

## Contents

<b>Editorial Foreword</b>	ix
<b>Acknowledgements</b>	xv
<b>Synopsis</b>	xvii
<b>Chapter One: Introduction</b>	1
Prologue	1
Context	2
Nature of the investigation	3
Aim and scope of the study	4
Sub-technical, non-technical or semi-technical vocabulary	5
<b>Chapter Two: Semi-technical vocabulary in Computer Science</b>	29
Introduction	29
The creative aspects of semi-technical vocabulary	30
The rhetorical dimension	31
Category 1: Allusion	32
Category 2: Figurative usage	33
Category 3: Neologism	33
Category 4: Semantic change	34
Category 5: Word play	36
The morphological dimension	36
Category 1: Derivation	36
Category 2: Compounding	37
Category 3: Blending	37
Category 4: Abbreviation	38
Category 5: Phrasal-verb formation	38
<b>Chapter Three: Semi-technical vocabulary in the comprehension of CS texts</b>	41
Introduction	41
Test designs	41
Test 1: Metaphor in CS texts	41

Test 2: Lexical modification in CS texts	43
Test 3: Allusions and facetious words in CS texts	44
Discussion of results	44
<b>Chapter Four: Methodology of experimentation</b>	<b>49</b>
Introduction	49
Research techniques and test design	49
Decontextualised word-level glossing exercise (DWLGE)	50
A multiple-choice questionnaire survey (MCQS)	50
Hands-on computer tasks (HOCTs)	51
Thinking-aloud technique (TAT)	52
Retrospective interview (RI)	54
Selection of subjects	55
Selection of semi-technical vocabulary	56
Selection of dictionaries	62
Selection of the electronic glossary of semi-technical vocabulary	62
Selection of venues	63
Selection of time	63
Experimentation	64
Pilot study	64
Revision of the research design after the pilot study	64
MCQS with/without TAT and RI	65
HOCTs with/without TAT and RI	67
Comparison of results between the GLDs and the EGSV	69
Summary	73
Conclusions	73
Limitations	75
Recommendations for future research	75
<b>Appendix 1</b>	
Test 1, Stage 1. Individual metaphors	79
<b>Appendix 2</b>	
Test 1, Stage 2. Words in context	81
<b>Appendix 3</b>	
Test 1. Supplementary information	83
<b>Appendix 4</b>	
Test 3, Stage 2. Allusions and facetious words	87

<b>Appendix 5</b>	
Test 3. Supplementary information	89
<b>Appendix 6</b>	
Decontextualised word-level glossing exercise 1 (DWLGE 1)	93
<b>Appendix 7</b>	
Decontextualised word-level glossing exercise 3 (DWLGE 3)	95
<b>Appendix 8</b>	
A test of vocabulary understanding (main study)	97
<b>Appendix 9</b>	
Hands-on computer tasks. Instructions	105
<b>Appendix 10</b>	
Thinking-aloud Technique. Notes	109
<b>Appendix 11</b>	
Thinking-aloud Technique. Advice sheet	111
<b>Bibliography</b>	113
<b>Editorial postscript</b>	119



## Editorial Foreword

I was indisposed recently, and had to cancel one of my classes. The following weekend, I received this [verbatim!] e-mail from one of the students, explaining why he himself would not be in the next class:

I am now suffering the same problem what you had in the last. I think the disease store in my body for a long time and now explose. Just like the virus in the computer. I think I need to see the doctor today and take a rest to defect the disease.

This student's use of analogy illustrates that, for him, it is the computer virus<sup>1</sup> which is the reality, and which has become institutionalised in his language; the medical virus is now the living metaphor (cf. Goatly 1977: 31–5).

Metaphor is pervasive in the language of Computer Science (CS). Not only is the distinction between the reality and the metaphor often blurred, even for native speakers of English, but it is compounded for non-native speakers in that they may, in many cases (as in the, not untypical, example cited), be unaware that there is a metaphor at work at all. The underlying theme of the present study by Jacqueline Lam is the effect of metaphor in CS on non-native English-speaking students' reading comprehension in this domain.

