the three dictionaries chosen for the study. These items were paired with a number of very high frequency words (nouns listed at level one by Hindmarsh). In the

following list, the high frequency words are written in capital letters beside each of the 30 test items.

Table 3.1: Words from Nation's "University Word List" used in Study 3

Group 5:

category FAMILY
 client CLOCK
 duration HOUSE
 enlighten CHILD
 err MONEY
 gravity KNIFE
 homogeneous MAN
 incorporate EXERCISE
 intersect RIVER
 perpetrate WINTER
 parenthesis HOTEL
 retard FOOD
 rudimentary QUESTION
 subtle SCHOOL
 symptom BOOK
 trait CAR
 trivial HAT
 version JOB

Group 6:

abnormal SHOE
 agitate TELEVISION
 civic FATHER
 clarify DOCTOR
 collide WINDOW
 compute GLASS
 controversy LETTER
 hierarchy TEACHER
 identical HEAD
 interact POLICEMAN
 interlock CHAIR
 interlude GIRL

In Miller and Gildea's experiment (1987) subjects merely reproduced example sentences from the dictionary entries, which made it impossible to ascertain whether they had understood the meaning and use of the target words. Target words were linked with high frequency words in my study so that subjects would have to create a new context to accommodate them both. The high-frequency words chosen for the experiment were intentionally context-neutral, so that they did not provide clues to target word meaning, or encourage a false understanding of the target words. It was felt that if words which were semantically linked to the target words were included, the task would become too easy for subjects, who would simply employ grammatical devices to link the two words, without giving any clear indication of what the target word meant to them.

There is a precedent for using paired words in a productive vocabulary test. Corson (1983) also tested subjects' ability to use target words by eliciting sentences containing

both a target word and another more common word. However, in Corson's method the common words were chosen because they suggested the semantic area in which the target word was to be used. Corson's target words were polysemous, and his keywords really served as pointers towards more specialised areas of use.

There were three versions of the pilot test, offering subjects the opportunity to access entries from OALD, LDOCE or COBUILD for each of the 30 words. Because of hardware limitations it was impossible to reproduce all the typographical detail of the original entries in LDOCE, OALD and COBUILD. The dictionary entries on the computer screen did not include phonetic transcriptions, COBUILD extra column information, or derivational forms listed at the end of the entries, but did include all other information, including examples, derivational forms and word class changes.

Procedure

Eighteen subjects took part in the pilot study. They were all enrolled on an intensive English program, prior to registering as postgraduate students at Warwick University. Data from one of the subjects was discounted at the analysis stage, because the mean look-up time was almost twice as long as that for any of the other subjects in the experiment, and was felt to be atypical.

Subjects were first tested using the Eurocentres Vocabulary Size Test (EVST) (Meara and Jones 1990). It was felt that vocabulary size served as an approximate indicator of general English language proficiency, and as the EVST test is computerised it had the added advantage of being quick and easy to administer at the beginning of each experimental session. Subjects were then presented with 30 experimental trials. In each trial a target word and a high-frequency word appeared on the computer screens, and subjects were asked to use both of these words to create a sentence. If the subject did not know the target word s/he could access a dictionary entry for that word by pressing

ENTER on the keyboard. It was anticipated that all subjects would know the high frequency words. Thirty pairs of words were presented to each subject in this way.

Each test file in the program recorded personal details of each candidate, any access to definitions and the length of time spent reading those definitions, and the sentences produced by the subjects. Most subjects took longer than I had expected to complete the test, and two subjects did not complete the full set of thirty items.

Results of the pilot study

Only those sentences written after the subjects had consulted a definition were analysed.

These sentences were rated using a 5 point categorization. The categories were:

- 0** **meaningless in context/word omitted/no sentence**
- 1** **clear that meaning at least partially understood, but wrong part of speech**
- 2** **meaning not entirely clear from context; right part of speech**
- 3** **meaningful but with marked syntax or lexical collocation**
- 4** **entirely meaningful in context**

Only those sentences rated 4 were counted as correct.

Table 3.2 shows the data for each dictionary group, The table shows mean vocabulary test score (EVST), the mean number of items looked up (look up), the mean time taken to look up each item (time) and the mean number of correct sentences produced after look-up (correct).

Table 3.2: Data from the pilot study

	LDOCE	OALD	COBUILD
EVST	5450	5575	5550
S.D	0947	1547	1645
look up	20.6	19.1	24.6
S.D	0.9	9.0	6.0

	LDOCE	OALD	COBUILD
time	23.0	45.0	33.6
S.D	7.6	12.8	15.5
correct	8.0	10.0	8.6
S.D	3.7	6.1	4.9

These data were subjected to a series of one-way analyses of variance. The analyses failed to show any significant differences between the three groups on EVST, look up, and correct [$F < 1.0$]. There was, however, a significant difference in mean look-up time between LDOCE and OALD users; LDOCE users read the dictionary entries considerably faster [$F(2,10) = 3.61, p < .005$].

The data were also analysed for correlation between:

- a) vocabulary test score and mean look-up time
- b) vocabulary test score and number of correct sentences
- c) mean look-up time and number of correct sentences.

Vocabulary score did not correlate significantly with mean look-up time or with the number of correct sentences, nor was there any significant correlation between mean look-up time and the number of correct sentences.

Discussion of the results of the pilot study

The data showed that OALD users took longest to read dictionary definitions, whereas LDOCE users took least time.

However, it had to be borne in mind that this was only a pilot study. The results of only seventeen subjects, some of whom did not complete the test, were analysed, and this did not constitute a large enough sample to be representative. The OALD user group was particularly heterogeneous, with high standard deviations; this suggested that the variation noted in this study might not be evident in a different sample.

Moreover, certain non-representative effects may have been created by the test itself. Some data was lost because subjects accidentally pressed the wrong keys, and some test items generated ambiguous sentences, or proved very difficult to create sentences for. The test program was adjusted to eliminate these problems in the second study.

3.2.2. Changes to the test program

It was decided to reduce the number of test items for the main study. The 30-item pilot test was time-consuming; it normally took about one hour to complete, but could take up to two hours. To this was added the ten to fifteen minutes needed for the preliminary vocabulary test. The advantage of testing the readability of a larger number of dictionary entries had to be offset against the disadvantage of testing fewer subjects. It was difficult to persuade large numbers of subjects to sacrifice more than an hour of their time, and it was also difficult to fit lengthy testing into timetabled sessions.

In order to decide which words to delete from the second version of the test, I conducted an item analysis, counting the number of times each word had not been looked up, had been used in a 2-rated sentence, or had not been used to create a sentence at all.

I assumed that items which were frequently looked up, and which tended to be used in sentences which could easily be marked as either correct or incorrect, would provide the most data relevant to the assessment of the readability of dictionary definitions.

When subjects did not look up a word, this did not necessarily mean that they already knew the word; the sentences they produced reveal that some of them only **thought** they knew the word, and in fact could not use it appropriately. However, the performance of subjects who do not read definitions is not relevant to this study, and items which many subjects either know or think they know can be eliminated without risking the loss of much data. I therefore chose to omit from the second version of the test those 5 items

which 7 or more of the 18 subjects did not look up: CATEGORY, DURATION, CLIENT, ABNORMAL, and IDENTICAL.

Type 2 sentences were those in which the meaning of the target word was not entirely clear from context, although grammatically acceptable and not obviously deviant. A certain amount of relevant data would be lost if target words used in many 2-rated sentences were eliminated, because we can predict that if these items appeared in future versions of the test they would be looked up, and the time taken to read their definitions could be recorded. However, these items would provide little relevant data regarding subjects' understanding of the definitions, because it is impossible to judge from a 2-rated sentence whether the subject knows what the word means. It seemed particularly difficult to make the meaning of adjectives clear in the context of a single sentence; subjects tended to place them, predicatively or attributively, in relation to the given common word to form sentences such as *I bought this trivial hat* and *Where did you buy those abnormal shoes?* (both rated 2).

I chose to omit from the second version of the test those 5 items which occurred in 6 or more 2-rated sentences: HOMOGENEOUS, SUBTLE, TRAIT, TRIVIAL and HIERARCHY.

In some cases subjects did not produce a sentence after having read the definition for a particular item. This failure to produce a sentence was in itself an interesting piece of data; it suggested that the subject did not understand the definition, and/or did not learn from the dictionary entry how the word could be used. 8 out of the 18 subjects failed to find a sentence for PARENTHESIS, and an examination of the sentences the other 10 subjects suggested that the definitions of PARENTHESIS are difficult to understand; only 4 sentences merited a 4-rating. However, in subsequent discussion some of the subjects who had failed to produce sentences claimed that they did in fact know the

meaning of PARENTHESIS, but found it impossible to create a sentence including the word. PARENTHESIS may be a special case in this respect, as it denotes an entity widely recognised but seldom referred to - neither COBUILD or LDOCE give examples of its use, although LDOCE explains the meaning of PARENTHESIS by referring the user to an example sentence with some words placed between parentheses.

The only other item which a large number of subjects failed to create sentences for was INTERLOCK. This item also generated three sentences with a 2 rating, and a number which were difficult to categorise, such as YOU CAN INTERLOCK THE CHAIR WITH THESE WIRE and THE CHAIR INTERLOCK WITH TWO CLUNCHES. Two subjects did not reach the end of the test, and so I lacked complete information about the last few items, INTERLOCK amongst them.

Removing PARENTHESIS and INTERLOCK from the test would not only reduce the number of items, but also considerably reduce the amount of time it took to administer the test. Items which were not looked up were usually dealt with quite quickly by the subjects; it was far more time-consuming to look up a word and then struggle to write down a sentence, before finally abandoning the attempt.

After removing the items discussed above, the second version of the test was 18 items long.

I also made two smaller changes to the test, by pairing COMPUTE with HALF rather than GLASS (which was misread as CLASS by some), and adjusting the program so that subjects could reaccess dictionary entries while writing, and would not pass on to the next item by mistake.

3.2.3. The main study

Having made the necessary changes to the test program, I proceeded with a larger-scale study to investigate the effect of different dictionary defining styles on productive dictionary use.

Subjects

52 subjects participated in this experiment. All were overseas students studying in Britain; some were taking preessional programmes before embarking on British university degree courses at Birmingham University, and the remainder were enrolled on summer English language programmes at Aston University, Coventry Technical College and Henley College of Further Education in Coventry. The subjects came from a variety of language backgrounds. National groupings were as follows: Japan 18, Italy 7, Taiwan 5, Thailand 4, Germany 3, France 3, Bangladesh 2, Spain 2, Turkey 2, Ethiopia 1, Greece 1, Indonesia 1, Iran 1, Sri Lanka 1, Peru 1.

Procedure

A modified version of the program used in the pilot study was used. Subjects were presented with 18 experimental trials. In each trial a target word and a high frequency word appeared on the computer screen, and subjects were asked to use both these words to create a sentence. Subjects were given the opportunity to access definitions from OALD, LDOCE or COBUILD, as before.

The following data was recorded for each subject: EVST score, definition type, number of words looked up, time taken to consult each definition and sentences containing each of the 18 keywords.

It was decided that data from subjects with EVST scores lower than 2000 should not be considered in the analysis. This reduced the number of subjects by one, to 51.

A oneway analysis of EVST score by group was conducted to ensure that OALD users, LDOCE users and COBUILD users were of equivalent ability. The composition of the three groups was not found to differ by this measure [$F(2,48) = .5259$ $p < .5$].

712 look-ups were subsequently analysed. Three independent judges were asked to decide on the acceptability of target word use in the sentences in terms of meaning, collocation and syntax. Each sentence was marked correct if two or more of the three judges rated the target word use as acceptable, and was marked incorrect if two or more of the three judges rated the target word use as unacceptable. Sentences judged incorrect were later examined to try to identify the causes of error.

Results

Table 3.3 summarises the data obtained from Study Three regarding vocabulary size (EVST), the number of words subjects looked up (look up), the mean time taken to read the dictionary entry (time), the mean number of correct sentences and the mean percentage of correct sentences (percent).

Table 3.3: Data from Study Three

	LDOCE (16)	OALD (19)	COBUILD (16)
EVST	5438	4932	5128
S.D	1444	1270	1667
look up	12.5	14.6	13.8
S.D	4.8	4.1	3.2
time	36.0	45.7	52.3
S.D	24.7	19.3	30.0
correct	5.1	4.4	5.7
S.D	2.7	1.9	3.0
percent	44.7	34.9	42.6

No significant difference was found between the three groups regarding the number of words they looked up ($F(2,48) = 1.2156$ $p = .3055$) the average time they took to look up words ($F(2,48) = 1.7563$ $p = .1836$) and the number of correct sentences they produced ($F(2,48) = .9617$ $p = .3895$). Thus the variation in the defining styles of the three dictionaries had produced no observable effect on the behaviour of the users.

Standard deviations for **time** were very high. I can find no obvious explanation for this, but two possible influential factors may be the fact that the Study Three subjects were newly arrived in Britain, and the fact that they were tested in a range of institutions. It had thus been impossible for Study Three subjects to establish a "class pace" in the way that subjects who all attended the same course of study might do.

The proportion of incorrect sentences was very high across all three groups. The majority decision of the panel of three judges was that 443 (63%) of the 701 sentences written after consulting a definition contained an unacceptable use of the target word. This figure should be interpreted as a conservative estimate; the judges found it extremely difficult to decide whether target word use was acceptable or not. In many cases the meanings expressed were probably not what the subjects had intended to express, and the judges felt that some of the sentences they marked as acceptable were written without real understanding of the target words.

Three categories of error were used to describe the incorrect sentences:

- 1) Failure to create a sentence with the target word
- 2) Semantic error
- 3) Usage error, eg concerning transitivity, countability or morphology.

These categories were treated as mutually exclusive, but many sentences classed as containing semantic errors also contained errors of usage. Semantic errors ranged from total misunderstanding of word meaning, for example:

The window collides the entrance

to errors which reflected considerable understanding of the word's basic meaning, but ignorance of connotation or lexical collocation, for example:

The doctor tries to clarify her illness to her.

Assigning the target word to the wrong word class (with or without morphological adjustment) was counted as a usage error in cases where word meaning was essentially unchanged, but was categorised as a semantic error in those cases where the subject had partially or totally failed to understand the word's meaning, as in:

The girl was interluded in the story

and

My father works as a civic in our town.

Table 3.4 below shows the distribution of the three categories of error according to dictionary group, and shows that the majority of errors fell into category 2 - semantic errors.

Table 3.4: Percentage of incorrect sentences related to error category and definition type

Error type:	1	2	3
LDOCE	4.5	39.4	8.7
OALD	5.7	46.5	9.7
COBUILD	3.8	40.9	8.4

There were no significant differences in error frequency across groups for categories 1 and 3, but the OALD group had a significantly higher number of category 2 errors [$F(2,51) = 3.9018, p = .0265$].

Having tested the effect of different defining styles on dictionary-using behaviour, my next objective was to examine the reasons why the subjects used the target words incorrectly after consulting a dictionary entry. The data provided a certain amount of insight into the causes of lexical and grammatical errors. One of the most predictable findings was that the target words themselves had an effect upon the type of error produced. Errors of grammar and usage were commonest in cases where the target word was associated with more unusual syntactic structure, regardless of definition type. Thus there was a high proportion of errors amongst sentences containing the target words COLLIDE and INTERACT, because subjects did not apply the rules governing the use of plural and singular subjects and WITH. It was also found that verbs amongst the target words were particularly vulnerable to word class conversion errors; in 38 out of the 58 instances of word class change in the data the change was from verb to noun or adjective, while there were only three instances of target word nouns or adjectives being converted to verbs. Subjects may have chosen to convert verbs to syntactically less complex word classes because this relieved them of the need to deal with valency patterning. Thus word class conversion can be regarded as a kind of avoidance strategy.

Semantic errors were also more frequent with some types of target word than with others, and were particularly common where the target word had a limited range of lexical collocations. For example, a very small range of things can be PERPETRATED or CLARIFIED, and so these two words attracted a high proportion of errors assigned to category 2.

Errors were also caused by subjects' failure to apply grammatical and collocational information available in the dictionary entries. Grammatical information seemed to have had little influence over language production; OALD and LDOCE users could refer to codes, such as *n.*; *v.*; *adj.*; *Tn*; *usu attrib*; and *esp passive*, but they produced no fewer

category 3 errors and word class conversion errors than COBUILD users, who were not given the COBUILD grammar codes from the extra column. Perhaps subjects ignored grammatical information because they could not understand the abbreviations, or perhaps they simply lacked the background in English grammar necessary to apply the information. Subjects may have felt that collocational information in the dictionary entries was unimportant or optional because it is often given in parenthesis; the dictionaries also make frequent use of "etc" to end lists of lexical collocations, from which subjects may have inferred that the range of collocating words was wider than it really was.

In some cases, however, crucial collocational information was not available in the dictionary entry, and this lack may also have been responsible for errors. None of the dictionaries warned of selection restrictions with CLARIFY, for example, and only one of the 29 subjects who looked this word up managed to avoid an error. In contrast all three dictionaries suggested collocations for PERPETRATE, and 12 out of the 40 subjects who looked up PERPETRATE went on to use it appropriately.

Having identified some contributing factors to error in my data, my final research objective was to determine whether the kidrule strategy was in use amongst my adult, non-native speaker subjects. The 3 point categorization system was not a useful tool for this, as the kidrule strategy can manifest itself through errors assigned to any of the three categories. A subject might, for example, pick a word from the definition which was synonymous with the target word but which did not share the target word's valency pattern, and thus produce a grammatical error (category 3). Alternatively the subject might pick a non-synonymous word from the definition, or one which was only synonymous in certain contexts, and thus produce a lexical error (category 2). Kidrule could even result in a category 1 error if the subject forgot to insert the target word, and

composed a sentence containing a segment of the definition instead. Thus the only way of determining the influence of kidrule was to operate the kidrule process in reverse, and substitute for the target word an appropriate segment from the definition the user originally consulted. In cases where acceptable sentences could be produced by this means, I could guess (but not prove) that the kidrule strategy had been employed.

I assume that Miller and Gildea also used this method to determine the number of kidrule sentences in their data. It will be clear, however, that categorizing the errors in this way is a rather uncertain process, and at times depends more on intuition than on objective assessment. It may be misleading to quantify what cannot be proven, and for this reason I did not subject to statistical analysis my own estimates of kidrule error.

Kidrule influence was most obvious in cases where a subject had picked out a segment from the dictionary entry which was not in any way synonymous with the target word, thus producing nonsense. There were a number of instances of this in the data. For example one user seems to have picked on the word *different* in the LDOCE entry for VERSION:

1 a slightly different form, copy or style of an article

and thus produced the sentence:

I will begin new job that is version.

Another seems to have focused on *going across* in the OALD entry for INTERSECT:

1 divide (sth) by going across it

and therefore wrote:

We must intersect the river for arrive village.

I puzzled over the sentence:

Doctors clarify medicals before using them

for some time before I realised that by *medical*s the subject probably intended **medical instruments**, and the sentence was probably the result of picking out the words *clear*, *pure* and *heating* from the second definition of CLARIFY in LDOCE:

2. to make (a fat, esp. butter) clear and pure, esp by gentle heating.

Table 3.5 illustrates this process with a number of other sentences from the data:

Table 3.5: sentences formed by selecting non-synonymous segments from the dictionary entry

Definition	Segment	Sentence
controversy (about/over sth) public discussion or argument....	argument	I explain you my controversy with Tom in the letter
If something retards a process or development, it causes it to happen more slowly....	develop(ment)	If children eat much food, they retard very fast
interact 1 (with sth) act or have an effect on each other....	have an effect on	Policeman shoots thief with a magnum which interacts people who hear it
interact 1 (with sth) act or have an effect on each other....	act	The policeman teach way for a tourist by interacting
interact to have an effect on each other or something else by being or working close together....	have an effect on	Policeman's activities can interact the society to a great extent

<u>Definition</u>	<u>Segment</u>	<u>Sentence</u>
a controversy is a discussion or argument about an action or proposal that many people do not approve of...	action	many letters which request the controversy were sent to the prime minister
clarify (cause (sth)to) become clear or easier to understand....	understand	He is a doctor in Economics, so he clarified Economics problem in this country
civic 1 of a town or city; municipal: a civic function, eg the opening of a new hospital by the mayor of a town...	opening	My father participated the civic of the new school in our town
Civic is used to describe 1 people or things that have an official or important status in a particular town or city...	official/ important	My father is civic in his company My father is civic of his company My father went to a civic city, London

There were 50 cases such as these, where the subject chose a non-synonymous segment from the dictionary entry. However there was an almost equal number of cases (48) where subjects chose segments with meanings which are presented as synonymous with that of the target word. Although the adoption of this strategy can result in sentences just as ill-formed as those produced by randomly focusing on a familiar word, these sentences do possess their own logic. The LDOCE definition of INTERSECT, for example:

to be in such a position as to cut across (each other or something else)

justifies the sentence:

Last week I intersected the river with a small boat

and the first LDOCE definition of CLARIFY:

1. to make clearer and easier to understand esp. by explaining and giving more details..

gives grounds for such sentences as:

The doctor clarified his condition

Doctors cannot clarify this odd disease

and the heartfelt:

Doctors hardly clarify their talks to the patient.

In Table 3.6 below are a number of other errors which seem to be the result of over-reliance on just one of the definitions in the dictionary entry. In many cases the writer has also ignored grammatical information which was available, but the resulting sentences are for the most part comprehensible, and suggest partial understanding of the meanings of the target words.

Table 3.6: sentences formed by selecting synonymous segments from the dictionary

entries		
<u>Definition</u>	<u>Segment</u>	<u>Sentence</u>
retard v [Tn](fml) 1 make (sth) slow or late: retard the mechanism, eg of a clock . retard the spark, eg of an engine. 2 slow the progress or development of (sb/sth);hinder: Lack of sun retards plant growth.	make (sth) slow or late	Because the food is retarded he get angry I am always retarding to eat foods I am two hours waiting for food, it is retarding too much
retard v [T] esp. fml or tech. to delay, esp. in development, cause to happen later than usual or expected: cold weather retards the growth of the crops.	to delay	As the train retards, I will not be able to buy the food If the harvest is retarded this year there will not be enough food Strikes often retard the distribution of food

<u>Definition</u>	<u>Segment</u>	<u>Sentence</u>
If two or more people or objects collide , they hit one another violently after one or both of them have been moving very quickly....	hit (violently)	A car collided the window He collided his head on the window Windows should not be collided
collide v [1 (with)] 1 to crash violently: The two planes collided (with each other) in midair....	to crash (violently)	The window collides when the thieves jumped into the room
1 A symptom is something wrong with your body or with the way that it works, that is taken as a sign of illness....	something wrong with your body	Reading books at a dark place gives you some symptoms The medical book tells me my body has symptoms If a book falls in your head you will certainly feel the symptoms
perpetrate v [T] fml to do (something wrong or criminal; be guilty of:....	be guilty	I am very perpetrated to let young Tom stay outside in this winter day The woman in the red dress perpetrated for drug-related offences
perpetrate v [Tn] (fml or joc) (a) commit (a crime etc): perpetrate a dreadful outrage. (b) be guilty of (a blunder, an error, etc):...	be guilty (of an error)	He is perpetrating for not putting on a sweater in winter

<u>Definition</u>	<u>Segment</u>	<u>Sentence</u>
enlighten v [T] to cause to understand deeply and clearly, esp. by making free from false beliefs: Peter thought the world was flat until I enlightened him \ an enlightening experience	to cause to understand deeply and clearly	The Lord Buddah was enlightening the truth

In my data there was just one sentence where the subject had failed to include the target word, but had included a segment of the definition instead. The word SIN appears to be a substitute for the target word ERR in:

It is not a sin to have a lot of money

and also appears in the OALD entry for ERR:

1 (a) make mistakes; be wrong . (b) do wrong; sin.

