

Foreword

This book revolves around two main themes. One is vocabulary assessment methods, the other vocabulary use research by means of corpus analysis and computational linguistics. The chapters are based on individual papers which were presented either at a workshop at Stockholm University in May 2010, or at a thematic panel at the 20th Eurosla Conference in Reggio Emilia in September 2010. We felt that these conference contributions offered some new insights into L2 vocabulary research and consequently decided to compile them into a book that would present recent L2 vocabulary research and suggest some new directions in the field.

Different ways of assessing vocabulary reflect different conceptualizations of vocabulary knowledge. Vocabulary knowledge can be viewed as the number of words a person knows (hence, there are tests of vocabulary size, e.g. Nation & Beglar, 2007), the amount of information a person has about a particular word (deep knowledge tests measure how well certain words are known, e.g. Wesche & Paribakht, 1996), how a word associates with other words (e.g. Read, 1993), and the speed with which words are retrieved (Laufer & Nation, 2001). Lexical richness in free production has been measured by lexical profiles (e.g. Laufer & Nation, 1995; Bardel, Gudmundson & Lindqvist, 2012). Some of the chapters in the book discuss problems of these measurement methods and make suggestions for refinements and additions (Cobb; Gyllstad; Lindqvist et al.).

The introduction of language corpora, corpus analysis techniques and other computer analyses into second language research has made it possible to conduct studies on sizeable and varied samples of spontaneous linguistic productions. Cross-corpora comparisons and new types of analyses can be performed that provide new insights into lexical knowledge and its development in a second language. Some of the chapters of the book reflect these developments in lexical research. These chapters analyze the vocabulary found in learners' performance in speaking (Lindqvist et al.) or in writing (Levitzky-Aviad & Laufer; Tono).

Besides being concerned with these two overarching themes, the chapters also focus on a number of central issues in vocabulary research. One such issue is the role of word frequency, which is a recurrent factor when measuring lexical richness and is discussed from different points of view in some of the chapters (Cobb; Levitzky-Aviad & Laufer; Lindqvist et al.).

Another central issue is the relationship between knowledge of single words and multi-word units, which is addressed in detail by Henriksen, who sees collocational knowledge as part of communicative competence. Even very advanced learners seem to have difficulty with mastering this kind of knowledge fully, as Levitzky-Aviad and Laufer found. Their data shows in fact that students improved over time as far as measures of single words were concerned, but not with respect to multi-word units. Knowledge of multi-word units is normally considered to be indicative of deep knowledge, a construct that is discussed thoroughly in Gyllstad's chapter.

Yet another fundamental theme in vocabulary acquisition research pertains to the differences between learning and using oral and written vocabulary. The studies in this book examine data from written and spoken language, some focussing on production, some on comprehension. The differences in lexical sophistication between spoken and written modes are discussed by Lindqvist et al. and by Milton. Milton also points out that the correlations between vocabulary size scores and listening skills are generally weaker than the correlations with the written skills of reading and writing, and suggests some possible explanations for this difference. As regards written production, Tono's chapter addresses the important issue of vocabulary errors as correlates of proficiency level, and analyzes the kinds of errors characterizing different proficiency levels in academic essays.

Below is a brief summary of the chapters.

Henrik Gyllstad, in his chapter *Looking at L2 vocabulary knowledge dimensions from an assessment perspective – challenges and potential solutions*, notes how the recent upsurge of interest in L2 vocabulary and L2 vocabulary assessment has been followed by a situation where a large number of knowledge constructs are proposed and investigated. As Gyllstad points out, the development of competing definitions and perspectives is part and parcel of any flourishing academic domain, but still, it is a problem if constructs are given very different interpretations from study to study. Taking the fundamental constructs of vocabulary breadth and depth (Anderson & Freebody, 1981) as a point of departure, and drawing on some subsequent critical work on their viability and use, Gyllstad discusses some of the basic assumptions underlying these constructs. In particular, he emphasizes that empirical data on the learning and assessment of lexical items larger than single words, e.g. phrasal verbs, collocations and idioms, raise questions as to where to draw the line between breadth and depth. The author ends his paper by presenting suggestions for potential remedies.

Multi-word units are further discussed in Birgit Henriksen's contribution, *Research on L2 learners' collocational competence and development – a progress report*. According to previous studies, mastery of formulaic sequences – includ-

ing collocations – is a central aspect of communicative competence, which enables the speaker to process language both fluently and idiomatically and to fulfil basic communicative and social needs. In light of studies that show that collocational competence is acquired late and often not mastered very well by L2 language learners, Henriksen discusses the features of learners' collocational competence and the problems in its development. Different research approaches to investigating L2 learners' collocational development are discussed with a focus on the dynamic non-linear models of Larsen-Freeman (1997, 2006), which view language development as a complex process, allowing for individual variation resulting from language use conditions and the choices made by individual learners.

In his paper *Measuring the contribution of vocabulary knowledge to proficiency in the four skills*, James Milton examines how vocabulary knowledge relates to the ability to perform in the four language skills of reading, writing, listening and speaking in a foreign language. According to Milton, the recent insight that vocabulary is essential to language learning has led to systematic ways of describing and testing vocabulary knowledge, allowing researchers to model the growth of a foreign language lexicon across the various stages of language development. As pointed out by Milton, there is an increasing body of research supporting the idea that vocabulary knowledge and performance in a foreign language are linked and this chapter aims at making the nature and extent of this link clearer, investigating different aspects of word knowledge and different communicative skills. It has been acknowledged for some time that vocabulary knowledge is a good predictor of general proficiency in a foreign language. However, most research on this relationship has been conducted with measures of vocabulary size only, and within the realm of reading skills only. Strong correlations between receptive vocabulary size tests and reading comprehension tests have been found. A feature of recent work in vocabulary studies has been to try to investigate more fully the links between lexical knowledge and learner performance, and to investigate the scale of the contribution which vocabulary, in all its dimensions, can make to a variety of communicative skills in foreign language performance. Milton concludes that the studies he reviews show a moderate to strong relationship between vocabulary measures and the ability to read, write, listen, and it seems also speak, in the foreign language.

