

**EUROSLA MONOGRAPHS SERIES 2**

**L2 VOCABULARY ACQUISITION,  
KNOWLEDGE AND USE  
new perspectives on  
assessment and corpus analysis**

**EDITED BY**

**CAMILLA BARDEL**

University of Stockholm

**CHRISTINA LINDQUIST**

University of Uppsala

**BATIA LAUFER**

University of Haifa

**EUROPEAN SECOND LANGUAGE ASSOCIATION 2013**

## **Eurosla monographs**

Eurosla publishes a monographs series, available open access on the association's website. The series includes both single-author and edited volumes on any aspect of second language acquisition. Manuscripts are evaluated with a double blind peer review system to ensure that they meet the highest qualitative standards.

### **Editors**

Gabriele Pallotti (Series editor),  
University of Modena and Reggio Emilia

Fabiana Rosi (Assistant editor),  
University of Modena and Reggio Emilia

© The Authors 2013  
Published under the Creative Commons  
“Attribution Non-Commercial No Derivatives 3.0” license  
ISBN 978-1-300-88407-1

First published by Eurosla, 2013  
Graphic design and layout: Pia 't Lam

An online version of this volume can be downloaded from [eurosla.org](http://eurosla.org)

## *Table of contents*

Foreword <i>Camilla Bardel, Christina Lindqvist and Batia Laufer</i>	5
Looking at L2 vocabulary knowledge dimensions from an assessment perspective – challenges and potential solutions <i>Henrik Gyllstad</i>	11
Research on L2 learners' collocational competence and development – a progress report <i>Birgit Henriksen</i>	29
Measuring the contribution of vocabulary knowledge to proficiency in the four skills <i>James Milton</i>	57
Frequency 2.0: Incorporating homoforms and multiword units in pedagogical frequency lists <i>Tom Cobb</i>	79
A new approach to measuring lexical sophistication in L2 oral production <i>Christina Lindqvist, Anna Gudmundson and Camilla Bardel</i>	109
Lexical properties in the writing of foreign language learners over eight years of study: single words and collocations <i>Tami Levitzky-Aviad and Batia Laufer</i>	127
Automatic extraction of L2 criterial lexico-grammatical features across pseudo-longitudinal learner corpora: using edit distance and variability-based neighbour clustering <i>Yukio Tono</i>	149
About the authors	177

## **Editorial board**

Cecilia Andorno, University of Pavia  
Dalila Ayoun, University of Arizona  
Camilla Bardel, Stockholm University  
Alessandro Benati, University of Greenwich  
Sandra Benazzo, Université Lille 3  
Giuliano Bernini, University of Bergamo  
Camilla Bettoni, University of Verona  
Marina Chini, University of Pavia  
Jean-Marc Dewaele, Birkbeck College, UCL  
Anna Giacalone Ramat, University of Pavia  
Roger Gilabert, University of Barcelona  
Gisela Håkansson, Lund University  
Henriëtte Hendriks, University of Cambridge  
Martin Howard, University College Cork  
Gabriele Kasper, University of Hawai'i at Mānoa  
Judith Kormos, Lancaster University  
Folkert Kuiken, University of Amsterdam  
Maisa Martin, University of Jyväskylä  
James Milton, Swansea University  
John Norris, Georgetown University  
Lourdes Ortega, Georgetown University  
Simona Pekarek-Doehler, Université de Neuchâtel  
Manfred Pienemann, University of Paderborn  
Leah Roberts, University of York  
Jason Rothman, University of Iowa  
Michael Sharwood Smith, Heriot-Watt University Edinburgh  
Nina Spada, University of Toronto  
Richard Towell, University of Salford  
Danijela Trenkic, University of York  
Ada Valentini, University of Bergamo  
Ineke Vedder, University of Amsterdam  
Christiane von Stutterheim, Heidelberg University  
Johannes Wagner, University of Southern Denmark

## *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 20<sup>th</sup> 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.

# *Looking at L2 vocabulary knowledge dimensions from an assessment perspective – challenges and potential solutions*

Henrik Gyllstad  
Lund University

The heightened interest in L2 vocabulary over the last two or three decades has brought with it a number of suggestions of how vocabulary knowledge should be modelled. From a testing and assessment perspective, this paper takes a closer look at some of these suggestions and attempts to tease out how terms like model, dimension and construct are used to describe different aspects of vocabulary knowledge, and how the terms relate to each other. Next, the two widely assumed dimensions of vocabulary breadth and depth are investigated in terms of their viability for testing purposes. The paper identifies several challenges in this regard, among others the questionable assumption that multi-word units like collocations naturally belong in the depth dimension, and problems that follow from the complex and often ill-defined nature of the depth dimension. Suggestions for remedies are provided.

## **1. Introduction**

Ever since Meara (1980) pointed out the then Cinderella-like status of vocabulary some three decades ago, the field of foreign and second language vocabulary (L2)<sup>1</sup> has seen a formidable explosion in terms of activity and the number of studies published. The dramatic yet welcome increase in research on vocabulary over the last 30 years has brought with it an increase also with regard to terminology. A striking example of the plethora of terms that may exist for a single concept, arguably some having more or less central meanings than others, can be seen in Wray's (2002) account of terms used to describe aspects of formulaicity, presented as Figure 1. As Wray points out, even though there are clear cases of conceptual duplication across the terms used, there are also cases of terms shared across different fields that do not refer to the same thing. Whether

---

1 Henceforth, the abbreviation L2 will be used to denote both a second and a foreign language.

this proliferation of terms relates to a parallel proliferation of constructs is a crucial issue for research on language testing and assessment.

**Figure 1.** Terms used to describe aspects of formulaicity (taken from Wray, 2002: 9).

amalgams – automatic – chunks – clichés – co-ordinate constructions – collocations – complex lexemes – composites – conventionalized forms – F[ixed] E[xpressions] including I[dioms] – fixed expressions – formulaic language – formulaic speech – formulas/formulae – fossilized forms – frozen metaphors – frozen phrases – gambits – gestalt – holistic – holophrases – idiomatic – idioms – irregular – lexical simplex – lexical(ized) phrases – lexicalized sentence stems – listemes – multiword items/units – multiword lexical phenomena – noncompositional – noncomputational – nonproductive – nonpropositional – petrifications – phrasemes – praxons – preassembled speech – precoded conventionalized routines – prefabricated routines and patterns – ready-made expressions – ready-made utterances – recurring utterances – rote – routine formulae – schemata – semipreconstructed phrases that constitute single choices – sentence builders – set phrases – stable and familiar expressions with specialized subsenses – stereotyped phrases – stereotypes – stock utterances – synthetic – unanalyzed chunks of speech – unanalyzed multiword chunks - units

Cronbach and Mehl define a construct as “some postulated attribute of people, assumed to be reflected in test performance” (1955, p. 283). A person may at any time possess such an attribute, either fully or to some degree, or not possess it. A complementary definition of the term construct is supplied by Chapelle, who states that “a construct is a meaningful interpretation of observed behavior” (1998, p. 33). If applied to the field of vocabulary assessment, then a test-taker’s scores on a vocabulary test constitute the observed behaviour that is to be interpreted meaningfully, and by extension, the scores are assumed to indirectly reflect some kind of mental ability or knowledge, in this case knowledge about words. A construct is thus a form of knowledge or an ability that can be observed and/or measured, and as such it is of course essential to the scientific study of any kind, since it enables a scientific community to label knowledge and/or abilities, to define clearly what they are, to potentially break them down into several interrelated sub-abilities, and to relate them to other constructs. However, contention is no doubt part and parcel of any thriving academic discipline, and although the evolution of a construct often involves competing definitions and perspectives, it becomes problematic in the long run if constructs are not clearly and properly defined, and if some degree of homogeneity is not reached. In the worst case it could hinder a further understanding of the field.

In the remainder of this paper, I will first take a look at some of the central terminology used for describing knowledge and abilities in the field of L2 vocabulary acquisition, primarily from a testing and assessment perspective. I will discuss how the terminology is used, identify potential problems, and suggest remedies to these when possible. I will then discuss the origins and applications of the influential and widely-used dimensions of vocabulary breadth and depth, particularly in relation to some of the challenges that researchers face when using these for assessment purposes. In doing this, I will also propose remedies to overcome some of the more persistent challenges.

## **2. Central terminology used in research on L2 vocabulary acquisition and assessment – models, dimensions and constructs**

As was pointed out in the previous section, the heightened interest in L2 vocabulary has entailed an increase in the number of constructs that have been proposed and used. Recent examples connected to vocabulary size tests, i.e. tests of the number of words in a language for which a learner has at least a basic form-meaning knowledge, are written receptive vocabulary size (Meara & Buxton, 1987), controlled productive vocabulary size (Laufer & Nation, 1999) and aural receptive vocabulary size (Milton & Hopkins, 2006). These three examples have a parent construct ('vocabulary size') as a common denominator, but are more specific by adding terms that narrow the construct down even further, e.g. 'receptive', 'productive', 'aural', and 'written'. This is obviously a good thing, as the added specificity makes it clearer what kind of knowledge is targeted. Interestingly, even though the notion of construct is arguably very central when describing vocabulary knowledge and its assessment, the term itself is not always used specifically in the literature. Instead, the term *dimension* often appears when L2 vocabulary researchers discuss acquisition and assessment matters. Here are some examples of 'dimensions' proposed in the literature on L2 vocabulary acquisition.

- Henriksen (1999), in describing a model of lexical development:
  - a) partial to precise knowledge, b) depth of knowledge, and c) receptive to productive use ability.
- Meara (2005), in describing a model of lexical competence/skill:
  - a) vocabulary size, b) vocabulary organization, and c) vocabulary accessibility.
- Daller et al. (2007), in describing a learner's vocabulary knowledge in "lexical space":
  - a) lexical breadth, b) lexical depth, and c) lexical fluency.

The first thing to note about the three proposals is that they all assume three dimensions, perhaps either true to a geometrical definition of space assuming length, breadth and depth, or simply giving support to the proverb that says that all good things come in threes. As to the first dimension (a) of the three models, it could be seen to deal with the same underlying process, namely the building of a repository of vocabulary items. What is characteristic of this dimension is that it has more to do with quantity than quality. Learners are shown to know  $x$  number of words, but this knowledge is minimally seen as a basic form-meaning mapping. Meara's (2005) *vocabulary size* and Daller et al.'s (2007) *lexical breadth* are very similar in this sense, whereas my understanding of Henriksen's (1999) *partial to precise knowledge* dimension is that she refers to the development of individual word knowledge, and that she emphasizes that the acquisition process is not an all-or-nothing activity. There are differences among authors as regards the second dimension (b), too. Daller et al. see *lexical depth* largely from a word knowledge framework perspective. Based on Nation's (2001) (see Table 2) descriptive approach to what aspects are involved in knowing a word, depth is seen as those aspects that go beyond the basic form-meaning mapping, e.g. concepts and referents, associations, collocations and constraints on use. Meara's second dimension is called *vocabulary organisation*, and it is conceptually different to that of Daller et al. Meara envisages vocabulary organisation as the structured, lexical network that makes up a learner's mental lexicon. The focus here is on the links between words in this network and on how, from a more holistic perspective, they can inform us about the network as a whole. The fundamental difference between these first two approaches will be further discussed later on in this chapter. Henriksen's dimension, called *depth of knowledge*, may sound closer to that of Daller et al., but in fact she discusses it more in terms of network building in line with Meara's conception of vocabulary organisation. When it comes to the third dimension (c), the versions proposed by Daller et al. and Meara are conceptually close. The former call it *lexical fluency* and state that it is intended to define "how readily and automatically a learner is able to use the words they know and the information they have on the use of these words" (Daller et al., 2007, p. 8). This may involve the speed and accuracy with which word forms can be recognised receptively or retrieved for expressing targeted meanings when speaking or writing (productive vocabulary). Meara's version, called *vocabulary accessibility*, is said to have to do with "how easily you can manipulate the words you know" (Meara, 2005, p. 271), which is likely to imply both receptive and productive aspects, even though Meara's development of tests of this dimension has focused largely on receptive recognition skills. Henriksen's version is called *receptive to productive use ability*, which is argued to be a continuum, describing "levels of access or use ability" (1999, p. 314). Thus, there is a clear conceptual overlap between the three dif-

ferent versions, but it is also evident that the authors describe these dimensions in different ways and propose different ways to operationalise them.

The use of the term *dimension* raises the question as to what the relation is between this term and the term *construct*. It seems that in some cases in the literature construct and dimension are used more or less synonymously, whereas in other cases they are used hierarchically in a hyponymic relation, with dimension as a hypernym and construct as its hyponym. There are also cases of the converse relation, for example in Henriksen (1999), where construct is the superordinate (hypernym) term and dimension the subordinate (hyponym). Another term that is used in this context is *model*. Hierarchically, a model can be seen as a set of propositions that clarify how different constructs relate to each other. Meara (2005) talks about his three dimensions as being part of a model of vocabulary skills, while Henriksen (1999) proposes a model of lexical competence. Daller et al. (2007) do not use the term model when discussing their multi-dimensional space, but it is interesting to note that the name of the volume in which their text is published is called *Modelling and Assessing Vocabulary Knowledge*. The terms *model*, *dimension* and *construct* might be seen as co-existing at different hierarchical levels, albeit with some restrictions. Thus, I would like to propose that a model may consist of several dimensions, which in turn may comprise various constructs. A dimension can also be a construct, so long as type of knowledge or ability referred to is clearly defined – and by extension – measurable through some sort of test or assessment. If it is not, then the use of dimension rather than construct is more suitable. Furthermore, a dimension can consist of several constructs, just as a construct in principle can be divided into two or more ‘sub-constructs’. An example of this would be the dimension of vocabulary size, which can also be said to be a construct. In order to accommodate more detailed descriptions of vocabulary knowledge, e.g. *aural receptive vocabulary size* (Milton & Hopkins, 2006) or *controlled productive vocabulary size* (Laufer & Nation, 1999), it is possible to treat these as two sub-constructs within the construct (and dimension) of vocabulary size. From an assessment perspective, researchers ought to define constructs with precision. One way of doing this is by following Bachman’s (1990, p. 40-45) three-stage analysis:

- a. the construct needs to be defined theoretically;
- b. the construct needs to be defined operationally;
- c. procedures must be established for the quantification of observations.

The theoretical definition (a) is a specification of the relevant characteristics of the ability we intend to measure, and its distinction from other similar constructs. If there are several subcomponents to a construct, then the interrela-

tions between these must be specified. When it comes to the operational definition of the construct (b), this process involves attempts to make the construct observable. To a great extent, the theoretical definition will govern what options are available. For example, the theoretical definition of the construct ‘listening comprehension’ suggests an operationalisation as a task in which information must be decoded aurally in some fashion. With respect to the third stage (c), our measurement should be quantified on a scale. If applied to vocabulary depth (see the section below), with many subcomponents argued to be part of this construct, it is then very important to try to pin down how they relate to each other. To the best of my knowledge, this has not been done. On a theoretical level, Schmitt (2010b) has intuitively hypothesized how the different word knowledge aspects of Nation’s (2001) framework (see Table 2) relate to each other developmentally, but these hypotheses need to be empirically tested.

Having discussed the use of terminology in L2 vocabulary knowledge modelling, I will now turn to discussing the viability of two of the most influential dimensions in the field, vocabulary breadth and vocabulary depth, in order to see if they can be treated as constructs.

### **3. Vocabulary breadth and vocabulary depth: two influential dimensions and some inherent issues and challenges**

#### *3.1. The definitions of vocabulary breadth and depth*

Two of the most prominent dimensions used in L2 vocabulary research are ‘vocabulary breadth’ and ‘vocabulary depth’. As was made clear in the previous section, competing terms exist (e.g. ‘size’ instead of ‘breadth’), but the breadth and depth terminology can be traced back to a paper by Anderson and Freebody (1981), where breadth and depth are referred to as “aspects”. For now, this term will be used as in the authors’ original wording. I will later come back to how it relates to dimension and construct. Anderson and Freebody use the two aspects in a discussion about the role of vocabulary knowledge in reading comprehension, and they state clearly at the beginning of their paper that what they are interested in is “knowledge of word meanings” (1981, p. 77). This is how they define the two aspects (Anderson & Freebody, 1981, pp. 92-93)

It is useful to distinguish between two aspects of an individual’s vocabulary knowledge. The first may be called “breadth” of knowledge, by which we mean the number of words for which the person knows at least some of the significant aspects of meaning. ... [There] is a second dimension of vocabulary knowledge, namely the quality or “depth” of understanding. We shall



assume that, for most purposes, a person has a sufficiently deep understanding of a word if it conveys to him or her all of the distinctions that would be understood by an ordinary adult under normal circumstances.

These two aspects of vocabulary knowledge have indeed been influential and widely used. Not surprisingly, though, they have also been the subject of some criticism.