This seems to be further proof of the influence of kidrule in my data, and the confusion of ERR with SIN may also account for other errors, such as:

Human being errs when see money

and

Money enable people to make erring thing.

Although it was impossible to establish exactly what had passed through the subjects' minds as they read the dictionary entries, kidrule seemed to explain about a quarter of the incorrect sentences in my data. Most of the sentences which had not been produced through the kidrule strategy contained the same kinds of grammatical and collocational error as appeared in some kidrule sentences, but were identifiable because it was not possible to correct them by substituting a segment of the dictionary entry for the target word. The majority of errors where the sense of the target word deviated wildly from its

accepted meaning were the products of kidrule; on the whole, where the subject had not adopted the kidrule strategy, the target word was used in the correct semantic area.

Thus it seems that there were three main factors responsible for errors in my data:

1. subjects' disregard for grammatical and collocational information available to them in the dictionary entry
2. subjects' ignorance of collocational and usage rules needed to avoid production errors, but unstated in the dictionary entry
3. application of the kidrule strategy.

A further factor, however, seemed to be operating in a small number of errors which were not the result of grammatical or collocational deviation, or the equation of the target word with a segment of the dictionary entry. In these errors, as in some kidrule errors, the target word was used in a completely inappropriate sense, but the cause of error could be traced to confusion with some phonologically or orthographically similar word. We can guess, for example, that this type of confusion lay behind:

It is easy to perpetrate the weather in winter

because PERPETRATE sounds somewhat like PREDICT. It may also have been the reason why one subject wrote:

Snow perpetrates in winter

perhaps confusing PERPETRATE with PRECIPITATE.

In a particularly striking case, one Japanese subject seemed to have confused CRIME with CLIMB in the COBUILD definition for PERPETRATE:

If someone perpetrates a crime or other harmful or immoral act, they successfully commit it...

and then to have struggled to make some sort of sense of the collocation by creating:

The man perpetrate to crime a mountain in winter.

Individually or in combination, the four factors identified above seemed to account satisfactorily for almost all the incorrect sentences in my data. Just a handful remained unexplained. In these sentences, subjects' preconceived notions of word meaning seem to have simply overridden all the information in the dictionary entry. In some sentences a target word appears to be assigned a completely new meaning for no obvious reason. Thus COLLIDE is (apparently) used to mean BLOCK UP in two separate sentences in my data:

I can't see through the window because it collided by the books

and

The window collides the entrance.

Likewise nothing in the entry for INTERACT offers a clue to what the subject intended by it in the following sentence:

The policeman interacted me and told me not to make strange noises with my tongue.

The exact causes of these errors was not recoverable from my data; perhaps they were the result of a kidrule strategy that failed, because the chosen segment in the dictionary entry had been misinterpreted, alternatively the cause may lie in the subjects' language background, and be the result of confusion between cognates or orthographically similar words.

3.3. Conclusions

The findings of Study Three provide strong evidence that there is little difference in intelligibility between the three major dictionaries. Apparently neither the restricted

LDOCE defining vocabulary nor the COBUILD folk definitions make dictionary reading quicker or more successful.

It is possible that some differences between the behaviour patterns of different groups may have been obscured because of individual differences between members of the groups. Although the composition of each group did not differ significantly in terms of vocabulary size, individuals within each group varied greatly in language proficiency, and the groups had not been controlled for other possibly influential factors such as language background and study experience. It seems likely that speakers of certain languages will be advantaged when reading the dictionary entries, as will subjects who have received more extensive training in dictionary use. A possible relationship between background and dictionary-using proficiency clearly needed to be explored.

The findings also suggested that adult non-native speakers employed the kidrule strategy, but it appeared that two different kinds of kidrule strategy were in use: one random, and testimony to learners' misreading of dictionary entries, the other considered, and testimony to the misleading nature of many of those entries. About half the cases of kidrule error I identified resulted in nonsense, because the subjects had picked out segments from the dictionary entry which were not in any way synonymous with the target word. The remaining kidrule errors, however, possessed a certain logic because the subjects had chosen (at random or knowingly) segments of the dictionary entry which were presented as synonymous. In many cases the dictionary entry did not provide sufficient information for the subject to avoid this type of error.

Those defects in learners' dictionary definitions which Jain (1981) objected to - the emphasis on one-word synonyms, and the failure to provide adequate information on selection restrictions - were still noticeable in the editions of the dictionaries used in this study, and were probably partly responsible for the second type of kidrule strategy error

in my data. Unfortunately the differences in procedure between Miller and Gildea's study and Study Three make it impossible to directly compare their results; in Study Three, for example, only sentences produced after look up were analysed, whereas in Miller and Gildea's study every sentence was counted, regardless of whether the subject knew the word in advance. It would appear, however, that adult English learners are using their dictionaries with a little more success than native-speaker school children; my data did not suggest quite such a hit-and-miss approach to dictionary consultation as that described by Miller and Gildea, and many of the errors I recorded reflected at least partial understanding of target word meaning. Mitchell and Miller and Gildea document errors which had been caused by merely reading the final words of an entry, but few such errors occurred in my data. Data from Study Three gave every indication that subjects were reading and reproducing elements from two parts of the dictionary entry: the definition, and the example sentences that usually followed the definition.

Thus Study Three closed one line of enquiry, but opened several more. I was now satisfied that none of the three learners' dictionaries was substantially worse or better than the other two at helping learners to encode new words. This meant that I could remove one variable from my remaining experiments and proceed with just a single dictionary as a source of definitions.

One question for further research concerned whether the presence of example sentences made the look up task a much longer one, and whether examples increased the productive effectiveness of dictionary entries. It was also clear that another variable would have to be recognized and controlled. I had become aware of the possibility that subjects from different parts of the world might approach the sentence writing task in different ways. In Study Three there was great variation between subjects on an individual level, and it is possible that some of these differences could be accounted for by culturally determined

attitudes. For example, it seems likely that subjects from some parts of the world might favour a painstaking approach to their task; they might look up more target words, and deliberate more before composing their sentences. Other nationalities might prefer risk-taking, guessing at the meaning of some words and quickly reaching decisions regarding the meaning and use of the words they looked up. Cultural factors, such as the degree of creativity expected of learners, and the value placed on accuracy as opposed to fluency, seem likely to play a part in determining dictionary-using behaviour, just as they play a part in determining behaviour in the language classroom. First language knowledge, and the presence of cognates or words borrowed from English in the mother tongue, also seem likely to be an influential factor in determining the success of dictionary use; I had found examples in my data of phonological confusion caused by first language influence, and some of the errors I had failed to explain may have been caused by false equivalencies between target and first language words.

Study Three had not been designed to monitor the differences between subjects from different backgrounds, but the high standard deviations between subjects within each of the three dictionary user groups may have been a reflection of the nationality mix within these groups, and their differing responses to the task. Could it be that the most important factor affecting productive success was not the learners' dictionary consulted, but the language and cultural background of the user? By comparing the results of user groups from different backgrounds, it might be possible to identify behavioural patterns undetectable when subjects were grouped by dictionary type. I therefore proceeded, in my next study, to investigate the effects of cultural and first language variation on productive dictionary use.

Chapter Four

The effect of language background and culture on productive dictionary use

4.1. Introduction

Study Three indicated that variation in the defining style of the three major learners' dictionaries did not significantly affect their readability or their productive effectiveness. This closed one line of enquiry for me, and enabled me to proceed with just one of the learners' dictionaries as a source in future studies. My conclusions also suggested two further lines of enquiry; one concerning the background of the dictionary user, the other concerning the form of the dictionary entry. The first of these is investigated in the study reported in this chapter, and the second line of enquiry is pursued in Chapter Five.

The data from Study Three suggested that the background of the dictionary user might have greater influence on productive success than the learners' dictionary chosen for consultation. In this respect the indications matched my own intuition; I think most EFL teachers will agree that learners from some parts of the world tend to have better dictionary skills, and that certain types of dictionary misreading can be linked to certain language backgrounds. Few prior studies have investigated this phenomenon, however, and no research has systematically compared the productive monolingual dictionary use of representative samples of subjects from different cultures.

Those few studies which have looked at the effect of culture on dictionary use tend to acknowledge the influence of the user's first language. Ard (1982) chose subjects from three different language backgrounds (Japanese, Arabic and Spanish) for close observation, and came to the conclusion that:

while the nature of bilingual dictionaries makes it unlikely that students will often find acceptable words to use in compositions, the success rate depends on the native language background of the students. Students

from languages "close" to English ... are more likely to be successful.

(1982:2)

Ard's sample was too tiny to be representative, and he himself calls for further research on a larger scale, but his finding that people with different language backgrounds have different approaches to dictionary use, and possibly different dictionary needs, finds support in the work of Meara and English (1988). In this study lexical errors taken from a corpus of Cambridge First Certificate examination papers were assigned to six categories. It was found that the distribution of error types varied markedly from one language to another, which led the researchers to the conclusion that their monolingual learners' dictionary (LASD) is far more effective with some languages than with others; Swahili speakers, for example "are more than three times as likely to meet a dead end than are Finnish speakers of about the same level" (p8).

Further support for the view that language and culture are important factors is provided by Bogaards (1990, 1992), who focussed on just one aspect of look-up strategy - the dictionary users' choice of search word when looking up multi-word idioms - and noted that French and Dutch dictionary users exhibited very different look-up behaviour:

il existe des comportements typiquement français ou néerlandais, différents de ceux que manifestent les étrangers. Il est donc permis de croire que les choix que font les sujets dépendent dans une large mesure de leur langue maternelle.

(1990:94)

One major study in this area, however, reports no influence of language and culture on dictionary use. Battenburg (1991) dismisses the possibility that there is "wide variation in the reported behaviour of dictionary users ... grouped according to their native language backgrounds" (p 89). Using questionnaire data to investigate the frequency of consultation of different dictionary types and dictionary information types, he found "no significant patterns" in the reports from different language

groups, and came to the conclusion that language learners' use of dictionaries was largely unaffected by their mother tongue and culture.

Perhaps Battenburg failed to find a connection between dictionary use and first language because he depended on the subjects' own reported behaviour, rather than direct observation. Significant patterns may also have failed to emerge because his subjects were not picked to represent language backgrounds in equal proportion. In Study Four I decided to monitor dictionary consultation and language production after dictionary consultation, using relatively large subject groups which were balanced in terms of size, educational experience and level of study, but which were very different in culture and language background.

4.2. Study Four

This study aimed to compare the productive dictionary use of two culturally distinct groups of subjects by asking the following questions:

1. Do subjects from different language backgrounds differ in the number of words they look up?
2. Do subjects from different language backgrounds differ in the time they take to consult dictionary entries?
3. Do subjects from different language backgrounds differ in the number of acceptable sentences they produce after dictionary consultation?

It also attempted to investigate errors in productive dictionary use by analyzing incorrect sentences with the following questions in mind:

1. Do subjects from different language backgrounds produce different kinds of error?
2. Can first language influence account for some of these errors?
3. Can cultural differences account for some of these errors?

For this study, the tests were administered in the home countries of the subjects. The subjects were 51 Portuguese undergraduates studying English at tertiary level in Portugal, and 44 Malaysian undergraduates studying English at tertiary level in Malaysia. Both groups of subjects were studying in Faculties of Education, and

intended to become English teachers. The two groups were chosen because they were identical in age, educational level and language learning purpose, but came from very different backgrounds in terms of language and culture. English is a foreign language in Portugal, but English and Portuguese are both Indo-European languages and share many cognate words. In Malaysia, on the other hand, English is the second language, but the national language Bahasa Malaysia belongs to a completely different language family (Malayo-Polynesian).

Subjects were tested to establish their vocabulary size, and were then required to create sentences with the eighteen target words, in the manner described in Chapter Three. In this study, however, all subjects had access to the same dictionary entries, taken from LDOCE.

Sentences produced after look-up were rated for appropriacy by three independent judges, as in Study Three, but in this study a rating scale from one (completely inappropriate) to six (completely appropriate) was substituted for the cruder distinction between "correct" and "incorrect" required of the judges in Study Three. Judges were asked to ignore spelling mistakes, and comment only on the appropriacy of the target words, rather than complete sentences. By averaging the ratings of the three judges an appropriacy score was calculated for each instance of target word use in the data.

The sentences produced by the subjects were then categorized according to error type, and the distribution of error types across the two groups was compared.

Results

Table 4.1 below summarises the initial findings for the two groups, in terms of mean vocabulary size (EVST), mean number of words looked up (number), mean number of seconds taken to read the dictionary entry (time) and mean sentence score (score).

Table 4.1: A comparison of the performance of the Portuguese and Malaysian groups

	Portuguese(51)	Malaysian(44)
EVST	5279	6781
S.D	1263	1379
number	12.90	16.00
S.D	3.34	2.44
time	24.66	31.75
S.D	14.21	12.74
score	3.52	3.09
S.D	0.68	0.60

The two groups were significantly different in all respects: despite the fact that the Malaysians had a larger vocabulary size [$t = 5.54$, $df = 93$, $p < .01$], they performed worse in all the other variables - they looked up more words [$t = 5.08$, $df = 93$, $p < .01$], they took longer to read the dictionary entries [$t = 2.55$, $df = 93$, $p < .01$], and they gained lower scores for the sentences they produced [$t = 3.20$, $df = 93$, $p < .01$]. This suggests that, despite the fact that their level of English was more advanced, the Malaysians were less confident and less efficient dictionary users.

The finding that the Malaysian group achieved lower sentence scores had to be regarded with caution, however, because, like the correct/incorrect marking system used in Study Three, the 6-point scale proved to be a rather crude measure. The average scores gave some rough indication of the level of acceptability of the group's sentences, but they were not entirely reliable because raters did not achieve a high degree of agreement and consistency. The system required raters to judge the gravity of errors, and raters clearly differed in the degree of importance they attached to correct grammatical collocation, lexical collocation, word formation and even overall sentence construction (which they were asked to disregard). Raters' scores also proved internally inconsistent - the same type of error was sometimes given a different score on different occasions by the same rater. Attitudes to errors tended to evolve as the

raters progressed through the data, and this was probably because they were not trained for the rating task, nor did they discuss the task initially.

I therefore decided that, in order to examine the causes of error in the data, it would be necessary to categorise the sentences systematically. The new categories were derived from typical patterns of target word use in a sample of 100 sentences taken equally from the two groups. These patterns were listed and subsequently "collapsed" to create a minimum number of mutually exclusive categories. The four categories finally identified are defined with examples below; they are similar to the codes used in the pilot study described in 3.2.1., but I believe that they describe the data more accurately because they were devised after closer examination of the sentences subjects produced.

Type One - superficially well-formed in terms of the grammar of English (or containing only inflectional errors). The normal interpretation of the target word makes sense, although it is not always certain that the meaning has been fully understood (ie the sentence may be "covertly idiosyncratic" (Corder 1971)).

Examples:

Bad weather retards the growth of food.

My father is strongly aware of his civic obligations.

Children believe in Santa Claus until people enlighten them.

Type Two - superficially well-formed in terms of the syntax of English (or containing only inflectional errors), but the normal interpretation of the target word does not make sense in context - the sentence is "overtly idiosyncratic" (Corder 1971).

Examples:

His boat was already in the middle of the river when I intersected him.

Having lots of books is a symptom of reading very much.

He cut himself with a knife with gravity.

Type Three - not well-formed in terms of the syntax of English because of idiosyncratic rules regarding one or more of the following: transitivity; voice; number; grammatical collocations; compounding.

Examples:

He collided against the window when he was pushed.

Sandra's version about that job is different than mine.

The policemen interacted together.

Yesterday I bought a book about Malaria symptom

Type Four - not well-formed in terms of the syntax of English because of idiosyncratic word class conversion rules and/or word formation rules, such as incorrect or absent derivational forms.

Examples:

That was a very controversy letter.

Like money too much is an awful err.

He feels agitates when he was asked about televisions programme.

The film on the television was agitate.

Comments on the categorisation system

Type two was "collapsed" from two original categories - a) syntactically well-formed sentences where the target word did not make sense, and b) syntactically well-formed sentences where the lexical collocations of the target word were idiosyncratic. In practice it proved impossible to differentiate between the two categories with most words in the data - idiosyncratic lexical collocations were symptomatic of misunderstanding of target word meaning, as can be seen, for example, from the typical sentences *The food is retarded* and *winter perpetrates colds*.

The possibility of including a type five - where the normal interpretation of the target word makes sense, but the inflectional ending is inappropriate - was considered and

rejected. This type of error is not caused by the subjects' failure to apply information in the dictionary entry, nor is it caused by the dictionary's failure to supply information about the target word, and therefore it was not really relevant to my study. Moreover the inclusion of a fifth type, where the target word makes sense, would necessitate a sixth type, where there are inflectional errors and the target word does not make sense. I therefore decided to ignore errors in the data which reflected incomplete mastery of English sentence structure, rather than a lack of understanding of target word meaning and use. I classed as type one sentences such as: *The river intersect my father's farm* (missing third person 's'), *Children must be enlighten by their parents* (missing past participle ending) and *The girls have lunch during interlude* (missing article). However the problem of how to classify incorrectly formed target words was never entirely resolved (see points 4 and 5 below).

Problems with the categorisation system

The following problems with the categorization system remained:

- 1) The *exact* meaning of some of the target words is open to dispute. In particular, it is unclear whether, when objects COLLIDE, one of them can remain static. I classed sentences as type one whenever it was possible to conceive of a context where the sentence would be acceptable. Thus, for example, *The window collided with me* was classed as type one, despite its obvious strangeness.
- 2) It was also difficult to decide the extent to which truth value should influence categorisation. This particularly affected the word INTERSECT, which was linked with the common word RIVER, giving rise to a number of sentences of the type *The two rivers intersect*. These were categorised as type one, although it is probably a geographical impossibility for rivers to intersect.
- 3) In some cases the decision to class word use as type two or type three seemed to be of little help in explaining the root cause of the error, and obscured the similarity between two slightly different constructions. In type two errors the meaning attached

to the target word was idiosyncratic, but the word was placed in a syntactically acceptable context. In type three errors the syntax of the target word was idiosyncratic. However, there were cases where grammar and meaning seemed to be inextricably linked, and the use of the target word was incorrect in terms of both syntax and meaning. In these cases, the sentence was placed in type three. Thus the very common construction *The doctor clarified my illness* was placed in type two, while the equally common *The doctor clarified me about my illness* was placed in type three (because of its idiosyncratic valency patterning) - the meaning attached to CLARIFY seems to be the same in both cases. Likewise sentences such as *The food is retarding and I am very hungry* (type three - intransitive use of RETARD) and *Food is always retarding him* (type two - inappropriate lexical collocations) both appear to be the result of the same basic mistake, ie. the equation of RETARD with the more "core" word DELAY, which can be used both transitively and intransitively.

4) Target words with inflectional errors were placed in class one if the word made sense in context and there were no other syntactic errors, while target words with incorrect or absent derivational forms were placed in type four. However, there seemed to be little real difference between careless errors of the missing third person "s" variety, and careless errors where -ed or -ing is missing from a derived adjectival form.

5) Although I made the decision to ignore errors of article use and plural inflection on the grounds that they could not be rectified by applying information of the kind found in dictionary entries, in practice it was occasionally impossible to be certain whether such errors were caused by carelessness, by lack of knowledge of English sentence structure, or by a failure to apply information about the grammatical behaviour of the target word. For example, I placed *The policeman has interact with the public* in type one, on the assumption that the writer had simply omitted the past participle -ed ending, either through ignorance of the structure, or through carelessness. The sentence makes sense once the -ed ending is added, although it may still be "covertly

idiosyncratic". However, the writer may have been using INTERACT as a nominal form, in which case the grammatical information in the dictionary entry had been ignored, and the sentence should have been placed in type four. On a very few occasions such as this the categorisation system relied on subjective judgement.

The reliability of the categorisation system

All 1,356 sentences produced after dictionary consultation were categorised. A sample of 100 sentences was then categorised by another professional in the field, to test for reliability. There was 73% agreement between raters, and, as expected, a high degree of statistical correlation [$\Phi = 1.09$]. The greatest area of disagreement was in the coding of type one and type two sentences; eleven sentences which I had placed in type one were placed in type two by the second marker, and six sentences which I had placed in type two were placed in type one by the second marker. 100% reliability for the categorisation system was not achieved for three main reasons:

- 1) a few of the sentences in the data were highly deviant, and were grammatically and semantically ambiguous
- 2) the raters disagreed over the acceptable use of ERR; as it is seldom used except in semi-proverbial phrases, the second rater placed in category two almost every use of ERR which I had placed in category one
- 3) the categorisation system called for close attention on the part of the rater; I was familiar with the scheme (having devised it myself), the second marker found it difficult to remember, and occasionally failed to apply the correct criteria.

The opinions of the second assessor were taken into consideration in the final allocation of sentences to categories. Sentences with ERR which were originally placed in category one were retained in category one, but a number of other sentences were recategorized after careful consideration of the categorization criteria.

Table 4.2 below summarises the distribution of sentence types across the two groups. All figures are percentages of the total number of sentences produced after look-up by each group.

Table 4.2: A comparison of the performance of the Portuguese and Malaysian groups in terms of sentence type distribution.

	Portuguese(51)	Malaysian(44)
Type one	52.66	46.91
S.D	15.60	18.77
Type two	23.43	22.78
S.D	14.09	10.97
Type three	14.46	18.93
S.D	11.24	10.51
Type four	09.82	11.65
S.D	10.87	08.72

The Malaysian group produced slightly fewer type one and type two sentences, and slightly more type four sentences. These differences were not great enough to be statistically significant. However the Malaysian group did produce a significantly greater number of type three sentences [$t = 1.99$ df 93 $p < .05$]. This suggests that they were failing to process the grammatical information supplied within the dictionary entry for each target word. The Portuguese subjects may have been able to interpret and apply the dictionary information more successfully - alternatively, they may have possessed some of the relevant knowledge already, and have had less need of the LDOCE grammar codes.

This statistical analysis obscures the fact that for some words the difference in performance between the two groups was much greater than for others, and that for some words the success of the two groups was reversed. Table 4.3 below gives a detailed breakdown of the performance of the two groups for each of the target words. It can be seen that differences between the two groups were particularly marked for the target words PERPETRATE, VERSION, COLLIDE and RETARD, and that

seven of the eighteen target words were used more successfully by the Malaysian group, despite their overall lack of success.

Table 4.3: Use of the target words by Portuguese and Malaysian subjects.