The following chapter, *Frequency 2.0: Incorporating homoforms and multiword units in pedagogical frequency lists*, is written by Tom Cobb, who developed the French version of the Lexical Frequency Profile (LFP) and the LFP tool towards new technical solutions. As Cobb remarks, a condition for the survival of the data-driven approach to language learning is the development of language corpora and accessible software tools that make close language inspection feasible

in language learning contexts. The growing acceptance of frequency as a decisive factor for learning has given further support to the LFP method. However, Cobb argues that the data-driven approach must now take on new challenges. First, larger corpora and techniques of analysis should reveal both the extent of homography in existing frequency lists, as well as the means for handling it. Second, larger corpora also reveal the existence of multiword units of such high frequency as to suggest their official inclusion in standard lists. Cobb's chapter reports on how ways forward on both these fronts are developed technically in order to obtain more fine-grained LFP analyses.

In the next paper, *A new approach to measuring lexical sophistication in L2 oral production*, Christina Lindqvist, Anna Gudmundson and Camilla Bardel also discuss the frequency-based perspective to vocabulary acquisition. The authors describe the elaboration of a method designed to measure lexical sophistication in spoken French and Italian as second languages, the Lexical Oral Production Profile (LOPP). The method was developed in a series of studies on Swedish learners' oral production of the two languages. In the first version, the method relied purely on frequency, and measured the proportion of high-frequency vs. low-frequency words, very much in line with the LFP (Laufer & Nation 1995). In the second version of the method, factors other than frequency were taken into account: thematic vocabulary and cognate words. By integrating these aspects into the lexical profiler, it no longer only relies on the distinction between low-frequency and high-frequency words, but on a division between basic and advanced words. The new version of the method proved to provide more homogeneous results within groups than the previous one. The authors further discuss lexical profiling in general and, in a similar vein as Cobb, propose to include additional information in frequency lists such as multiword units and homographs. A further issue brought up is how to treat instances of non target-like use in lexical profiling. Finally, possible areas of use of the tool are discussed. Apart from using it for research purposes, it can for example be used in a pedagogical setting.

In *Lexical properties in the writing of foreign language learners over eight years of study: single words and collocations* Tami Levitzky-Aviad and Batia Laufer used the Israeli Corpus of Learner English to examine the progress in vocabulary use over 8 years of learning. They used the LFP to analyse 290 written samples (200 words each) of learners of four proficiency levels that corresponded to grades 6, 9, 11 and university. The compositions of these proficiency groups were compared on lexical richness - the proportion of frequent to non frequent vocabulary, lexical variation - type-token ratio and the number of multi-word units - habitually occurring lexical combinations characterized by restricted co-occurrence of elements. They also tested learners' productive vocabulary knowledge by

a vocabulary size test. Results showed a significant improvement in the active *knowledge* scores across all stages of English learning, but not in the *use* of vocabulary. A significant increase in the use of infrequent vocabulary and in lexical diversity was found only with university students. As for the number of multi-word units in the compositions, no significant differences were found between the proficiency groups. In light of this limited progress, recommendations are made for further investigations into the effect of different pedagogical approaches to the teaching of foreign language vocabulary.

The last chapter is Yukio Tono's study *Automatic extraction of L2 criterial lexico-grammatical features across pseudo-longitudinal learner corpora: using edit distance and variability-based neighbour clustering*. The aim of this study is to identify lexico-grammatical features of English as L2, which could serve as criteria for distinguishing different proficiency levels. A corpus of Japanese-speaking learners of English, the JEFLL Corpus, was created, consisting of spontaneous, timed, in-class essays by more than 10,000 participants. The data was gathered cross-sectionally from school year 7 to 12. In order to extract the criterial features across proficiency levels, the whole JEFLL Corpus was corrected by a teacher, and two sets of data were prepared: the original vs. the corrected versions. They were aligned at sentence level and compared against each other digitally; in this way the differences within sentences were extracted automatically. Three different types of error candidates were identified: (i) omission, (ii) addition, and (iii) misformation. The data shows that the errors related to verbs serve as more salient criterial features for the early stages of learning while lexical choice errors characterize the later stages. The results also indicate that there is a clear pattern of development in how nouns and verbs are modified by elements such as modals, prepositional phrases and subordinate clauses. Methodological and pedagogical implications of the study are discussed.

We would like to express our gratitude to the participants at the two meetings on vocabulary acquisition held in Stockholm and Reggio Emilia. We also thank the reviewers of this volume, as well as the series editor Gabriele Pallotti, the editorial assistant, Fabiana Rosi, and the language editor Françoise Thornton-Smith, who proofread the final version of the manuscript.

February 2013

Camilla Bardel, Stockholm
Christina Lindqvist, Uppsala
Batia Laufer, Haifa

References

- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.
- Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of lexical sophistication in advanced learners' oral production. Vocabulary acquisition and use in L2 French and Italian. *Studies in Second Language Acquisition*, 34(2), 1-22.
- Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 141-165.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590-619.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-329.
- Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: are they related? In S. Foster-Cohen & A. Nizegorodcew (Eds.), *EUROSLA Yearbook Yearbook 1* (pp. 7-28). Amsterdam: Benjamins.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371.
- Wesche, M., & Paribakht, S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13-40.