Firstly, as was pointed out by Read in his account of the term depth (2004), Anderson and Freebody's definitions leave us with a number of unclear terms. For example, in relation to "depth", it is not clear what is meant by "distinctions". Also, it raises the question as to what "an ordinary adult" is and what "normal circumstances" are. My own reading of Anderson and Freebody (1981) is that what they mean by distinctions when outlining the depth aspect is in effect meaning distinctions. This is arguably clear in the passage following the one where breadth and depth are initially defined (Anderson & Freebody, 1981, p. 93):

[...] the meaning a young child has for a word is likely to be more global, less differentiated than that of an older person. With increasing age, the child makes more and more of the adult distinctions.

The interpretation that the term "distinctions" refers to meaning distinctions is furthermore strengthened by a later passage, where a study by Gentner (1975) is reported. In this study, children were asked to act out, with the help of dolls, transactions based on given directions involving verbs like *buy*, *sell*, *spend*, *give* and *take*. The children acted out *buy* and *sell* as if they were essentially *take* and *give*, thus disregarding the money transfer element that is inherent in the meaning of the former verbs. It could thus be argued that what Anderson and Freebody originally meant by vocabulary depth was the repertoire of meanings and subtle sense distinctions that a word can convey. However, in Read's (2004)

**Table 1.** The application of the term depth in L2 vocabulary acquisition research (based on Read, 2004: 211-212).

1. Precision of meaning	(the difference between having a limited, vague idea of what a word means and having a much more elaborated and specific knowledge of its meaning)
2. Comprehensive word knowledge	(knowledge of a word, not only its semantic features but also orthographic, phonological, morphological, syntactic, collocational and pragmatic characteristics)
3. Network knowledge	(the incorporation of the word into a lexical network in the mental lexicon, together with the ability to link it to – and distinguish it from – related words)

account of how the term depth had been operationalised up to the early 2000s, there are three applications of the term. The additional two are seen as points 2 and 3 in Table 1.

It is clear from the above descriptions that it is only the first application called ‘Precision of meaning’ that is consistent with how Anderson and Freebody (1981) originally defined depth of word knowledge. The second operationalisation outlined by Read is that of comprehensive word knowledge. Here, as the name implies, a sizeable number of aspects are involved in knowing a word. One of the most recent and influential descriptions of such aspects is that of Nation (2001), shown here as Table 2. It is beyond the scope of this paper to go into a detailed description of Nation’s framework, but one thing is relevant. Typically, the aspects called ‘spoken’ and ‘written’ under the heading ‘Form’, together with ‘form and meaning’ under the heading ‘Meaning’ are seen as breadth aspects, whereas the remaining ones in the table are usually considered depth aspects. This means that knowledge of word parts, word associations, grammatical functions and collocations are usually considered depth of word knowledge aspects, an assumption I will return to later in this chapter.

**Table 2.** Description of “what is involved in knowing a word”, from Nation (2001: 27).

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concepts and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use (register, frequency)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

R = receptive knowledge, P = productive knowledge

The third operationalisation according to Read is network knowledge. The assumption behind network knowledge is that newly learned words are stored in a network of already known items. One of the main proponents of this interpretation is Paul Meara and associates (see e.g. Meara & Wolter, 2004; Wolter, 2005; Meara, 2006), but Henriksen subscribes to this view as well, as we saw earlier in this chapter.

### 3.2. *Critical views of breadth and depth*

A point of criticism that has been levelled at the use of breadth and depth has to do with their being fundamentally different constructs, and thus not really comparable. For example, Meara and Wolter (2004) have argued that vocabulary breadth, or vocabulary size, as they prefer to call it, is a construct that is a measure of a learner's entire vocabulary, since scores on a particular number of words are extrapolated to give an indication of an overall size score, given that the selection of test items is valid. As such, vocabulary size is not a characteristic of individual words. Vocabulary depth, on the other hand, is typically seen as a characteristic of individual words, where extrapolation is not possible, or at least very difficult.

Even though vocabulary breadth (or vocabulary size) is not without its problems as a construct, it has desirable measurement characteristics. With its ratio scale, assessment scores start at zero and range up to thousands, even tens of thousands for advanced learners of a language. One of the inherent problems with vocabulary size, however, is linked to the old question of what a word is. In order to try to come up with estimates of someone's vocabulary size, it is important to decide and state clearly if the unit of counting in word frequency lists is word form, lemma or word family. Of course, except perhaps for beginner learners, it normally makes sense to work with lemmatized lists. Once learners have understood the inflectional system of a language, especially for receptive knowledge, they can fairly straightforwardly link different forms of a verb (*play, plays, playing*) or a noun (*house, houses*) together, at least when it comes to non-morphologically complex languages like English. Another approach to word frequency lists is to use the concept of word families. Word families are normally defined as "a headword, its inflected forms, and its closely related derived forms" (Nation 2001, p. 8). Even though it makes some sense to use word families from a learning burden point of view, it is questionable to assume that once a member of a word family is known, all the other members will be known too, perhaps without ever having seen some of them. Bogaards (2001) has rightly warned against this assumption (see also Cobb & Horst, 2004; Schmitt & Zimmerman, 2002), lamenting the fact that no empirical evidence has been presented to properly support its validity. Bogaards uses this example in his criticism, arguing that the following uses of the form *level*, as a consequence, should then not be problematic to L2 learners in terms of understanding (2001, p. 322-323):

- (a) a high level of radiation
- (b) on a level with
- (c) a level teaspoon
- (d) have a level head
- (e) to level a part of the town
- (f) death is a leveler
- (g) a leveling staff
- (h) an unlevel surface

It is clear that the polysemy and the derivational patterns of the form *level*, as illustrated in (a) – (h) above, may still pose a problem to learners of English, just like Bogaards implies. However, it should be noted that it might be the case that receptive understanding is still easier than productive knowledge in this regard. Thus, understanding the concept of *an unlevel surface*, in the sense that the prefix *un-* negates the adjective *level* in the context of surfaces, is arguably more straightforward than being able to produce a derivative word form expressing that same meaning. For example, how should a learner know which prefix to use for negating *level* from the range of alternatives, for example *in-*, *dis-*, *non-* or *un-*?

### 3.3. Two specific challenges to the viability of breadth and depth

In addition to the points of criticism accounted for above, there are two further challenges to the constructs of vocabulary breadth/size and depth, namely:

- a) the ubiquity of lexical items larger than one single orthographic word,
- b) the multi-faceted nature of the depth construct.

The first challenge is the ubiquity of lexical items larger than one single orthographic word. Below, a number of examples of such items, all part of the vocabulary of English, are juxtaposed with a single orthographic word.

<i>break</i>	single orthographic word
<i>break up</i>	phrasal verb
<i>lunch break</i>	compound noun
<i>break a record</i>	collocation
<i>break a leg</i>	idiom

The first three examples should be fairly uncontroversial, but the difference between a collocation and an idiom is perhaps not so straightforward. In this analysis, the sequence *break a leg* is an idiom since it is not possible to understand its meaning by adding up the meanings of the individual components,

i.e. *break a leg* is non-compositional. However, this sequence can also evoke a more literal reading, to denote the fracture of a bone that someone might suffer in an accident. In this reading, the sequence would be what Howarth (1996) refers to as a free combination. Likewise, the sequence *break a record* has two possible readings, too. One of them denotes the more literal process of someone destroying a vinyl record, as played on turntables. This would then also be called a free combination. However, the other reading would be called a collocation, since one of the components (words) of the sequence is used in a figurative, de-lexical, or technical sense, in this case the verb *break*. It stands to reason that lexical items like these are very important for second language learning. The point here is that some of them behave like single orthographic words – certainly the compound noun, but arguably the phrasal verb and perhaps the collocation and idiom as well. If this is the case, then they should be made part of the vocabulary inventory and included in a frequency list where single orthographic words would reside jointly with multi-word items (see Cobb, this volume and Henriksen, this volume). As a case in point, Shin and Nation (2008) have presented an analysis, based on the 10-million-word spoken part of the British National Corpus (BNC), in which as many as 84 collocations occurred with such high frequency that they would make it into the top 1,000 single word types of the spoken corpus. It should be noted here that Shin and Nation’s use of the term collocation mainly resides in one of two traditions of collocation research, called the frequency-based tradition, the other being the phraseological tradition (see Nesselhauf, 2004; Gyllstad, 2007; Barfield & Gyllstad, 2009 for accounts of these). The 84 collocations of the first frequency band include for example *you know*, *I think*, and *come back*. Furthermore, as many as 224 collocations would make it into the second 1,000 word type band of the corpus (see Table 3). As argued by Shin and Nation (2008), a large number of collocations would qualify for inclusion in the most frequent single word bands, if no distinction was made between single words and collocations. This argument seriously challenges the construct of vocabulary size.

**Table 3.** The number of collocations that would potentially qualify into single word frequency bands of English (table taken from Shin & Nation, 2008: 345).

Collocations	84	224	259	324	387
		(308)*	(567)*	(891)*	(4698)*
Single word frequency bands	1 <sup>st</sup> 1000	2 <sup>nd</sup> 1000	3 <sup>rd</sup> 1000	4 <sup>th</sup> 1000	5 <sup>th</sup> 1000

\* The number in brackets shows the cumulative number of collocations.

If we accept the assumption that lexical items such as collocations are part of everyone's vocabulary, then we need to start thinking of ways of incorporating lexical items larger than single words into measures of vocabulary size. The reason why this has not yet been done is probably because it is fraught with all sorts of problems. It is very likely that the vocabulary size construct based on single orthographic words will maintain its validity for years to come because of its desirable measurement characteristics. However, attempts at creating measures of vocabulary size where the nature of word usage – as illustrated by Shin and Nation's study – is addressed should be well on their way (see e.g. Martinez & Schmitt, 2012, and chapter by Cobb, this volume).

Another consequence of this discussion is that it is not clear whether collocations and collocation knowledge should reside in the vocabulary depth construct. For many researchers who follow Nation's (2001) descriptive framework of word knowledge (see Table 2), aspects except for basic form and meaning knowledge are typically treated as depth components (see e.g. Read, 2000; Jiang, 2004; Milton, 2009; Schmitt, 2000, 2010a). In my own work on developing English collocation tests (Gyllstad, 2007, 2009), I have been reluctant to call my two test formats – COLLEX and COLLMATCH – depth tests. Both test formats are receptive recognition measures of verb + noun collocations such as *pay a visit*, *do justice* and *keep a diary*. The reason for my reluctance is that I have not seen any convincing arguments yet for why they should be measures of depth. True, if one subscribes to the idea that any test that measures either form knowledge or form-meaning knowledge of single words is a size test, and everything else is a depth test, then it follows that collocation tests would be depth tests. However, I think this is an over-simplification.

This is also clearly connected to the second major challenge to the dichotomy breadth/depth: the multi-faceted nature of the depth construct, as it is conventionally used. Typically, the following aspects of word knowledge are listed under the heading depth, in its comprehensive word knowledge interpretation:

- meaning knowledge beyond the most frequent, dictionary-based meaning of a word
- word associations
- collocations
- word parts
- grammatical functions

These aspects of depth are quite disparate, which makes the definition of depth as a single construct and its subsequent operationalisation very difficult. As Milton (2009) rightly points out, depth has not been sufficiently and unambiguously defined (Milton, 2009, p. 150):

The difficulties in measuring qualities, such as depth, start with the definitions of this quality. We lack clear, comprehensive and unambiguous definitions to work with and this challenges the validity of any test that might fall within this area. [...] Without a clear construct, it is impossible to create a test that can accurately measure a quality whatever that quality is.

I have two additional points to make here. First of all, the coining of depth as a dimension has been valuable in pushing the thinking and theorizing in the field forward. However, it only makes sense to call it a dimension; as a construct, it is arguably far too vague and elusive. Secondly, one important approach to ascertaining the viability of a construct is through empirical investigation, and the most straightforward way of doing this is through correlation studies. A considerable number of studies have indeed been carried out to investigate the relation between breadth and depth (e.g. Qian, 1999; Nurweni & Read, 1999; Vermeer, 2001; Meara & Wolter, 2004; Wolter, 2005; Gyllstad, 2007). Qian (1999) used the Vocabulary Levels Test (VLT) (Nation 2001) as a size measure and found correlations between scores on that test with scores on the Word Associates Test (WAT) (Read, 1993, 1998) as a depth measure at  $r = .82$ , based on data from 74 L1 Korean and L1 Chinese ESL college and university students, predominately 18-27 year-olds. Nurweni and Read (1999) administered both a receptive vocabulary size measure and a WAT format depth measure to 350 L1 Indonesian ESL first-year university students, and they observed a correlation of  $r = .62$  for the whole group. In a subsequent analysis, in which the 350 students were subdivided according to scores on a general proficiency exam, they observed a correlation of  $r = .81$  for high level students (10% of the whole group);  $r = .43$  for mid level students (42% of the whole group); and  $r = .18$  for low level students (48% of the whole group). Vermeer (2001), testing 50 L1 and L2 Dutch kindergarten 5-year-olds, arrived at correlations ranging between  $r = .70$  and  $.83$  between a receptive vocabulary size measure and an association task depth measure. Meara and Wolter (2004) found a modest level of correlation between scores on a test of overall vocabulary size and scores on a vocabulary depth test ( $r < .3$ ), based on data from 147 Japanese learners of English. This depth test, called V\_Links, is argued to be a test of lexical organisation, following the lexical network interpretation of depth (Read, 2004). The result was taken as support for the view that size and organisation are “more-or-less independent features of L2 lexicons” (Meara & Wolter, 2004, p. 93). Wolter (2005), putting different versions of V\_Links to the test, found similarly low, or even inverse (though not significant), correlations with vocabulary size. Wolter concludes that there is evidence to suggest that vocabulary organisation, as measured by V\_Links (versions 2.0 and 4.0), and vocabulary size may develop orthogonally (2005, p. 208).

On balance then, except for the studies by Meara and Wolter, breadth and depth seem to correlate highly with each other, which raises questions about their viability as independent constructs. Based on his own investigations of breadth and depth, Vermeer concluded that (2001, p. 222):

Breadth and depth are often considered opposites. It is a moot point whether this opposition is justified. Another assumption is that a deeper knowledge of words is the consequence of knowing more words, or that, conversely, the more words someone knows, the finer the networks and the deeper the word knowledge.

Vermeer's caveat is thus that one should not assume *a priori* that breadth and depth are poles.

In order to illustrate in detail some of the challenges implied by using size and depth empirically, I will briefly account for a study (taken from Gyllstad, 2007) which aimed at finding validation support for two tests of collocation, the aforementioned COLLEX and COLLMATCH tests. The purpose was to see whether the collocation tests gravitated more towards vocabulary size or vocabulary depth when correlated with tests widely assumed to be size and depth tests, respectively. Scores from 24 Swedish learners of English on five different tests were gathered. The learners ranged from upper secondary school students to third term university students. The five tests used are shown in Table 4. The analysis yielded very high correlations between the test scores from vocabulary size (VLT) and vocabulary depth (WAT) at  $r = .93$ . The collocation tests (COLLEX, COLLMATCH) correlated at  $r = .90$  with vocabulary size (VLT) and at  $r = .85-.90$  with the vocabulary depth measure (WAT).

**Table 4.** Tests used in a validation study investigating how collocation knowledge relates to the vocabulary size and depth constructs (based on Gyllstad, 2007).

Test	Brief description	Source
COLLEX	A 50-item test of receptive collocation knowledge	Gyllstad (2007)
COLLMATCH	A 100-item test of receptive collocation knowledge	Gyllstad (2007)
Vocabulary Levels Test (VLT)	Version 1, 150 items (vocabulary size)	Nation (2001); Schmitt (2000)
Word Associates Test (WAT)	A 320-item test (vocabulary depth)	Read (1998)
CAE Reading Comprehension Test	43 items	Cambridge ESOL Examination



The question is, what does all this tell us? The collocation tests correlated highly with vocabulary size and almost equally highly with vocabulary depth. At the same time, the size and depth measures in turn correlated highly with one another. A common way of interpreting high correlations is to assume that the variables that are involved are closely related or even the same thing. From a testing perspective, Norbert Schmitt (personal communication) has argued for the fact that every size test is in fact also a depth test. What he seems to mean by this is that for any given word in a size test, test-takers must have some sort of depth of word knowledge of that word in order to fulfill the test task. This presupposes, of course, a view of depth where word knowledge starts with a rather incomplete and partial level of knowledge, for example mere form recognition or very tentative and uncertain meaning knowledge. Most researchers, however, assume that basic form-meaning knowledge is part of the vocabulary breadth/size knowledge construct, and that depth is what comes beyond this basic knowledge.

An analysis that could shed light on the potential difference between the assumed constructs is multiple linear regression (see Bachman, 2004). It would for example be possible to try to estimate how much of the variation in a set of reading comprehension scores can be explained by vocabulary size scores. Then, as a second step, the variable of vocabulary depth would be entered into the regression model in order to ascertain whether the percentage of explained variance would increase. If that is the case, then vocabulary depth could be argued to bring an added, unique contribution to the variance in reading comprehension scores. As a case in point, Qian (1999) found that his measure of depth of vocabulary knowledge added a further 11% to the prediction of reading comprehension scores, over and above the prediction afforded by vocabulary size. A final remark that needs to be made here, though, is that we must look critically at the test instruments themselves. For example, in my own study (Gyllstad, 2007) and several of the studies reported above, including that of Qian (1999), a version of the Word Associates Test (WAT) (Read, 1993, 1998) was used. Some of the words featuring in the WAT are fairly low-frequency items, and vocabulary size is therefore suspected to have a considerable influence on test-takers' performance. A closer look at some of the words featured in the specific WAT test version used in Qian (1999) and Gyllstad (2007) confirms this. For example, target words like *ample*, *synthetic* (both 6K), and *fertile* (7K), together with associate words like *cautious* (5K) and *plentiful* (8K) are clearly not high-frequency words. This confounds the two variables and arguably explains at least part of the observed high correlations between vocabulary size and vocabulary depth scores.