Sentence Type:	1	2	3	4	Total
ENLIGHTEN					
Portuguese	82%	5%	5%	8%	38
Malaysian	63%	27%	10%	-	30
ERR					
	1	2	3	4	Total
Portuguese	62%	-	2%	36%	45
Malaysian	44%	5%	21%	30%	43
GRAVITY					
Portuguese	43%	22%	27%	8%	37
Malaysian	47%	3%	47% ¹	3%	34
INCORPORATE					
Portuguese	67%	25%	7%	-	27
Malaysian	50%	26%	17%	7%	42
INTERSECT					
Portuguese	58%	33%	8%	-	36
Malaysian	40%	37%	21%	2%	43
PERPETRATE					
Portuguese	45%	33%	12%	8%	49
Malaysian	12%	46%	24%	17%	41
RETARD					
Portuguese	20%	33%	33%	15%	46
Malaysian	44%	28%	14%	14%	43

Sentence Type:	1	2	3	4	Total
-----------------------	----------	----------	----------	----------	--------------

RUDIMENTARY

Portuguese	87%	3%	-	10%	30
Malaysian	88%	5%	-	9%	43

SYMPTOM

Portuguese	48%	46%	7%	-	46
Malaysian	40%	46%	14%	-	35

VERSION	1	2	3	4	Total
----------------	----------	----------	----------	----------	--------------

Portuguese	74%	9%	18%	-	34
Malaysian	41%	32%	24%	2%	41

AGITATE

Portuguese	62%	19%	5%	14%	21
Malaysian	68%	5%	10%	17%	41

CIVIC

Portuguese	52%	40%	-	5%	42
Malaysian	34%	37%	-	29%	38

COLLIDE

Portuguese	58%	-	39%	3%	38
Malaysian	37%	5%	47%	11%	38

CLARIFY

Portuguese	18%	29%	47%	6%	17
Malaysian	14%	24%	62%	3%	37

COMPUTE

Portuguese	31%	63%	-	6%	35
Malaysian	48%	45%	3%	5%	40

Sentence Type:	1	2	3	4	Total
CONTROVERSY					
Portuguese	45%	27%	-	27%	22
Malaysian	59%	37%	-	4%	27
INTERACT					
Portuguese	56%	16%	19%	9%	43
Malaysian	56%	22%	-	22%	41
INTERLUDE					
	1	2	3	4	Total
Portuguese	48%	10%	30%	12%	50
Malaysian	56%	14%	7%	23%	43

In order to investigate first language influence in the data, it was also necessary to consider the extent to which the target words were related to words in Malay and Portuguese. Table 4.4 below shows where relationships exist between the English target words and their Portuguese and Malaysian equivalents. The asterisks indicate the group with the higher percentage of type one sentences for each target word; where no asterisk is given the difference between the two groups was less than 5%.

Table 4.4: Relationships between the target words and words in Portuguese and Bahasa Malaysia.

	Portuguese	Bahasa Malaysia
ENLIGHTEN	.*	-
ERR	errar*	-
GRAVITY	gravidade	graviti
INCORPORATE	incorporar*	-
INTERSECT	.*	-
PERPETRATE	.*	-
RETARD	retardar	.*

	Portuguese	Bahasa Malaysia
RUDIMENTARY	rudimentar	-
SYMPTOM	sintoma*	simtom
VERSION	versao*	versi
AGITATE	agitar	.*
CIVIC	civico*	sivik
CLARIFY	clarificar	-
COLLIDE	colidir*	-
COMPUTE	computar	.*
CONTROVERSY	controversia	kontroversi*
INTERACT	-	(ber)interaksi
INTERLUDE	interludio	

It can be seen that fourteen of the eighteen target words had Portuguese cognates, while only six of the target words had been borrowed into Malay. It should be noted, however, that the **noun** *komputer* exists in Bahasa Malaysia, and Malay versions of two other target words - *intersek* and *interlud* - occasionally creep into the speech of bilinguals, although these forms remain very rare. (Educated Malaysians frequently switch between their first language and English while speaking, using lexical items from English as stop-gaps when they cannot recall the words they require in Malay, and vice-versa (Omar 1987). This makes it difficult to establish the full extent of English borrowings into Bahasa Malaysia.)

Discussion of results

The three primary research questions were all answered in the affirmative. The two groups, similar in age and educational level, differed significantly in the number of words they looked up, the time they took to consult the dictionary entries, and the acceptability of the sentences they produced after dictionary consultation. These

findings conflict with Battenburg's finding that native language and cultural background do not affect dictionary use; instead they support the view that there are culturally typical ways of consulting a dictionary.

Three factors are probably responsible for these evident differences in dictionary use: the English language learning background of the subjects, cultural attitudes to task completion, and the proximity of English to the mother tongue.

Whereas the Portuguese subjects had learnt English as a foreign language in the classroom, and had been given very few opportunities to use English communicatively outside class, for the Malaysian subjects English was virtually a second language; many university textbooks were available only in English, and their university education was partially English-medium. This meant that the two groups had a history of acquiring English vocabulary by very different means. The Malaysians had larger lexicons, but they also had far less formal experience of vocabulary learning; by and large they had picked up words by continual exposure to the language. The Portuguese subjects, on the other hand, had learnt most English words through translation exercises and dictionary use.

It may be that the Portuguese subjects had greater prior experience of dictionary use; this would help to explain why they read the entries more quickly, and it would also help to explain why they interpreted the entries more successfully. Greater familiarity with the grammar codes used in learners' dictionaries, and possibly greater familiarity with the grammatical concepts encoded in the dictionary entry, may have contributed to the Portuguese subjects' relative success with the grammar of the target words.

However, the possibility that the Portuguese were more practised in dictionary use does not explain why they chose to look up fewer words despite their inferior vocabulary knowledge. One probable reason for their more confident behaviour is

that they recognized more cognates among the target words. I also suspect, but cannot prove, that the faster and self-assured Portuguese approach, and the more thoroughgoing Malaysian approach were, at least partially, culturally determined. Speed of task completion is probably more highly valued in Portugal than in Malaysia, and looking up words inevitably takes time.

The proximity of Portuguese to English doubtless favoured the Portuguese subjects to a certain extent. Portuguese and English share many lexical and grammatical features, whilst Bahasa Malaysia belongs to an entirely different language family. However, the Portuguese group's advantage in this respect is not so straightforward as might first appear, for the following reasons:

1) cognates are not always recognised as such. Studies such as those of Horsella and Sindermann (1983) and Moss (1992) show that learners often fail to notice the formal similarities between target language words and first language words; in Moss's study, Spanish-speaking students' overall average cognate recognition was only about 60%.

2) Learners may assign an identical meaning, collocational range and syntactic patterning to the cognate target word, when in fact it differs from the word in the learner's first language in one or more of these respects.

3) Speakers of languages unrelated to English often have access to the meaning of unfamiliar English words, either because they are already familiar with another European language, or because there are English borrowings in their first language. In this study, the Malaysian group did not know any European language other than English, but were familiar with many words of English origin which had been borrowed into Bahasa Malaysia. Moreover, whereas Portuguese speakers may be unaware of the relationship between a word in their own language and a word in English, Malaysian speakers are often conscious of the "Englishness" of borrowed words, because most have not been assimilated into the language sufficiently to admit

the word formation processes common to native words. (The Malay noun *interaksi*, from INTERACTION, is exceptional in its admission of the verb-forming prefix *ber-*).

As can be seen in Table 4.4 above, where a related word existed in one language but not in the other, this did not always place the language with the related word at an advantage. ERR, INCORPORATE and COLLIDE were used more successfully, but RETARD, AGITATE and COMPUTE were used less successfully by those who had a related word in their first language. Little difference in success rate between groups was recorded for three other words which only existed as cognates in one of the two languages: RUDIMENTARY, CLARIFY and INTERACT.

However, where related words existed in both languages the Portuguese group tended to be more successful. Both groups produced a similar percentage of type one sentences for GRAVITY, but while the Malaysians created better sentences with CONTROVERSY, the Portuguese showed greater mastery of three words: SYMPTOM, VERSION and CIVIC. In one case, at least, incomplete correspondence between the English word and the first language cognate was responsible for inappropriate language use. The Malaysian word *sivic* is usually used nominally, as the name of a school subject, and 29% of the Malaysian sentences for CIVIC were placed in category four (wrong part of speech). *Sivic* also occurs adjectivally in Bahasa Malaysia in the expression *kesedaran sivic*, which is commonly translated within Malaysia as "civic-minded", but this translation was coded as type two when it appeared in the experimental data because the English judges thought the use inappropriate.

Most interesting were results for the three cases where no related word existed in either language - ENLIGHTEN, INTERSECT and PERPETRATE - because these target words tested each group's ability to interpret dictionary information without the

influence of prior lexical knowledge. In all three cases the Portuguese were substantially more successful at interpreting the dictionary entry.

4.3. Conclusions

The first and most clearly demonstrable conclusion to be drawn from Study Four is that background does affect the success with which learners use dictionaries for productive tasks. Although we cannot be sure what factors contribute to this effect, it seems possible that the proximity of English to the mother tongue, culturally determined attitudes to task completion, and prior experience of dictionary use may all play a part in determining the frequency, speed and success with which learners look up words.

The results also point to the possibility that learners who have learnt English by more formal means, with recourse to reference books, may be better able to interpret dictionary entries than learners who have acquired much of their language knowledge naturally, through extensive contact with English speakers and English texts. In Study Four the formal learners made better use of grammatical information in dictionary entries, and also read entries at greater speed. How generalizable this finding might be to other groups of learners remains a topic for further investigation, however.

Finally, from my investigation into the influence of cognates and related words, I conclude that learners rarely approach the look-up task with a completely open mind. When subjects looked up target words in Study Four, they were, in the majority of cases, matching dictionary information against lexical knowledge already acquired in the first language. In some cases they may also have been matching dictionary information against recollections of previous encounters with the target words in context. This is probably even more true of spontaneous dictionary consultation, because learners are normally unlikely to look up a word that they do not already partially know, or think they know, or that they have not encountered in context. In

this matching process, prior knowledge sometimes overrides conflicting information available in the dictionary entry. There is also the possibility that this prior knowledge is sometimes responsible for the learner's failure to assimilate all the supporting information in the dictionary entry. Productive dictionary use can be likened to a juggling act, where the learner has to activate both old and new data for a given word, but may be unable to call into play all the available information, all at once.

This image may help to explain kidrule, the strategy discussed in Chapter Three whereby learners work with only a segment of the dictionary entry, ignoring other segments which are vital to productive success. The image also led me to wonder what degree of detail would be most appropriate to learners consulting a dictionary to check the meaning of unknown words. If dictionary users are already burdened with prior expectations regarding word behaviour, what sort of dictionary information will be of greatest benefit to them, and what sort of information will they tend to ignore?

Chapter Five

The role of examples in productive dictionary use

5.1. Introduction

The investigations reported in this chapter evolved directly from conclusions reached in Study Three, and research questions that arose from that study. The data from Study Three indicated that both components of the dictionary entry - the definition and the illustrative examples - were read by subjects and used by them to inform their own written work. However, although words and phrases reproduced in the subjects' sentences bore witness to the fact that they *were* reading the examples, Study Three provided no means of assessing whether they were benefiting from what they had read.

Clearly, examples increase the length of a dictionary entry, and learners' dictionaries without examples would be both cheaper and more handy to carry around. In these respects a learners' dictionary without examples would be preferable to one which contained them as an additional feature. An additional advantage to the example-free dictionary might be that the entries could be read more quickly; intuitively this seemed likely, yet I could find no experimental evidence to support the assumption.

On the other hand, the illustrative example is generally regarded as an extremely positive and useful feature. An abundance of examples is one of the distinguishing features of learners' dictionaries, and lexicographers and reviewers write convincingly of the value of dictionary examples as aids to successful language production.

Drysdale (1987), for example, claims that a well-constructed example can perform a wide range of functions, such as distinguishing meanings, illustrating grammatical patterns, and showing typical collocations. Cowie (1989) also justifies examples in learners' dictionaries on the grounds that they help in the decoding process by clarifying meaning and distinguishing between related meanings, and in the encoding process by indicating grammatical patterns, acceptable collocations and native stylistic norms.

Other lexicographers extol examples in a similar vein:

Illustrative quotation can convey a great deal of information about collocation, variety of usage (degree of formality, humorous or sedate context), connotation (affective implications), grammatical context (if a verb, does it take a direct object?) and, of course, designative meaning.

(Landau 1984:166)

A good balance between exemplification, metalinguistic labelling and explanation is more helpful than anything else for production and comprehension. Enlarging editorial notes is dangerous. Examples remain chiefly responsible for the achievement of a good translation.

(Marello 1987:226-227)

Like a picture an example can be worth a thousand words of definition.... examples can supplement and extend the definition, often with great economy of means. An extra burden is placed on the definition without an example. Furthermore, in definitions lacking examples it is more difficult for a user to bring to bear his or her knowledge of a word in context to help clarify meanings.

(Creamer 1987:243)

Such comments imply that, without the benefit of examples, dictionary users would produce far more productive errors; examples, it would seem, help learners to avoid just such errors as I had noted in Studies Three and Four - errors of meaning, collocation, and grammar. How was it then that my subjects had produced these errors, when they had access to examples for the target words? Perhaps the lexicographers' views derive from a consideration of what *should* happen, when a skilled dictionary user consults a skillfully chosen example. Such views may not take into account what *does* happen, when ordinary dictionary users, with a tendency to misread dictionary entries, consult examples which do not adequately reflect all the lexical features which they need to know.

The small amount of evidence available from testing and observing dictionary use is far less positive about the role of examples. Black (1986) found no significant difference between comprehension test scores for words defined with and without examples, and although Miller and Gildea (1985) found that native speaker children

produced more acceptable sentences when they had access to examples, they were ultimately unhappy about their experimental method, and concluded that "this result could not be interpreted to mean that the children learnt more about the meaning of a word from illustrative sentences than they learnt from definitions" (1987:90).

The problem with Miller and Gildea's method was that, when their subjects were asked to write sentences with the aid of examples, many simply reproduced the examples they had been given. Black's study, which involved a multiple-choice decoding task, was flawed because subjects seemed to have simply guessed the correct answer in a high proportion of cases. Both studies also ignored the possibility that subjects might already have known some of the look-up words.

In my Studies Three and Four subjects were permitted to write sentences for the target words without consulting the dictionary entries, if they wished. If subjects chose to look up target words, it was therefore highly likely that they lacked adequate prior knowledge of these words. I also aimed to prevent subjects from simply repeating example sentences from the dictionary entry by requiring them to include both the target word and a given high frequency word in their sentences.

Nevertheless, subjects in my studies still used dictionary examples as a source for collocations and turns of phrase. Table 5.1 lists all the occasions when subjects in Study Four used dictionary examples in their own sentences, produced after look up. The Table differentiates between the Portuguese subjects (Port), and the Malaysian subjects (Malay), and shows the number of members of the two groups who reproduced one or more lexical words other than the target word from the examples in the LDOCE entry for the target word.

Table 5.1: The influence of dictionary example sentences and phrases on sentences produced in Study Four

Target word and examples	Port.	Malay.
ENLIGHTEN Peter thought the world was flat until I enlightened him!	4 (11%)	8 (27%)
an enlightening experience	1 (3%)	4 (13%)
ERR To err is human. (old saying)	1 (2%)	1 (2%)
It's better to err on the side of caution (=to be too careful, rather than not careful enough)	1 (2%)	6 (14%)
GRAVITY Anything that is dropped falls towards the ground because of the force of gravity.	*	*
He doesn't seem to understand the gravity of the situation.	12 (33%)	1 (3%)
INCORPORATE They incorporated her suggestions into their plans.	2 (7%)	2 (5%)
The new plan incorporates the old one.	1 (2%)	-
a new desktop computer incorporating an electronic mail facility	1 (4%)	-
INTERSECT intersecting paths/lines	-	6 (14%)
PERPETRATE to perpetrate a crime/a fraud	15 (31%)	-
(fig., humor) It was the managing director who perpetrated that frightful statue in the reception area.	3 (6%)+	9(22%)+

Target word	Port.	Malay.
RETARD Cold weather retards the growth of food	9 (19%)	16 (37%)
RUDIMENTARY I have only a rudimentary knowledge/grasp of chemistry.	6 (20%)	13 (30%)
Their road-building equipment is fairly rudimentary.	-	-
a rudimentary airfield	-	-
Ostriches have rudimentary wings.	-	-
SYMPTOM The symptoms don't appear until a few days after you're infected.	-	-
Yellow skin is a symptom of jaundice.	1 (2%)	-
The lower production levels are a symptom of widespread dissatisfaction among the workforce.	3 (7%)	-
He recognizes the symptoms, but refuses to admit that he has marital problems.	1 (2%)	3 (9%)
VERSION This dress is a cheaper version of the one we saw in that shop.	-	-
The two newspapers gave different versions of what happened.	7 (21%)	24(57%)
the accepted version of events	2 (6%)	1 (2%)
earlier/later versions	-	-
Did you read the whole book or only the abridged version?	1 (3%)	-
an English version (=translation) of a German play	-	2 (5%)

Target word	Port.	Malay.
AGITATE He became quite agitated when he was asked about his criminal past.	4 (19%)	16 (39%)
to agitate for cheaper school meals	-	1
CIVIC The president's visit was the most important civic event of the year.	10 (24%)	11 (28%)
civic duties/pride	16 (38%)	5 (13%)
CLARIFY Will you clarify that statement?	1 (6%)	-
When will the government clarify its position on equal pay for women?	-	1 (3%)
COLLIDE The two planes collided (with each other) in midair.	-	3 (8%)
The President collided with Congress over his budget plans.	-	-
(no examples given for COMPUTE)		
CONTROVERSY The lie detector tests have been the subject of much controversy.	-	3 (11%)
recent controversies surrounding his appointment to the Cabinet	-	1 (4%)
INTERACT The two ideas interact.	6 (14%)	2 (5%)
INTERLUDE a brief interlude of democracy before a return to military rule	2 (4%)	4 (9%)
total	120 (17%)	143 (21%)

* Some words in the LDOCE example for GRAVITY (FALL, GROUND and DROPPED) were so central to the meaning of GRAVITY, and collocate so commonly with it, that they were used even by those subjects who did not access the dictionary entry, and it is impossible to judge the extent to which the example influenced subjects' choice of words.

+ In the case of the second example for PERPETRATE, subjects did not reproduce lexical items, but rather the structure "It was the who"

From the totals it can be seen that there are 120 instances of Portuguese subjects reproducing elements from example sentences and phrases in the LDOCE entries. 17% of all Portuguese sentences produced after dictionary consultation contained such elements. The Malaysian group borrowed words from the examples slightly more frequently, in 143 cases (21%), yet the subjects in this group were also shown to be more unsuccessful dictionary users. Perhaps they depended more on dictionary examples because they were less confident of their ability to create their own sentences with the target words, yet many of the sentences with borrowed elements were given a low acceptability rating by the panel of judges. It was impossible to say whether dependence on dictionary examples was helping subjects, who might otherwise have been unable to make sense of the target words, or hindering subjects, who were wasting time on a part of the entry that might confuse them, and were gaining false information from examples regarding target word meaning and range of use.

Study Five provided the opportunity to examine the effect of examples more closely.

5.2. Study Five

This study aimed to investigate the value of examples in a learner's dictionary. The best type of dictionary entry is, presumably, one which can be quickly absorbed, and which can be put to effective practical use. I therefore asked two research questions concerning these two attributes of the dictionary entry:

- 1) Do definitions with examples take longer to read than definitions without examples?
- 2) Are definitions with examples more helpful in productive dictionary use than definitions without examples?

The subjects for this study were 40 adult non-native speakers studying English in Britain (at Warwick University, Henley College of Further Education and Eurocentres Cambridge). They came from a wide variety of cultural and language backgrounds, both European and non-European.

The same procedure was adopted as in Study Four, but this time there were two versions of the main test. In each version, example sentences and phrases for half the target words had been removed.

In version A examples for the first nine target words were removed, but examples for the last nine words were retained. In version B examples for the first nine target words were retained, but for the last nine words were removed.

The t-test was used to analyse EVST score by group, to ensure that the two groups A and B were of equivalent ability. The composition of the two groups was not found to differ significantly by this measure. The mean score for group A was 5432, and for group B was 4719 [$t = 1.39$, $p = .168$]. This slight difference in means would not affect results, because each subject in group A and group B looked up some dictionary entries with examples, and some entries without. The main analysis thus involves a within subject comparison.

All the sentences produced by the subjects after look up were analysed and categorised as "appropriate" (type one), or "inappropriate" (types two, three and four), using the categorisation system I had developed for Study Four. In type one sentences the clause surrounding the target word is superficially well-formed in terms of the grammar of English (or contains only inflectional errors), and the normal interpretation of the target word makes sense. In type two sentences the normal interpretation of the target word does not make sense, and in sentence types three and

four the clause surrounding the target word is not well-formed in terms of the grammar of English.

Results for sentences produced by groups A and B after consulting entries with examples were compared with results for sentences produced after consulting entries without examples. The t-test was used to establish whether there were any significant differences in look-up time and correctness.

Results

Table 5.2 below summarises the findings regarding look up time and correctness.

Table 5.2: A comparison of subjects' performance with and without access to dictionary examples.

	Without examples	With examples
Time	31.17	33.93
S.D.	18.75	17.11
% Correct	42.18	52.04
S.D.	23.80	26.63

Look up time was slightly longer for dictionary entries with examples, but this difference did not prove significant [$t = -.69, p = .494$]. Although subjects produced a higher percentage of correct sentences after access to dictionary entries with examples, this also proved non-significant [$t = - 1.75, p = .085$]. There was no evidence that subjects produced more accurate sentences when they were provided with illustrative examples.

Discussion of results

These results do not confirm the opinion of lexicographers regarding the value of examples, and neither do they meet the expectations of the subjects themselves, who, when I explained the purpose of the experiment to them, invariably expressed the belief that the findings would show significantly higher scores for words where the full dictionary entry, with examples, had been available.

I can think of three possible explanations for the apparent failure of examples to improve the appropriacy of the sentences in this study. The first concerns the quality of the dictionary examples, the second concerns the ability of the subjects to process the information in the examples, and the third concerns the appropriacy of the research tool.

The quality of the dictionary examples

In my review of the three major learners' dictionaries in Chapter Three I refer to the debate over whether dictionary examples should be taken directly from a citation corpus, or whether they should be made up by lexicographers. The Oxford Advanced Learner's Dictionary has always preferred the latter approach, on the grounds that more information regarding word use can be squeezed into an invented example than would naturally occur in a single example from a corpus. Such an approach relies very heavily on the lexicographer's intuition regarding what is typical, and even what is acceptable; the lexicographer's idiosyncracies can result in examples which are judged as very odd by others. One study of criteria for error gravity, for example, where subjects were asked to rate the appropriacy of a series of sentences from various sources, found that eighteen out of twenty native speakers judged as erroneous an OALD example sentence *The boy went off in a faint* (Hughes and Lascaratou 1982). Clearly, if the examples themselves are odd-sounding and atypical, it is likely that learner language which reproduces elements of those examples will be judged inappropriate by native speakers.

The COBUILD approach, on the other hand, favours example sentences taken from a corpus of authentic native-speaker texts. Once again the lexicographer's intuition is required to select appropriate examples from a range of recorded instances of use, but the lexicographer will also be guided by the frequency of occurrence of different structures and collocations, and will usually try to select examples which reflect common patterns of use. COBUILD examples have been criticised for the difficulty of the language, and their frequent references to people, places and events which are only explicable in terms of their original context.

LDOCE lexicographers also make use of a corpus of naturally occurring text, but they often modify the original citations by removing unnecessary words and obscure references. In this respect they steer a middle course between the two opposing approaches taken by OALD and COBUILD. The LDOCE example sentences for the eighteen target words used in my studies did not contain difficult language, but they did not always seem to present typical patterns of use. Moreover they shared with examples in OALD and COBUILD a deficiency which does not seem to have been recognised by those who write in praise of dictionary examples: the examples do not indicate to the user which collocational and structural features are obligatory, and which are optional.