Metaphors are cognitive tools which determine communicative reference points, provide frames of orientation, and – although some may find this tendentious – shape our understanding (cf. Paprotté & Dirven 1985; Ortony 1993; Stern 2000):

Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature. ... Our concepts structure what we perceive, how we get around in the world, and how we relate to other people. Our conceptual system thus plays a central role in defining our everyday realities. ...the way we think, what we experience, and what we do every day is very much a matter of metaphor. (Lakoff & Johnson 1980: 3)

---

<sup>1</sup> The use of the term *virus* in a computer context was coined by David Gerrold in his science fiction novel, *When Harlie was one* (1972). "A computer virus is a self-replicating program containing code that explicitly copies itself and that can *infect* other programs by modifying them or their environment such that a call to an infected program implies a call to a possibly evolved copy of the virus. Many people use the term *virus* to cover any sort of program that tries to hide its possibly malicious function, or tries to spread into as many computers as possible." (Jalobeanu 2000)

Lam shows how words of the general lexical stock become ‘technologised’ through metaphor in specific contexts, in what has been variously termed ‘non-technical’, ‘sub-technical’, ‘semi-technical’ or ‘academic’ lexis. She gives a useful chronological review of a number of sources with applied linguists’ definitions of ‘semi-technical’ etc., followed by a discussion of the creative aspects of such vocabulary in CS. Allying the criterion of frequency of occurrence in specialised texts (cf. Yang 1986) with semantics – i.e. whether the words selected have more or less restricted meanings (cf. King 1978; Coxhead & Nation 1999) – and basing her data on the Hong Kong University of Science and Technology (UST) CS Corpus,<sup>2</sup> she investigates some of the problems this type of vocabulary presents to students in their reading in CS.

Lam’s discussion of the creativity of semi-technical vocabulary in CS covers rhetoric and morphology. Within the former, she exemplifies from the UST CS Corpus features of allusion, figurative usage (including anthropomorphism and animation), neologisms, semantic change (including specialisation, generalisation and contextual transfer) and pun; and within the latter, derivation, compounding, blending, abbreviation and phrasal verb formation. She discusses some of the witty figurative expressions (known rhetorically as *conceits*) found in CS writing.

Of particular interest are allusions, the implied or indirect historical, mythological, topical, literary, personal or other cultural references, made under the assumption that they will be recognised and shared by the reader. An example of a common allusion in computing is the ‘Trojan horse’, malicious or harmful code contained inside an apparently benign, even useful, program in such a way that it can gain control, and even destroy the file allocation table on a hard disk.<sup>3</sup> Students may know what

---

<sup>2</sup> Cf. James & Davison (1992), James *cum al.* (1994), James (1996). This corpus was originally compiled to inform the design of materials to assist first-year university students in reading CS texts (cf. Tao 1994). The project has spawned further work at UST in syn-tax (Ng & Lam 1994) and lexicology (James 1993; Tong 1993; Li & Pemberton 1994), and elsewhere, in lexicography (Lynch 1999).

<sup>3</sup> The earliest citation in the *Oxford English dictionary* of *Trojan horse* in this context is dated 1974. The allusion is to Homer’s *Iliad*: during the Trojan War (?c. 1250 BC), fought to recover Helen, wife of King Menelaus of Sparta, who had been abducted by Paris, son of the Trojan king, Priam. Troy was besieged for nine years by the Spartans and their Greek allies, with a fleet of over 1,000 ships. At last, they built a large wooden horse, to offer to the Trojans as a gift. They left the horse, inside which they had hidden a contingent of warriors, outside the gates of Troy, and made as if to leave in defeat. The Trojans took it within their walls. During the night, the Greek soldiers emerged, and opened Troy’s gates from the inside, to allow their compatriots to enter and overrun the city.

a Trojan horse program is, but may not realise why it is so named. This may not be important, because although the allusion is interesting, and possibly a useful mnemonic, it is not necessarily significant. Lam selected, however, two cultural allusions from texts in the UST CS Corpus – *Robin Hood* and *Las Vegas* – the understanding of which was necessary to grasp the contextual meanings in given texts. She shows how these names were recognised by Hong Kong students at the word level, but were hardly understood at the sentence level – that is, the relevance of the allusions was lost on them, and the texts therefore misunderstood.<sup>4</sup>

Another frequent source of difficulty for Hong Kong students is the rhetorical convention of *anthimeria* (*functional shift*, or *conversion*), in which one part of speech is substituted for another without affixation or other formal process of derivation. Anthimeria is a powerful developmental force in English, particularly in the realm of computing, as in e.g.

it may be necessary to turn off the interrupt feature of the printer

where *interrupt* functions as an adjective, or

the machine should be capable of 208M multiplies per second

with *multiplies* functioning as a noun. Lam's study demonstrates that such unfamiliar syntax – which may not be cited in conventional sources, and which, indeed, may run counter to what students have been taught about English – can prove extremely difficult, nay disconcerting, to non-native speakers.