#### 4. Concluding remarks

In this chapter, I have discussed the terminology used in modelling vocabulary knowledge, especially in relation to assessment purposes. In particular, the uses and referents of terms like *model*, *dimension* and *construct* have been addressed. Although a certain degree of terminological variation is bound to exist in all scientific disciplines, rigour and consensus are equally desirable. I have proposed that a distinction be made between dimension and construct, and that constructs must be defined clearly following procedures suggested by e.g. Bachman (1990). Furthermore, by taking a closer look at the two influential dimensions of vocabulary breadth and depth, I have argued that vocabulary depth has been valuable in furthering the thinking in the field, but its ill-defined, cover-all nature makes it inappropriate as a construct to be used in assessment procedures. I have also highlighted some of the inherent problems of using breadth and depth in vocabulary assessment, such as the ubiquitous existence of multi-word units and the question of their potential inclusion in the breadth dimension.

#### *Author's note*

I would like to thank two anonymous reviewers, the volume editors and the series editor for valuable comments and suggestions.

#### 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.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. (2004). *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Barfield, A., & Gyllstad, H. (2009). Introduction: Researching second language collocation knowledge and development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language - Multiple interpretations* (pp. 1-18). Basingstoke: Palgrave Macmillan.
- Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 23(3), 321-343.
- Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 15-38). Amsterdam: Benjamins.

- Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In L. Bachman & A. Cohen (eds.), *Interfaces between second language acquisition and language testing research* (pp. 32-70). Cambridge: Cambridge University Press.
- Cronbach, L.J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). Editors' introduction: Conventions, terminology and an overview of the book. In H. Daller, J. Milton, & J. Treffers-Daller (Eds), *Modelling and assessing vocabulary knowledge* (pp. 1-32). Cambridge: Cambridge University Press.
- Gentner, D. (1975). Evidence for the psychological reality of semantic components: The verbs of possession. In D. A. Norman, D.E. Rumelhart & the LNR research group (Eds), *Explorations in cognition* (pp. 211-246). San Francisco: Freeman.
- Gyllstad, H. (2007). *Testing English collocations: Developing tests for use with advanced Swedish learners*. PhD Thesis. Lund: Lund University. Available on-line at: <http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=599011&fileOId=2172422>
- Gyllstad, H. (2009). Designing and evaluating tests of receptive collocation knowledge: COLLEX and COLLMATCH. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language - Multiple interpretations* (pp. 153-170). Basingstoke: Palgrave Macmillan.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317.
- Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Lexicographica Series Maior 75. Tübingen: Max Niemeyer.
- Jiang, N. (2004). Semantic transfer and development in adult L2 vocabulary acquisition. In P. Bogaards & B. Laufer (eds.), *Vocabulary in a second language* (pp. 101-126). Amsterdam: Benjamins.
- Laufer, B. & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51.
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied linguistics*, 33(3), 299-320.
- Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics; Abstracts* 13, 221-246.
- Meara, P. (2005). Designing vocabulary tests for English, Spanish and other languages. In C. Butler, M. Gómez-González & S. Doval Suárez (Eds), *The dynamics of language use* (pp. 271-286). Amsterdam: Benjamins.
- Meara, P. (2006). Emergent properties of multilingual lexicons. *Applied Linguistics*, 27(4), 620-644.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142-154.
- Meara, P., & Wolter, B. (2004). V\_Links: Beyond vocabulary depth. In D. Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), *Angles on the English-speaking world 4* (pp. 85-96). Copenhagen: Museum Tusulanum Press.

- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *The Canadian Modern Language Review*, 63(1), 127-147.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nesselhauf, N. (2004). What are collocations? In D.J. Allerton, N. Nesselhauf, & P. Skandera (Eds.), *Phraseological units: Basic concepts and their application* (pp. 1-21). Basel: Schwabe.
- Nurweni, A. & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes*, 18(2), 161-175.
- Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56(2), 282-308.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371.
- Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (ed.), *Validation in language assessment* (pp. 41-60). Mahwah, NJ: Lawrence Erlbaum.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 209-227). Amsterdam: Benjamins.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2010a). *Researching vocabulary – A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, N. (2010b). Key Issues in teaching and learning vocabulary. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Del Mar Torreblanca-Lopéz, (Eds.) *Insights into non-native vocabulary teaching and learning* (pp. 28-40). Bristol: Multilingual Matters.
- Schmitt, N. & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly*, 36(2), 145-171.
- Shin, D. & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal*, 62(4), 339-348.
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(2), 217-235.
- Wolter, B. (2005). *V\_Links: A new approach to assessing depth of word knowledge*. Unpublished PhD thesis. University of Wales, Swansea.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

# *Research on L2 learners' collocational competence and development – a progress report*

Birgit Henriksen  
University of Copenhagen

The focus of this article is L2 collocational research. Collocations, i.e. frequently recurring two-to-three word syntagmatic units (e.g. *soft noise*, *tolerance for*) are a subset of formulaic sequences. Mastery of formulaic sequences has been described as a central aspect of communicative competence, enabling the native speaker to process language both fluently and idiomatically and to fulfil basic communicative needs. It has been argued that collocational competence is equally important for L2 learners. Nevertheless, this is a language phenomenon which is said to be acquired late and which is often not mastered very well by even fairly competent L2 language learners. This paper provides an extensive overview of L2 collocational research carried out from 1990 to 2011, presenting the main findings from a large number of studies in an attempt to discuss whether L2 learners do have problems in relation to developing collocational competence, and if so why. The second half of the paper focuses on the different approaches used in collocational research, looking at the specific challenges researchers may be faced with in relation to describing L2 collocational competence, use and development.

## **1. Introduction**

The seminal works by Pawley and Syder (1983), Nattinger and DeCarrico (1992) and Lewis (1993) have drawn language researchers' and teachers' attention to the frequency and importance of formulaic sequences (FSs), i.e. recurring lexical chunks in language use. A range of different types of FSs have been identified: idioms (*if life deals you with lemons make lemonade*), figurative expressions (*to freeze to the spot*), pragmatic formulas (*have a nice day*), discourse markers (*let me see now*), lexicalized sentence stems (*this means that...*), and collocations (*rough crossing*, *remotely clear*), which are the focus of this article. Mastery of FSs is a central aspect of communicative competence (Barfield & Gyllstad, 2009b; Nation, 2001; Schmitt, 2004; Wood, 2010; Wray, 2002), enabling the native speaker to process language both fluently and idiomatically (Pawley & Syder, 1983) and to fulfil basic communicative needs (Wray, 2002).

Moreover, memory and the ability to chunk language into units play an importance role in language use and learning (Ellis, 2001; 2003; 2005). Hoey (2005) has also argued for the facilitating processing effects in terms of lexical priming for recurrent lexical units.

Mastery of FSs is also important for L2 learners. During the last two decades, we have witnessed an increasing focus in SLA research and in second and foreign language teaching publications both on FSs in general and more specifically on collocations (e.g. Barfield & Gyllstad, 2009a; Granger & Meunier, 2008; Lewis, 2000; Schmitt, 2004; Wood, 2010). The central role of FSs in language knowledge and the benefits of mastering language chunks in relation to fluency and native-like selection are important reasons for focusing on formulaic language, including collocations (see Nation, 2001, pp. 317-318).

Collocations are frequently recurring two-to-three word syntagmatic units which can include both lexical and grammatical words, e.g. verb + noun (*pay tribute*), adjective + noun (*hot spice*), preposition + noun (*on guard*) and adjective + preposition (*immune to*). Many of the studies on collocations have shown that even high-level learners seem to experience problems in relation to using and developing L2 collocational knowledge (e.g. Arnaud & Savignon, 1997; Nesselhauf, 2005; Revier & Henriksen, 2006). Researchers wanting to explore L2 collocational knowledge, use and development may however also be faced with a number of serious challenges (Henriksen & Stenius Stæhr, 2009). The aim of this paper is to provide a progress report on L2 collocational research to see if we can find empirical support for the more general claim that collocations are a problem area for L2 language learners, and to discuss whether researchers are faced with specific challenges when describing L2 learners' collocational development and use.

A number of central issues taken up in the studies will be addressed: how can collocations be defined? Why do L1 and L2 learners need to develop collocational competence? Do L1 and L2 learners differ in their use and development of collocations? Is it problematic if L2 learners' knowledge and use of collocations differ from those of L1 users? Which types of collocations have been studied and which research instruments have been used? Can specific research challenges be identified? The final section will outline some of the more general issues raised by the collocational research reviewed, i.e. issues which should be taken into consideration in future studies.

## 2. Defining and identifying collocations

A key issue in collocational research is the question of defining and identifying collocations. It is generally agreed that collocations are a subset of FSs. Researchers have proposed various taxonomies which identify, distinguish and

classify different types of FSs, using a number of criteria (e.g. Boers & Lindstromberg, 2009; Koya, 2005). Nesselhauf (2005) discusses in detail different potential defining criteria, and Nation (2001) outlines 10 different scalar criteria: frequency of co-occurrence, adjacency, grammatical connectedness, grammatical structure, grammatical uniqueness, grammatical fossilization, collocational specialization, lexical fossilization, semantic opaqueness and uniqueness of meaning. Many researchers place FSs on a continuum with collocations as an intermediate category (for an alternative classification see Warren, 2005). Nattinger and DeCarrico (1992) outline three distinguishing criteria between idioms, collocations and free combinations: flexibility, compositionality and productivity. Cowie and Howarth (1996) argue that collocations can be distinguished from the other types of FSs by being characterized as institutionalized, memorized, restricted and semantically opaque units. Laufer and Waldman (2011) use the criteria of restricted co-occurrence and relative transparency of meaning. Howarth (1998, p. 24) stands out by focusing on the function of collocations, defining them as “combinations of words with a syntactic function as constituents of sentences (such as noun or prepositional phrases or verb and object constructions).”

An often quoted (e.g. Wray, 2002), but very illustrative example of a collocation is the adjective + noun unit *major catastrophe*. If we look at other possible options for adjectives found in a thesaurus, covering more or less the same semantic content as *major*, the following near-synonyms will often be listed: *big, large, great, huge, substantial, enormous, vast, gigantic, and colossal*. The Oxford collocations dictionary (Deuter, 2002) offers *big, great, and major* as preferred collocates, but none of the other conceivable adjectives. Many of these are potential options on the reference level, but are less appropriate on the pragmatic level of conventionalized, i.e. standard, language use. Other often cited contrastive examples are *strong coffee* vs. *powerful car* and *blonde hair* vs. *light paint*.

Two major traditions have been adopted in relation to identifying collocations (see Barfield & Gyllstad, 2009; Granger & Pacquot, 2008; Gyllstad, 2007; Nesselhauf, 2005). Firstly, the frequency-based view which identifies collocations on the basis of the probability of occurrence of their constituent words, often in large language corpora. Secondly, the phraseological view which is based on a syntactic and semantic analysis of the collocational unit, using some of the criteria mentioned above, such as degree of opacity, syntactic structure and substitutability of word elements. The advantage of using the corpus approach is that it employs objective criteria such as frequency, range and collocational span. However, a data-driven approach focuses on performance and not competence (Howarth, 1998) and disregards central questions of memory storage and language processing. By not including a semantic analysis, this procedure may lead to the identification of recurring lexical bundles that native speakers would not

classify as collocational unit, i.e. the chunks may have little psycholinguistic validity for the language users (e.g. *and the* and *of a*). On the other hand, the more subjective phraseological approach only identifies chunks with clear semantic relations between the constituents, and fails to report the actual frequency of use of the collocations. Some of these collocations may be fairly low in frequency and may therefore not constitute the most suitable targets for L2 learning and teaching (*judicial organ*, *ruggedly handsome*). Many researchers now apply both procedures, initially identifying the frequently occurring combinations in a large corpus through statistical measures (see Schmitt, 2010, p. 124-132 for a detailed presentation) and subsequently including and excluding specific combinations on the basis of an analysis of the word pairs identified. Using the computational approach as a starting point makes it possible to distinguish between collocations of varying frequency of use.

Following Gyllstad (2007), collocations can be viewed as both 1) lexical units, i.e. instances of language use which can be identified in written or spoken production and 2) associative mental links between words in language users' minds. A number of researchers have studied the psycholinguistic validity of FSs (e.g. Columbus, 2010; Durrant, 2008, 2009; Ellis, Simpson-Vlach, & Maynard, 2008), substantiating the fact that the different types of units identified in language data may indeed be seen as independently represented chunks in the mental lexicon. The question of psycholinguistic validation of FSs, including collocations, is important in relation to establishing useful inventories for the learning and teaching of collocations (see e.g. Durrant, 2009).

So far, it has been assumed that collocations are arbitrary structures, i.e. conventionalized combinatory options preferred by native speakers. However, as pointed out by Boers, Eyckmans, and Stengers (2006) and Boers and Lindstromberg (2009) this is not the case for all FSs, including collocations; in other words some collocations are motivated rather than arbitrary. Some collocations may be semantically motivated and can be traced back to specific etymological sources (e.g. *weeding out*), whereas others are formally motivated e.g. based on alliteration and assonance (*tell a tale*, *say a prayer*, *seek + solace*, *solitude*, *a solution* and *support*, *do + damage*, *a degree* and *a doctorate*). Arbitrary collocations can primarily be identified on the basis of frequency of occurrence in the language input, whereas the motivated collocations can also be identified on the basis of semantic or formal criteria via analysis (see also Walker, 2011). Based on a number of experiments (see again Boers et al., 2006 for an overview), Boers and his colleagues have argued that this difference between arbitrary and motivated collocations may influence the learnability of different types of collocations and thus the teaching approaches to be adopted. As discussed, one useful pathway to acquiring arbitrary collocations may be via rote learning approaches, whereas the motivated collocations may be learnt through the use of insightful, analytic



learning approaches, thus enabling L2 learners to benefit from the increased cognitive involvement connected with the processing of these collocations.

Different categories of FSs have been identified. Fewer attempts have been made to classify collocations systematically into different subcategories. As we have seen, some collocations are grammatical (sometimes referred to as 'colligations', see Gyllstad, 2007, p. 25), others lexical. Some collocations may differ in their degree of fixedness, transparency and arbitrariness. The degree of semantic transparency is a central variable used to distinguish between different types of collocations. If the learner knows the meaning of the two lexical items included, the collocation *major catastrophe* is fully transparent, and can therefore be understood through a process of decoding the two lexical elements in their literal sense. This is also the case with a verb + noun collocation like *take the money*. Other collocations are less straightforward, being either semi-transparent (*take a course*) or non-transparent (*take sides*). The meaning of the semi-transparent collocation is not decoded as easily as the literal counterpart, but is on the other hand not likely to be as salient as the non-transparent collocation which is idiomatic and cannot be understood on the basis of its constituents. Consequently, it has been argued that primarily the semi-transparent collocations will cause problems for language learners and should therefore be the main focus of L2 research and teaching (Nesselhauf, 2003; 2005). Many FSs have specific pragmatic functions as speech acts, discourse markers or conversational up-takers, playing an important role in social interaction. However this is not the case for most collocations which are composite units (Howarth, 1998) fulfilling a referential function (e.g. *major catastrophe, tell a tale*) as syntactic phrases. Some of the collocations are semantically motivated; others are formally motivated, whereas others again seem to be arbitrary combinations which have become the preferred lexical choice. Finally, many collocations are low in frequency; especially those that have high mutual semantic coherence (e.g. *preconceived notions*). All of these aspects may have an influence on the frequency, salience and learnability of the individual collocations.

### 3. L1 and L2 language users' need for collocational competence

It has been widely argued (e.g. Boers et al., 2006; Boers & Lindstromberg, 2009; Durrant, 2008; Lorenz, 1999) that collocational competence is important for language production and reception, enabling both the L1 and L2 language user: 1) to make idiomatic choices and come across as native-like; 2) to process language fluently under real-time conditions (Columbus, 2010; Ellis et al., 2008); 3) to establish 'islands of reliability' (Dechert, 1983; Raupach, 1984) which enable the language user to channel cognitive energy into more creative

production; 4) to disambiguate meaning of polysemous words, e.g. the verb *commit* in the following collocational contexts: *commit a crime*, *commit oneself*, *commit to memory*; and 5) to understand connotational meaning (what Sinclair, 2004 has described as semantic prosody), e.g. the fact that the verb *cause* is often associated with negative connotations as in *cause an accident*.