Two LDOCE example sentences which seemed particularly misleading were *The two ideas interact* (exemplifying the use of INTERACT) and *It was the managing director who perpetrated that frightful statue in the reception area* (exemplifying the use of PERPETRATE). Although the examples do illustrate some possible syntactic patterns for the two verbs, other information that a well-constructed example sentence is expected to supply - information regarding meaning and collocational range - was partially or completely lacking in both. With these examples the lexicographer seems to have fallen into the trap which Drysdale warns against:

there is a temptation, when preparing school or general dictionaries, to avoid the obvious examples, the set phrases and the clichés, and to stretch students' minds and experience of the language by encouraging them to be creative.

1987:221

The collocation of IDEAS with INTERACT seems to have been an attempt at atypical, creative use, although strangely it also occurs in an OALD example sentence - *ideas that interact*. It sounds so odd that whenever it was reproduced in the subjects' own sentences raters tended to judge it as inappropriate. OALD and COBUILD both exemplify INTERACT much more successfully by distinguishing two separate meanings: one collocating with inanimate nouns in the sense of "have an effect on each other's development or condition", (they give examples from a scientific context for this sense) and the other collocating with animate nouns in the sense of "act together or co-operatively" or "communicate and work together". LDOCE, however, gives only one sense and one example for INTERACT, thereby suggesting that the verb only collocates with inanimate nouns.

The LDOCE example provides some grammatical information (the fact that the verb is intransitive and takes a plural subject) but does not indicate the possibility of a subject in the singular, followed by WITH. These are limitations, but they will not in fact lead directly to productive errors, because the dictionary user is led to believe that the collocational and structural possibilities are fewer than is actually the case. If users follow the structural and collocational guidelines provided they will probably produce error-free sentences. Unfortunately the LDOCE example offers virtually no guidance as to the meaning of INTERACT; the sentence is too abstract and decontextualised to make any real sense. For learners to use INTERACT communicatively in sentences of their own, they will have to concentrate on the definition, rather than on the example.

LDOCE provides only one example for INTERACT, but two for PERPETRATE, the first of which - *to perpetrate a crime/a fraud* - presents more typical collocations, and the second of which - *It was the managing director who perpetrated that frightful statue in the reception area* - is more idiosyncratic. This second example is intended to illustrate the humorous use of the word, thus presumably "helping the user to compose according to native stylistic norms". The result was probably more misleading than helpful, however; subjects did not attempt to convey humour in their own sentences, and although none of them collocated PERPETRATE with STATUE, the example probably encouraged the view that PERPETRATE had a far wider collocational range than is in fact the case. Thus one subject produced *Last year winter perpetrated many horrible storms* and another *it is perpetrated by eating icecreams at winter*, both of which were judged inappropriate.

Although a number of subjects borrowed from the example the construction *it was thethat....* (which is not governed by the choice of PERPETRATE as verb), few recognised that PERPETRATE requires an animate subject (information which is essential to productive success). The dictionary user has no means of knowing which parts of an example can freely vary and which parts are fixed; presumably some subjects took the structure *it was thethat....* to be an obligatory or typical feature of PERPETRATE. This would be quite understandable, as some examples do illustrate a fixed structure of which the target word forms part. The two examples for the target word ERR, for instance, illustrate the only two structures in which ERR is likely to occur in modern English - *To err is human*. (old saying) and *It's better to err on the side of caution* (=to be too careful, rather than not careful enough). The label (old saying) and the change of typeface for **err on the side of** are intended to indicate the fixed nature of these expressions. Most subjects did not avail themselves of this information, but one or two who had clearly been receptive went on to produce:

It is better to err than spend a lot of money

and

Paul robbed a bank but the err is human.

Only the first was counted as acceptable.

Although examples are not a very successful means of circumscribing collocational range, they might help to clarify tendencies if several examples were placed together, each with collocations from the same semantic area, or sharing essential lexical features. This is not a practice followed by LDOCE, presumably because of lack of space. CLARIFY, for example, is given two object collocations - *statement* and *position* - both of which seem natural and typical, but which are unfortunately sufficiently different in meaning to justify further unacceptable collocations in learner language production, such as *doubt* and *illness* in my data. Likewise only one example for PERPETRATE indicates the necessity for an animate subject; in the first example the verb is in the infinitive.

The number of example sentences provided by LDOCE for each target word ranged from none (for COMPUTE) to five (for VERSION), but no policy was apparent to account for this variation. Subjects did not produce a greater number of errors when they were not allowed access to the five examples for VERSION, which is a relatively simple word from the syntactic point of view, but I felt that errors might have been prevented if certain target words with more complex grammatical and collocational restrictions had been exemplified more fully. Only one or two examples were given for CLARIFY, COLLIDE, INTERACT, INTERSECT, PERPETRATE and RETARD, yet these were the words that caused greatest problems to subjects in all my studies.

The ability of the subjects

From the findings from Study Four I concluded that first language and cultural background played a part in determining the success of productive dictionary use. In Study Four the Portuguese learners were more successful than the Malaysian learners

of English; they read dictionary entries more quickly, and produced more appropriate sentences after dictionary consultation.

My subjects in Study Five came from a variety of language backgrounds, and this is reflected in their mean scores, which tended to fall between those of Malaysian and Portuguese groups, with higher standard deviations. EVST scores in Study Five were similar to those in Study Three, and lower than those for either of the two groups in Study Four; reading times for entries with examples were also slightly faster than in Study Three, but slower than for either group in Study Four. The percentage of acceptable sentences produced after accessing entries with examples was somewhat higher than average for both Studies Three and Four, however, and close to that of the Portuguese group.

The broad similarity of the data from Study Five and the data from Studies Three and Four suggests that my subjects' productive dictionary use was fairly typical of advanced learners of English. Their normal patterns of behaviour and relative overall success in the productive task suggest that their failure to benefit significantly from the dictionary examples was not due to any unusual defects in their dictionary-using skills.

Nevertheless the data reveals that these subjects were not receptive to information available to them in the dictionary examples. It was evident that they resisted the influence of structural models, as can be illustrated by examples from the data for COLLIDE. LDOCE provides two example sentences for this target word which encapsulate a number of pieces of information regarding its grammatical behaviour - *The two planes collided (with each other) in midair*, and *The President collided with Congress over his budget plans*. These examples show that COLLIDE can be used as an intransitive verb, that it can take a plural or a singular subject, animate or inanimate, and that a second agent can be linked to the verb by means of WITH. It

would appear that subjects generally ignored the help that the example sentences offered, however. They frequently used COLLIDE transitively in their own sentences:

The window was collided by the car

A car collided my house so many windows were broken

My car collided the window of a candy shop

The window was broken because it was collided with the stone

and they employed grammatical collocations other than

WITH:

The basketball collide on the window of my house

The stone collided into the windows.

As I pointed out in the preceding section, examples do not prohibit the use of alternative structures; the examples show what is possible, not what is impossible. It may be the case, therefore, that subjects were conscious of the patterns illustrated in the two example sentences, but decided to freely experiment with other structures which might or might not be acceptable. If this were the case, it would imply that subjects were more concerned to express their own ideas than to achieve absolute accuracy. This attitude might reflect the influence of communicative language teaching, which may, in some of its manifestations, sacrifice accuracy in an attempt to foster fluency and self-expression.

An alternative possibility is that the subjects lacked sufficient grammatical knowledge to recognize what kind of information the dictionary examples were making available to them. This is borne out by informal conversations held with some of the subjects when they had finished their tests, in which it became clear that concepts of transitivity and grammatical collocation were very poorly understood. Once again, the influence of communicative language teaching may be responsible in part for my subjects' lack of language awareness, because communicative teaching tends to

prioritize other requirements for successful communication over the kind of overt grammatical knowledge required to get the most out of a dictionary example.

The appropriacy of the research tool

One further explanation for the apparent failure of dictionary examples to improve the language production of my subjects lies in the method of data analysis chosen for this experiment. The coding system I adopted may not have been a sufficiently fine instrument to detect partial growth in the understanding of word meaning and use as a result of exposure to examples. Subjects who displayed some understanding of word meaning might still produce sentences coded as inappropriate if their word knowledge was less than complete.

There is some indication that subjects derived benefit from access to examples in the patterns of error across the two groups. For example, ten subjects in group B who were given access to the definition only, used CIVIC as a noun in sentences such as:

His father is the most famous author and civic in United Kingdom

My father is Coventry civic.

Only three subjects in group A, who had access to examples, did the same.

However, although subjects in group A managed to avoid using CIVIC as a noun, a number of them still produced inappropriate sentences, such as:

Susan's father who is a Department Head of Engineering will visit an important civic city at the end of this year.

The slight superiority displayed by the group with access to examples for CIVIC was not sufficient to produce an overall significant difference between the two groups, but it does suggest that the presence of examples is capable of influencing results in certain cases. In other cases, subjects may have acquired word knowledge through access to examples, which they were as yet unable to express productively. This study

did not investigate the long-term effects of access to dictionary examples, nor the perceptions of the dictionary users, although, as previously stated, in post-experiment discussions my subjects expressed very positive views regarding the value of examples.

5.3. Conclusions

In this study, dictionary examples were not found to significantly affect the success of productive dictionary use. Lexicographers and writers on lexicography see the potential of the example as a conveyor of linguistic information, but it would appear that the examples used in this study did not always live up to this potential, nor did the subjects always recognize it, where it existed.

It may be, however, that the examples in my study were serving a purpose my experimental method could not fully detect. Results for sentences produced with the aid of examples were better than for sentences produced with the definition only. The difference was not significant, but standard deviations were also relatively high, and it is possible that a more positive result could have been obtained from a larger sample of subjects, or a more homogeneous group.

My method of judging the appropriacy of the sentences also depended heavily on grammatical and collocational acceptability; it could not accurately assess whether examples had played some part in developing subjects' understanding of word meaning, as opposed to word behaviour. We know from the questionnaire based studies reported in Chapter One that word meaning is the most important type of information learners seek in their dictionaries, and word grammar is given relatively low priority. Subjects were, perhaps, getting what they wanted from examples, and improving their understanding of word meaning in ways inaccessible to the measuring instrument.

Chapter Six

Conclusion

6.1. A summary of the findings and their implications

The five experiments reported in this thesis investigate the interface between EFL dictionary, user, and task. Studies One and Two examine receptive dictionary use during English language tests, while Studies Three, Four and Five look at the way English language learners produce target words in context sentences after dictionary consultation.

The first two studies found that those subjects who consulted dictionaries during an EAP reading comprehension test tended to take longer to complete the test, but did not achieve significantly higher test scores. In these studies, dictionary use was monitored under conditions similar to those in which dictionary use might naturally take place; indeed, many tertiary institutions regard dictionaries as essential tools for certain kinds of language work, and allow candidates access to dictionaries during examinations. The results appear to suggest that test candidates derive no real benefit from their dictionaries, an impression partially borne out by a closer analysis of the words subjects looked up, and the treatment of these words in the dictionaries. It was found that the dictionaries were used most frequently to check the meanings of words which were not essential for the correct answering of the test questions, and it was also found that the dictionaries often provided insufficient information about the meaning of those words that candidates really did need to know.

Such findings have implications for language teachers and dictionary makers.

Subjects apparently lack the ability to distinguish between essential and non-essential textual information - a necessary reading skill regardless of whether the reader employs a dictionary or not. To distinguish what is essential requires the reader to consider text structure and his or her own reading purpose, and it is an important

preliminary to contextual guessing. Methods of language teaching which treat the text as a repository of language items obscure the real-life need to focus on some parts of the text and skim over others; training in the choice of which unknown words to look up, which to guess, and which to ignore would perhaps enable examination candidates to benefit from dictionaries to a greater extent.

The fact that the dictionary entries themselves did not always clarify meanings intended in the texts may in part have been due to the slightly technical nature of the texts, which were taken from the *New Scientist* rather than from a "general English" source. Although LDOCE and OALD are aimed at a broad market of advanced learners, a great many learners at this level are studying or intend to study in the medium of English; such learners typically read texts of the *New Scientist* type, while studying their own subject specialisms, in the EAP classroom, and when taking EAP tests. A closer consideration of "user typology" (Hartmann 1985) might increase dictionary makers' awareness of the kinds of texts dictionary users typically read, and thus improve the match between entry information and readers' needs.

The results of Studies One and Two do not automatically imply, however, that readers trained by current methods will fail to benefit from the consultation of current EFL dictionaries. The test itself was a further factor in the studies, and may have obscured the value of dictionaries to readers because EAP test designers typically avoid testing candidates on their lexical knowledge. Studies One and Two examined dictionary using behaviour in a natural context - the EAP test - but the findings may not be generalisable to other natural reading contexts such as textbook study or even examinations in the subject specialism, where the examiner is assessing the candidate's ability to interpret text meaning, rather than the exhibition of reading skills applicable to a range of texts. In some real-life reading contexts interpretation of the entire text hinges on the learner's understanding of one or two previously unknown lexical items. In such cases, I believe, dictionary use would be bound to

benefit the reader - provided that the reader looked up the appropriate words, the dictionary provided appropriate meanings, and the reader could understand the dictionary entries.

Study Three set out to compare the intelligibility of Oxford Advanced Learner's Dictionary, Longman Dictionary of Contemporary English, and Collins COBUILD English Language Dictionary. The findings indicated that users of the three dictionaries behaved in much the same way and with a similar degree of productive success; little difference was found between the three user groups regarding the number of words they looked up, the average time they took to read dictionary entries, and the number of correct sentences they produced after look-up, although OALD users were found to make more semantic errors. Comparative studies of this sort are rare, and although earlier editions of LDOCE and OALD were tested for perceived intelligibility by MacFarquhar and Richards (1983) I am not aware of any prior comparison which includes COBUILD, the dictionary which departs most radically from traditional native-speaker dictionary defining style. The objectives of Study Three were limited, and the study was not designed to compare user preferences or the long-term benefits of the three different defining styles. It is therefore possible that the defining styles influenced user behaviour in ways which the study could not record. The existing evidence suggests that LDOCE and COBUILD are of equal value, however, while OALD provides only marginally less support for the user.

Whereas choice of learners' dictionary appears to have little effect on user behaviour, the findings of Study Four suggest that first language background and culture greatly influence the frequency, speed and efficiency of dictionary use. The Malaysian subjects and the Portuguese subjects who took part in Study Four were similar in age and educational level, but they differed significantly in the number of words they looked up, the time they took to consult the dictionary entries, and the acceptability of

279

the sentences they produced. In the light of these findings a case can be made for adapting dictionary design to meet regional needs. Etymological information or a simple system of flagging cognate headwords might benefit the learner reader, although in my studies of productive use cognates often encouraged error because of collocational and syntactic variations between the first and foreign language words. Thus the generally more successful Portuguese group had more problems than the Malaysians with certain target words such as RETARD and AGITATE, which are cognate with Portuguese words but which have not been borrowed into Bahasa Malaysia. For productive use a more helpful approach might be to present errors commonly made by the user group for which the dictionary is intended, within the dictionary entry, or possibly in an extra column. Understandably, publishers aim for as large a market as possible and will hesitate before including information which delimits a dictionary's readership, but monolingual learners' dictionaries which address the needs of a specified language group might ultimately justify the initial investment, just as popular bilingual dictionaries do. Advanced level EFL dictionaries could elaborate on the example set by the elementary-level **Longman Learner's Dictionary of English inglese/italiano italiano/inglese**, which does carry some warnings regarding first language-related lexical errors. The comparison of first and foreign language lexical behaviour as an aid to productive dictionary use is rare at present, however; the semi-bilingual **Harrap's English Dictionary for Speakers of Arabic**, for example, simply translates each term at the end of the monolingual entry, with no mention of possible variations in collocational range, syntactic behaviour and register.

First language differences were apparently not the only cause of variance in the dictionary-using behaviour of the Malaysian and Portuguese groups in Study Four. The Portuguese group were more successful in their use of those target words for which no related word existed in either language, and overall the Malaysian group produced a significantly higher number of type three (grammatical) errors. Despite

200

the fact that their vocabulary size scores were lower, it would appear that the Portuguese had better dictionary-reading skills than the Malaysians, and could interpret grammatical information in the dictionary entries more successfully. One possible reason for this is that in Portugal English is a foreign language, and therefore tends to be acquired in the language classroom with the aid of dictionaries and reference books, while in Malaysia English is a second language, and learners can gain considerable knowledge of the language through exposure to it outside the classroom. Teachers of dictionary skills in an ESL context should bear in mind the possible mismatch between their learners' vocabulary size and their formal knowledge of vocabulary behaviour. Like the majority of native speakers, ESL learners may lack the metalanguage to describe such concepts as valency, countability and register, despite being able to employ their existing wordstore appropriately. Although complete mastery of EFL dictionary coding systems eludes most of us (see West 1987), the concept and the coding of an essential feature such as transitivity can be taught effectively. Training of this kind would have greatly improved the acceptability of the Malaysian group's sentences.

In Study Five it was found that the presence or absence of example sentences and phrases in the dictionary entries had little immediate effect on productive word use. Subjects did not take significantly less time to read dictionary entries from which the examples had been removed, and they did not produce a significantly greater number of correct sentences when they had access to entries complete with examples. This finding was perhaps the most surprising of the findings from all five studies, because it not only challenged received opinion regarding the value of dictionary examples, but it also conflicted with the expectations of the subjects themselves, who anticipated far greater success with access to examples than without. The data suggested that once again three factors were at play: the user, the dictionary, and the task. Users often failed to assimilate collocational and grammatical information expressed within the example, but LDOCE's provision of examples was also inconsistent; some target

words with a predictable range of use were exemplified more fully than others whose collocational and grammatical behaviour was highly restricted, and some examples were atypical or encouraged false assumptions about the range of use of the target word. The task itself also perhaps failed to reveal an increase in word knowledge which could not be immediately expressed productively. It is possible that the limited contexts of use provided by the dictionary examples were insufficient in themselves to noticeably improve language production, but nevertheless provided an initial stock of knowledge which could be stored in the learners' mental lexicons and built upon in subsequent encounters with the target words.

6.2. A critique of the experimental method employed in Studies Three, Four and Five

The method of data collection employed in Studies Three, Four and Five was innovative. Subjects interacted with a computer, and the computer program recorded whether or not they accessed a dictionary entry, and how long they spent reading it. The program also controlled the amount of dictionary information available to the subject.

This way of collecting information about dictionary use is reminiscent of the method envisaged, but not put into practice, by Hatherall (1984). Hatherall felt that data gathered by a computer would be more accurate and more complete than that collected in the standard way, using dictionaries in book form and a human observer:

if the dictionary user is himself looking up data in a computer rather than a book, his behaviour can be monitored with ease, at least in terms of what and when (how often). Wholly reliable information in these two areas should prove invaluable in also explaining how and why.

A further advantage of using a computer to monitor subjects' behaviour is that it is unobtrusive, and therefore does not interfere with normal reading and writing processes. This is the main reason why Hulstijn (1993) used computers in his study of learners' reading and inferencing strategies.

Although our studies were developed independently, Hulstijn's methodology was remarkably similar to mine. Like me, he tested his subjects' vocabulary size, and set them a language task with the option of looking up any words they did not know on a computer screen. Hulstijn, however, was concerned primarily with reading behaviour; he set his subjects reading tasks, used the computer to monitor the order in which they looked up words in texts, and offered them access to translations rather than monolingual dictionary entries.

As far as I am aware, Hulstijn's is the only study apart from my own to use computers to monitor the process by which learners complete language tasks. The technique may become more popular, however, because computer based methodology seems to combine many of the advantages of test-based research with those of observation-based research. Large numbers of subjects can be monitored, as in test-based research, and some or all of the data may be easily quantifiable (for example the number of words looked up, the time taken to look up the words, and each subject's vocabulary score). At the same time, as in an observation-based study a detailed record of each subject's behaviour can be made, which can subsequently be analysed to gain information regarding the quality of individual dictionary consultations.

Information of the type that earlier studies of dictionary use failed to obtain can easily be gathered by using a computer. In the test-based research into dictionary use reviewed in 1.2., many researchers could not establish with any certainty the extent to which dictionaries had been used to answer test questions, and some researchers even had difficulty establishing whether or not a dictionary had been used at all (for

example Tono 1989, Atkins and Knowles 1990, Luppescu and Day 1993). In some studies the titles of the dictionaries used by individual subjects do not appear to have been recorded (Tono 1988, Luppescu and Day 1993). The larger the number of subjects involved, the more difficult it becomes for a human observer to monitor dictionary use in detail.

In the smaller-scale observation-based studies, researchers had a better chance of recording details of dictionary use during a given task, but none of the studies reviewed in 1.3. provides precise information in all the areas considered in my studies: dictionary type, dictionary entry information, and quantity of use in terms of frequency and length of time.

By choosing the target words for the production task, and by providing the appropriate dictionary entries on the computer screen, I was able to control the type and quantity of dictionary information available to each subject. The time taken to read each entry could be recorded precisely, and the experimental method thus solved a problem experienced by earlier researchers investigating the effects of dictionary look-up, who monitored overall task completion time, but did not have exact information on how the time had been spent (Bensoussan Sim and Weiss 1984, Neubach and Cohen 1988, Luppescu and Day 1993).

Although the task assigned to subjects in my studies did not reflect real-life language activity to the extent of some (but not all) the tasks in the test-based and observation-based studies reviewed in 1.2. and 1.3., it did allow subjects to choose whether or not they wished to access dictionary information. In this respect it replicated normal dictionary-using behaviour; subjects only looked up words when they needed to in order to complete the task. In some studies (Miller and Gildea 1985, Black 1986) subjects were required to look up words regardless of whether they knew them already or not, so the data collected did not necessarily reflect new information gained from dictionary consultation.

Of course, dictionary look-up was simplified for my experiment; subjects did not have physical contact with a dictionary in book form, and they did not have to search for the appropriate entry. I do not feel that this invalidates my findings, as my studies were designed to investigate the way subjects interpret dictionary entries, rather than their word-finding skills.

It must be acknowledged, however, that the computerised format prevented subjects from accessing grammatical information relating to the target word entries in other parts of the dictionaries, or in the extra column in the case of the COBUILD users in Study Three. Subjects were also unable to look up words other than the target words; if given access to a dictionary in book form they may, for example, have looked up words used within the target word entries. It is impossible to judge the extent to which these activities would have taken place, had a dictionary in book form been available to the subjects. Certainly Miller and Gildea (1985), who appear to have used dictionaries in book form when they set a similar task, make no mention of extensive dictionary searches. Nevertheless it remains a possibility that the subjects' performance would have improved if they had been allowed access to the complete dictionary (and the extra column, in the case of COBUILD users). My findings therefore only provide evidence of the effect on language production of information found within the dictionary entry; they do not reflect the effect of dictionary use in conditions where subjects might look up grammatical codes and related words, and might indeed have access to a variety of further aids, such as bilingual dictionaries, grammar books, and informants.

In two further respects my studies are also open to criticism, suggesting that the experimental method may require further refinement. The first, and simpler, problem, relates to the choice of target words and high frequency "prompt" words for the studies.

The target words were taken from the central bands of Nation's University Word List, without any regard for their meaning or syntactic behaviour. The original list was then systematically reduced by eliminating common words, words with more than one dictionary entry or with excessively long entries, and technical terms. The list was reduced still further on the basis of piloting, by the elimination of words which did not elicit revealing data (see Chapter Three). In retrospect, it might have been useful to balance the list at an early stage by choosing target words with representative patterns of syntactic behaviour. This would have provided data on the relative value of entries for words from different word classes. Also, the choice of ERR as a target word was probably a mistake, because although it did provide interesting data (it was often used as a nominal form), its rather archaic flavour and restricted range of use meant that even those subjects who had perfectly understood its meaning and syntax produced odd-sounding sentences. This created problems at the analysis stage, when raters found it difficult to decide whether the sentences were appropriate or not. (Of course, it could be argued that the dictionaries themselves are responsible for conveying all stylistic constraints to the productive dictionary user, and therefore that all odd-sounding sentences are the result of unsuccessful look-up.)