Other lexical and morphological means of novel word creation often used in CS terminology, and sources of potential confusion to students, include words imported from general sources (e.g. *bookmark*, *director*, *folder*, *time bomb*) or specific sources (*virus*, *infection*), neologisms (phrases such as *logic bomb*, or word formations such *cybercrime*, *cybernetics*, *cyberspace*); people's names used as technical terms (e.g. *Bernoulli*),

---

<sup>4</sup> Cf. May (2001: 12): “[Is] it a good idea to go around dropping allusions ...? Does it really make a description more vivid by stealing it from someone else? Isn't it, more often, a lazy way of talking or writing – or, worse still, just a way of showing off? ... Allusions can still be used well – but they have to be used with wit and imagination. They must be surprising, not as dull and obvious as they usually are.” Even the obviousness, of course, is lost on a non-native learner.

onomatopœia (e.g. *beep*, *click*), sound symbolism (e.g. *bug*), back formation (e.g. *download*, *input*), and slang (e.g. *spam*). Through her experiments with first-year students at UST, Lam shows that such features can cause learners severe problems in the comprehension of CS-related texts. Since it is unlikely that teaching materials (even those based on a specialised corpus, however representative) could contain all the coinages that may be met in discipline-specific texts, students need to develop strategies for dealing with the unpredicted as much as to be able to cope with the predictable (cf. Aston 1996).

As a first step in the development of appropriate learning materials for first-year university students encountering advanced CS texts, Lam proposes the creation of a bilingual electronic glossary,<sup>5</sup> based on data from a specialised corpus. There is some evidence (cf. Reinking & Rickman 1990) that secondary students receiving computer instruction of difficult text words with electronic text score more highly on vocabulary measures than do students reading printed pages with dictionaries or glossaries; Lam demonstrates the effectiveness of the use of an electronic glossary with tertiary students. The present report will inform the compilation of a collection of just such glossaries in a number of domains.

Gregory James

#### References

- Aston, G. 1996. What corpora for ESP? Paper presented at an ESP conference, University of Pavia, November. [Online.] Available at: [www.sslmit.unibo.it/guy/pavesi.htm](http://www.sslmit.unibo.it/guy/pavesi.htm).
- Coxhead, A. & Nation, P. 1999. The specialised vocabulary of English for academic purposes. *Wellington Occasional Papers in Applied Linguistics* 1, 31–55.
- Flowerdew, L. & Tong, A. K. K. (Eds). 1994. *Entering text*. Hong Kong: Language Centre, Hong Kong University of Science and Technology.
- Gerold, D. 1972. *When Harlie was one*. New York NY: Ballantine.
- Goatly, A. 1977. *The language of metaphors*. London & New York NY: Routledge.
- Jalobeanu, M. 2000. The Net abuses and education. Paper presented at the 4th Romanian Internet Learning Workshop, Miercurea Ciuc, June–July. [Online.] Available at: [cd.ubbcluj.ro/rilw/2000/papers/jalobeanu\\_paper1.html](http://cd.ubbcluj.ro/rilw/2000/papers/jalobeanu_paper1.html).
- James, G. 1993. Postscript to Berry's discussion of 'unique' adjectives and the definite article. In Pemberton & Tsang, pp. 189–95.
- \_\_\_\_\_. 1996. Extending corpus data in compiling a learner's dictionary. In E. F. Kotze (Ed.), *SA linguistics* 95 (pp. 57–64). Port Elizabeth: University of Port Elizabeth Press.
- James, G. & Davison, R. 1992. Dimensions of difficulty in English for science: A corpus-based investigation. In N. Bird & J. Harris (Eds), *Quilt and quill: Achieving and maintaining quality in language teaching and learning* (pp. 316–24). Hong Kong:

---

<sup>5</sup> For a study on the effectiveness of bilingual vs monolingual presentation of vocabulary, see Smallwood (2000).