It has also been pointed out that FSs fulfil basic communicative and social needs (Wray, 2002). Since many collocations are primarily referential units and therefore do not have the same sociocultural function as many of the pragmatic phrases discussed by Nattinger and DeCarrico (1992), this may have an impact on the saliency and frequency of occurrence of many collocations, as well as L2 learners' motivation to acquire collocations compared to the more pragmatically oriented FSs. However, as argued by Warren (2005), collocations should be viewed as multi-word lexical items with form-meaning pairings which are associated with specific situations or phenomena and thus can be seen as lexical items that fulfil important communicative functions.

Through extensive exposure to L1 input in various contexts and co-texts, native speakers will have developed strong associative links (Ellis, 2001; 2003; 2005) between constituents in preferred collocations. In the on-going process of L1 acquisition, the native speaker will also have acquired knowledge of the meaning of the collocation and knowledge of the use restrictions of a particular collocation. As will be discussed below, the fact that the L2 learner does not have the same repeated and extended exposure to L1 input may, however, have important implications for L2 development and use of collocations.

#### 4. Main findings from the L2 studies

The results from the L2 studies reviewed here will be discussed in relation to the four main questions mentioned in the introduction. Due to the number of studies on collocations, this overview is, however, not exhaustive. For a discussion of some of the studies not included here see Koya (2005) (Japanese studies), Pei (2008) (Chinese studies), Fan (2009) and Laufer and Waldman (2011). Finally, it has not been possible to include newer articles published in 2012.

Two types of collocations have been the focus of investigation: lexical collocations, i.e. possible syntagmatic combinations between nouns, verbs, adjectives and adverbs (e.g. *foul play*, *take sides*, *truly happy*) and grammatical collocations, i.e. collocations which include prepositions (e.g. *hand over to*, *present with*, *important for*).

Many researchers have focused on lexical verb+noun collocations (e.g. Bahns & Eldaw, 1993; Barfield, 2003; Bonk, 2001; Chan & Liou, 2005; Eyckmans, 2009; Gyllstad, 2007; Howarth, 1996; Koya, 2005; Laufer &

Girsai, 2008; Laufer & Waldman, 2011; Peters, 2009; Revier & Henriksen, 2006), often looking at the restricted, semi-transparent collocations which are hypothesized to pose a special challenge for language learners (e.g. Nesselhauf, 2003, 2005; Revier, 2009). Another focus area has been the lexical adjective+noun combination (e.g. Jaén, 2007; Li & Schmitt, 2010; Peters, 2009; Siyanova & Schmitt, 2008). Some researchers delimit their scope of investigation to one type of collocation; others include two types, whereas others include a range of collocational structures in their studies (e.g. Barfield, 2009; Fan, 2009; Fayez-Hussein, 1990; Gitzaki, 1999; Hoffman & Lehmann, 2000; Groom, 2009; Keshavarz & Salimi, 2007; Prentice, 2010; Skrzypek, 2009; Ying & O'Neill, 2009).

#### *4.1. Do native and non-native speakers differ in their use of collocations?*

Many of the studies compare the productive use of collocations by native and non-native speakers (e.g. Bahns & Eldaw, 1993; Biskup, 1992; Fan, 2009; Farghal & Obiedat, 1995; Granger, 1998; Howarth, 1996, 1998; Lorenz, 1999; Nesselhauf, 2003). Not surprisingly, significant differences are found between the two groups (see Fan, 2009 and Laufer & Waldman, 2011 for an extensive overview of these studies). The NNSs often use fewer collocations (Laufer & Waldman, 2011) and a more restricted range of collocates (Fan, 2009), under-using types found in L1 data (Granger, 1998) and overusing other types (Lorenz, 1999). Fan also reports L1 use of a range of informal collocations, types of collocations not found in the L2 data. Other studies, however, document elements of native-like use, especially of highly frequent lexical units (e.g. Jiang, 2009). For example, 45% of the learner collocations analysed by Siyanova and Schmitt (2008) were central, appropriate collocations; a figure which matched the L1 data. These findings could be explained by the fact that we are dealing with high level learners' command of frequent and strongly associated word combinations. Generally, however, the studies tend to show that both second and foreign language learners do differ in their productive use of collocations compared to native speakers, both quantitatively in terms of the number and types of collocations used, as well as qualitatively in terms of error-free use. This is not surprising, however, and matches the general findings for other aspects of SLA, including the use of single-word lexical items.

Looking more closely at the quality of the collocations produced, infelicitous or erroneous use of collocational structures in L2 language use has been found (e.g. Laufer & Waldman, 2011; Nesselhauf, 2005; see again Pei, 2008 for a review of the Chinese studies). Many studies have reported the influence of L1 transfer on L2 collocational use (e.g. Bahns & Eldaw, 1993; Biskup, 1992; Fan, 2009; Granger, 1998; Jiang, 2009; Nesselhauf, 2003), showing that many

L2 learners tend to rely on using L1 translation equivalents (congruent collocations). Wang and Shaw (2008), however, have found that the tendency to transfer is dependent on the relative closeness perceived between the informants' NL and TL, and that other intralingual factors may also influence collocational use, a result which mirrors research findings on transfer for other aspects of SLA. Using acceptability judgement tests, Leśniewska and Witalisz (2007), could not find a clear indication of L1 influence for their advanced learners, i.e. the informants did not seem to reject or accept collocations on the basis of L1 congruence. It is argued that more advanced L2 learners may be able to function independently of the L1. The influence of the L1 will be taken up again below in connection with a discussion of the development and use of L2 collocations. L2 learners also underuse some collocations and seem to overuse other collocations compared to L1 users (e.g. Jiang, 2009), using the same collocations repeatedly in their production instead of choosing between various potential options (e.g. Lorenz, 1999). The favoured constructions could, in line with Hasselgren (1994), be described as 'collocational teddy bears'. In relation to underuse, Farghal and Obiedat (1995) found that L2 learners tend to use lexical simplification strategies, e.g. synonymy.

The study by Koya (2005) is one of the few studies which include both a receptive and productive test of collocational knowledge, documenting that the learners' receptive knowledge is broader than their productive knowledge. Laufer and Waldman (2011) also stress that L2 learners seem to experience problems in using collocations productively, not in their receptive understanding of the collocations. Again, these results are not surprising, and match the general SLA findings for other areas of language use, e.g. single-word vocabulary use.

#### *4.2. Is it problematic if L2 learners' knowledge and use of collocations differ from those of L1 users?*

Some of the studies have investigated the relationship between collocational knowledge and general language skills, reporting correlations between collocations and general proficiency as well as writing skills (Al-Zahrani, 1998) and between L2 learners' performance on collocational tests and cloze tests assessing general language proficiency (Keshavarz & Salimi, 2007). Similar results are found in some of the Chinese studies (Pei, 2008). Contrary to the results reported by Bahns and Eldaw (1993) and Koya (2005), Gyllstad (2007) found a correlation with vocabulary size. All these studies show that L2 learners' collocational knowledge is in some way related to language proficiency. One could therefore assume that lack of collocational knowledge and deviating use of collocations may be problematic for L2 learners. A correlation is, however, not the

same as a causal relation and a number of other important factors will also influence L2 learners' language performance.

As shown, L2 collocational use does deviate from L1 use, both quantitatively and qualitatively. Wray (2002, p. 74) has stressed the need of L2 learners to master FSs in order to identify with the target language community. However, if we view L2 use from a lingua franca perspective, native-like attainment and selection may not necessarily be the goal for L2 development compared for example to communicative efficiency. Howarth (1998) points out that infelicitous collocational choices made by L2 learners should in fact be viewed more positively as instances of risk-taking behaviour, arguing that these are indications that the interlanguage users are employing various communication strategies (e.g. experimentation, transfer, analogy and repetition) in order to cope communicatively.

The use of FSs, including collocations, is very genre-specific. Mastery of collocations may be a hallmark of certain types of academic writing which emphasize clarity, precision and lack of ambiguity (Howarth, 1998). If, as argued, collocations function as central composite syntactic units for clause level production, lack of collocational knowledge may be expected to have a negative effect on L2 performance not just productively for the L2 learner, but also receptively for the receiver, if central referential units are misunderstood. Apart from leading to unfortunate misunderstandings, advanced non-native speakers' collocational deviations may at least signal a lack of academic expertise. Moreover, the study by Millar (2011) has documented that malformed L2 collocations, both in terms of lexical misselection of a constituent and misformation of the collocation, lead to an increased processing burden for native speakers in terms of slower reading speed. But again, some of the same receptive processing effects could also be hypothesized for other aspects of language use, e.g. heavily accented L2 speech or word stress errors.

Most researchers working with FSs have argued that language users draw on a large inventory of ready-made FSs to supplement creative language production (e.g. Ellis et al., 2008; Erman & Warren, 2000; Hoey, 2005; Pawley & Syder, 1983) and that this facilitates language processing. Looking at the processing advantages of FSs for both native and non-native speakers, the findings of the earlier experimental studies by Schmitt and his colleagues (Schmitt Grandage, & Adolphs, 2004; Schmitt & Underwood, 2004; Underwood, Schmitt, & Galpin, 2004) are, however, very mixed. In a later study, Conklin and Schmitt (2008) did find significant processing advantages for FSs in literal as well as non-literal use for both native and non-native speakers. As discussed (Columbus, 2010; Weinert, 2010), these mixed results may be due to the methods employed or the types of FSs tested, influenced by factors such as frequency, familiarity, recency and context – aspects which may be expected to play a significant role in a usage-based account of language use and language acquisi-

tion (Weinert, 2010, p. 11). None of these earlier processing studies focuses directly on collocations, but the recent study by Columbus (2010), which included restricted collocations, reports faster processing for all three types of FSs tested over compositional control sentences. The evidence of certain processing advantages of FSs – including collocations – seems to be mounting.

#### *4.3. What characterizes L2 collocational development?*

Many of the studies document that collocational competence develops very slowly and unevenly (e.g. Groom, 2009; Laufer & Waldman, 2011). Even so-called ‘very advanced learners’ who are fairly competent in other aspects of English (e.g. morpho-syntax) often experience problems in using appropriate collocations (e.g. Arnaud & Savignon, 1997; Biskup, 1992; Farghal & Obiedat, 1995; Laufer & Waldman, 2011). This may point to the need to redefine the notion of ‘advanced learners’, if many high-level learners do indeed fail to master such prevalent and crucial aspects of language use.

As reported by Pei (2008), a number of the Chinese studies found an increase in use of collocations from beginners to more advanced learners. Gitzaki (1999), Bonk (2001), Gyllstad (2007) and Revier (2009) also reported an increase in collocational development across proficiency levels, whereas Bahns and Eldaw (1993) failed to establish a difference across learner groups. Laufer and Waldman (2011), who looked at collocational use across 3 proficiency levels, found some development for their advanced learners, but even these learners produced deviant collocations compared to L1 use. The advanced learners who used more collocations than the other learner groups were also inclined to produce more errors. Again, these results are in line with the findings for other aspects of L2 development. Moreover, some of the studies show differential development across various types of collocations, which emphasizes the need to look more specifically at the categories (e.g. lexical and grammatical) or even subcategories of collocations studied, as well as the relative frequency of the collocations targeted.

Gyllstad (2007) argues that a period of 4-6 months could not give his students of English at university level sufficient TL exposure which could lead to a measurable increase in the students’ collocational knowledge. Nesselhauf (2003; 2005) also found that increased exposure to the L2 only seemed to improve L2 collocational knowledge slightly. The group results from the Li and Schmitt study (2010) also showed little increase over the 12-month period studied. These findings have, however, been contested by the research carried out by Groom (2009) who argues that the results are much dependent on the operationalization of the construct of collocational knowledge and the way the data analysis is handled. Nesselhauf analysed her data on the basis of a phraseologi-

cal approach to collocations, whereas Groom applied a more frequency-based approach, using two frequency-based measures of collocations in his analysis. Groom (2009) found that his intermediate and advanced data contained more 'lexical bundles' than the L1 data analysed. Normally we would expect native speakers to outperform L2 learners, so this seems to be a counterintuitive finding. However, as argued (Groom, 2009), L1 users have a larger repertoire of options to choose from and therefore show more lexical variation in their choice of collocations. Consequently, the L1 data contains fewer instances of the same lexical units. Groom (2009) hypothesises that fewer instances of the same constructions found in the L2 data over time may therefore in fact be an indication of collocational development, i.e. learning could be described as a downwards adjustment to native-like use.

Yamashita and Jiang (2010) and Wolter and Gyllstad (2011) have looked more closely at the role of the L1 for collocational development and use. Yamashita and Jiang (2010) used an acceptability judgement task to investigate L1 influence on collocational development for both second and foreign language learners. Not surprisingly, the second language learners scored better than the foreign language learners. Comparing both error rate scores and reaction time scores for collocations with L1 equivalents (congruent collocations) and without L1 equivalents (non-congruent collocations), they found that the foreign language learners did better on both scores for the congruent collocations, whereas the second language learners only did significantly better on the error rate scores for the congruent collocations. This might suggest that both the L1 and the amount of exposure influence L2 collocational development. Wolter and Gyllstad (2011) have also looked at the influence of L1 intralexical knowledge on the creation of collocational links in the L2 mental lexicon. Using priming tasks and a receptive test of collocational knowledge (the COLLMATCH test, see Gyllstad, 2007), it was found that collocations with L1-L2 equivalents were processed much faster than non-congruent collocations. Moreover, their informants also scored better on the L1 equivalents in the receptive test. Both results seem to confirm that links in the mental lexicon between the L1 and L2 may play an important role in L2 collocational development and use.

#### *4.4. Why do L2 learners have problems in relation to using and developing their collocational competence?*

It is an underlying assumption in the research literature that the L2 learner - when developing collocational competence - needs to go through the same developmental processes described in most single-word vocabulary acquisition research. This entails that the learner must be able to 1) recognize collocations, i.e. notice and delineate them in the input; 2) understand the meaning and func-

tion of the collocations, i.e. create form-meaning and form-function mappings; 3) understand collocation use restrictions, i.e. expand knowledge of use; 4) choose between different collocational options, i.e. distinguish between collocations in the lexical network; and 5) develop collocational fluency in order to access the collocation with ease. In relation to all these aspects, collocational competence must develop both receptively and productively. The development of collocational competence is thus, like single-word learning, a very complex and cumulative process, demanding enormous amounts of varied language exposure and rich conditions for consolidation through repetition and language use.

Different reasons for why even fairly 'advanced' L2 learners may fail to develop sufficient collocational competence have been put forward. Many of these suggestions are, however, tentative explanations offered by the researchers without direct empirical support. Firstly, the conditions afforded for L2 language development, especially in FLA situations, may not be beneficial for successful L2 collocational development to take place, primarily because L2 learners do not get sufficient exposure in varied contexts and co-texts to be able to recognize and process collocations as recurring lexical units (Durrant & Schmitt, 2010). Moreover, collocations are less frequent than many single-word lexical items that make up the collocation. Consequently, the process of forging and strengthening associative links between the constituents in the collocation by repeated priming will be severely hampered, i.e. the initial traces of associative learning may be lost if the links are not strengthened through repeated exposure (Durrant & Schmitt, 2010).

Secondly, it has been claimed (e.g. Barfield, 2009; Gyllstad, 2007; Wray, 2002) that L2 learners tend to focus on individual words – both receptively and productively, i.e. apply a word-focused approach, and therefore fail to notice recurring chunks in the input. Due to a range of social and cognitive factors, L2 learners do not process the collocations holistically, i.e. they do not draw on a bank of ready-made lexicalized routines like the L1 language user. Instead they rely more on the open-choice rather than the idiom principle (Erman & Warren, 2000; Sinclair, 1991), using language creativity as a starting point for language production, i.e. constructing collocations on the basis of the semantic reference of the individual lexical items, reassembling the collocational unit when the communicative need arises (see Wray, 2002, pp. 205-213). This view has, however, been contested by Durrant and Schmitt (2010), who have shown that advanced L2 learners acquire collocations through an implicit process of associative learning similar to the holistic approach adopted by L1 learners. They argue that L2 learners' problems with acquiring collocations are not due to a non-formulaic approach to learning, but are most likely a product of lack of sufficient L1 exposure and thus a failure to create associative links between the constituents of the collocations.



Thirdly, many literal collocations may not cause comprehension problems, if the learners know the meaning of the individual components of the collocation (Warren, 2005). However, collocations differ in their semantic transparency and may therefore be more or less comprehensible for the L2 learner. Moreover, some collocations are not salient and therefore not noticed as readily as other units by the L2 learner. Finally, we do not know if separate lexical entries are established for collocations - and if so, how these differ from and are associated to the lexical entries for individual lexical items that make up the collocation. It is also not clear whether - or how - this may psycholinguistically affect access routes to the collocations. As shown, many L2 learners produce collocations through a process of L1 transfer. We do not, as yet, know whether the same process of going via the L1 lexical entry takes place when learners decode collocations in their L2 and how this may affect L2 learners' comprehension of collocations (see Wolter & Gyllstad, 2011).

L2 learners may also lack awareness of collocations as lexical units (Ying & O'Neill, 2009) and therefore fail to notice them in the input. Moreover, some L2 learners do not focus on acquiring depth of knowledge of already known words, but they concentrate on learning new words (Ying & O'Neill, 2009), i.e. they see the collecting of new single words as the hallmark of good vocabulary development. Finally, due to the fact that many collocations primarily have a referential function, learners may not be as motivated to notice and acquire collocations compared to the FSs with a more clear pragmatic and thus immediate social and interpersonal function.