The high frequency "prompt" words were included in the task so that subjects would be forced to create original sentences for the target words, rather than merely repeating segments from the dictionary entry. They were considered to be neutral, core words which would neither aid or mislead subjects in their interpretation of the target words, but there remains a possibility that subjects were influenced by them, and sought a semantic connection between these words and the target words. Possible ways of reducing the effect of this influence in further studies might be to vary or rotate the high-frequency words, or to provide subjects with more information about the design of the test, stating that the words are randomly paired. The first of these options would complicate the experimental design considerably, however, while the

second option might not prove effective because subjects might not understand the information they were given.

The second and more serious problem with the methodology of the studies concerns the coding systems used to investigate the appropriacy of subjects' sentences. Several systems were tried, but none proved entirely successful. An examination of prior studies involving error gravity rating shows that researchers have tended to work with doctored or specially picked examples (James 1977, Hughes and Lascaratou 1982, Sheorey 1986). James (1977) believed that it was inadvisable to rate any sentence that was not entirely self-contained, because of the danger of misinterpreting the meaning; for an error to be judged, it had to be recognizable in no further context than the sentence it occurred in. My data could not be selected or doctored without invalidating my findings, but it was clear at the analysis stage that some sentences could be interpreted in more way than one, and others were so syntactically deviant that the target word meaning was entirely obscured.

Studies which have attempted to categorise undoctored sentences produced after dictionary consultation (Miller and Gildea 1985, Meara and English 1988) do not seem to have put their coding systems to the test by comparing their scores with those of an external rater. This is understandable, given that the coding systems are fairly complex, and an external rater (if one is to be found) may prove less reliable than the researchers themselves, who know the data well. I found that the raters who judged error gravity and categorised errors in my data did not always concur, and this, according to Sheorey (1986), is entirely to be expected: "individual teachers of ESL, regardless of their native language, tend to evaluate errors or error types differently". One possible solution to the problem of inter-rater variation might be to employ large numbers of raters, and identify overall rating trends. For these studies, however, such an approach would have proved impracticable, as large numbers of competent raters

287

with a background in Applied Linguistics and English Language Teaching were not readily available.

Despite the fact that individual raters varied in their interpretation of error gravity and in their application of the categorisation system, the rating systems did succeed in providing a broad general picture of the level of appropriacy of sentences produced under different conditions. Subjects' behaviour could then be examined more closely through discussion of the possible causes of individual errors, so a more detailed picture of productive look-up behaviour could be built up.

6.3. Towards a theory of productive dictionary use

Because all my findings were derived from data obtained under controlled experimental conditions, it is necessary to generalise from these findings in order to build up a picture of the way EFL learners read and interpret dictionary entries under normal circumstances. The task in Studies One and Two was closest to ordinary dictionary-using experience, because the subjects could freely choose to look up whatever words they wished whenever they wished, in an environment where dictionary use is often permitted. Nevertheless even in these studies the task did not necessarily elicit the same kind of reading and dictionary using behaviour that occurs normally. In Studies Three, Four and Five new words were presented to subjects without any surrounding context, something that rarely happens during the vocabulary acquisition process. Under normal conditions, learners will only look up words in a monolingual dictionary if they already know (or think they know) something about them. They may encounter the look-up word in one or more contexts, and form a partial opinion about its meaning and use before consulting the dictionary (a strategy investigated by Hulstijn (1993)). Alternatively they may start their search by looking up an L1 word in a bilingual dictionary, and use their monolingual learners' dictionary to confirm the bilingual dictionary information (as attested by Tomaszczuk (1979)).

They are most unlikely to start their search from "cold", and progress immediately to productive use, as my subjects were required to do.

However, my data provides evidence that, even in such an unnatural-seeming situation, learners tend to treat the look-up task as a process of matching and comparing what they read in the dictionary with their own prior knowledge, rather than as a process of acquiring totally new and unexpected information.

Three forms of prior knowledge were available to my subjects:

- 1) inferred knowledge from previous encounters with the target word.
- 2) knowledge of the meaning and syntactic behaviour of a cognate or related word in the L1
- 3) knowledge of the meaning and syntactic behaviour of a word or phrase used within the dictionary entry, which could be substituted for the target word.

Of these three types of knowledge, the first was probably the most useful, but given the difficulty level of the target words it was also the type of knowledge my subjects were least likely to possess. Some of the errors in my data appeared to be the result of confusing the target word with a known word (perhaps because of orthographical or phonological similarities), and allowing prior knowledge of that word to override the information in the dictionary entry. Learners might spend several seconds reading the dictionary entry, but reject it in favour of conflicting information they believe themselves to possess already.

My subjects' use of target words often appeared to be influenced by the second type of prior knowledge - knowledge of a cognate or a related word. Although the dictionary entry in its entirety might not match the information the subjects possessed regarding the meaning and behaviour of the L1 word, it was often possible for them to isolate a segment of the entry which suggested that the L1 and the L2 word behaved

in an identical way. Errors which seem to be the result of focussing on only one synonymous segment of the entry can be classed as kidrule errors, but EFL learners are probably encouraged in their use of this strategy when they find they can match a segment from the target word entry with the L1 cognate of the target word.

In cases where subjects had no prior knowledge of the target word, kidrule also seemed to be employed as a strategy for arriving at word meaning and use. Learners might form a working hypothesis about the behaviour of the new word on analogy with a word that they already know in the dictionary entry (the third type of prior knowledge).

A variation of the kidrule strategy might be to derive a familiar concept from segments of the dictionary entry; Miller and Gildea observed this strategy being used by their subjects for deriving information about word meaning from example sentences; they suggest that such a strategy might also occur naturally in the vocabulary acquisition process:

When used with illustrative sentences the LUCAS task seems to provide more insight into children's thinking than we had anticipated. Certainly, something more than a kidrule strategy is at work here. The students cannot simply search through an illustrative sentence to find a familiar word; they must first abstract a familiar concept from the unfamiliar word's context of use, and only then apply kidrule. Which suggests that perhaps the kidrule strategy is simply the second half of the more general strategy that children use to pick up new words by hearing them used

(1985:24)

It would seem that dictionary look-up strategies and natural vocabulary acquisition strategies are not unconnected. In both cases, the new word and concept is linked to familiar words and concepts, even if the link is a tenuous one which results in partial or total misunderstanding of the real meaning of the new word. At this stage both receptive and productive use of the new word will cause problems for the learner; the

false or incomplete equation between the old and new will result in interlanguage errors. Once the link has been made, however, more information may be gradually added to the new entry in the learner's mental lexicon - information gathered from further encounters with the word in context, and perhaps also from further consultations of the same dictionary entry.

Meara comments on the gradual process of natural vocabulary acquisition:

It seems that words are absorbed slowly over time, and that only gradually do they become fully integrated into the learner's personal stock of words, when he can use them with the same sort of fluency that characterises the words he uses in his native language"

(1980:227)

Likewise Béjoint points out that "the [dictionary] user does not progress at once from ignorance to total knowledge" (1988:139). The sentences in my data attest the failure of subjects to assimilate the full range of productive information available to them. Regardless what quantity of information is available in the dictionary entry, dictionary consultation seems to add only one or two more information components to whatever foundation the learner already possesses.

This conclusion is not a negative one. Learners do appear to acquire word knowledge through dictionary consultation, but often not in sufficient quantity to enable productive use of a new word after a single consultation. Further research is needed to investigate the long term effects of the "dictionary habit", and the processes by which dictionary consultation contributes to the gradual acquisition of word knowledge. My subjects, who often produced such odd and unacceptable sentences in my studies, may have become more receptive to the use of the target words in context, and may have more readily acquired further word knowledge, once the process of acquisition had been activated by looking the words up.

6.4. Questions that remain to be answered

The findings discussed in this chapter suggest many further research questions that need to be addressed. For some types of investigation the computer-based method of data gathering adopted in Studies Three, Four and Five seems appropriate, while other types of investigation demand a departure from this approach, and the use of more open-ended data-gathering techniques.

The computer-based experimental method entails the manipulation of the variables of user type, dictionary type and entry information structure to find combinations that are significantly more or less successful. These variables could, in theory, be played off against each other until the best matches were found; by so doing it might ultimately be possible to prescribe the best type of dictionary with the ideal quantity and type of entry material for a given population of learners.

As far as dictionary type is concerned, a first step in this direction might be to establish whether EFL dictionaries are really more useful to the EFL learner than native speaker dictionaries. When designing Study Three I assumed that EFL dictionaries would be more accessible and useful to my subjects, and I therefore did not include a native-speaker dictionary in the study. Béjoint (1981) did not share this assumption, however, and although Bogaards (1991) provides evidence that French learner's dictionaries are better vocabulary learning tools than French native-speaker dictionaries, and MacFarquhar and Richards (1983) found that learners preferred OALD and LDOCE to Webster's New World Dictionary, the productive value of native-speaker dictionaries in an EFL context still requires investigation.

The dictionary variable could also be manipulated by including bilingual or monobilingual dictionary entries, although this would entail considerable changes to the experimental design, as both the English and the first language entries would need to be made available to the subjects. Moreover, bilingual dictionary users writing in

English would normally consult their dictionary via the first language, so the existing test prompt of target word plus high frequency word would not elicit natural productive bilingual dictionary use.

Having dismissed existing LDOCE examples as useful aids to language production in Study Five, an interesting variation on this study would be to compare sentences produced after access to KWIC concordance output as opposed to standard dictionary entries. The possibility of exposing learners to multiple examples of target word use on a computer screen was proposed by Miller and Gildea (1987) before the advent of classroom concordancing, and now that concordance programs such as Longman Mini concordancer and Oxford Microconcord are readily available, the relative advantages of dictionaries and concordance output as aids to productive language use are even more worthy of investigation, especially because concordances can be derived from corpora which represent the kind of texts a specified group of learners is most likely to need to produce.

The standard learners' dictionary entry could also be manipulated in a number of ways other than simply removing the example sentences and phrases. On the assumption that a EFL dictionary entry should contain all and only the information that the learner needs, an expanded entry could be progressively reduced to the point where learners begin to produce a significantly greater number of lexical errors. Such a point might be reached at different stages for each target word - a possibility which suggests that target words might be chosen according to a different system in future studies, so that the researcher could group them according to such factors as their syntactic behaviour and conceptual complexity. It might thus be possible to arrive at an optimum entry length for different categories of words. Complete data for every conceivable word category is perhaps too ambitious a target, or at best a very long-term goal, but in the short term it might prove possible to discover the ideal quantity of information for the productive use of certain verb patterns, for example. It is

unlikely that the optimum entry length would prevent dictionary consultation resulting in productive error, but the optimum length would constitute the point beyond which further dictionary information would cease to significantly improve learners' language production. Findings of this sort would be of immediate practical value to lexicographers, who might save dictionary space by reducing the length of entries for certain categories of words, while increasing the information available for other categories.

This kind of research would obviously need to be linked to studies of representative groups of dictionary users, as the optimum entry length may vary according to the varying types and varying amounts of existing word knowledge possessed by each user group. The dictionary user variable could be manipulated in a number of different ways. The significant differences between Malaysian and Portuguese dictionary use recorded in Study Four were ascribed to the effects of an ESL and an EFL education, yet my hypothesis that ESL and EFL dictionary users possess different levels of dictionary-using skill clearly needs to be tested against other ESL and EFL populations. The behaviour of speakers of cognate and non-cognate languages, and of Romance language speakers versus Germanic language speakers could also be usefully compared, while further variables include subject specialism, study skills experience and pre-test dictionary skills training. The effect of all these on dictionary-using behaviour could be investigated using the data-gathering method employed in Studies Three, Four and Five.

One more user variable, the effect of personality on dictionary use, may prove more difficult to investigate by existing methods. Having observed that the Portuguese subjects tended to work faster, and take greater risks, while the Malaysian subjects were slower and more cautious, I would like to investigate these indicators of dictionary-using success in greater detail. This line of enquiry may remain closed for the time being, however, as existing English and American personality tests do not

provide reliable results for non-native speakers belonging to an alien cultural background.

Although a huge range of research enquiries can be pursued by means of the computer-based data-gathering device used in my studies, there remain some questions regarding EFL dictionary use which are more appropriately addressed by other means. In particular the long-term retention of word knowledge after dictionary consultation is worthy of investigation. In my Study Three, little difference was found between the behaviour of users of LDOCE, OALD and COBUILD, yet it is possible that dictionary consultation produced effects my data-gathering method could not register. Bogaards (1991) found that the success of bilingual dictionary users and DFLE users was reversed when subjects were tested again after a lapse of fifteen days. A similar test might reveal differences in the success rates of LDOCE, OALD, and COBUILD users. Similarly, the presence or absence of examples might be found to affect learner word knowledge in the long-term rather than the short term.

Finally, the opinions of the dictionary users themselves would appear to be a rich source of insight into the perceived value of various types of dictionary entry. Opinions gathered by formal and structured means, such as questionnaires, largely prevent the expression of information in areas unanticipated by the researcher, but in the course of the experiments reported in this thesis many subjects commented spontaneously on their understanding of word meaning, their expectations from the dictionary entry, and the thinking behind their language production. I have tried to use the insights gained from these encounters to inform the conclusion of this thesis - but I feel that EFL dictionary users have plenty more to tell us, provided that we are willing to listen.

Bibliography

- Abbot V, Black J and E Smith**, 1985. The representation of scripts in memory. *Journal of Memory and Language*. 24 179-199
- Ager D** (ed), 1988. Written Skills in the Modern Languages Degree. London: AMLC in association with CILT
- Ahmed M**, 1989. Vocabulary learning strategies. in: P Meara (ed). Beyond Words. British Studies in Applied Linguistics 4. London: CILT 3-14
- Aitchison J**, 1987. Words in the Mind. Oxford: Blackwell
- Alexander R**, 1984. Fixed expressions in English: reference books and the teacher. *English Language Teaching Journal* 38(2) 127-134
- Alexander R**, 1985. Phraseological and pragmatic deficits in advanced learners of English: Problems of vocabulary learning? *Die Neueren Sprachen* 84(6) 613-621
- Ard J**, 1982. The use of bilingual dictionaries by EFL students while writing. *ITL Review of Applied Linguistics* 58 1-27
- Atkins B**, 1985. Monolingual and bilingual learners' dictionaries: a comparison. in: R Ilson (ed). Dictionaries, Lexicography and Language Learning. ELT Documents 120. Oxford: Pergamon Press 15-24
- Atkins B, Lewis H, Summers D and J Whitcut**, 1987. A research project into the use of learners' dictionaries. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985 Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 29-43
- Atkins B and F Knowles**, 1990. Interim Report on the Euralex/AILA Research Project into Dictionary Use. in: T Magay and J Zigany (eds). BudaLEX '88. Proceedings from the 3rd EURALEX Congress. Budapest: Akademiai Kiado. 381-392
- Atkinson R**, 1975. Mnemotechnics in Second-Language Learning. *American Psychologist* 30 821-828
- Baker M**, 1988. Sub-Technical vocabulary and the ESP teacher: an analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language* 4(2) 91-105
- Bailey R**, 1986. Dictionaries of the next century. in: R Ilson (ed). Lexicography: an emerging international profession. Manchester: University Press 123-137
- Barnhart C**, 1962. Problems in Editing Commercial Monolingual Dictionaries. in: F Householder and S Saporta (eds). Problems in Lexicography Publication twenty-one of the Indiana University Research Centre in Anthropology, Folklore and Linguistics. Bloomington: Indiana University 161-181
- Battenburg J**, 1989. A Study of English Monolingual Learners' Dictionaries and their Users PhD dissertation Purdue University
- Battenburg J**, 1991. English Monolingual Learners' Dictionaries: a user-oriented study Lexicographica Series Maior 39. Tübingen: Niemeyer Verlag

Bauer L, 1980. Review: Longman Dictionary of Contemporary English. *RELC Journal* 11(I) 104-108

Bauer L, 1981. Review: Chambers Universal Learners' Dictionary. *RELC Journal* 12(2) 1981 100-102

Baxter J, 1980. The dictionary and vocabulary behaviour: a single word or a handful? *TESOL Quarterly* 14(3) 325-336

Béjoint H, 1981. The foreign student's use of monolingual English dictionaries: a study of language needs and reference skills. *Applied Linguistics* 2(3) 207-222

Béjoint H and A Moulin, 1987. The place of the dictionary in an EFL programme. in: A Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985 Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 97-114

Béjoint H, 1988. Psycholinguistic evidence and the use of dictionaries by L2 learners. in: M Snell-Hornby (ed). ZurLEX'86 Proceedings papers read at the EURALEX international congress. Tübingen: Francke Verlag 139-148

Béjoint H, 1990. Monosemy and the dictionary. in: T Magay and J Zigany (eds). BudaLEX '88. Proceedings from the 3rd EURALEX congress. Budapest: Akademiai Kiado. 13-26

Béjoint H, 1989. The teaching of dictionary use: present state and future tasks. in: F Hausmann, O Reichmann, H Wiegand and L Zgusta (eds). Wörterbücher Dictionaries Dictionnaires. Berlin: Walter de Gruyter 208-215

Benson M, 1989. The collocational dictionary and the advanced learner. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art. Anthology Series 23. Singapore: SEAMEO Regional Language Centre 84-93

Bensoussan M, 1983. Dictionaries and tests of EFL reading comprehension. *English Language Teaching Journal* 37(4) 341-345

Bensoussan M and B Laufer, 1984. Lexical guessing in context. *Journal of Research in Reading* 7(1) 15-32

Bensoussan M, Sim D and R Weiss, 1984. The effect of dictionary usage on EFL test performance compared with student and teacher attitudes and expectations. *Reading in a Foreign Language* 2(2) 1984

Berrian R, Metzler D, Kroll N and G Clark-Meyers, 1979. Estimates of imagery, ease of definition and animateness for 328 adjectives. *Journal of Experimental Psychology: Human Learning and Memory* 5(4) 435-447

Black A, 1986. The effects on comprehension and memory of providing different types of defining information for new vocabulary: a report on two experiments conducted for Longman ELT Dictionaries and Reference Division. Cambridge: MRC Applied Psychology Unit (unpublished internal report)

Bogaards P, 1990. Où cherche-t-on dans le dictionnaire? *International Journal of Lexicography* 3(2) 79-102

Bogaards P, 1990. Deux langues, quatre dictionnaires. in: F Dolezal, A Kucera, A Rey, H Wiegand, W Wolski and L Zgusta (eds). Lexicographica International Annual for Lexicography 6. Tübingen: Niemeyer Verlag 162-173

Bogaards P, 1991a Word frequency in the search strategies of French dictionary users. in: F Dolezal, A Kucera, A Rey, H Wiegand, W Wolski and L Zgusta (eds). Lexicographica International Annual for Lexicography 7. Tübingen: Niemeyer Verlag 202-212

Bogaards P, 1991b Dictionnaires pédagogiques et apprentissage du vocabulaire *Cahiers de Lexicologie* 59, 1991-2. 93-107

Bogaards P, 1992. A la recherche de collocations dans le dictionnaire de langue étrangère. in: Lexicologia e Metalexigrafia Actas do XIX Congresso Internacional de Linguística e Filologia Romanicas, Unniversidade de Santiago de Compostela 1989. 176-185

Boo H and R Carter, 1989. Vocabulary, culture and the dictionary. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art Anthology Series 23. Singapore: SEAMEO Regional Language Centre. 172-183

Bramki D and R Williams, 1984. Lexical familiarization in Economics text, and its pedagogical implications in reading comprehension. *Reading in a Foreign Language* 2(1) 169-181

Brazil D, 1987. The Transcription of Pronunciation in Dictionaries and its implications for spelling. *Journal of the Simplified Spelling Society* 1 5-10

Broeders T, 1987. The treatment of phonological idioms. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds. 1-3 April 1985. Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 246-256

Brown A, Downing J and J Sceats, 1971. Words Children Want to Use: A Handbook to Chambers Young Set Dictionary. Edinburgh: Chambers

Brown J, 1979. Vocabulary: learning to be imprecise. *Modern English Teacher* 7(1) 25-27

Bujaz Z, 1975. Testing the performance of a bilingual dictionary based on topical current texts. *Studia Romanica et Anglica Zagrebiensia* 39 193-204

Burger H, 1991. Roget, Darwin and Bohr: Parallelizing Evolutionary Diversification with Conceptual Classification. *International Journal of Lexicography* 4(1) 23-67

Carney F, 1988. Review: The Collins COBUILD English Language Dictionary. *Practical English Teaching* March 1988 49

Carter R, 1983. You look nice and weedy these days: lexical associations, lexicography and the foreign language learner. *Journal of Applied Language Study* 1(2) 172-189

Carter R, 1987. Vocabulary and second/foreign language teaching. *Language Teaching Abstracts* 20(1) 3-16

Carter R, 1989. Review: Collins Cobuild English Language Dictionary. *English Language Teaching Journal* 43(2) April 150-152

Chapman L (ed), 1980. The Reader and the Text Oxford:Heinemann

258

Clarke D and I Nation, 1980. Guessing the Meaning of Words from Context: strategy and techniques. *System* 8(3) 211-220

Cook V (ed), 1986. Experimental approaches to second language learning. Oxford: Pergamon Press

Corder P 1967 The significance of learners' errors. *IRAL* 5(4). reprinted in: J Richards (ed), 1974. Error analysis: perspectives in second language acquisition. London: Longman 19-27

Corder P, 1971. Idiosyncratic dialects and error analysis. *IRAL* 9 (2) reprinted in: J Richards (ed), 1974. Error analysis: perspectives in second language acquisition. London: Longman 158-171

Corson D, 1983. The Corson measure of passive vocabulary. *Language and Speech* 26(1) 3-20

Cowie A P, 1978. The place of illustrative material and collocations in the design of a learner's dictionary. in: P Strevens (ed). In Honour of A S Hornby. Oxford: Oxford University Press 127-139

Cowie A P, 1981. Lexicography and its pedagogic implications: an introduction. *Applied Linguistics* 2 203-206

Cowie A P, 1987. Syntax, the dictionary and the learner's communicative needs. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985. Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 183-192

Cowie A P (ed), 1987. The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985 Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag

Cowie A P, 1989a. Learners' dictionaries - recent advances and developments. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art Anthology Series 23. Singapore: SEAMEO Regional Language Centre 42-51

Cowie A P, 1989b. The language of examples in learners' dictionaries. in: G James (ed). Lexicographers and Their Works Exeter Linguistic Studies volume 14. Exeter: University Press 55-65

Craik F and E Tulving, 1975. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104 268-294

Creamer T, 1987. Beyond the definition: some problems with examples in Chinese and English bilingual dictionaries. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985 Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 238-245

Crow J and J Quigley, 1985. A semantic field approach to passive vocabulary acquisition for reading comprehension. *TESOL Quarterly* 19(3) 497-513

Crystal D, 1986. The ideal dictionary, lexicographer and user. in: R Ilson (ed). Lexicography: an emerging international profession. Manchester: University Press 72-81

Dagut M, 1977. Incongruities in lexical "gridding" - an application of contrastive semantic analysis to language teaching. *IRAL* 15(3) 221-229

Davies A, Howatt A and C Criper (eds), 1984. Interlanguage. Edinburgh: Edinburgh University Press

Dechert H and U Sandrock, 1986. Thinking-aloud protocols: the decomposition of language processing. in: V Cook (ed). Experimental approaches to second language learning. Oxford: Pergamon 111-126

Deno S, 1968. Effects of words and pictures as stimuli in learning language equivalents. *Journal of Educational Psychology* 59(3) 202-206

Diab T, 1989. The role of dictionaries in English for Specific Purposes: a case study of student nurses at the University of Jordan. in: G James (ed) Lexicographers and their Works. Exeter Linguistic Studies volume 14. Exeter: University Press 74-82

Diab T, 1990. Pedagogical Lexicography: a case study of Arab nurses as dictionary users. Lexicographica Series Maior 31. Tübingen: Niemeyer Verlag.