- Education Department.
- James, G. with Davison, R., Cheung, A. H. Y. & Deerwester, S. 1994. *English in Computer Science: A corpus-based lexical analysis*. Hong Kong: Longman Asia.
- King, R. J. 1978. Vocabulary and the use of context in sci tech English. *TESOL Newsletter* 12, 4–21.
- Lakoff, G. & Johnson, M. 1980. *Metaphors we live by*. Chicago IL: University of Chicago Press.
- Li, E. S. L. & Pemberton, R. 1994. An investigation of students' knowledge of academic and subtechnical vocabulary. In Flowerdew & Tong, pp. 183–96.
- Lynch, D. 1999. *The concise dictionary of computing and information technology*. Dublin: Blackhall Publishing.
- May, D. 2001. From Mekon to Morpheus. *Viewpoint* [Review of A. Delahunty, S. Dignen & P. Stock (Eds), *The Oxford dictionary of allusions*. Oxford: Oxford University Press, 2001.] *The Times*, 26.4.2001, p. 12.
- Ng, I. W. Y. & Lam, P. M. Y. 1994. Modality in computer science text. In Flowerdew & Tong, pp. 82–8.
- Ortony, A. (Ed.). 1993. *Metaphor and thought*. 2nd edition. Cambridge: Cambridge University Press.
- Paprotté, W. & Dirven, R. (Eds). 1985. *The ubiquity of metaphor: Metaphor in language and thought*. Amsterdam & Philadelphia PA: John Benjamins (Amsterdam Studies in the Theory and History of Linguistic Science. Series IV: Current issues in linguistic theory 29).
- Pemberton, R. & Tsang, E. S. C. (Eds). 1993. *Studies in lexis*. Hong Kong: Language Centre, Hong Kong University of Science and Technology.
- Reinking, D. & Rickman, S. S. 1990. The effects computer-mediated texts on the vocabulary learning and comprehension of intermediate-grade readers. *Journal of Reading Behavior* 22, 4, 395–411.
- Stern, J. J. 2000. *Metaphor in context*. Cambridge MA: MIT Press.
- Smallwood, I. 2000. The effects of monolingual and bilingual presentations of lexical items on vocabulary acquisition. In L. Y. O. Mak & M. M. L. Keung (Eds), *Changing languages: Language education in the era of transition* (pp. 98–115). Hong Kong: Language Centre, Hong Kong University of Science and Technology.
- Tao, P. K. 1994. Comprehension of non-technical words in science: The case of students using a foreign language as the medium of instruction. *Research in Science Education* 24, 322–30.
- Tong, K. S. T. 1993. From 'single parent' to 'bound pairs': The secret life of computerese. In Pemberton & Tsang, pp. 196–214.
- Yang, H. Z. 1986. A new technique for identifying scientific/technical terms and describing science texts. *Literary and Linguistic Computing* 1, 2, 93–103.



## Acknowledgements

If I have seen farther it is by standing on the shoulder of giants  
Isaac Newton

On my working desk in the Language Centre at UST sits a penholder, on which is inscribed “Teachers help us grow”. The four-word sentence resonates inside me: indeed many of my teachers have helped me grow. I owe all of them my gratitude, but the one to whom I want to express my deepest thanks on this occasion is Professor Gregory James, who opened up my career path in 1993. Without his generosity, my research into semi-technical vocabulary would not have been possible, and without his efficiency and insistence this book would not have existed.

My sincere thanks also go to Reinhard Hartmann, my thesis supervisor, for his patience and guidance in the whole project; to Tom McArthur for extending the horizon of my awareness of word-formation and kindling all over again my interest in the subject matter; and to Tony Cowie, my external examiner, for his critical yet constructive and useful comments.

Acknowledgements are due to all my colleagues in the UST Language Centre, in particular, John Milton, who spent time discussing research problems with me and inspired me with the different software programs he has developed; Edward Li Siu-leung, who shared with me his research material and his knowledge of test design; Ian Smallwood, who introduced me to various computer programs he had designed; Richard Pemberton, Keith Tong Sai-tao and Lynne Flowerdew for their articles; and Ada Fan Shui-fun for her substantial contribution in preparing the manuscript.

I should also like to offer a special word of thanks to Giuseppe Mario Saccone, Jon Mills, Allen Ng Fai-lun, Eddie Lee and Terence Chow Chi-wai for their help at different stages of the project.

This book is dedicated to all my family members, in particular my elder sister, Florence Lam Kam-fong, and my younger brother, Aslan Lam Kam-chung, both of whom have given me valuable professional advice.