As pointed out by Fan (2009), the problems L2 learners experience with collocations in production may also be directly related to the problems the L2 learners have in accessing their general L2 grammatical and lexical knowledge. Fan's learners are clearly hampered by the complexity of syntax and lexis in the written on-line elicitation task used and thus experience difficulties in producing collocations. Fan (2009) argues that the studies which investigate collocations in isolation fail to show this production effect due to the elicitation procedures used.

Viewed from a formal teaching perspective, some of the problems L2 learners experience may be teaching induced. Many teachers tend to focus on individual words (e.g. in glosses and tasks) and often lack useful materials for raising learners' awareness of collocations. Koya (2005) compared the collocations included in language teaching textbooks with collocations in English corpora and found that target use collocations are underrepresented in the textbooks, and the ones included occur with relatively low frequencies. Moreover, if collocations are targeted in the teaching programme, these are often presented in isolation due to the decontextualized approaches used. Finally, Laufer and Waldman (2011) hypothesize that the problems which even advanced L2 learn-

ers experience with collocations may in fact be caused by the use of communicative approaches to teaching, arguing that a more form-focused approach to teaching should be adopted.

Some studies have looked at the effect of teaching on L2 learners' collocational knowledge, focusing specifically on awareness raising activities. The Chinese studies on teaching reported by Pei (2008) show positive effects of teaching collocations to L2 learners. Eyckmans (2009) found that noticing activities can improve learners' awareness of syntagmatic links. This result has, however, been contested in a more recent study of chunk learning (Stengers et al., 2010) which showed no positive effect of teacher-led noticing activities compared to the control groups. Ying and O'Neill (2009), Peters (2009) and Barfield (2009) also describe different approaches to collocations in language teaching, emphasizing the need to raise L2 learners' awareness of collocations, for example of the contrastive differences between collocations and the need to draw learners' attention to the collocations with no direct translation equivalence between the L1 and the L2 (see also Bahns, 1993). Laufer and Girsai (2008) looked at the benefits of form-focused instruction, stressing the need to adopt a teaching approach to collocations based on contrastive analysis and the use of translation. Webb and Kagimoto (2011) investigated the learning effect of the number of collocates presented with the node word, the position of the node word in relation to the collocate and the presentation of synonymous collocations together in the same teaching set. They found that increasing the number of nodes for the same collocate benefited learning, whereas the presentation of synonymous collocations affected learning negatively. The relative position of the collocational constituents did not seem to have an effect. Based on a corpus study focusing on a number of different semantic and pragmatic features of collocations, Walker (2011) has suggested that the use of concordance data may support learning, making the process more meaningful and memorable to the learners. In a teaching study, Chan and Liou (2005) did find positive effects of using a concordancing approach to the teaching of collocations. Handl (2009) has also raised the issue of presentation of collocations in learner dictionaries in order to help learners identify the collocations they need. However, L2 learners often have no knowledge of collocation dictionaries or other potential resources for working with collocations independently.

## **5. Research Approaches to Investigating Collocational Competence and Development**

Let us now shift the focus to different research approaches employed in the studies reviewed and discuss the challenges researchers are faced with when investigating L2 learners' collocational knowledge, use and development. An

overview is given in table 1. Again, the list is not exhaustive and does not include some of the studies reviewed by Pei (2008) and Koya (2005) and some of the studies mentioned in Fan (2009).

**Table 1.** Overview of the research methods used

<b>Methodologies</b>	<b>Studies</b>
<b><i>Written and oral on-line tasks</i></b>	
Written corpora, essays	Chi et al., 1994; Howarth, 1998; Granger, 1998; Gitsaki, 1999; Lorenz, 1999; Kazubski, 2000; Nesselhauf, 2003; Revier & Henriksen, 2006; Wang & Shaw, 2008; Siyanova & Schmitt, 2008; Bell, 2009; Durrant & Schmitt, 2009; Fan, 2009; Prentice, 2010; Li & Schmitt, 2010; Laufer & Waldman, 2011
Oral production	Prentice, 2010
<b><i>Off-line elicitation</i></b>	
Written translation tasks from L1 to L2	Biskup, 1992; Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Gitsaki, 1999; Koya, 2005; Webb & Kagimoto, 2011
Gap fill tasks: Cloze tests and fill-in-the-blank tests	Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Herbst, 1996; Arnaud & Savignon, 1997; Gitsaki, 1999; Shei, 1999; Hoffman & Lehman, 2000; Bonk, 2001; Durrant, 2008; Durrant & Schmitt, 2010; Revier, 2009; Prentice, 2010
Multiple choice tasks, matching and judgement	Fayez-Hussein, 1990; Granger, 1998; Bonk, 2001; Mochizuki, 2002; Honsun, 2005; Gyllstad, 2007; Leśniewska & Witalisz, 2007; Siyanova & Schmitt, 2008
Recognition task	Barfield, 2003; Gyllstad, 2007
Association task	Barfield, 2009; Fitzpatrick, 2012
<b><i>On-line reaction tasks</i></b>	
Eye movement task	Underwood et al., 2004; Columbus, 2010
Self-paced reading	Conklin & Schmitt, 2008; Millar, 2011
Recognition task with reaction time	Siyanova & Schmitt, 2008; Yamashita & Jiang, 2010; Wolter & Gyllstad, 2011

Three general types of elicitation tools have been used (Siyanova & Schmitt, 2008, p. 1) written on-line tasks, often in the form of essays produced by both NSs and NNSs and often collected in large data banks; 2) off-line elicitation tools in the form of productive translation tasks, cloze format tasks and association tasks as well as receptive multiple-choice and judgement tasks; 3) on-line reaction tasks

tapping into the processing of collocations in language use. As discussed by Fan (2009), especially the on-line productive tasks are very demanding, forcing the informants to concentrate on syntactic and lexical processing at the same time. The use of naturally generated on-line tasks may therefore have an impact on the findings of these studies compared to other elicitation methods.

The variety of study aims and approaches mirrors the research diversity found in general single-word vocabulary acquisition research. Not surprisingly, the use of different research instruments is related to the different research aims addressed in the studies. The different focus areas of the studies and the lack of homogeneity in the elicitation tools used, however, make comparisons across the research field difficult, complicating attempts to make any valid generalizations about L2 learners' collocational knowledge, use and development.

### *5.1. Research challenges*

As pointed out by Gyllstad (2007) and Granger (2009), a major challenge which makes comparisons across studies difficult is related to the different definitions of the construct of collocational knowledge. Whether a frequency-based or a phraseological view is used to identify collocations clearly leads to different types of units targeted and may be the cause of the varying results reported.

Secondly, most studies focus on lexical verb + noun and adjective + noun collocations. As mentioned above, the various types of collocations may differ in relation to frequency, saliency and learnability. As shown by Gitsaki (1999), lexical collocations may be acquired before grammatical collocations. When looking at the studies reviewed, there is, however, often a lack of control in the selection of the collocations targeted, both in relation to frequency, the degree of mutual semantic association between the constituents, the degree of restrictiveness and opacity, and as regards the length and directionality of the unit. Moreover, few studies distinguish between motivated and non-motivated collocations or look at the mutual translatability of the collocations between the informants' native and target language or the distance between the languages studied (see e.g. Wolter & Gyllstad, 2011). All of these factors may, as shown, influence the salience and learnability of the collocations and can therefore have an influence on the results found in the various studies.

Thirdly, it is not always clearly stated which aspect of collocational competence and which aspect of the developmental process is in focus, i.e. whether the research intends to tap into the initial process of recognition of the collocational unit, the process of mapping meaning or function on to form, expansion of knowledge of use restrictions of the unit, or the development of receptive and productive fluency. As argued by Laufer and Waldman (2011) L2 learners may primarily be experiencing problems in production of collocations.

Moreover, many of the researchers employ elicitation procedures developed for their specific study, using task types and testing instruments that have not been validated or piloted extensively. Some researchers have therefore carried out extensive work on developing receptive (Eyckmans, 2009: DISCO; Gyllstad, 2007: COLLEX and COLLMATCH) and productive (Revier, 2009: CONTRIX) standardized tools for measuring collocational knowledge, which will make comparison across studies with the same research aims a more attainable goal in the future.

Furthermore, many of the studies are descriptive and lack a developmental focus, looking at the product of learning rather than the process of acquisition. Most of the developmental studies are cross-sectional, and only very few longitudinal studies that follow the same learners have been conducted (e.g. Barfield, 2009; Bell, 2009; Fitzpatrick, 2012; Li & Schmitt, 2010). Finally, only a few studies on instructional effects have been carried out (Chan & Liou, 2005; Durrant, 2008; Laufer & Girsai, 2008; Stengers et al., 2010; Webb & Kagimoto, 2011; see also Pei, 2008).

Most of the studies have based their research on one elicitation procedure, and only a few studies have included two measures. The paper by Siyanova and Schmitt (2008) employs a multi-study approach, using three different elicitation techniques to explore L2 learners' collocational knowledge from different perspectives – focusing on L2 learners' productive use of collocations, their intuitions about collocational restrictions and their receptive processing of collocations. The research programme is unique in that it focuses on three different areas of collocational competence, studying both L1 and L2 informants. Unfortunately, different informant populations are used in the three sub-studies. So even if the study uses a triangulation approach, we have no way of knowing what kind of relation could be found between the three competence areas for the same informant.

Research on collocations in L1 has, not surprisingly, shown differences in collocational use across spoken and written language. Many of the L2 studies reported here, however, focus on written data and many studies examine, as discussed above, collocational knowledge in isolation, using different types of decontextualized, experimental techniques.

As shown above, research on L2 collocation use and development has increased tremendously during the last two decades. Many of the studies have empirically documented some of the problems L2 learners experience in relation to acquiring and using collocation competence. This short progress report has, however, also highlighted some of the conflicting results found. Even if many of the studies employ a quantitative design, some of these do not analyse very large amounts of data, only including small samples of collocations. Moreover, researchers focus on a few collocations or specific collocational types.

Much of the research conducted is exploratory, and researchers fail to use validated, standardized elicitation procedures. Some of the newer studies are, however, aimed at developing and validating instruments for measuring collocational knowledge. Finally, many of the studies focus on the state of the learners' collocation knowledge and use, and the studies that look at collocation development are primarily cross-sectional.

## **6. The Need for Following the Development of Individual Learners over Time**

Many of the collocational studies are based on L1 and L2 data extracted from large corpora. As pointed out by Laufer and Waldman (2011), the advantage of this approach is that large amounts of data can be examined across a variety of data sources and informant groups (across L2 proficiency levels or L1 vs. L2 data) with the use of concordance software. The disadvantage is, however, that only very few studies are longitudinal, tracing the same learners over time with the same tasks. Consequently, we often do not follow the use and development of collocation knowledge from the perspective of the individual learner.

Granger (2009, p. 65) argues that we “need to abandon the notion of the generic L2 learner and distinguish between different types of L2 learners and L2 learning situations”, stressing the need to look at variables that influence learner language such as the learner's L1 (e.g. Wolter & Gyllstad, 2011), degree of exposure (e.g. Groom, 2009) or proficiency level, as well as factors pertaining to the task such as medium, genre, or task type (e.g. Forsberg & Fant, 2010). Most of these factors have tended to be neglected in most L2 learner corpus research.

The need to study language development from a usage-based perspective as it unfolds for the individual learner, the need to take contextual factors into consideration and the need to allow for inter-learner and intra-learner variation in the results reported, echoes some of the very central assumptions about language learning outlined by Larsen-Freeman (1997; 2006) in her discussion of complex, dynamic non-linear models of language development. According to Larsen-Freeman, we need to abandon the ‘developmental ladder metaphor’ which views language development as a linear process which proceeds more or less neatly through a series of stages towards native-like attainment. As argued, the language system adapts to the changing contexts the learners are exposed to. Adaptation and fluctuation of the system dependent on the language use conditions of, and the choices made by, the individual learner should therefore be expected. Moreover, development in one subsystem of language may support or compete with development in another subsystem. Because language is viewed both as a cognitive and social resource embedded in a usage-based context, Larsen-Freeman argues that the L2 learn-

ers' identities, goals and affective states will influence their language use and consequently their language development.

The conflicting results found in some of the collocation studies reported earlier as well as the failure to report development over time in some of the studies may, as is often pointed out by the researchers themselves, be due to differences in the operationalization of the construct of collocational knowledge, the collocations targeted or the lack of sensitivity of the elicitation tools employed. One could, perhaps, also hypothesize that the results are an effect of the quantitative approach adopted and the reliance on learner corpus data in many of the studies. One could speculate whether a research approach which focuses more on individual learners and their differential development should be adopted to complement the quantitative approaches employed. Some learners for example choose to focus on learning new vocabulary items instead of developing depth of knowledge of already acquired lexical items (Ying & O'Neill, 2009). The orientation of learning resources in this way will most likely have a negative effect on the learner's acquisition of collocations, i.e. the competition between these two lexical 'subsystems' will be detrimental to the development of collocational competence.

L1 language learners develop collocational competence through extended exposure to their native language in varying contexts and co-texts. Repeated exposures create and strengthen associative links between the collocational constituents in the language learner's memory organisation, priming (Hoey, 2005) the learner to recognize and use the collocations as holistic units. Repeated exposure to collocations in varying contexts and co-texts is also a prerequisite for developing collocational competence for the L2 learner.

Words and collocations are by nature carriers of semantic meaning. If we exclude the most frequent 2000-3000 word families with very high text coverage and range, most lower-frequency words are related to specific topics, situations, genres, contexts and co-texts. Technical and special purpose contexts and language materials are classic examples of input rich in specialized vocabulary. The nature of the L2 language learners' contact with the target language will naturally influence the lexical items the learner encounters. For FL learners the selection of lexical items is most often under the control of the teacher and dependent on the materials introduced in the language classroom and highly limited by the time allotted to language learning. Additional, self-generated L2 input will often be dependent on the learners' personal interests and the special context situations the learners choose to engage in. We all have stories of learners who have a personal interest for example in internet role plays or computer games and therefore have an exceptionally well-developed vocabulary within these specialized areas. As pointed out by Nation (2001, p. 20) "One person's technical vocabulary is another person's low-frequency word". Hoey (2005, p.

14) also stresses the uniqueness of the individual learner's input and the problems of documenting the learning process.

All these observations are in themselves fairly trivial, but if we link the role of context and co-text in L2 input to the points raised by Larsen-Freeman (1997; 2006) in relation to how the individual language learners adapt and orient themselves to the communicative situations and the needs they experience, the question of frequency becomes extremely crucial. If we look at the frequency of the individual collocations in language input, it is clear that a collocation like *major catastrophe* is less frequent than the two words that make up the collocational unit. Or phrased differently, the likelihood of learners encountering the collocation repeatedly in input is smaller than encountering the individual words and is highly dependent on the type of input the learner encounters. In a small exploratory case study, Dörnyei, Durow, and Zahran (2004) investigated the effect of individual learner differences on the acquisition of FSs. Not surprisingly, they found that the individual learner's motivation, active interaction and social adaptation to the second language situation highly affected the learning outcome. This result might explain why a larger study of the acquisition of FSs which was based on whole-sample statistics failed to produce significant results.

Inspired by Larsen-Freeman's approach, Bell (2009) carried out a longitudinal study, describing "the messy little details" of lexical development which become apparent when looking more closely at one individual learner. As the case study shows, the data reveals instances of fluctuation and variability in the learner's lexical development similar to the scouting and trailing behaviour described by Larsen-Freeman. The learning path can be characterized as showing jagged development and fluctuating patterns of use with structures moving into prominence and/or disappearing. Moreover, Bell identifies the use of intermediate structures and results of competing sub-systems. The longitudinal studies by Barfield (2009) and Li and Schmitt (2010) are examples of case studies which follow individual learners' development of collocation knowledge over time. The in-depth analysis of the individual learners enables Barfield (2009) to describe how different learners approach the learning task, giving us interesting insights into how learners handle the challenges they meet and how they choose to organize their learning in relation to the contexts and needs they experience. Li and Schmitt (2010) also document in detail the inter- and intra-learner variation in the development of the four informants followed over a 12-month learning period. In a more recent study, Fitzpatrick (2012) tracks the changes in vocabulary knowledge of a single subject in a study abroad context by the use of word association data collected six times over an 8-month period. One of the focus areas in the study are the syntagmatic responses produced which give an insight into the developing productive collocational knowledge of the informant.

It is more than likely that collocational acquisition is much more idiosync-



cratic in nature and dependent on specific language use situations than single-word acquisition and therefore calls for more qualitative, case-study, longitudinal research approaches like the studies outlined above. Larsen-Freeman argues for the need to use both macro- and micro-level perspectives in SLA research in order to trace both the larger cross-learner patterns of interlanguage development and the developmental paths taken by the individual learner. One could argue that complementary research methodologies may be a fruitful path to pursue in future collocation research.