Dilley M and A Paivio, 1968. Pictures and words as stimulus and response items in paired associate learning of young children. *Journal of Experimental Child Psychology* 6 231-240

Dodd S, 1989. Lexicomputing and the dictionary of the future. in: G James (ed). Lexicographers and their Works. Exeter Linguistic Studies volume 14. Exeter: University Press 83-93

Dolezal F, Kucera A, Rey A, Wiegand H, Wolski W and L Zgusta (eds), 1990. Lexicographica International Annual for Lexicography 6. Tübingen: Niemeyer Verlag

Dolezal F, Kucera A, Rey A, Wiegand H, Wolski W and L Zgusta (eds), 1991. Lexicographica International Annual for Lexicography 7. Tübingen: Niemeyer Verlag

Drysdale P, 1987. The role of examples in a learner's dictionary. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985 Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 213-233

Duchastel P, 1978. Illustrating Instructional Texts. *Educational Technology* November 1978 36-39

Duchastel P and R Waller, 1979. Pictorial illustration in instructional texts. *Educational Technology* November 1979 20-25

Dunmore D, 1989. Using contextual clues to infer word meaning: an evaluation of current exercise types. *Reading in a Foreign Language* 6(1) 337-347

Ellegård A, 1978. On dictionaries for language learners. *Moderna Språk* 72 (3) 225-242

Evans M, Markowitz J, Ahlswode T and K Rossi, 1986. Digging in the Dictionary: building a relational lexicon to support natural language processing applications. Paper presented at the Computers in Language Research and Language Learning conference in Urbana October 25-26 1986

Faber H and A Maley (eds), 1980. Leserverstehen im Fremdensprachenunterricht. Munich: Goethe Institut

Fleming M, 1977. The Picture in Your Mind. *Audio Visual Communication Review* 25(1) 43-62

Fox G, 1987. The Case for Examples. in: J Sinclair (ed). Looking Up: an account of the COBUILD project in lexical computing. London/Glasgow: Collins 137-149

Fox G, 1989. A vocabulary for writing dictionaries. in: M L Tickoo (ed), Learners' Dictionaries: State of the Art Anthology Series 23. Singapore: SEAMEO Regional Language Centre 153-171

Galisson R, 1983. Image et usage du dictionnaire chez les etudiants (en langue) de niveau avance. *Etudes de Linguistique Appliquée* 49 5-88

Goodman K, 1969. Analysis of oral reading miscues: applied psycholinguistics. *Reading Research Quarterly* 5(1) 9-30

Gombrich E, 1972. The Visual Image. *Scientific American* 227 82-96

Gove P, 1969. About the dictionary. in: J Sanderson and W Gordon (eds). Exposition and the English Language. New York: Appleton-Century Crofts

Greenbaum S, Meyer C and J Taylor, 1984. The image of the dictionary for American College students. *Dictionaries: Journal of the dictionary society of North America* Part 6 31-52

Greig C, 1978. Review: The Longman Dictionary of Contemporary English. *Modern English Teacher* 6(5) 28

Gropper G L, 1963. Why is a picture worth a thousand words? *Audio Visual Communication Review* 11(4) 75-95

Grunwell P (ed), 1988. Applied Linguistics in Society. London: CILT

Hammerly H, 1974. Primary and secondary associations with visual aids as semantic conveyors. *IRAL* 12 119-125

Hanks P, 1987 Definitions and Explanations. in: J Sinclair (ed). Looking Up: an account of the COBUILD project in lexical computing. London/Glasgow: Collins 116-136

Hartmann R (ed), 1979. Dictionaries and Their Users. Exeter Linguistic Studies volume 4. Exeter: University Press

Hartmann R, 1981. Review Article - Dictionaries, Learners, Users: Some issues in lexicography. *Applied Linguistics* II(3) 297-303

Hartmann R (ed), 1984. LEX'eter '83 Proceedings: Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983. Lexicographica Series Maior I. Tübingen: Niemeyer Verlag

Hartmann R, 1985. Contrastive Text Analysis and the search for equivalence in the bilingual dictionary. Symposium on Lexicography II Tübingen: Niemeyer Verlag 121-132

Hartmann R, 1987. Four perspectives in dictionary use: a critical view of research methods. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985 Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 11-28

- 301
- Hartmann R**, 1989. What we (don't) know about the English language learner as a dictionary user: a critical select bibliography. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art Anthology Series 23. Singapore: SEAMEO Regional Language Centre 213-221
- Hartmann R**, 1989. Sociology of the dictionary user: hypotheses and empirical studies. in: F Hausmann, O Reichmann, H Wiegand and L Zgusta (eds) Wörterbücher Dictionaries Dictionnaires. Berlin: Walter de Gruyter 102-111
- Hartmann R**, 1989. The dictionary as an aid to foreign language teaching. in: F Hausmann, O Reichmann, H Wiegand and L Zgusta (eds). Wörterbücher Dictionaries Dictionnaires. Berlin: Walter de Gruyter 181-189
- Hartmann R**, 1992. State of the art article: Lexicography, with particular reference to English learners' dictionaries. *Language Teaching* 25(3) 151-159
- Hatherall G**, 1984. Studying dictionary use: some findings and proposals. in: R Hartmann (ed). LEX'eter '83 Proceedings: Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983 Lexicographica Series Maior I. Tübingen: Niemeyer Verlag 183-189
- Hausmann F, Reichmann O, Wiegand H and L Zgusta** (eds), 1989. Wörterbücher Dictionaries Dictionnaires. Berlin: Walter de Gruyter
- Herbst T and G Stein**, 1987. Dictionary using skills : a plea for a new orientation in language teaching. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985. Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 115-28
- Hill C**, 1985. Alternatives to Dictionaries. in: R Ilson (ed). Dictionaries, Lexicography and Language Learning ELT Documents 120. Oxford: Pergamon 115-122
- Hindmarsh R**, 1980. Cambridge English Lexicon. Cambridge: Cambridge University Press
- Hornby A** (ed), 1948. A Learner's Dictionary of Current English. Oxford: Oxford University Press
- Horsella M and G Sindermann**, 1983. The transfer of cognate lexis: in search of facilitating factors in EST. *English for Specific Purposes* (Oregon State University) 79 8-16
- Householder F and S Saporta** (eds), 1962. Problems in Lexicography. Publication twenty-one of the Indiana University Research Centre in Anthropology, Folklore, and Linguistics. Bloomington: Indiana University Press
- Huang G**, 1985. The productive use of EFL dictionaries. *RELC Journal* 16(2) 54-71
- Hughes A and C Lascaratou**, 1982. Competing criteria for error gravity. *English Language Teaching Journal* 36(3) 175-182
- Hulstijn J**, 1993. When do foreign language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *The Modern Language Journal* 77

502

Hupka W, 1989. Wort und Bild Lexicographica Series Maior 22. Tübingen: Niemeyer Verlag

Ilson R, 1983. Etymological information: can it help our students? *English Language Teaching Journal* 37(1) 76-82

Ilson R (ed), 1985. Dictionaries, Lexicography and Language Learning. ELT Documents 120. Oxford: Pergamon Press

Ilson R, 1986. British and American Lexicography. in: R Ilson (ed). Lexicography: an emerging international profession. Manchester: University Press 51-71

Ilson R (ed), 1986. Lexicography: an emerging international profession. Manchester: University Press

Ilson R, 1987. Illustrations in Dictionaries. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985. Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 193-212

Iqbal Z, 1987. Aspects of the Learner's Dictionary with special reference to advanced Pakistani learners of English. PhD dissertation. University of Aston. Birmingham.

Jackson H, 1988. Words and Their Meaning. London: Longman

Jacobs G, 1989. Dictionaries can help writing - if students know how to use them. ERIC ED 316 025

Jain M, 1981. On meaning in the foreign learner's dictionary. *Applied Linguistics* 2(3) 274-286

James C, 1977. Judgements of error gravity *English Language Teaching Journal* 31(1) 116-124

James G (ed), 1989. Lexicographers and their Works. Exeter Linguistic Studies volume 14. Exeter: University Press

Jenkins J, Neale D and S Deno, 1967. Differential memory for picture and word stimuli. *Journal of Educational Psychology* 1967 303-307

Jia M and J Lian, 1983. Shuangyu cidiande chatu (illustrations in bilingual dictionaries). *Cishu Yanjiu* (Lexicographical Studies) 4 19-25

Johns T, 1980. The text and its message: an approach to the teaching of reading comprehension for students of development administration. in: H Faber and A Maley (eds) Leserverstehen im Fremdensprachenunterricht. Munich: Goethe Institut 147-165

Jones G, 1985. Deep Dyslexia, Imageability, and Ease of Prediction. *Brain and Language* 24 1-19

Kellog G and M Howe, 1971. Using words and pictures in foreign language learning. *The Alberta Journal of Educational Research* 18 89-94

Kerim-Zade I and V Pavlov, 1989. The semantico-functional variability of words and the teaching of vocabulary to advanced EFL students. *Applied Linguistics* 10(4) 382-391

303

Kharma N, 1985. Wanted: a brand-new type of learners' dictionary. *Multilingua* 4(2) 85-90

Kipfer B, 1987. Dictionaries and the intermediate student: Communicative needs and the development of user reference skills. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985. Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 44-54

Kirkpatrick B, 1985. A lexicographical dilemma: monolingual dictionaries for the native speaker and for the learner. in: R Ilson (ed). Dictionaries, Lexicography and Language Learning. ELT Documents 120. Oxford: Pergamon 7-14.

Knowles F, 1988. The role of dictionaries in developing written skills. in: D Ager (ed) Written Skills in the Modern Languages Degree. London: AMLC in association with CILT 95-109

Knowlton J, 1966. On the definition of "picture". *Audio Visual Communication Review* 14(2) 157-183

Kopstein F and Rashal S, 1954. Learning foreign vocabulary from pictures versus words. *American Psychologist* 9 407-408

Kroll J and Potter M, 1984. Recognizing words, pictures, concepts: a comparison of lexical, object and reality decisions. *Journal of Verbal Learning and Verbal Behaviour* 23 39-66

Lamy M-N, 1985. Innovative Practices in French Monolingual Learners' Dictionaries as Compared to their English Counterparts. in: R Ilson (ed). Dictionaries, Lexicography and Language Learning ELT Documents 120. Pergamon Press 25-34

Landau S, 1984. Dictionaries: the art and craft of lexicography. Cambridge: Cambridge University Press

Laufer B and M Bensoussan, 1982. Meaning is in the eye of the beholder. *English Teaching Forum* 20(2) 10-13

Laufer B, 1985. A case for vocabulary in EAP reading comprehension materials. in: A Cornu, J Vanparijs, M Delahaye and L Baten (eds) Beads or Bracelet or, how do we approach LSP? Selected papers from the 5th European Symposium on LSP. Oxford: Oxford University Press 284-291

Laufer B, 1989. What percentage of text-lexis is essential for comprehension? in: C Laurén and M Nordman (eds) Special Language: from humans thinking to thinking machines. Avon: Multilingual Matters 316-323

Laufer B, 1990. "Sequence" and "order" in the development of L2 lexis: some evidence from lexical confusions. *Applied Linguistics* 11(3) 281-296

Laurén C and M Nordman (eds), 1989. Special Language: from humans thinking to thinking machines. Avon: Multilingual Matters

Lemmens M and H Wekker, 1991. On the relationship between lexis and grammar in English learners' dictionaries. *International Journal of Lexicography* 4(1) 1-14

Liu Z, 1983. Cishu yao you hao chatu (The need for good illustrations in lexicographical works). *Chishu Yanjiu (Lexicographical Studies)* 3 142-149

Loughridge B, 1990 Which Dictionary? London: The Library Association.

Lupescu S and Day R, 1993. Reading, Dictionaries and Vocabulary Learning. *Language Learning* 43(2) 263-287

MacFarquhar P and J Richards, 1983. On dictionaries and definitions. *RELC Journal* 14(1) 111-124

MacKay R, 1986. Review: The Words You Need. *English Language Teaching Journal* 40(1) 75-76

Magay T, 1988. On some problems of the bilingual learner's dictionary. in: M Snell-Hornby (ed) ZurILEX'86 Proceedings: papers read at the EURALEX international congress. Tübingen: Francke Verlag 171-177

Magay T and J Zigany (eds), 1990. BudaLEX '88: Proceedings from the 3rd EURALEX congress. Budapest: Akademiai Kiado.

Maingay S and M Rundell, 1987. Anticipating learner's Errors - Implications for dictionary writers. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985. Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 128-135

Marello C, 1987. Examples in contemporary Italian bilingual dictionaries. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985 Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 224-237

Martin R, 1989. L'exemple lexicographique dans le dictionnaire monolingue. in: F Hausmann, O Reichmann, H Wiegand and L Zgusta (eds). Wörterbücher Dictionaries Dictionnaires. Berlin: Walter de Gruyter 599-607

McCarthy M, 1984. A new look at vocabulary in EFL. *Applied Linguistics* 5(1) 13-22

McArthur T, 1986. Worlds of Reference: Lexicography, Learning and Language from the Clay Tablet to the Computer. Cambridge: Cambridge University Press

McArthur T, 1989. The background and nature of ELT learners' dictionaries. in: M L Tickoo (ed) Learners' Dictionaries: State of the Art Anthology Series 23. Singapore: SEAMEO Regional Language Centre 52-64

McDavid R and A Duckett, (eds). 1975 Lexicography in English. *Annals of the New York Academy of Sciences* vol 211

Meara P, 1980. Vocabulary acquisition: a neglected aspect of language learning. *Language Teaching and Linguistics Abstracts* 13 221-246

Meara P, 1984. The study of lexis in interlanguage. in: A Davies, A Howatt and C Criper (eds). Interlanguage. Edinburgh: University Press 225-235

Meara P (ed), 1989. Beyond Words. London: CILT

Meara P and F English, 1988. Lexical errors and learners dictionaries. ERIC ED 654 321

Meara P and G Jones, 1988. Vocabulary size as a placement indicator. in: P Grunwell (ed). Applied Linguistics in Society. London: CILT 80-87

Miller G and P Gildea, 1985. How to misread a dictionary *AILA Bulletin* 1985 13-26

Miller G and P Gildea, 1987. How Children Learn Words. *Scientific American* September 86-91

Mitchell E, 1983a. Formative Assessment of Reading. Working Paper 20 Search-do reading: (2) Using a dictionary - a preliminary analysis. Aberdeen: College of Education

Mitchell E, 1983b. Formative Assessment of Reading. Working paper 21 Search-do reading: (3) Difficulties in using a dictionary. Aberdeen: College of Education

Moles A, 1972. Teoria informazionale dello schermo *Versus* 2(3) 29-40

Mondria J-A and M Wit-de Boer, 1991. The Effects of Contextual Richness on the Guessability and the Retention of Words in a Foreign Language. *Applied Linguistics* 12(2) 249-267

Moon R, 1987. Monosemous words and the dictionary. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985. Lexicographica, Series Maior 17. Tübingen: Niemeyer Verlag 173-182

Moss G, 1992. Cognate recognition: its importance in the teaching of ESP reading courses to Spanish speakers. *English for Specific Purposes* 11(2) 141-158

Naganuma K, 1978. The history of the Advanced Learner's Dictionary: A S Hornby, ISED and Kaitakusha, Tokyo. in: P Strevens (ed). In Honour of A S Hornby. Oxford: Oxford University Press 11-13

Nation I, 1989. Dictionaries and language learning. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art. Anthology Series 23. Singapore: SEAMEO Regional Language Centre 65 -71

Nation I, 1990. Teaching and Learning Vocabulary. New York: Newbury House.

Nesi H, 1987. Do Dictionaries Help Students Write? in: T Bloor and J Norrish (eds). Written Language. British Studies in Applied Linguistics 2. London: CILT 85-97

Nesi H, 1989. How many words is a picture worth? A review of illustrations in dictionaries. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art. Anthology Series 23. Singapore: SEAMEO Regional Language Centre 124-134

Nesi H, 1991. English Dictionaries in the Classroom: a critical survey. *CLIE Working Papers* Number 12 ISSN 0964-8275

Nesi H, 1992. Go away and look it up! *Revue de Phonétique Appliquée* 103-104. Mons: Université de Mons-Hainaut 195-210

Nesi H and P Meara, 1991. How using dictionaries affects performance in multiple choice EFL tests. *Reading in a Foreign Language* 8 (1) 631-643

Nesi H and P Meara, 1994. Patterns of misinterpretation in the productive use of EFL dictionary definitions. *System* 22 (1) 1-15

Neubach A and A Cohen, 1988. Processing Strategies and Problems Encountered in the Use of Dictionaries. *Dictionaries: Journal of the Dictionary Society of North America* 10 1-19

- O'Brien E and J Myers, 1985.** When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory and Cognition* 11 12-21
- Omar A, 1987.** Malay in its sociocultural context. Kuala Lumpur: Dewan Bahasa dan Pustaka, Ministry of Education.
- Ott C, Blake R and D Butler, 1975.** The effect of interactive image elaboration on the acquisition of foreign language vocabulary. *Language Learning* 23(2) 197-206
- Ostyn P, Vandecasteele M, Deville G and P Kelly, 1985.** Towards an optimal programme of FL vocabulary acquisition. in: A Cornu, J Vanparijs, M Delahaye and L Baten (eds). Beads or Bracelet or, how do we approach LSP? Selected papers from the 5th European Symposium on LSP. Oxford: Oxford University Press 1985 292-304
- Paivio A and A Yarmey, 1966.** Pictures versus words as stimuli and responses in paired associate learning. *Psychonomic Science* 5 (6) 235-236
- Paivio A, Yuille J and S Madigan, 1968.** Concreteness, Imagery and Meaningfulness Values for 925 Nouns. *Journal of Experimental Psychology Monograph Supplement* 76 (1) Part 2 1-25
- Paivio A, 1971.** Imagery and verbal processes New York: Holt, Reinhart and Wilson
- Pawley A and Syder F H, 1983.** Two puzzles for linguistic theory: nativelike selection and nativelike fluency. in: J Richards and R Schmidt (eds). Language and Communication. London: Longman 191-226
- Perkins K and S Brutton, 1983.** The effects of word frequency and contextual richness on ESL students' word identification abilities. *Journal of Research in Reading* 1983 6(2) 119-128
- Perkins K and S Brutton, 1988.** Using a facet design to assess the effects of the form, the source, and the frequency of information necessary to answer ESL reading comprehension questions. *Language Learning* 38(2) 171-185
- Piotrowski T, 1989.** Monolingual and bilingual dictionaries: fundamental differences. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art. Anthology Series 23. Singapore: SEAMEO Regional Language Centre 72-83
- Pugh A and J Ulijn, 1984.** Reading for Professional Purposes. Oxford: Heinemann
- Quirk R, 1975.** The social impact of dictionaries in the UK. in: R McDavid and A Ducket (eds). Lexicography in English. New York: Annals of the New York Academy of Sciences 211 76-88
- Reif J, 1987.** The development of a dictionary concept: an English learner's dictionary and an exotic alphabet. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985. Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 146-158
- Rey-Debove J, 1970.** Le Domaine du Dictionnaire. *Langages* 19 3-34
- Richards J (ed), 1974.** Error analysis: perspectives in second language acquisition. London: Longman

Richards J, 1976. The role of vocabulary teaching. *TESOL Quarterly* 10(1) 77-89

Richards J and R Schmidt (eds), 1983. Language and Communication. London: Longman

Roberts R, 1984. Contextual Dictionaries for Languages for Special Purposes. in: A Pugh and J Ulijn (eds). Reading for Professional Purposes. Oxford: Heinemann 154-164

Robinson C, 1978. Review: Oxford Student's Dictionary of Current English. *Modern English Teacher* 6(5) 27-28

Rossner R, 1985. The Learner as Lexicographer: using dictionaries in second language learning. in: R Ilson (ed). Dictionaries, Lexicography and Language Learning ELT Documents 120. Oxford: Pergamon 95-102

Rudzka B, Channell J, Putseys Y and P Ostyn, 1981. The Words You Need. London/ Basingstoke: Macmillan

Rudzka B, Channell J, Putseys Y and P Ostyn, 1985. More Words You Need London/ Basingstoke: Macmillan

Rundell M, 1988. Changing the rules: why the monolingual learner's dictionary should move away from the native-speaker tradition. in: M Snell-Hornby (ed). ZurLEX'86 Proceedings: papers read at the EURALEX International congress. Tübingen: Francke Verlag 127-138

Ruse C, 1978. An EFL Dictionary. *Modern English Teacher* 6(5) 26

Salager F, 1983. The lexis of fundamental medical English: classificatory framework and rhetorical function (a statistical approach). *Reading in a Foreign Language* 1 (1) 54-64

Salager F, 1984. Compound nominal phrases in scientific-technical literature: proportion and rationale. in: A Pugh and J Ulijn (eds). Reading for Professional Purposes. Oxford: Heinemann 137-145

Sanderson J and W Gordon (eds), 1969. Exposition and the English Language. New York: Appleton-Century Crofts

Scholfield P, 1979. A non-standard dictionary definition schema. in: R R K Hartmann (ed). Dictionaries and Their Users. Exeter Linguistic Studies vol 4. Exeter: University Press 54-62

Scholfield P, 1980. Explaining meaning by paraphrase: problems and principles. *RELC Guidelines* 3 24-37

Scholfield P, 1981. Writing, vocabulary errors, and the dictionary. *RELC Guidelines* 6 31-40

Scholfield P, 1982a. The role of bilingual dictionaries in ESL/EFL: a positive view. *RELC Guidelines* 4(1) 84-98

Scholfield P, 1982b. Using the English dictionary for comprehension. *TESOL Quarterly* 16 (2) 185-194

Seliger H and E Shohamy, 1989. Second Language Research Methods Oxford: Oxford University Press

- Selinker L**, 1972. Interlanguage *IRAL* 10(3) reprinted in: J Richards (ed). Error analysis: perspectives in second language acquisition. 19-27
- Senior N**, 1984. French Picture Dictionaries and Word Books for Small Children. *The French Review* 57(6) 757-765
- Sheldon L** (ed), 1987. ELT Textbooks and Materials: Problems in Evaluation and Development ELT Documents 126. London: Modern English Publications
- Sheorey R**, 1986. Error perceptions of native-speaking and non-native-speaking teachers of ESL *English Language Teaching Journal* 40(4) 306-312
- Sinclair J** (ed), 1987. Looking Up: an account of the COBUILD project in lexical computing. London/Glasgow: Collins
- Sinclair J**, 1989. Uncommonly common words. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art Anthology Series 23. Singapore: SEAMEO Regional Language Centre. 135-152
- Snell-Hornby M**, 1987. Towards a learners bilingual dictionary. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985. Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 159-170
- Soekemi**, 1989. How to use a dictionary? in: M L Tickoo (ed). Learners' Dictionaries: State of the Art Anthology Series 23. Singapore: SEAMEO Regional Language Centre 204-211
- Stark M**, 1990. Dictionary Workbooks: A critical evaluation of dictionary workbooks for the foreign language learner. Exeter Linguistic Studies volume 16. Exeter: University Press
- Stauffer R, Abrams J and J Pikulski**, 1978. Diagnosis, Correction and Prevention of Reading Disabilities. New York: Harper and Row
- Stein G**, 1989. Recent developments in EFL dictionaries. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art Anthology Series 23. Singapore: SEAMEO Regional Language Centre 10-41
- Stein G**, 1991. Illustrations in Dictionaries. *International Journal of Lexicography* 4(2) 99-127
- Stock P**, 1988. The structure and function of definitions. in: M Snell-Hornby (ed). ZurLEX'86 Proceedings: papers read at the EURALEX International congress. Tübingen: Francke Verlag 81-90
- Strevens P** (ed), 1978. In Honour of AS Hornby. Oxford: Oxford University Press
- Summers D**, 1988. English language teaching dictionaries: past present and future. *English Today* 14 10-16
- Swales J**, 1983. Vocabulary work in LSP - a case of neglect? *Bulletin CILA* 37. 21-33
- Tadros A**, 1987. Review of COBUILD. *ESPMENA Bulletin* No 24. 20-26
- Thompson G**, 1987. Using Bilingual Dictionaries. *English Language Teaching Journal* volume 41 (4) 282-286

309

Thorndike E, 1928. The Psychology of the School Dictionary. *Bulletin of the School of Education*, Indiana University. Bloomington IV/6 24-31. Reprinted in *International Journal of Lexicography* 4(1) 15-22

Tickoo M (ed), 1989. Learners' Dictionaries: State of the Art. Anthology Series 23 Singapore: SEAMEO Regional Language Centre.