Jacqueline Lam Kam-mei



## Synopsis

Research and teaching are like the computer's central processing unit (CPU) and monitor: neither can serve well in its own right. Although there is an increasing amount of research in the field of vocabulary acquisition, not much has focused on how the nature of lexis used in textbooks affects students' comprehension of a particular text genre. This study is therefore devoted to examining a group of vocabulary items, termed 'semi-technical vocabulary', found in Computer Science (CS) related literature.

Chapter One presents the threefold aim of this study: to establish a lexical category in CS texts of 'semi-technical vocabulary'; to investigate whether this vocabulary adversely affects ESL students' understanding of such texts; and to assess the extent to which a corpus-based subject-specific glossary of semi-technical vocabulary is a more efficient consulting tool for students than a general learner's dictionary (GLD).

The experimental work throughout the study relates to students at the Hong Kong University of Science and Technology (UST), many of whom have problems in interpreting texts related to science and technology. Particular problems in the area of vocabulary are addressed. First, such words cause confusion to learners when they have a special meaning in scientific or technical contexts; second, specialists in the scientific or technical field may assume that students would already have understood this category of words, and hence seek to transfer complex concepts to the learners without realising the danger that learners may be totally ignorant of the extended meaning of a particular word in a designated ST context; third, the assumption of an identical interpretation of this category of words conceived both by the teachers and students may cause miscommunication between the two parties when a concept in science and technology is concerned; fourth, because semi-technical vocabulary is not a focus of consideration in dictionary arrangement, learners often find it difficult to obtain the accurate meanings of the words in ST contexts.

Chapter Two considers the criteria by means of which a vocabulary item can be classified as semi-technical. Two broad lexical categories are developed that will sufficiently describe the vocabulary items selected for the purpose of this study: rhetorical (covering *allusion*, *figurative usage*, *neologism*, *semantic change* and *word play*) and morphological (covering *derivation*, *compounding*, *blending*, *abbreviation* and *phrasal-verb for-*

mation). These are based on a typology adapted from Tom McArthur's *Oxford companion to the English language* (1992), taken together with three recent works on word formation: Laurie Bauer's *English word-formation* (1983), Randolph Quirk *et al.*'s *A comprehensive grammar of the English language* (1985), Sidney Greenbaum's *The Oxford English grammar* (1996) and such other works as the *UST Computer Science corpus* (1994), *Running Word 6 for Windows* (1994), *The new hacker's dictionary* (1993) and *The jargon file (Version 4.0.0: 24 July 1996)*, together with various dictionaries of computer terms.

Chapter Three gives a detailed account of some significant linguistic features in the language of CS, including the manipulation of metaphor, lexical modification and allusions. These stylistic devices of semi-technical vocabulary create comprehension problems in CS contexts. A cohort of 120 students studying on the Language Enhancement Course at UST in the Fall semester of 1995 was chosen to examine the validity of such an assumption. The proposal of a glossary of semi-technical vocabulary, that helps students understand CS-related texts is justified.

Chapter Four describes the methodology used in both the pilot and the main study, involving three different survey formats and two verbal data collection techniques. These are: (1) a decontextualised-word-level glossing exercise containing vocabulary items of the type being investigated; (2) a questionnaire in a multiple-choice question test format; (3) a list of hands-on-computer tasks to be performed in the Microsoft Word 6.0 program; (4) a thinking-aloud technique; and (5) a retrospective interview. Altogether 525 subjects at UST were involved in the empirical studies (80 in the pilot study and 445 in the main study).

Suggestions are added for the revision of the methodology in the main study, including modifying the test and survey design, improving the research design and selecting the dictionaries. The propositions formulated in Chapter 1 are confirmed, and conclusions are drawn suggesting possible applications of the results and findings in two areas: implications for computer studies; and implications for ESP teaching and ESP lexicography. Since the study was conducted in 1994–5, when computers were still not widely used in Hong Kong secondary schools, it is suggested that: (1) the same types of study should be repeated not only among first-year undergraduates but also final-year secondary school students; (2) a wider selection of semi-technical vocabulary test items than in the UST CS corpus and the Microsoft Word Manual should be incorporated into the surveys; (3) data from such investigations should shed light on the

lexical features of texts in this subject area, and help in the compilation of useful study instruments, such as specialist glossaries.