## 7. Rounding off

This research overview has shown that native and non-native speakers do differ in their use of collocations both quantitatively and qualitatively, and this holds for advanced L2 learners as well. It has been found that malformed L2 collocations may have negative effects on the processing speed for the recipients. Collocations, however, primarily fulfil a referential function and lack of collocational knowledge therefore might not lead to potential pragmatic failure in the same way, i.e. have the same social and interpersonal consequences as infelicitous use of some of the other types of FSs. On the other hand, collocations are conveyers of precise semantic information, so incorrect use of collocations may potentially lead to misunderstandings, and the failure to use them appropriately may signal lack of expertise and knowledge.

The development of collocational knowledge is slow and uneven and productive mastery clearly lags behind receptive use. But, as argued by many researchers, collocations are more low-frequent than the words that make up the collocations, and learners therefore mostly lack sufficient exposure to collocations to create, strengthen and maintain the associative links between the constituents.

Many conflicting findings have also been reported. This may in part be caused by the lack of clarity and agreement in the research field in relation to the underlying theoretical assumptions regarding the conceptualization of collocational knowledge and development. This naturally affects the type of research questions asked, the identification and selection of collocations targeted for investigation and the research approaches adopted. Moreover, the methodological problems identified in the review make it difficult to outline any valid generalizations across the many studies carried out. The findings show that learning and ability for use are affected by a number of factors pertaining specifically to the types of collocations targeted, their frequency, degree of semantic transparency and the context of learning. Researchers are therefore faced with a number of challenges in relation to language target selection criteria. Moreover, learners' awareness of collocations, their motivation to focus on

these and the teaching conditions afforded for acquisition to take place differ immensely, pointing to the need to combine macro-level, quantitative studies looking at large corpora of L1 and L2 language use and development with micro-level, qualitative case studies of the collocational competence and acquisitional patterns of the individual language learner.

None of these results is, however, surprising, and matches the general SLA findings for other areas of language use, e.g. single-words and other types of FSs. We therefore need to ask whether and, if so, in which way collocations are radically different from other types of FSs or single-word items. Are there specific obstacles related to learning collocations, e.g. factors such as transparency, saliency or function, which make them more difficult to learn or is it merely a matter of lack of exposure due to their frequency which hinders sufficient uptake and consolidation? Does the fact that learners often already have knowledge of the individual words that make up collocations hinder or facilitate learning? Can we transfer our knowledge and assumptions about the knowledge, use and development of single-words and FSs to research on collocations or should other models and approaches be adopted? It has been found that collocations are processed holistically as lexical units and that L2 learners tend to transfer collocational knowledge from their L1, but we still know little about the types of lexical entries formed for collocations, the links between lexical entries for single words and collocations, the links between lexical entries in the L1 and the L2, and the routes the language user takes in processing them. All these aspects may have an impact on the L2 learners' knowledge, use and development of collocations and are fruitful avenues of research. The newer studies carried out by Bell (2009), Wolter and Gyllstad (2011) and Fitzpatrick (2012) for example present some very promising research directions to take, which may help us find answers to some of these questions.

### *Acknowledgements*

I would like to express my gratitude to the editors, the two anonymous reviewers and Henrik Gyllstad and Brent Wolter for their comments on the paper.

### **References**

- Al-Zahrani, M. S. (1998). *Knowledge of English lexical collocations among male Saudi college students majoring in English at a Saudi university*. Ph.D. UMI, Ann Arbor, MI.
- Arnaud, P. J. L., & Savignon, S. J. (1997). Rare words, complex lexical units and the advanced learner. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp.157-173). Cambridge: Cambridge University Press.

- Bahns, J. (1993). Lexical collocations: a contrastive view. *English Language Teaching Journal*, 47(1), 56-63.
- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21(1), 101-114.
- Barfield, A. (2003). *Collocation recognition and production: Research insights*. Tokyo: Chuo University.
- Barfield, A. (2009). Exploring productive L2 collocation knowledge. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in language learners: Papers and perspectives in honour of Paul Meara* (pp. 95-110). Clevedon: Multilingual Matters.
- Barfield, A., & Gyllstad, H. (Eds.). (2009a). *Researching collocations in another language: Multiple interpretations*. Basingstoke: Palgrave Macmillan.
- Barfield, A., & Gyllstad, H. (2009b). Introduction: Researching L2 collocation knowledge and development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 1-20). Basingstoke: Palgrave Macmillan.
- Bell, H. (2009). The messy little details: A longitudinal case study of the emerging lexicon. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in language learners: Papers and perspectives in honour of Paul Meara* (pp. 111-127). Clevedon: Multilingual Matters.
- Biskup, D. (1992). L1 influence on learners' renderings of English collocations. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp.85-93). London: Macmillan.
- Boers, F., Eyckmans, J., & Stengers, H. (2006). Motivating multiword units. Rationale, mnemonic benefits and cognitive style variables. *EUROSLA Yearbook* 6, 169-190.
- Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Basingstoke: Palgrave Macmillan.
- Bonk, W. J. (2001). Testing ESL learners' knowledge of collocations. In T. Hudson & D. Brown (Eds.), *A focus on language test development: Expanding the language proficiency construct across a variety of tests* (pp. 113-142). Honolulu: University of Hawaii.
- Chan, T. P., & Liou, H.C. (2005). Effects of web-based concordancing instruction of EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, 18(3), 231-250.
- Chi, M. L., Wong, P. Y., & Wong, C. P. (1994). Collocational problems amongst ESL learners: a corpus-based study. In L. Flowerdew & A. K. Tong (Eds.), *Entering text* (pp. 157-165). Hong Kong: University of Science and Technology.
- Columbus, G. (2010). Processing MWUs: Are MWU subtypes psycholinguistically real? In D. Wood (Ed.), *Perspectives on formulaic language: acquisition and communication* (pp. 194-212). London/New York: Continuum.
- Conklin, C., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than non-formulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.
- Cowie, A. P., & Howarth, P. (1996). Phraseological competence and written proficiency. In G. M. Blue & R. Mitchell (Eds.), *Language and education (British studies in applied linguistics II)* (pp. 80-93). Clevedon: Multilingual Matters.

- Dechert, H. W. (1983). How a story is done in a second language. In Færch, C. & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 20-60). London: Longman.
- Deuter, M. (2002). *The Oxford collocations dictionary* Oxford: Oxford University Press.
- Dörnyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In Schmitt N. (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 87-106). Amsterdam: Benjamins.
- Durrant, P. (2008). *High frequency collocations and second language learning*. (Unpublished doctoral dissertation). The University of Nottingham, Nottingham.
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28(3), 157-169.
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics*, 47(2), 157-177.
- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163-188.
- Ellis, N.C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-68). Cambridge: Cambridge University Press.
- Ellis, N.C. (2003). Constructions, chunking and connectionism: the emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition*. Oxford: Blackwell.
- Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27(2), 305-352.
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 41(3), 375-396.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text* 20(1), 29-62.
- Eyckmans, J. (2009). Towards an assessment of learners' receptive and productive syntagmatic knowledge. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 139-152). Basingstoke: Palgrave Macmillan.
- Fan, M. (2009). An exploratory study of collocational use by ESL students: A task-based approach. *System*, 37(1), 110-123.
- Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics*, 33(4), 315-31.
- Fayez-Hussein, R. (1990). Collocations: The missing link in vocabulary acquisition amongst EFL learners. In J. Fisiak (Ed.), *Papers and studies in contrastive linguistic: The Polish English contrastive project*. (Vol. 26, pp.123-136). Poznan: Adam Mickiewicz University.
- Fitzpatrick, T. (2012). Tracking the changes: vocabulary acquisition in the study abroad context. *The Language Learning Journal*, 40(1), 81-98.
- Forsberg, F., & Fant, L. (2010). Idiomatically speaking: the effects of task variation on formulaic language in highly proficient users of L2 French and Spanish. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 47-70). London/New York: Continuum.

- Gitzaki, C. (1999). *Second language lexical acquisition: A study of the development of collocational knowledge*. San Francisco: International Scholar Publications.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and application* (pp. 145-160). Oxford: Oxford University Press.
- Granger, S. (2009). Learner corpora: A window onto the L2 phrasicon. In A. Barfield & H. Gyllstad (Eds.) *Researching collocations in another language: Multiple interpretations* (pp. 60-65). Basingstoke: Palgrave Macmillan.
- Granger, S., & F. Meunier (Eds.). (2008). *Phraseology. An interdisciplinary perspective*. Amsterdam: Benjamins.
- Granger, S., & Pacquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective* (pp. 27-49). Amsterdam: Benjamins.
- Groom, N. (2009). Effects of second language immersion on second language collocational development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 21-33). Basingstoke: Palgrave Macmillan
- Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Lund: Lund University.
- Handl, S. (2009). Towards collocational webs for presenting collocations in learners' dictionaries. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 69-85). Basingstoke: Palgrave Macmillan.
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237-258.
- Henriksen, B., & Stenius Stæhr, L. (2009). Processes in the development of L2 collocational knowledge: A challenge for language learners, researchers and teachers. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 224-231). Basingstoke: Palgrave Macmillan.
- Herbst, T. (1996). What are collocations: sandy beeches or false teeth? *English Studies* 77(4), 379-393.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Hoffmann, S., & Lehmann, H. M. (2000). Collocational Evidence from the British National Corpus. In J. M. Kirk (Ed.), *Corpora Galore: Analyses and Techniques in Describing English*. Amsterdam: Rodopi.
- Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Tübingen: Narr.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24-44.
- Jaén, M. M. (2007). A corpus-driven design of a test for assessing the ESL collocational competence of university students. *International Journal of English Studies*, 7(2), 127-147.
- Jiang, J. (2009). Designing pedagogic materials to improve awareness and productive use of L2 collocations. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 99-113). Basingstoke: Palgrave Macmillan.

- Kazubski, P. (2000). Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: A contrastive, corpus-based approach. Available on-line at <http://main.amu.edu.pl/przemka/research.html>
- Keshavarz, M. H., & Salimi, H. (2007). Collocational competence and cloze test performance: a study of Iranian EFL learners. *International Journal of Applied Linguistics*, 17(1), 81-92.
- Koya, T. (2005). *The acquisition of basic collocations by Japanese learners of English*. (Unpublished doctoral dissertation) Waseda University, Japan. Available on-line at <http://dspace.wul.waseda.ac.jp/dspace/bitstream/2065/5285/3/Honbun-4160.pdf>
- 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.
- Lauffer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29(4), 694-716.
- Lauffer, B., & Waldman, T. (2011). Verb-noun collocations in second-language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647-672.
- Leśniewska, J., & Witalisz, E. (2007). Cross-linguistic influence and acceptability judgments of L2 and L1 collocations: A study of advanced Polish learners of English. *EUROSLA Yearbook* 7, 27-48.
- Lewis, M. (1993). *The lexical approach*. Hove: Language Teaching Publications.
- Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach*. Hove: Language Teaching Publications.
- Li, J., & Schmitt, N. (2010). The development of collocations use in academic texts by advanced L2 learners: a multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 23-46). London/New York: Continuum.
- Lorenz, T. R. (1999). *Adjective intensification – learners versus native speakers: A corpus study of argumentative writing*. Amsterdam: Rodopi.
- Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, 32(2), 129-148.
- Mochizuki, M. (2002). Explorations of two aspects of vocabulary knowledge: Paradigmatic and collocational. *Annual Review of English Language Education in Japan*, 13, 121-129.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). Collocations in a learner corpus. *Studies in Corpus Linguistics (Vol. 14)*. Amsterdam: Benjamins.

- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.
- Pei, C. (2008). Review of empirical studies on collocations in the field of SLA. *Celea Journal*, 31(6), 72-81
- Peters, E. (2009). Learning collocations through attention-drawing techniques: A qualitative and quantitative analysis. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 194-207). Basingstoke: Palgrave Macmillan.
- Prentice, J. (2010). På rak sak: Om ordförbindelser och konventionaliserade uttryck bland unga språkbrukare i flerspråkiga miljöer. *Göteborgstudier i nordisk språkvvetenskap* 13. Göteborg: Intellecta Infolog.
- Raupach, M. (1984). Formulae in second language speech production. In H. W. Dechert, D. Möhle & M. Raupach (Eds.), *Second language production* (pp. 114-137). Tübingen: Narr.
- Revier, R. L. (2009). Evaluating a new test of whole English collocations. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 125-138). Basingstoke: Palgrave Macmillan.
- Revier, R. L., & Henriksen, B. (2006). Teaching collocations. Pedagogical implications based on a cross-sectional study of Danish EFL. In M. Bendtsen, M. Björklund, C. Fant & L. Forsman (Eds.), *Språk, lärande och utbildning i sikte* (pp. 191-206). Pedagogiska fakulteten Åbo Akademi Vasa.
- Schmitt, N. (2004). (Ed.). *Formulaic sequences: Acquisition, processing and use*. Amsterdam: Benjamins.
- Schmitt, N. (2010). *Researching vocabulary. A vocabulary research manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In Schmitt, N. (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 127-149). Amsterdam: Benjamins.
- Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In Schmitt, N. (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 173-189). Amsterdam: Benjamins..
- Shei, C. C. (1999). A brief review of English verb-noun collocation. Available on-line at <http://www.dai.ed.ac.uk/homes/shei/survey.html>.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. (2004). *Trust the Text: Language, corpus and discourse*. London: Routledge.
- Sivanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429-458.
- Skrzypek, A. (2009). Phonological short-term memory and L2 collocational development in adult learners. *EUROSLA Yearbook*, 9, 160-184.
- Stengers, H. F., Boers, F., Eyckmans, J., & Housen, A. (2010). Does chunking foster chunk uptake? In De Knop, S., F. Boers & T. De Rycker (Eds.), *Fostering language teaching efficiency through cognitive linguistics* (pp. 99-117). Berlin/New York: Mouton de Gruyter.

- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In Schmitt N. (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 153-172). Amsterdam: Benjamins.
- Walker, C. P. (2011). A corpus-based study of the linguistic features and processes which influence the way collocations are formed: Some implications for the learning of collocations. *TESOL Quarterly*, 45(2), 291-312.
- Wang, Y., & Shaw, P. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal*, 32, 201-232.
- Warren, B. (2005). A model of idiomaticity. *Nordic Journal of English Studies*, 4, 35-54.
- Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning. *Applied Linguistics*, 32(3), 259-276.
- Weinert, R. (2010). Formulaicity and usage-based language: linguistic, psycholinguistic and acquisitional manifestations. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 1-20). London/New York: Continuum.
- Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, 34(4), 430-449.
- Wood, D. (2010). *Perspectives on formulaic language: Acquisition and communication*. London/New York: Continuum.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, 44(4), 647-668.
- Ying, Y., & O'Neill, M. (2009). Collocation learning through and 'AWARE' approach: Learner perspectives and learning process. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 181-193). Basingstoke: Palgrave Macmillan.



# *Measuring the contribution of vocabulary knowledge to proficiency in the four skills*

James Milton  
Swansea University

This chapter examines the way vocabulary knowledge relates to the ability to perform communicatively in a foreign language and in particular the ability to perform in the four language skills of reading, writing, listening and speaking. It reviews recent research designed to investigate the way vocabulary knowledge and performance inter-relate. There is a tradition of research which demonstrates that measures of vocabulary knowledge are particularly good predictors of performance in the four skills, and recent research suggests that when measures of different dimensions of vocabulary knowledge are combined this predictive-ness can be enhanced. Large vocabularies, and speed and depth of vocabulary knowledge, appear indispensable to the development of good performance in any language skill and it is now possible to enumerate the scale of vocabulary that is needed for the CEFR levels of communicative performance.

## **1. Lexical knowledge and language learning**

A feature of the English language literature on language learning and language teaching methodology over the last 60 years or so is the way vocabulary as a subject for teaching has been side-lined. It receives little attention in much of the literature on second language acquisition as a general process (e.g. Mitchell & Myles, 2004; Lightbown & Spada, 2006). It is almost entirely absent from major books on the syllabus and theory of language teaching (O'Dell, 1997, p. 258). Wilkins (1972, p. 109) suggests this may have been a product of the development of structural approaches to linguistics after the Second World War and the way that, in these approaches, vocabulary could be reduced to the minimum needed to illustrate the structural content. However, the absence of vocabulary is notable even after structural approaches to language teaching became unfashionable and were replaced by communicative and other approaches. Definitive works in these areas either omit to mention the topic entirely, as in Littlewood (1983), or dismiss the subject as one which is unsystematic and incidental at best to language learning, as in Brumfit (1984). It is true that at an academic level there is much renewed interest in the subject but,

as Schmitt (2008) notes, the insights gained have failed to make their way into the mainstream literature on language pedagogy. An example of the prevailing attitude to vocabulary in pedagogy can be seen in the comment by Harris and Snow that “few words are retained from those which are ‘learned’ or ‘taught’ by direct instruction ... [and learners] extend their vocabulary through sub-conscious acquisition” (Harris & Snow, 2004, pp. 55-61). With this attitude, the explicit teaching of vocabulary, and the systematic organisation of vocabulary in the curriculum, is not a priority.