Tickoo M, 1989. Which dictionaries and why? Exploring some options.in: M L Tickoo (ed). Learners' Dictionaries: State of the Art Anthology Series 23. Singapore: SEAMEO Regional Language Centre 184-203

Todd-Trimble M, 1983. Review: Longman Dictionary of Scientific Usage. *The ESP Journal* 2(2) 176-179

Tomaszcyk J, 1979. Dictionaries: users and uses. *Glottodidactica* 12 103-119

Tomaszcyk J, 1987. FL Learners' communication failure: implications for pedagogical lexicography. in: A P Cowie (ed). The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985. Lexicographica Series Maior 17. Tübingen: Niemeyer Verlag 136-145

Tomaszcyk J, 1988. The bilingual dictionary under review. in: M Snell-Hornby (ed). Zurilex '86 Proceedings, Papers read at the EURALEX International Congress. Tübingen: Francke Verlag 289-298

Tono Y, 1988. Assessment of the EFL learner's dictionary using skills. *JACET Bulletin* 19 103-126

Tono Y, 1989. Can a dictionary help one read better? On the relationship between EFL learners' dictionary reference skills and reading comprehension. in: G James (ed). Lexicographers and their Works Exeter Linguistic Studies 14. Exeter: University Press 192-200

Underhill A, 1985. Working with the monolingual learners' dictionary. in: R Ilson (ed). Dictionaries, Lexicography and Language Learning ELT Documents 120. Oxford: Pergamon Press 103-114

Webber N E, 1978. Pictures and words as stimuli in learning foreign language responses. *Journal of Psychology* 98 57-63

West M, 1935. Definition Vocabulary. Bulletin 4. Toronto: University of Toronto Department of Educational Research

West M, 1953. A General Service List of English Words. London: Longman

West R, 1987. A Consumers Guide to ELT Dictionaries. in: L Sheldon (ed). ELT Textbooks and Materials: Problems in Evaluation and Development. ELT Documents 126. London: Modern English Publications 55-75

Whitcut J, 1986. The training of dictionary users. in: R Ilson (ed). Lexicography: an emerging international profession. Manchester: University Press 111-122

Whitcut J, 1987. Feature review: A Jungle of Words. *Practical English Teaching* September 46-47

Widdowson H G, 1978. Teaching Language as Communication. Oxford: Oxford University Press

Williams R, 1980. Lexical familiarization in content area textbooks in: L Chapman (ed). The Reader and the Text. Oxford: Heinemann 49-59

Williams R, 1984. A cognitive approach to English nominal compounds. in: A Pugh and J Ulijn (eds). Reading for Professional Purposes. Heinemann 146-153

Worth S, 1975. Pictures can't say ain't. *Versus* 12 85-108

Yang J and W Jiamin, 1983. Yuwen cidian de chatu (Illustrations in language dictionaries). *Cishu Yianju* (Lexicographical Studies) 4 13-18

Yorio C, 1980. Conventionalized language forms and the development of communicative competence. *TESOL Quarterly* 14(4) 433-443

Yorkey R, 1979. Review: Longman Dictionary of current (sic) English *TESOL Quarterly* 13(1) 393-403

Zgusta L, 1988. Lexicography Today. An annotated bibliography of the theory of lexicography Lexicographica Series Maior 18. Tübingen: Niemeyer Verlag.

Zgusta L, 1989a. Ornamental pictures in dictionaries. in: G James (ed). Lexicographers and their Works. Volume 14 Exeter Linguistic Studies. Exeter: University Press 215-223

Zgusta L, 1989b. Idle thoughts of an idle fellow, or, vaticinations on the learners' dictionary. in: M L Tickoo (ed). Learners' Dictionaries: State of the Art. Anthology Series 23. Singapore: SEAMEO Regional Language Centre 1-9

Appendices

Appendix 2.1.

UNIVERSITY OF WARWICK

READING COMPREHENSION TEST (TR4): SEPTEMBER 1988

Time : 1 hour (Part 1: 10 minutes/Part 2: 50 minutes)

NAME
COURSE (from October 1988)
DEPARTMENT (from October 1988).....

PART 1 (10 MINUTES)

Read the following two passages. You have ten minutes reading time. Then you will answer some questions about the passages. As you read, underline any words you would like to look up in a dictionary.

A Car for the Third World

Over the past few years, a revolution in vehicles for the Third World has been masterminded from a small workshop in Lancaster. Around 80 per cent of the world's roads are rutted, potholed dirt tracks. In these conditions, existing transport is inappropriate, and largely too expensive.

With that in mind, a British designer, Tony Howarth, set about rethinking the motor car. The result is the Africar, which will go into full production in the next few months. Initially, the car will be produced only in Britain, but the idea is that the car can be manufactured profitably almost anywhere in the developing world. Local content could be as much as 90 per cent.

Apart from its galvanised steel roll-cage and door frames, the Africar is made almost entirely from plywood-reinforced plastic (PRP). The materials used are both strong and light - at least 30 and 15 per cent lighter than steel and glass fibre respectively. Laminated plywood components are slotted together "like a giant

20 jigsaw puzzle" says Howarth, then sealed with a non-toxic epoxy resin. Howarth says that the body and chassis are corrosion-proof and dent-proof, while the panels can be easily repaired or replaced if the car is damaged. Howarth has designed the car in a variety of
25 configurations. Customers can choose from 4-wheel saloons to 6-wheel ambulances and 8-wheel tippers and tankers.

All four wheels on the saloon are independently sprung using Dunlop Hydragas units. These allow 30 centimetres
30 of vertical wheel movement, which is combined with a good 30 centimetres of ground clearance. The car uses an air-cooled, super-charged two-stroke engine. Three different sizes will be fitted to different models. All are capable of running on petrol, diesel, alcohol or gas.
35 The mechanics are basic, with gear-driven components, as opposed to valves and belts, which have a knack of going wrong.

Three prototype cars took part in a 10,000 kilometre test run from the Arctic to the Equator through six climatic
40 zones. According to Howarth, "the engines ran as efficiently at 50C as they did at minus 40C". He also points out that even on the soft shifting sands and dunes of the Algerian Sahara, the vehicles managed around 9.5 kilometres to the litre.

45 One of the major benefits of the Africar is that it can be manufactured in the developing world, and not just from a kit. The company claims that a manufacturing plant could be fully operational within 15 months of an initial feasibility study. But although interest in the
50 concept is widespread, international aid and relief agencies seem unwilling to take the plunge as customers. Some enterprising entrepreneurs have signed dealerships in Ghana, Sudan, Malawi, Mexico and Qatar, but one aid agency said that it was loath to buy and export vehicles

55 to any country where they are not already in production
and where parts may not be available. In spite of that,
the company could be on to a winner. In the West,
vehicle production is stagnant. Yet there is a vast
unexploited market in the developing countries
60 - over two-thirds of the world population owns just 2 per
cent of the world's vehicles. Howarth is confident: "We
are offering a vehicle that even in a basic prototype
form was giving a better ride, better adhesion, better
load-carrying capacity and better fuel consumption on and
65 off the road, than any existing vehicle", he said.

#####

A Fast Two-Seater Car

Edmund Jephcott, a British inventor, has come up with an
economical and stable car for two, called the Micro.
Jephcott has built a prototype car that will seat two
people, one behind the other. Now he wants a British
70 company to exploit this idea.

Government statistics show that most cars carry only one
or two people on 86 per cent of journeys (and 95 per cent
in towns). Underoccupied large cars waste both energy
and road space.

75 According to Jephcott, the main reason that no-one has
built a truly efficient two-seater is that the car is
unstable when cornering, unless the passengers are
sitting side-by-side. But this offers few advantages in
terms of drag and weight, unless the car is built
80 inconveniently low.

Jephcott's design overcomes the problem by using a narrow
body with an automatic tilting mechanism, so that two
people can sit safely in tandem, even on sharp corners.
The tilt mechanism, which was developed in conjunction

85 with "a major automotive components producer", means that
a car is highly stable while the narrow body cuts drag to
a minimum. Jephcott says that the car's low drag and
weight make it suitable for electric propulsion. The
vehicle has been assessed by the Motor Industry Research
90 Association and Jephcott now hopes that British firms
will be interested in taking the project through to
commercial manufacturing.

#####

UNIVERSITY OF WARWICK
READING COMPREHENSION TEST : SEPTEMBER 1988

NAME
.....

PART 2 : (50 MINUTES)

For each of the following questions, choose the answer nearest in meaning to the ideas in the passages you have just read. Mark the answer (A, B, C or D) on the answer sheet. Choose ONE answer only for each question.

If you have a dictionary, draw a circle around the words in the passages that you look up.

1. The main topic of both of the passages is:
 - A. the expense of running a car
 - B. original designs for new cars
 - C. the development of new transport systems
 - D. the need for better engine and body design

2. The main point of the first paragraph (lines 1 to 6) is to suggest that:
 - A. British cars are very expensive
 - B. some highly intelligent businessmen work in Lancaster
 - C. roads in the third world are very bad
 - D. most existing cars are not suitable for the Third World

3. In the second paragraph (line 13) "Local content" refers to the:
 - A. material to be used in manufacture
 - B. number of workers required
 - C. money needed for development
 - D. profit manufacturers will make

4. Lines 18-20 state that "laminated plywood components are slotted together like a giant jigsaw puzzle". This is so that:
 - A. it can be manufactured anywhere
 - B. any damaged part can be replaced
 - C. people can choose different designs
 - D. the car is corrosion and dent proof

5. In paragraph 4 (lines 28-37) which of the following features of the car is NOT mentioned:
- A. the body
 - B. the suspension
 - C. the fuel
 - D. the engine
6. The word 'configurations' in line 25 is most likely to mean:
- A. strengths and weaknesses
 - B. components
 - C. shapes and sizes
 - D. colours and forms
7. The word 'just' in line 46 could be replaced by:
- A. alone
 - B. right
 - C. only
 - D. already
8. The main purpose of paragraph 5 (lines 38-44) is to show that:
- A. the cars are suitable for all kinds of climate
 - B. are easy to drive even in the desert
 - C. are very economical with fuel
 - D. can cover large distances quickly
9. Why are some aid agencies reluctant to buy the car?
- A. Because they are unwilling to take the plunge
 - B. Because they are loath to export vehicles
 - C. Because there may be a shortage of spare parts
 - D. Because entrepreneurs have signed agreements
10. The last paragraph of A Car for the Third World:
- A. evaluates the design of the car
 - B. analyses car ownership in Africa
 - C. compares car production in Britain and the Third World
 - D. estimates the future success of the Africar
11. Jephcott's main objective in the design of the Micro, the car described in Fast Two-seater Car is to:
- A. carry only one or two people
 - B. save energy and road space
 - C. allow passengers to sit side-by-side
 - D. develop a tilt mechanism

12. The main problem with earlier designs for a two-seater car was:
- A. the danger of fire from electric propulsion
 - B. they were underoccupied
 - C. the expense of manufacture
 - D. the risk of the car turning over when cornering
13. The expression 'in conjunction with' in lines 84-85 means:
- A. in co-operation with
 - B. in competition with
 - C. in co-ordination with
 - D. in cahoots with
14. The word 'assessed' (line 89) is most likely to mean:
- A. marked
 - B. criticised
 - C. measured
 - D. judged
15. The words 'The car' (line 76) refer to:
- A. the prototype for the Micro
 - B. a truly efficient two-seater
 - C. any ordinary two-seater car
 - D. all cars in general

PLEASE GIVE IN ALL PAPERS

APPENDIX 2.2

Data from Study Two reported in Chapter Two, showing individual comprehension test scores and completion time in minutes for subjects who used dictionaries (DICU = group 1), subjects who chose not to use dictionaries (DICNO = group 2), and subjects who were not permitted to use dictionaries (NOTAV = group 3).

Grp	ID	Score	Time	Grp	ID	Score	Time
1	01	15	25	3	01	09	16
1	02	10	25	3	02	11	18
1	03	11	28	3	03	12	18
1	04	12	30	3	04	10	18
1	05	10	30	3	05	10	20
1	06	13	30	3	06	15	21
1	07	14	31	3	07	11	22
1	08	08	33	3	08	14	23
1	09	09	33	3	09	09	23
1	10	11	35	3	10	13	24
1	11	13	35	3	11	13	24
1	12	12	35	3	12	11	24
1	13	12	35	3	13	14	25
1	14	10	35	3	14	10	26
1	15	12	35	3	15	07	26
1	16	10	35	3	16	09	26
1	17	06	35	3	17	09	27
1	18	11	37	3	18	14	28
1	19	12	37	3	19	14	28
1	20	11	40	3	20	15	28
1	21	17	40	3	21	07	28
1	22	10	40	3	22	07	29
1	23	12	43	3	23	12	29
1	24	07	45	3	24	09	30
1	25	10	47	3	25	09	30
1	26	10	49	3	26	08	30
1	27	09	50	3	27	08	30
1	28	09	52	3	28	10	31
1	29	08	57	3	29	11	31
				3	30	08	32
				3	31	13	36
2	01	14	25				
2	02	13	35				
2	03	12	40				
2	04	08	43				
2	05	08	49				

APPENDIX 3.1

Data from the pilot study reported in Chapter Three, showing group (1=LDOCE, 2=OALD and 3=COBUILD), EVST score, the number of words looked up, average look-up time in seconds and average sentence score.

Group	ID	EVST	Number	Time	Sentence score
1	001	5550	20	016	06
2	002	4500	28	052	19
1	003	6650	16	018	04
3	004	5600	29	024	10
2	005	8650	04	036	04
3	006	3550	29	051	04
1	007	3350	28	036	10
2	008	6500	11	066	05
3	009	5450	24	039	09
1	010	5500	26	022	14
2	011	4700	20	040	09
3	012	4650	28	042	16
3	013	8500	13	012	04
2	015	4300	29	030	17
2	016	4800	23	046	09
1	017	6350	11	018	09
1	018	5300	23	028	05

APPENDIX 3.2

Data from Study Three reported in Chapter Three, showing group (1=LDOCE, 2=OALD and 3=COBUILD), EVST score, the number of words looked up, average look-up time in seconds and average sentence score

Group	ID	EVST	Number	Time	Sentence score
1	001	7750	08	024	04
2	002	7500	12	026	04
3	003	4500	18	057	05
1	004	6450	17	036	05
2	005	3350	16	027	04
3	006	3600	17	065	03
1	007	6500	12	043	05
2	008	3550	18	054	08
1	009	3600	13	112	09
1	010	4450	18	019	01
3	011	7550	13	019	11
1	012	5350	16	016	07
2	013	5350	14	093	05
3	014	5550	12	021	08
2	015	3350	14	032	03
3	016	5700	09	024	06
2	017	4200	17	063	04
3	018	5400	12	028	06
1	019	7550	08	015	05
2	020	4700	13	011	07
3	021	2450	18	116	08
1	022	6450	08	012	08
2	023	4650	18	042	07
3	024	3200	16	074	10
3	025	5450	12	045	08
1	026	6250	15	038	09
2	027	6450	14	027	04
3	028	3600	18	038	08
1	029	4600	09	014	02
2	030	6350	01	050	01
3	031	6400	11	095	06
2	032	4800	16	055	06
3	033	6600	15	064	04
1	034	6550	03	025	02
3	035	8600	09	025	01
1	036	3700	14	047	08
2	037	5400	18	055	02
3	038	4500	13	049	04
1	039	6200	06	041	01
2	040	6650	10	032	07
3	041	5500	11	025	03
1	042	4350	18	026	05
2	043	5500	12	040	03
3	044	3450	18	093	00
1	045	3600	17	056	04
2	046	2800	18	064	04
2	047	3600	15	030	02
2	048	5450	18	061	05
2	049	5400	18	043	06
1	051	3650	18	053	07
2	052	4650	17	065	03

APPENDIX 3.3

A sample of sentences produced by OALD users for Study Three.

The numbers in the left margin indicate the number of milliseconds each subject spent reading the dictionary entry. A zero indicates that the subject did not look up the word.

- 13280 I am enlightening some children about English culture
- 00570 I passed him erred money because it was forein notei
- 81890 I think knives chainge gravity
- 32740 Exercises incorporate examinations.
- 96220 This road and that railway are intersecting the River Thames
- 42450 (NO SENTENCE)
- 4120 I am always retarding to eat foods.
- 18900 I don't want to answer rudimentary questions.
- 35430 I bought the book about symptoms.
- 36420 I am looking for the job of part time version.
- 111770 I was agitated very much when I watched on television.
- 0 In near days, the hurt of civic is getting farher.
- 43390 My doctor said I was clarifying soon.
- 27030 The window collides the enterance.
- 25810 The time is coming when I was computing the test of half.
- 21690 I wrote the letter to him, the letter was shown my controversy.
- 15430 We interacted policemen on the road.
- 72120 Many boys and girls enjoyed a one hour interlude yesterday.

A sample of sentences produced by OALD users for Study Three (continued).

- 33340** I used to jump high walls when I was a child.
- 126740** You are erres if you believe that money get happiness.
- 136330** If you leave a knife on the air it will drop down because of the gravity.
- 58440** You have to incorporate this exercise to the test.
- 69640** The river intersects the big greenfield.
- 70080** Strong winds perpetrate a lot of destroys in winter.
- 88310** If you eat a lot of food you will retard the progress of slimming.
- 53330** The rudimentary questions made me confused.
- 4620** This medical book includes all the symptoms of the illness.
- 57890** The 3.0 version of this computer program will help you in your job.
- 19720** The violent films on television cause agitate to the children.
- 32740** My father participated the civic of the new school in our town.
- 51520** Doctors must be clarified to their patiens.
- 52350** I was standing at the window and I saw the car when collided with the van.
- 48720** I compted the half of the scores.
- 24600** I wrote you in my letter about the controversy in Parliament.
- 26260** The policeman interacting on us.
- 27080** I met this girl on the interlude of the performance.

A sample of sentences produced by OALD users for Study Three (continued).

- 178900** Children are usually enlightened at school.
- 10710** If you are counting money it is very bad to err.
- 48720** To kill somebody with a knife it has got a very big gravity.
- 0** These exercises have been incorporated in our test.
- 55200** This river is intersected by a lot of trees.
- 32790** (NO SENTENCE)
- 18940** I am two hours waiting for the food it is retarding too much.
- 17080** That question about the life after death was rudimentary.
- 0** This book is based on people's symptoms.
- 0** What I have just explained you is just a small version of my job.
- 5210** Television programs are usually very agitate.
- 13950** My father works as a civic in our town.
- 27900** Yesterday I collided with the window so today I have a big pain in my head.
- 10930** I have computed for a half an hour and it isn't finished yet.
- 13450** Yesterday I wrote a letter about the controversy of yugoslavia.
- 6870** The policeman interacted me and told me not to make strange noises with my tongue.
- 13570** The girl asked for an interlude, she had to go to the toilet.

APPENDIX 3.4

A sample of sentences produced by LDOCE users for Study Three.

The numbers in the left margin indicate the number of milliseconds each subject spent reading the dictionary entry. A zero indicates that the subject did not look up the word.

- 76010 The child thought the bears can to talk until I enlighten him.
- 74980 It is better to err with the money when you go away.
- 30540 He understand the gravity when you use a knife wrong.
- 9340 He incorporate the new exercise to his notebook.
- 13080 (NO SENTENCE)
- 16310 The winter perpetrates to the fruits.
- 18840 The food arrived with retard.
- 7470 The question had a rudimentary answer.
- 4010 This symptom is in your book.
- 9830 He told me a lot of version about his job.
- 7740 The television show agitate events.
- 10160 My father is a civic man.
- 10220 The doctor clarifies my ill.
- 3360 The window collides with the wall.
- 2470 He result the half of the compute.
- 6810 My letter was very controversy.
- 21420 The policeman interact with the fireman.
- 32190 (NO SENTENCE)

A sample of sentences produced by LDOCE users for Study Three (continued).

- 0 Children enlighten our hope.
- 15870 It is my errs to borrow you money.
- 0 Knife is a suitable word to describe the word of gravity.
- 7250 He incorporated many exercises for us.
- 8890 There is an intersecting path near the river.
- 13240 It is easy to perpetrate the weather in the winter.
- 12580 We need retard the food for the winter.
- 27410 The question is still rudimentary.
- 0 Book is the symptom of knowledge.
- 0 We can see the version of this job.
- 8620 There are many television programs agitating our children.
- 0 Father always teaches us about something about civic town.
- 0 The doctor clarified everything for us.
- 14610 There are two windows collided with each other.
- 0 I computed half of the examination.
- 0 There is a controversy issue in the letter.
- 0 The policeman interacted the case carefully.
- 21210 The girl was interluded in the story.

A sample of sentences produced by LDOCE users for Study Three (continued).

- 67280** Children belives in Santa Clause until their mothers enlighten them.
- 7360** Money is important. don't err when you pay and don't lose it.
- 0 If you threw a knife it will fall down because of the gravity.
- 0 The exercise incorporated a listening exercise and a piece of writing.
- 0 The two rivers intesect near Rome.
- 0 In winter the thieves perpetrated a plan to rob the money.
- 19830** If you don't eat your food your growing will retard.
- 0 The question was asked in a rudimentary way.
- 7750** I bought a book that explains the symptom of aids.
- 0 For his job he uses the last version of DOS.
- 0 I agitates when the television is out of work.
- 0 My father attended the civic school of music.
- 0 The doctor clarify his position during the discussion.
- 2640** The window collides when the thieves jumped in the room.
- 1370** I computed only af of the maths expression.
- 0 The controversy discussed in the letter was very interesting.
- 7140** The policemen interact during the new course.
- 7150** The girls danced during the interlude between the two period.

APPENDIX 3.5

A sample of sentences produced by COBUILD users for Study Three.

The numbers in the left margin indicate the number of milliseconds each subject spent reading the dictionary entry. A zero indicates that the subject did not look up the word.

- 36630 Children are enlightened by their parents and teachers.
- 71510 Don't err in using money.
- 57120 It is natural for a knife to fall down when you leave it alone begravity of the earth.
- 54820 The exercise is incorporated in the whole course.
- 82170 Two rivers intersect in the city.
- 48880 The murder had been perpetrated in the winter.
- 88710 Be careful about foods not to retard your health.
- 51020 He asked his teacher a very rudimentary question.
- 39000 The book says that a headache is the symptom of a cold.
- 81450 The editor's job is to prepare the publication of a new version of the book.
- 0 Some politicians worry whether the television programme agitates the public opinion concerning to the new tax problem.
- 0 The soldier's father was a great civic of Roma.
- 53120 The doctor clarifiedme explaining with basic words.
- 62560 All the windows of my car have been broken because it was collided by another car yesterday.
- 73930 More than half students are able to compute the fare at once.
- 58280 There is a serious controversy about the means to express Japanese language in Roman alphabets.
- 253860 Policemen's works must interact with people's life.

**A sample of sentences produced by COBUILD users for Study Three
(continued).**

- 76230 The girl is taking an interlude in the next room.
- 50700 The children does need more enlightens.
- 70740 Sometimes people are in err, but it should not about money.
- 84800 The knife is dropped by gravity.
- 66300 I was doing incorporate draw to something of exercise
- 55960 Somewhere all the river intersected to sea.
- 92330 The man perpetrate to crime a mountain in winter.
- 37410 Some people retard to cook food recentry.
- 17960 There were some rudimentary questions.
- 78590 The book written about symptom of flu.
- 56960 Her mother agitate about child's eyesight, because the child love watching television so long time.
- 79980 My father was determined to carry his civic responsibilities.
- 66020 The doctor was clarify to tell about my illness.
- 106390 He almost collide with me about the new window.
- 62280 It was very tired to finish computing, even half.
- 36420 I fed up with to avoid with mother by the letter.
- 101840 The policeman interacted about that accident with citizen.
- 43220 After this interludes, the girl start to play the piano again.

**A sample of sentences produced by COBUILD users for Study Three
(continued).**

- 44490** Children are enlightened by comics.
- 24930** If you err in life, you can not make money.
- 3680** Also knives are subject to gravity.
- 0** Exercise incorporates theory and practice.
- 12250** The Thames river does not intersect with any other river.
- 78050** Winter always perpetrates after summer.
- 3570** Stomacache retards food digestion.
- 13240** I have a question about the rudimentary tools used by men 4000 years ago.
- 9720** If a book falls in your head you will certainly feel the symptoms.
- 0** I wark with computers, and in my job there are different versions of basic.
- 9450** She became so agitated watching that film on television.
- 12630** My father warks in the civic centre.
- 0** I want to clarify my position.... I am a doctor!
- 0** When the two cars collided, the windows of the cars exploded.
- 21310** My job is to compute half of the incoming data of this company.
- 0** There were too many controversy in his letter.
- 6200** He grassed his friend interacting with that policeman.
- 12420** After a short interlude, he went back to his girl.

APPENDIX 4.1

Data from Study Four reported in Chapter Four, showing group (1=Portuguese, 2=Malaysian), EVST score, the number of words looked up, average look-up time in milliseconds and average sentence score.

Group	ID	EVST	Number	Time	Sentence score
1	001	6400	13	1864	3.0
1	002	6400	12	1431	3.3
1	003	4500	09	1378	3.5
1	004	6500	14	2005	3.7
1	005	6450	09	4069	2.9
1	006	4300	15	2123	3.9
1	007	3550	14	2024	2.4
1	008	6450	12	2067	3.5
1	009	4400	17	1218	3.0
1	010	3700	17	1997	4.4
1	011	6500	15	2063	3.8
1	012	5600	13	3649	3.5
1	013	4500	14	2827	3.8
1	014	5900	09	1530	1.9
1	015	3500	16	2261	3.9
1	016	5400	15	1484	3.9
1	017	3600	11	2425	2.8
1	018	5400	17	1589	3.5
1	019	3650	15	3133	3.4
1	020	5450	13	1334	4.2
1	021	3600	18	5609	4.5
1	023	7600	16	2434	4.6
1	024	5450	08	3057	3.7
1	025	5500	15	2939	4.6
1	026	6450	18	1656	4.1
1	027	3600	17	1804	4.0
1	028	3500	14	2163	3.4
1	029	6400	16	4160	3.4
1	030	6450	10	1078	3.1
1	031	6400	10	3073	3.2
1	032	4450	12	1606	2.4
1	033	5450	17	3975	4.5
1	034	6600	04	1524	4.2
1	035	6500	14	4801	3.3
1	036	4400	08	1588	2.9
1	037	4400	12	4121	3.4
1	038	4400	13	3332	3.2
1	039	7550	09	2896	2.2
1	040	3450	08	2049	4.0
1	041	3500	13	1858	2.6
1	042	5500	12	2604	2.8
1	043	6500	18	2116	3.7
1	044	4600	15	2621	3.9
1	045	6500	10	1320	4.9
1	046	6450	13	1517	3.4
1	047	3250	18	2250	2.8
1	048	6480	11	1540	4.0
1	049	6460	07	1254	4.6
1	050	3560	11	1704	2.5
1	051	6500	08	9465	3.9
1	052	5650	13	1160	3.6

Group	ID	EVST	Number	Time	Sentence score
2	002	6400	18	6154	3.0
2	003	7500	12	2407	2.5
2	004	6450	18	5286	3.1
2	005	7550	18	5667	3.3
2	006	8800	14	1446	3.6
2	007	6550	07	2482	4.8
2	008	7600	12	2303	3.7
2	009	4500	18	3050	2.5
2	010	6550	15	3455	2.5
2	011	5600	15	2339	3.7
2	012	5650	16	2131	2.8
2	013	4350	18	3870	2.8
2	014	5550	17	4452	3.3
2	015	6650	18	2724	3.8
2	016	9650	16	2621	2.8
2	017	6500	18	5022	4.0
2	018	8550	18	1937	4.3
2	019	5450	18	3189	3.5
2	020	6400	18	3369	3.4
2	021	6300	15	2286	2.0
2	022	7550	17	4407	2.9
2	023	6550	18	4025	3.4
2	024	6600	16	3811	2.4
2	025	4500	16	3054	3.0
2	026	6900	14	1554	2.7
2	027	8900	14	1425	3.0
2	028	4600	17	4681	3.3
2	029	5600	16	2880	2.1
2	030	8400	15	1702	3.4
2	031	7550	17	2798	3.3
2	032	6500	17	4574	2.9
2	033	5300	17	5921	3.2
2	034	6700	16	4406	2.7
2	035	6350	17	1215	2.3
2	036	4800	15	2111	2.5
2	037	9750	18	2233	4.5
2	038	7800	17	2201	2.4
2	039	8700	17	1848	2.8
2	040	8800	09	2122	2.9
2	041	7700	18	3549	3.0
2	042	6600	18	2951	2.9
2	043	7550	18	3660	2.8
2	044	5600	13	3895	3.0
2	045	6500	15	2495	3.5

APPENDIX 4.2

A sample of sentences produced by the Malaysian group for Study Four

The numbers in the left margin indicate the number of milliseconds each subject spent reading the dictionary entry. A zero indicates that the subject did not look up the word.

- 58390 To enlighten children is more difficult than to enlighten adult.
- 21530 Human always err when they wanted to borrow money.
- 0 The gravity make the knife that you throw falls back on land.
- 11490 To incorporate with each other we must have a lot of exercise.
- 8180 Intersecting a river is very difficult.
- 23900 Winter is one of the perpetrate factor of his death.
- 16810 Poisoning is a retard of food which were left without covering it.
- 38120 Rudimentary question is very easy to answer.
- 6210 Reading books is a very good symptom.
- 32580 Every people have a different version on that job.
- 9830 He always become agitate when watching ghost story on television.
- 14280 Father is an important civic axample toward his children.
- 0 The doctor clarified that he cannot help his pacient any more.
- 16310 The stone which were thrown to the window have collided it.
- 5330 Only half of the sums can be compute.
- 0 The letter have become a great controversy on them.
- 25100 Public and policeman got to have a good interact with each other.

**A sample of sentences produced by the Malaysian group for Study Four
(continued).**

- 28610 The girls have an short interlude during their
resesst.
- 17410 The children thought that the school is only
for the adults until their teacher enlightens
them.
- 43610 Don't blame the money for your own errs.
- 12570 When you throw a knife up to the sky, it will
fall down to the grown because of the gravity.
- 30870 Ones can incorporates another while doing any
educational exercises.
- 17190 We have to intersect a few paths on our way to
the river.
- 64370 A project on making the winter's coat cannot be
done because the managing direct.
- 50150 Lack of food is the cause of crop's retarding.
- 27900 The questions cannot be answered because he has
a rudimentary ideas about it.
- 0 The symptoms of stealing books in the library
are getting worst.
- 13400 I can get a better version of job in that
department.
- 72450 I've watched a forum in the television where
the panels agitated on the current issues.
- 0 My father is a civic-minded person.
- 7750 The doctor clarified the patient as a dead
person.
- 16370 A bird collided with a window while flying.
- 20760 I've took about half an hour to compute the
exercises.
- 32850 The letter cause the controversy between the
two politicians.
- 14720 The policeman was **interacting** with the woman.
- 46360 The girl gave a brief interlude about water
sports during the 19th century.

**A sample of sentences produced by the Malaysian group for Study Four
(continued).**

- 0 That child is playing the toys with his mother.
- 98320 Nowadays most of the people do such err things to earn money.
- 88490 Exercises can incorporate us into the best of health.
- 11590 That deep and long river intersect Kampung Batu Tiga and Kampung Telujuh.
- 17960 Peter supposed to meet Jenny during winter, but unfortunately he couldn't make it so he feel perpetrate.
- 16640 All the donation food supposed to be arrive by this time but it is retards just because it is raining heavily.
- 11480 All the question that you create for me is rudimentary.
- 4720 The symptom that we are facing nowadays is lack of people reading books, magazine articles, even newspaper.
- 20930 Some of the people assume sing as their fits job and that kind of versions is very popular among the teenager.
- 18630 The news that been shown in the television about the murder makes him agitate to know further about it.
- 24820 My father is the civic of that country long time ago.
- 6150 The doctor had carify that Ali had this disease from long time ago.
- 16590 The window was collided when the it was brought here.
- 4840 Half of the students tried to compute that Math exercise.
- 0 The letter cause lots of controversy.
- 0 The policeman is interacting with the tourists.
- 17020 That girl interlude herself between her studies and sports practice.

APPENDIX 4.3

A sample of sentences produced by the Portuguese group for Study Four

The numbers in the left margin indicate the number of milliseconds each subject spent reading the dictionary entry. A zero indicates that the subject did not look up the word.

- 49710 Children are enlighten by their parents.
- 28240 Stole money from a bank is to err.
- 17800 The boy cut his hand and he did not understand the gravity of the situation.
- 9110 This exercise incorporate a strong force.
- 7750 The road intersects the river.
- 24220 In that winter Paul was perpetrated to five years of prison.
- 15060 The lack of food retards the development of the child.
- 13190 It was a rudimentary question.
- 92160 The production of books is a symptom of lack of interest on it.
- 13950 This job is a version of the other.
- 8790 That film on television agitates many people.
- 17350 My father is very civic.
- 0 The doctor clarifies the patient.
- 5650 The stone collided in the window
- 21750 The ycompute the result of half a cake.
- 20200 This controversy letter concerned many people.
- 0 Policeman was interacting that girls.
- 16860 In the interlude girls will go to eat.

**A sample of sentences produced by the Portuguese group for Study Four
(continued).**

- 27250** The child is enlightening his father
- 19060** The man erred in the count of his money.
- 7090** The gravity of the imen pull the knife.
- 11050** The exercises of phisical education
incorporates andebol.
- 4670** The "douro" river intersected with the "tejo"
river.
- 10330** The murderer is perpetrating a murder to the
winter.
- 18620** The food retards because there is no gas.
- 4330** This question is very rudimentary.
- 21520** I think this book brings symptom.
- 19280** That version of job in newspaper seems to be
good.
- 0** The television is agitating the children with
that film.
- 8240** The father civic's is to support his family.
- 3410** The doctor clarifies his patient about his
desease.
- 9670** The children collides trough the window.
- 14390** He is in the computer, computing for half an
hour.
- 1810** This letter has many controversies about the
marriage.
- 3350** The policeman interacts with the ideas of
another policeman.
- 23020** The girl made an interlude in her job.

**A sample of sentences produced by the Portuguese group for Study Four
(continued).**

- 43560 This child was enlighten by him.
- 23950 It is better to err than spend a lot of money.
- 0 It is very gravity using a knife.
- 0 I incorporate some exercises in the book.
- 8900 I intersect the river by boat.
- 6590 Last winter I perpetrated a crime.
- 12410 Every dinner I retard the food.
- 0 This question is very rudimentary.
- 42730 The book has a lot of symptoms.
- 2970 I have another version of the job.
- 0 The television has some agitated films.
- 8620 My father is a civic of Portugal.
- 0 The doctor clarifies my headache.
- 5930 Mary collided the window violently.
- 7580 I made only half of the compute.
- 0 The letter controversies my opinion.
- 0 The policeman interacted with the thief.
- 24170 The girl interlude her time by reading a book.

APPENDIX 5.1

Data from Study Five reported in Chapter Five, showing group (1=4a, 2=4b), subgroup (3=without examples, 4=with examples), EVST, average time taken to look up entries (in milliseconds), the number of correct sentences produced after look-up, and the percentage of correct sentences produced after look up.

Gp	Subgp	ID	EVST	Time	No	Cor	Percent
1	3	101	3800	5139	9	3	033
1	3	102	8650	5558	8	4	050
1	3	103	5650	7105	3	1	033
1	3	104	6550	5158	9	3	033
1	3	105	7700	1228	6	2	033
1	3	106	7600	1418	4	4	100
1	3	107	4500	3563	8	1	013
1	3	108	5550	4272	9	6	067
1	3	109	5400	3628	9	4	044
1	3	110	3350	9271	2	1	050
1	3	111	4400	5618	9	4	044
1	3	112	4400	2361	9	6	067
1	3	113	7500	3461	8	5	063
1	3	114	6450	1014	4	3	075
1	3	115	4500	3998	7	0	000
1	3	116	3550	1327	8	2	025
1	3	117	3530	2491	9	2	022
1	3	118	4500	1917	7	2	029
1	3	119	5700	6034	8	0	000
1	3	120	5350	2409	9	3	033
1	4	101	3800	3052	9	5	056
1	4	102	8650	3487	9	8	089
1	4	103	5650	5664	5	2	040
1	4	104	6550	1757	9	5	056
1	4	105	7700	1543	5	3	060
1	4	106	7600	0510	3	3	100
1	4	107	4500	3356	8	5	063
1	4	108	5550	3945	9	8	089
1	4	109	5400	3175	6	2	033
1	4	110	3350	4485	6	4	067
1	4	111	4400	2151	5	1	020
1	4	112	4400	1495	6	4	067
1	4	113	7500	1732	8	5	063
1	4	114	6450	0758	2	2	100
1	4	115	4500	3735	7	0	000
1	4	116	3550	2713	5	1	020
1	4	117	3530	2201	8	3	038
1	4	118	4500	1287	6	2	033
1	4	119	5700	5204	7	0	000
1	4	120	5350	3321	9	3	033

Gp	Subgp	ID	EVST	Time	No	Cor	Percent
2	4	201	2600	8238	9	3	033
2	4	203	6300	1952	7	1	014
2	4	204	6100	4131	9	3	033
2	4	205	4550	3303	7	4	057
2	4	206	2500	4770	9	4	044
2	4	207	3600	4364	6	2	033
2	4	208	4450	4395	9	4	044
2	4	209	6500	4628	3	3	100
2	4	210	4550	3666	9	5	056
2	4	211	4500	4833	7	2	029
2	4	212	6500	2414	2	2	100
2	4	213	3400	1934	6	5	083
2	4	214	8650	0963	4	3	075
2	4	215	6500	3819	7	5	071
2	4	216	5350	2859	5	3	060
2	4	217	5250	3992	9	6	067
2	4	218	5450	1905	5	2	040
2	4	219	3600	4784	7	2	029
2	4	220	5350	6337	9	4	044
2	4	221	3400	6877	9	4	044

2	3	201	2600	4243	9	4	044
2	3	203	6300	1134	8	4	050
2	3	204	6100	1353	9	4	044
2	3	205	4550	1856	9	4	044
2	3	206	2500	2748	7	3	043
2	3	207	3600	2607	8	1	013
2	3	208	4450	3100	9	5	056
2	3	209	6500	3118	3	2	067
2	3	210	4550	2842	9	2	022
2	3	211	4500	1759	9	5	056
2	3	212	6500	1029	3	2	067
2	3	213	3400	0484	8	2	025
2	3	214	8650	0937	3	2	067
2	3	215	6500	1500	1	1	100
2	3	216	5350	2699	5	1	020
2	3	217	5250	2284	9	3	033
2	3	218	5450	2197	3	2	067
2	3	219	3600	4146	6	2	033
2	3	220	5350	3824	8	1	013
2	3	221	3400	3863	9	1	011

APPENDIX 5.2

A sample of sentences produced with access to examples for Study Five

The numbers in the left margin indicate the number of milliseconds each subject spent reading the dictionary entry. A zero indicates that the subject did not look up the word.

- 59320 (NO SENTENCE)
- 83100 A larger amount of money can err the his whole life.
- 16640 The knife drops on the floor that is by the gravity.
- 79810 He incorporates our suggestions in his exercise.
- 4340 The end of the two rivers intersect together to form a lake.
- 60580 A big traffic jam is perpetrated by the careless mistakes in this winter.
- 69430 The Thyhoon retards the export and import of the food.
- 101610 This is the rudimentary question for Peter.
- 95570 This symptom is hard to find from the general books.
- 2960 They do not know what version of software needs to implement in the computer to do the job.
- 29930 Somebody was agitated to do strong complain after hearing the news from the television yesterday.
- 13730 His father is a civic servant.
- 2690 The doctor has already clarified to him how to take those medicines.
- 2420 The window was broken because it was collided with the stone.
- 1270 Half of the students here are busy with computing.
- 22070 The letter has described the controversy about the violence that happened yesterday.

**A sample of sentences produced with access to examples for Study Five
(continued).**

- 54110** Those people always interact with the policeman about the parking place.
- 169780** The girl has a goog planning during the interlude.
- 93210** The child was enlightened that the sun rises at east.
- 114630** Sometimes an accountant will also err for his money.
- 42940** To keep the children away from the knives is the gravity of safety.
- 12140** This exercise is incorporated with the final coursework.
- 0** The two rivers intersect each other.
- 55260** Peter perpetrated a murder at last winter.
- 16750** He asked those silly question because of rudimentary knowledge.
- 0** According to the symptoms that book may be the best selling one in this year.
- 32070** The difficult job is the second version.
- 15440** The television program agitate all the childern.
- 32510** My father is civic in Hong Kong.
- 8510** The doctor clarify to tell to the customers about the illness.
- 7690** The window was collided by the car.
- 0** The half size of the cake can be computed by the computer
- 28880** The letter talking about the controversy of our point.
- 12300** The policemen would be interacted form different counties.
- 38720** The girl paly interlude.

**A sample of sentences produced with access to examples for Study Five
(continued).**

- 20650** The child enlightened the answer after the teacher tell him.
- 24060** Money can make someone in an erring way.
- 0 Gravity makes the knife fall down from the desk into the ground.
- 0 The exercise needs to incorporate with another.
- 0 There are two rivers which intersected with each other.
- 23450** It is perpetrated by eating icecream at winter.
- 0 The food can be retarded until tomorrow.
- 14890** The question needs a rudimentary answer.
- 12200** The book teach us about the symptoms of lung cancer.
- 2800** Kelvin, you have to finish your job because the most update version of the software is comming.
- 12030** Television programs make somebody agitate in front of others.
- 13730** My father always intrested on the civic events of the city.
- 0 Please ask your doctor to clarify what he will going to do when facing your disease.
- 35980** The baskatball collide on the window of my house.
- 0 Half of my course work rely on computing question.
- 0 Please don't write a letter to the president talking so much controversy matter in the country.
- 4450** Policeman always interacts together because they know well.
- 8230** This girl like the interlude of the song very and very much.

APPENDIX 5.3

A sample of sentences produced without access to examples for Study Five

The numbers in the left margin indicate the number of milliseconds each subject spent reading the dictionary entry. A zero indicates that the subject did not look up the word.

- 41950 All children must be enlightened from fault beliefs.
- 8900 He seems likely to be erred with the money given.
- 7250 The knife drops to floor because of gravity force.
- 19940 You have to incorporate with your friends to do the exercise.
- 15820 The river A and river B are intersected at point C in the map.
- 45580 He was perpetrated to play sking in this area in the last winter.
- 34050 He retards to buy those food because the date is expired.
- 25540 The question requires to do rudimentary experiment.
- 17850 The book has describing the synton of aids.
- 0 His job is to produce different version of computer software.
- 30100 The serious accident shown on the television made everyone agitating.
- 42570 His father is the most famous author and civic in United Kingdom.
- 0 A doctor should clarify the diagnosis of his patient.
- 0 Two cars collided together and the windows were broken.
- 3290 Half of the results for the examination is already computed.
- 25710 The letter of appeal for the examination result is controversy at the board of examiners.
- 118370 The policemen interacted each other to ask for a higher salary.

**A sample of sentences produced without access to examples for Study Five
(continued).**

- 28720 The girl falled asleep at the interlude of the concert.
- 92380 All children enlighten a story after the old man explanation.
- 13960 The amoount of money was erred to return to customer.
- 48670 The knife drop on the ground due to gravity.
- 10830 The student incorporate to finish the homework.
- 4890 The river intersect at the west countryside.
- 15490 The winter perpetrate lately.
- 16370 The fast food retard to give us due to the cooker wrong.
- 11470 The rudimentray question cannot test the middle school student.
- 10160 The book with symptom sale a low price.
- 26200 Peter said the different version of the job nature to us.
- 73660 The people of Hong Kong agitates against the attitude of the Hong Kong grovernment change, when they are watching television.
- 23070 Father is the civic residence of Canada.
- 9070 Doctor must clarify the record of each patients to other doctoide the treatment.
- 8120 Windoe news collides the point of view of Time news about the raking of the school.
- 0 The faster computer computes the half processing time of the older one.
- 94370 Peter returns the letter which is about the controversy of my point.
- 3460 This strick interacts the status of the policeman.
- 67990 Girls like the interlude in the disco.
- 11210 Some children enlighten the significant fact of balanced diet.

**A sample of sentences produced without access to examples for Study Five
(continued).**

- 8680 A teller informs to the customer who errs to pay more money.
- 18230 Knife will be dropped if you do not hold it owing to the gravity.
- 7580 Different aspect of exersices shall be incorporated in this weekend camp.
- 0 There are a lot of rivers intersectin among themselves.
- 18400 In the winter, some thefts would perpetrate goods in a market because wear a long cloth.
- 19720 The party cannot lauch in time because some foods retard to be made.
- 19500 This question is rudimentary so that small child may be learned itself.
- 2860 The book is easy to be found because special symptom have beeviev.
- 41300 The job advertisement have different version on Friday and Saturday.
- 13020 Shaking television image agitated him, so he turned it off.
- 25540 My father is a good civic.
- 67670 We should clarify our symptoms to the doctor.
- 8240 Strong wind make the window collide.
- 2080 His time to compute this question is only half of Tom.
- 4780 Letter from last week is completely controversy to this week.
- 18610 Two patrolling policeman interact the robbery on the house.
- 24330 The girl have meeting after a long interlude.