In academic circles, the place of vocabulary in language learning has been significantly revised over the last decade and current academic thinking is very much at odds with much classroom and textbook practice. Far from being an element which is merely incidental to language learning, current thinking advocates that vocabulary may be crucial to the development of language performance overall. In a recent version of generative grammar, the Minimalist Program (Chomsky, 1995), the differences between languages are seen to be mainly lexical in nature and this leads Cook (1998) to suggest that the Minimalist Program is lexically-driven. The properties of the lexical items shape the sentence rather than lexical items being slotted into pre-existent structures. The task the language learner faces, therefore, is principally one of learning the vocabulary of the foreign language. The acquisition of vocabulary items in sufficient quantity triggers the setting of universal grammatical parameters. This approach is reflected in the Lexical Learning Hypothesis (Ellis, 1997) according to which vocabulary knowledge is indispensable to the acquisition of grammar.

One of the outcomes of the recent academic interest in vocabulary has been the development of ways for describing and testing vocabulary knowledge, which are both principled and systematic. Recently developed methods allow normalised data to be produced so the growth of a foreign language lexicon over the course of learning can be modelled. With this information it becomes possible to measure the contribution of vocabulary knowledge to language development and confirm whether the close relationship between vocabulary growth and language level exists in practice.

## **2. Dimensions of vocabulary knowledge**

A feature of our understanding of vocabulary, whether in a first or foreign language, is that knowledge of this aspect of language is multi-faceted. Potentially, there is a lot involved in knowing a word. The ancient Greeks, for example, clearly identified three elements of word knowledge: knowledge of aural and written forms and knowledge of the meaning of the word. For at least a century, too, a distinction has been made between receptive knowledge, and produc-

tive word knowledge. Some words, it seems, exist in the minds of language speakers primed for use and can be called to mind in speech or in writing easily and quickly. Other words are not used in this way but can, nonetheless, be called to mind for comprehension if they occur in the speech or writing of others. Each of these facets of knowledge can contribute to language performance in its own different way. A language user with extensive knowledge of words in their phonological form but no knowledge of the written form of words, for example, has the potential at least to speak and understand speech but no capacity for reading or writing. There is no definitive list of what comprises word knowledge and even native speakers will not know every facet of every word in their lexicon. In measuring vocabulary knowledge in order to assess how it impacts on overall language performance, therefore, decisions have to be made as to exactly what it is that is being measured.

The nearest thing we have to a definitive list of what it means to know a word is Nation's (2001) table shown in table 1. This table usefully encapsulates knowledge of the various forms of a word, the various aspects of meaning a word can carry with it, and the elements of use which are also part of word knowledge. Knowledge of form includes not just knowledge of the written and sound forms of a word but also knowledge of affixation, knowledge of the way extra parts can be added, or the ways in which a word can change, to reflect changes in its grammatical function or to add to its meaning. Knowledge of meaning includes not just knowledge of a core meaning, perhaps a link with a direct foreign language counterpart, but also the concepts, referents and associations, which a word may carry with it. Words in different languages often carry differences in nuances of meaning, which, if a learner is to perform fluently, may need to be known. And knowledge of use includes knowledge of the grammar of a word but also the way words like to behave in relation to each other. Some words like to occur in combination with other words, in particular idioms for example, and some words, like swear words, may be restricted in the occasions where they can be used appropriately, and this knowledge will also be needed if the language is to be used fluently and skilfully. Each facet of knowledge is sub-divided into receptive and productive knowledge.

This is a very useful and insightful list, and makes apparent just how much is involved in fully knowing a word. It is also clear that designing a test that can capture knowledge in all this diversity is scarcely practical. A single test could not possibly hope to encompass every aspect of knowledge described in this table. There is a further difficulty inherent in this table in that the various forms of knowledge are characterised but not precisely defined. In assessing knowledge of word parts, for example, it is unclear at what point the additions and changes to a word will form a new word rather than a derived form of an existing one. Nor is it clear, for example, how frequently a word must co-occur with another

word for a collocation to be created. But if vocabulary knowledge is to be tested and understood, then these are constructs which must be precisely defined.

**Table 1.** Description of “what is involved in knowing a word”, from Nation (2001: 27).

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concepts and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use (register, frequency)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

In order to reduce this complexity to manageable proportions, therefore, it has become common to think of vocabulary knowledge in terms of dimensions rather than a lengthy list of discrete elements. Each dimension can encompass a range of the separate elements in Nation’s list, which are linked in some way to form a single, larger entity. A common distinction, instigated by Anderson and Freebody (1981), is that between vocabulary breadth, that is, the number of words a learner knows regardless of the form they are known in or how well they are known, and vocabulary depth, which is how well or how completely these words are known. Neither of these terms is completely unambiguous. Vocabulary breadth, sometimes called vocabulary size, may be used to reflect a learner’s recognition vocabulary only: their ability to recognise the form of a word as a real word in the foreign language, and distinguish it from an artificially created non-word. The term may also be used to reflect a learner’s ability to recognise a word and link it to meaning or to a translation in the first language. Defining a vocabulary item like this entails a higher order of knowledge than defining it in terms of sheer recognition and it might be expected that

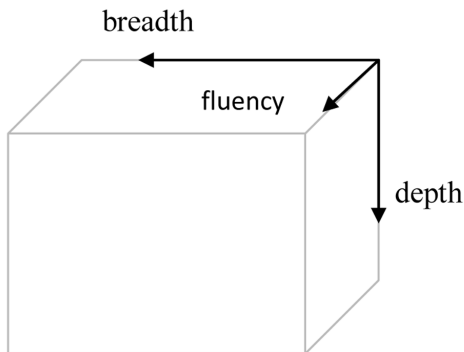
measurements of knowledge made using a higher order of knowledge criterion would be smaller than measurements made using a recognition requirement only. Notwithstanding these differences, vocabulary breadth has become sufficiently well acknowledged to be included in several well-established tests. Meara and Milton's (2003) X-Lex measures recognition knowledge of the most frequent 5000 lemmatised vocabulary items in a number of languages. Nation's (2001) Vocabulary Levels Test tests the ability of learners to recognise vocabulary items and link them to a definition among a selection of items drawn from a range of frequency bands and lists. Usefully, there is also a productive version of this test (Laufer & Nation, 1999).

Vocabulary depth is less well defined. It can be characterised in terms of knowledge of any of the several facets which Nation lists and which might involve knowledge about a word rather than just recognising it: associational knowledge, collocational knowledge, inflectional and derivational knowledge, knowledge of concepts and referents, and knowledge of constraints on use (Read, 2000). It has also been characterised in terms of movement along a continuum from partial to precise knowledge of a word (Henriksen, 1999). These approaches have been criticised because it is difficult to find a concept that holds together the variety of elements, which might fall into this category (Milton, 2009). However, Meara and Wolter (2004) can make a case for doing this by defining depth in terms of the number of links between words and the networks words can create. A word that is recognised as a word in a language, but where nothing more is known about it, has no links and is not networked. Once a meaning is attached to that form and some idea is gained as to how the word can be used, then it develops links with other words and begins to network and it does not matter whether these are grammatical or associational or collocational links. Words, which have this network of links around them can be said to be known more deeply than those, which are not known in this way. Depth of knowledge could be measured by counting the number of links in a word's network. While this is an interesting approach, there is an argument too that breadth and depth are, in effect, the same thing. As Vermeer (2001) points out, a condition of having a large network of links to a word is knowing a large number of words to make those links. A condition for developing depth in vocabulary knowledge will be to develop vocabulary breadth. The two are interdependent. As these points demonstrate, there is something of a contradiction in the way vocabulary specialists handle the distinction between breadth and depth. On the one hand, it can be assumed that breadth and depth will be closely related so that scores on tests of breadth can be used to validate newly created tests of depth (e.g. Gyllstad, 2007; this volume). On the other hand, it is still common to talk of vocabulary knowledge in terms of breadth and depth as two separate and contrasting dimensions which should not be closely connected.

Perhaps because the term is ill-defined, there is an absence of well-established and standardised tests in this field. Attempts, such as Wesche and Paribakht's (1996) Vocabulary Knowledge Scale (VKS), to fill this void are not without their difficulties and in us, as Wolter (2005) points out, they function as breadth rather than depth measures.

Meara (1996) adds a third dimension to breadth and depth by also characterising vocabulary knowledge in terms of the automaticity with which the words a person knows can be recognised and processed, or accessed for use in language. Daller et al. (2007) call this fluency and the presence of this dimension, and with it the attempt to characterise and measure the ability to activate what would otherwise be receptive knowledge, retains the productive and receptive distinction, which has proved so useful in measuring vocabulary knowledge. The presence of this third dimension allows Daller et al. to suggest a hypothetical, three-dimensional lexical space (Figure 1).

**Figure 1.** The lexical space: dimensions of word knowledge and ability (based on Daller et al., 2007: 8)



This hypothetical space allows learners with different types of vocabulary knowledge to be positioned differently in this space and systematically distinguished from each other. As Meara and Wolter (2004) explain, these distinctions might be used to explain how learners can behave differently from each other in their ability to perform in their foreign language. A learner with high vocabulary breadth but low fluency and low depth might be usefully distinguished from a learner with the same vocabulary breadth but higher fluency and depth. Although the number of words they know might be the same, the latter learner might be expected to be more skilful and able in using the foreign language in communicative situations. There is an increasing body of research evidence to support the idea that vocabulary knowledge and performance in a for-

eign language are linked (e.g. Alderson, 1984; Laufer, 1992; Laufer & Nation, 1999; Qian, 1999; Zimmerman, 2004) and it is the nature and extent of this link that this chapter intends to make more clear.

### **3. Performance in the four skills in a foreign language and vocabulary knowledge**

The goal for any foreign language learner is to use the language in some way. This may be for speech and casual conversation, or for translation of texts, or for study through the medium of the foreign language. It has become a commonplace in the assessment of language to consider language in terms of four separate skills: the receptive skills of reading and listening, and the productive skills of speaking and writing. In reality, of course, these distinctions are not so clear and the ability to read and listen fluently requires the learner to actively anticipate the language that is likely to occur and then monitor input to check that the possibilities which have been created are occurring. Nonetheless, the distinction is enshrined in formal and assessment schemes. The University of Cambridge Local Examinations Syndicate (UCLES) exams, such as the International English Language Testing System (IELTS) test, administer separate papers for each of these skills and devise separate grading schedules for them. The Council of Europe's (2001) Common European Framework of Reference for Languages (CEFR) hierarchy uses both global descriptors of language performance as a whole (p. 24), and descriptors separated into the four skills (pp. 26-27). These descriptors are couched in terms of performance of language rather than in terms of the language knowledge, which is likely to underlie performance. The example below of the CEFR's global descriptor for performance at C2 level illustrates this (Council of Europe, 2001, p. 24).

Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

There is an assumption that language knowledge, such as vocabulary knowledge, will develop in relation to language performance and this is reflected both in the wording of the descriptors for the four skills but also in the presence of levels criteria specifically for vocabulary range and vocabulary control in the CEFR document (p. 112). These two terms are not explicitly defined but range appears broadly to refer to the vocabulary size available to the learner, and con-

trol appears to be closer to vocabulary depth in that it refers to the accuracy and appropriateness of vocabulary selection and use. Table 2 presents the descriptors for vocabulary range.

**Table 2.** CEFR vocabulary range descriptors

<b>level</b>	<b>descriptor</b>
<b>C2</b>	Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.
<b>C1</b>	Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.
<b>B2</b>	Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution.
<b>B1</b>	Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events. Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.
<b>A2</b>	Has a sufficient vocabulary for the expression of basic communicative needs. Has a sufficient vocabulary for coping with simple survival needs.
<b>A1</b>	Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.

In terms of the dimensions of vocabulary knowledge described in the first section, it is clear that what is anticipated is that learners will grow an increasingly large lexicon as they progress through the levels. The ‘basic vocabulary’ at A1 Level becomes ‘a very broad lexical repertoire’ at C2 level. It is implied that only at the most advanced levels will developments in vocabulary depth be relevant. At C1 level ‘Good command of idiomatic expressions and colloquialisms’ is expected, and at C2 level ‘awareness of connotative levels of meaning’ is added. The use of expressions like command in the descriptors also implies that learners have these items available for use and that vocabulary knowledge has progressed along the fluency dimension as well as the breadth and depth dimensions. Implicit within this framework, therefore, is the understanding that a requirement of making progress in communicating through foreign language is acquiring greater volumes of vocabulary and acquiring ever greater sophistication and control in the use of this vocabulary. The Framework document, there-



fore, also suggests that it might be possible and useful for vocabulary size and depth measurements to be attached to the different levels.

There is some empirical evidence that links vocabulary breadth measures with the CEFR language levels. Milton (2010), shown in Table 3, provides EFL vocabulary sizes (out of the most frequent 5,000 lemmatised words in English) gained from over 10,000 learners in Greece taking both recognition tests of their vocabulary size and also formal UCLES exams at levels within the CEFR framework.

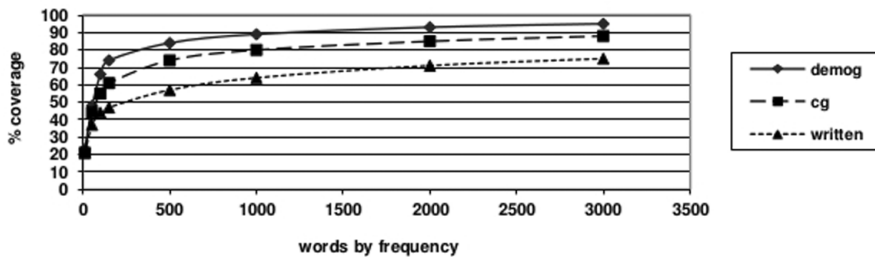
**Table 3.** Vocabulary size estimates, CEFR levels and formal exams (Milton, 2010, p. 224)

CEFR Levels	Cambridge exams	XLex (5000 max)
A1	Starters, Movers and Flyers	<1,500
A2	Kernel English Test	1,500 – 2,500
B1	Preliminary English Test	2,500 – 3,250
B2	First Certificate in English	3,250 – 3,750
C1	Cambridge Advanced English	3,750 – 4,500
C2	Cambridge Proficiency in English	4,500 – 5,000

While there is some individual variation around these ranges, Milton is able to conclude that “the assumption made in the CEFR literature, that as learners progress through the CEFR levels their foreign language lexicons will increase in size and complexity, is broadly true” (2010, p. 224). Variation may be explained by the way vocabulary knowledge and language performance are imperfectly linked. Learners with the same or similar vocabulary sizes – and remember these are based on knowledge of the 5,000 most frequent lemmatised words in English and so are not absolute vocabulary size estimates – may make different use of this knowledge to communicate more or less successfully. Milton and Alexiou (2009) report similar vocabulary size measurements for CEFR levels in French and Greek as foreign languages.

If vocabulary breadth predicts overall language performance well, then it might be expected that vocabulary breadth will link well also with the four separate skills. However, there are reasons for thinking that the oral skills, speaking and listening, will have a different relationship with vocabulary knowledge from the written skills, writing and reading. Figures for coverage (the proportion of a corpus provided by words in the corpus arranged in frequency order) in spoken and written corpora suggest that written text is typically lexically more sophisticated than spoken text. A comparison (Figure 2) of coverage taken from written and spoken sub-corpora of the 100 million word British National Corpus illustrates this (Milton, 2009, p. 58).

**Figure 2.** Coverage from written and spoken corpora in the BNC



The relationship between text coverage and comprehension (and by extension communication more generally) in a foreign language is now quite well understood. In this example the most frequent 2,000 lemmatised words in English provide about 70 % coverage of the *written* sub-corpus. This suggests that a learner with a vocabulary size at about this level would struggle to communicate at anything but the most basic level through the medium of writing. The same vocabulary size provides much more coverage in both the spoken sub-corpora, around 90 %. The *demographic* (demog in figure 2) sub-corpus is taken from samples of general conversation and the *context governed* (cg in figure 2) sub-corpus from examples of rather more formal and organised spoken language: lectures, court room exchanges and sermons. A figure of 95% coverage is often associated with full comprehension and even though this is probably an over-simplification Figure 2 does nonetheless suggest that a learner with a vocabulary size of about 2,000 lemmatised words might be quite communicative in speech and listening. It has been suggested (Milton et al., 2010) that the relationship between vocabulary size and performance in the written skills might, for all practical purposes, be linear and that up to the highest levels of performance greater vocabulary breadth is associated with better language performance. A learner's vocabulary would have to be well beyond the 5,000 words measured in this paper before the additional benefit of extra words ceased to add to comprehension. However, the relationship in the oral skills may not be linear and Milton et al. suggest that beyond a comparatively low level, 2,000 or 3,000 lemmatised words, there may be little benefit to performance in learning more vocabulary. Beyond this level, therefore, learners will experience diminishing returns in their oral performance for the time invested in learning vocabulary.

## 4. Examining the relationship between vocabulary knowledge and language proficiency

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 the relationship has been conducted with measures of vocabulary size only, and within the realm of reading skill only (Stæhr, 2008). Generally, such studies have found strong correlations between receptive vocabulary size tests and reading comprehension tests, ranging from 0.50 to 0.85, with learners from different proficiency levels (e.g. Laufer, 1992; Qian, 1999; Albrechtsen, Haastrup & Henriksen, 2008).

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 investigate the scale of the contribution which vocabulary, in all three of its dimensions, can make to a variety of communicative skills in foreign language performance. By extension, such research also tests the credibility of theories such as the Lexical Learning Hypothesis (Ellis, 1997), and contributes firmer evidence to the place that vocabulary should have in the structure of the foreign language learning curriculum, since in this view of learning it is vocabulary knowledge which drives learning in other aspects of language. However, the considerations above have suggested that the relationship between vocabulary knowledge and overall language skill may potentially be difficult to model and to measure. Different dimensions of vocabulary knowledge might need to be measured separately and their effects combined if the full nature of the relationship with language skill is to be seen. Further, it might be that the relationship will vary according to the level of the learner and the skills the learner needs. The following sections will examine particular pieces of research in this area, which illustrate the state of our knowledge and from which broader conclusions can be drawn.

### 4.1. *Stæhr (2008)*

In this paper Stæhr attempts to investigate the relationship between vocabulary size and the skills of reading, writing and listening among EFL learners in lower secondary education in Denmark. He further investigates the role played by the most frequent 2,000 word families in achieving pass grades in the tests of these skills.

Stæhr measures vocabulary size using the improved version of the Vocabulary Levels Test (VLT) made by Schmitt, Schmitt and Clapham (2001). The test consists of five separate sections, which represent four levels of word frequency and a similarly structured test of knowledge of Coxhead's Academic Word List

(Coxhead, 2000). However, the academic word level was excluded from Stæhr's study as not relevant for low-level learners. The test assesses learners' receptive knowledge of word meaning at the 2,000, the 3,000, the 5,000 and the 10,000 level, and the test results can thus give an indication whether learners master the first 2,000, 3,000, 5,000 or 10,000 word families in English. Although the VLT was originally designed as a diagnostic test intended for pedagogical purposes, researchers (e.g. Read, 2000; Schmitt et al., 2001) acknowledge its use as a means of giving a good guide to overall vocabulary size. Tests of language skills were assessed as part of the national school leaving examination. Reading and listening skill abilities were measured using pencil-and-paper multiple-choice tests. Writing ability was measured using the scores awarded for an essay task where the participants had to write a letter to a job agency applying for a job.

Stæhr's results indicate a correlation between vocabulary size and reading, which is comparable with the findings of other research mentioned above and suggests a strong and statistically significant relationship between the amount of vocabulary a learner knows in the foreign language and their ability to handle questions on a text designed to test their ability to fully comprehend the text. His analysis, using binary logistic regression, shows that as much as 72% of the variance in the ability to obtain an average score or above in the reading test is explained by vocabulary size (Nagelkerke  $R^2 = 0.722$ ). The results also illuminate the relationship with other language skills. The correlation between vocabulary size and both writing and listening ability is also statistically significant and reasonably strong. Stæhr suggests that 52% of the variance in the ability to obtain an average or above-average writing score is accounted for by vocabulary size (Nagelkerke  $R^2 = 0.524$ ), and that 39 % of the variance in the listening scores, in terms of the ability to score above the mean, is accounted for by the variance in the vocabulary scores (Nagelkerke  $R^2 = 0.388$ ). His interpretation of this is that this amount of variance is substantial. Even the contribution towards listening, the smallest in this study, is considerable, given the fact that it is explained by one single factor. This confirms the importance of receptive vocabulary size for learners in all three skills investigated.

Stæhr's findings further indicate the importance of knowing the most frequent 2,000 word families in English in particular and he suggests that knowledge of this vocabulary represents an important threshold for the learners of his study. Knowledge of this vocabulary is likely to lead to a performance above average in the listening, reading and writing tests of the national school leaving exam. The results seem to emphasize that the 2,000 vocabulary level is a crucial learning goal for low-level EFL learners and suggest that the single dimension of vocabulary size is a crucial determiner of the ability to perform in the three foreign language skills tested. The more vocabulary learners know, the better they are likely to perform through the medium of the foreign language.

#### 4.2. *Milton, Wade & Hopkins (2010)*

Even if it appears that a single dimension, vocabulary size, contributes hugely to a learner's ability to perform through a foreign language, this is still far from a complete explanation of the nature of the relationship. There is a gap, it appears, in vocabulary's explanatory power. Studies investigating the relationship between vocabulary and language proficiency almost never offer an insight into the relationship between vocabulary and speaking ability. Zimmerman (2004) is possibly the only study, which demonstrates that such a relationship exists. The correlations between vocabulary size scores and aural skills, such as listening, are generally weaker than the correlations with the written skills of reading and writing. This may be due to the way oral language is lexically less sophisticated than written language, but may also be due to the nature of the tests to measure vocabulary size, which are invariably delivered through the medium of writing. Milton et al. (2010) also point out that it is quite conceivable that word knowledge may be held in the lexicon in aural form only, and not necessarily in written form and, particularly if vocabulary learning comes about predominantly through oral input as Ellis (Ellis, 1994, p. 24) suggests. With regard to aural skills, therefore, the vocabulary tests may be tapping into vocabulary knowledge in the wrong form for the relationship to be made fully clear.

Milton et al. (2010) therefore conducted a study which investigated whether vocabulary knowledge can be held in different formats, written and/or aural, and whether the measurements of vocabulary size made using two different formats can better explain performance in all the four skills. To measure written vocabulary size they used X-Lex (Meara & Milton, 2003), where the words tested only ever appear in written form, and to measure vocabulary size in aural format they use A-Lex (Milton & Hopkins, 2006), which is designed as an equivalent of X-Lex but where words are just heard. To measure performance in the four skills they used IELTS sub-scores. Data was collected from 30 students at intermediate and advanced level, from a variety of national and language backgrounds, attending a pre-session course in UK. Milton et al. hypothesised that scores from the written vocabulary size test would correlate best with IELTS scores for reading and writing, that scores from the aural vocabulary size test would correlate best with the IELTS scores for speaking, and that a combination of the two vocabulary tests would best explain scores from the IELTS listening test where the test format involved words in both written and aural form.

Milton et al.'s results, like Stæhr's, show statistically significant correlations between vocabulary size and language performance measures, although the strength of the relationships is weaker. They also very strongly suggest that vocabulary knowledge need not be held in both written and aural form combined, and that words can be held in one form only. The correlation between the

two vocabulary size tests is moderate to poor at 0.41, even if the relationship is still statistically significant. Interestingly, it appears that elementary level learners have knowledge predominantly in aural form, while the more advanced learners tend increasingly to grow lexicons where words appear to be known through written form only (see also Milton & Hopkins, 2006; Milton & Riordan 2006). It seems that vocabulary size can predict oral skills comparably with written skills provided that vocabulary size is measured appropriately. The correlation between A-Lex and speaking scores (0.71) is very similar to the correlations observed between X-Lex and reading and writing scores (0.70 and 0.76).

Regression analysis suggests that vocabulary size can explain broadly similar amounts of variance in all the four skills. If the relationship is assumed to be linear, and one should bear in mind that for oral skills in particular this need not be the case, then between 40 % and 60 % of variance in sub-skills scores can be explained through the single variable of vocabulary size. Variance in the listening sub-test, which involves both reading questions and listening for answers, is best explained through a combination of the written and aural sub-scores. Analysis using binary logistic regression, used because the relationship may not be linear, produces comparable results explaining between 41% and 62% of variance in the ability to score grade 5 or above on the IELTS sub-tests. The fact that binary logistic regression explains more variance in the speaking scores (Nagelkerke  $R^2 = 0.61$ , Cox & Snell  $R^2 = 0.45$ ) than the linear regression (Adjusted  $R^2 = 0.40$ ) is tentatively suggested by Milton et al. as evidence that the relationship between vocabulary size and performance in tests of speaking skill is non-linear, although differences in the way these scores are calculated make this a highly subjective interpretation.

The significance of these results is to confirm the importance of the vocabulary size dimension in all aspects of foreign language performance. Vocabulary size, calculated appropriately, appears consistently to explain about 50% of variance in the scores awarded to learners for their performance in the sub-skills of language, including speaking skills where hitherto the relationship has been assumed to be less strong. The fact that, as in explaining listening sub-scores, measurements for different aspects of vocabulary knowledge can be aggregated to enhance the explanatory power of vocabulary in the four skills suggests that continuing to investigate the various dimensions of vocabulary knowledge may yield useful insights.

#### *4.3. Schoonen (2010)*

Recent work by Schoonen and his colleagues has investigated the influence of the dimensions of vocabulary knowledge additional to size on the development of language skills. They have tried to use a combination of size and depth and flu-

ency measures to better explain variation in performance in the language skills. These have been usefully summarised in a conference paper (Schoonen, 2010).

Data on the influence of lexical variables on reading comprehension and writing proficiency is drawn from a number of studies (including Schoonen et al., 2011), which includes vocabulary size and automaticity (or fluency) with other predictor variables. Vocabulary size was measured using the Vocabulary Levels Test (VLT) using the Schmitt et al. (2001) version. Speed was measured using two computer-delivered tests where speed of word recognition and speed of word retrieval could be measured. The other variables were collected using a metacognitive knowledge questionnaire and grammatical knowledge tests. It is acknowledged that one of the problems associated with this approach and the analysis of the data it produces is multicollinearity. Because the lexical variables and other factors such as grammatical knowledge may all be influenced by the frequency of their occurrence in natural language and the degree of exposure a learner has had to the foreign language, these variables may correlate closely with each other. Separating out the impact of individual variables from each other may be difficult.

The results show that vocabulary size and the vocabulary speed measures produce statistically significant and positive correlations with scores from the reading and writing tests. It is concluded that the predictiveness of vocabulary size can be enhanced when combined with their measures of speed or fluency. Lexical variables in this study explain in the range of 30% of the variance of reading and writing scores, slightly smaller than obtained in either of the Stæhr or Milton et al. studies. In this analysis, too, vocabulary is a good predictor of performance and the proportion of variance it explains is substantial.

#### *4.4. Vocabulary knowledge and the four skills*

Perhaps the most important conclusion that emerges from the research is the importance of vocabulary knowledge in being able to understand and communicate in a foreign language. The studies reported above, among others, demonstrate this clearly, showing a moderate to strong relationship between vocabulary measures and the ability to read, write, listen, and it seems also speak, in the foreign language. Generally speaking, the more words a learner knows, the more they are likely to know about them, and the better they are likely to perform whatever the skill. The single factor of vocabulary can explain up to 50% of the variance in performance in scores gained from tests of the four skills. This is a large figure, given that variation might be expected in learners' ability to apply their lexical knowledge - some are likely to be more skilled in using what they know than others. Nonetheless, because this close connection between vocabulary knowledge and skill exists, it is not perhaps surprising that

vocabulary sizes can be linked to language levels as those presented in the CEFR and that vocabulary size can be used as a reliable placement measure. The expectation that oral skills would not be so closely linked to vocabulary size has not emerged in these studies possibly because the measures of skill used relate to measures such as IELTS scores, which are rather academic and might favour a more linear relationship than would be the case if the skills were measured in a non-academic context. Unusually in the spoken register, the skills rewarded in the IELTS speaking sub-test may benefit from the more extensive use of infrequent vocabulary. This conclusion has emerged despite the clear evidence that in successful language performers words are held predominantly in the written form and have presumably been learned by reading rather than through oral interaction.

Stæhr (2008) has remarked that the explanatory power of vocabulary size in explaining variance in scores on language skills suggests that vocabulary size may be *the* determinant factor, pre-eminent among the other factors which may be at work in performing in and through a foreign language. Schoonen's findings, however, suggest that this may be an exaggeration, since size and other factors appear so closely linked and the importance of other variables exceeds vocabulary in his study. Nonetheless, vocabulary knowledge, and vocabulary size in particular, are clearly a very major contributor to success in language performance. It has emerged that knowledge of the most frequent 2,000 words, in particular, is an important feature in successful communication through a foreign language. There is a caveat here, in that the findings suggest that in oral skills the importance of vocabulary knowledge diminishes with increasing size rather faster than it does in skills that involve the written word. The reason for this is worth consideration and the best explanation available is that this is connected with coverage and differences in the way we handle written and spoken language. Corpora suggest that, in English language for example, the most frequent words in a language are even more frequent in spoken language than in written language. Adolphs and Schmitt's (2003) analysis of spoken data in CANCODE indicates that important coverage thresholds such as the 95% coverage figure for general comprehension might be reached with between 2,000 and 3,000 words; perhaps half the figure needed to reach the same threshold in written discourse.

The studies by Stæhr (2008), Milton et al. (2010) and Schoonen (2010) discussed above suggest that, because the dimensions of vocabulary knowledge are so closely linked, a single measure of vocabulary knowledge is likely, by itself, to be a good indicator of skill and level in a foreign language. Because vocabulary breadth in English is now easily measurable using reliable tests for which we have normalised scores, perhaps it is not surprising if vocabulary size or breadth has become particularly closely associated with performance in the four



skills. It seems from the studies above, however, that other dimensions also contribute to performance, perhaps as much as size, and that a combination of scores for size and depth, or size and speed, for example, can add up to 10% to the explanatory power of vocabulary knowledge in skills performance. Very crudely, the more sophisticated the measures of vocabulary knowledge, the more they are likely to explain variance in performance in the four skills, up to the level of around 50%. Beyond that point other factors will be needed to improve the explanatory power of any model. These could be knowledge factors, such as grammatical knowledge, or skill factors in the ability that users have in applying their knowledge when listening, reading, speaking or writing. This is clearly an avenue for further research.

The studies discussed above also allow us to reconsider the concept of lexical space explained at the outset of the chapter: the idea that learners can be characterised differently according to the type of knowledge they have of the words they know in their foreign language, and this can explain how they vary in performance. One interpretation why the depth and size dimensions correlate so well is that they are essentially the same dimension, at least until learners become very knowledgeable and competent and sufficient words are known for subtlety in choice or combination to become possible (see Gyllstadt, this volume). The convenient rectangular shape in Figure 1 is transformed into something much narrower at the outset of learning where lexical size is paramount, and becomes wider at the most competent levels where increased depth becomes a possibility and a potential asset. Co-linearity is noted by Schoonen who suggests another possibility (Schoonen, personal correspondence), that there will be an 'equal' development in all three dimensions, and all three will be strongly correlated, but this is probably a spurious correlation due to language exposure as common cause. Theoretically, it remains possible to have uneven profiles, including differences in breadth and depth, but to evaluate this experimental studies would be required where one dimension only is trained, for example speed, as in Snellings, Van Gelderen & De Glopper (2004).

#### *4.5. Vocabulary knowledge, theories of language learning, and implications for pedagogy*

At the outset of this chapter I suggested that there was a contradiction between much pedagogical theory and practice and recent SLA theories, as regards the importance and relevance of vocabulary knowledge to the process of acquiring proficiency in a foreign language. Current methods and approaches to language teaching fail to consider how vocabulary should be systematically built into the curriculum or suggest that this would not be appropriate assuming that the acquisition of vocabulary is merely incidental to the process of lan-

guage learning as a whole. Learners will not need a systematically constructed language lexicon of any particular size or content in order to achieve success in their second language (see for example Häcker 2008; Milton 2006). Contrary to this, recent theories, for example Ellis's Lexical Learning Hypothesis (Ellis, 1997), suggest that learning a large vocabulary is central to successful language learning and drives the learning of other aspects of language. I suggested that some of the recent research into vocabulary knowledge and performance in the four skills of language might illuminate this debate and provide better guidance for best practice in teaching and learning foreign languages.

If vocabulary were really marginal to the process of developing communicative ability in a foreign language, it might be expected that learners with large vocabularies and extensive lexical knowledge would, presumably, perform comparably with learners with much smaller vocabularies. If, however, the volume of vocabulary a learner knows is driving the acquisition of the other aspects of language and overall proficiency, then a much closer association might be expected. Learners with small or poorly developed vocabularies could not be as proficient nor as fluent in performing through the foreign language. It was suggested that because of the complexity in describing vocabulary knowledge completely and in measuring the variety of aspects that can be involved, the relationship between vocabulary and language learning might be difficult to capture and to measure. The research in this area suggests that the relationship between vocabulary and the development of skills in performing in the foreign language can be modelled and measured and several key features emerge.

Part of this modelling process can now make important distinctions in our understanding of the structure of the mental lexicon and the nature of the vocabulary needed to achieve communicative goals. It appears that word knowledge may be in phonological form or orthographic form and that important communicative goals are likely to be achieved with fewer words in speech than in writing. Surprisingly, it appears that a substantial volume of a highly fluent foreign language user's knowledge may reside in the realm of orthographic knowledge only. Speakers of languages using alphabetic systems of writing, where the spelling clearly relates to the pronunciation, still manage, it seems, to avoid storing this information or storing it correctly and so fail to recognise by sound words they can recognise in writing. Such a possibility has been discussed by Suarez and Meara (1989) and Segalowitz and Hulstijn (2005), who suggest that advanced learners develop a direct route to meaning from the written form, cutting out any intermediate phonological form, but there has been little to demonstrate that this can occur before. It implies that high fluency is linked to high literacy and the ability to access large amounts of particularly written text, to access the necessary infrequent words, and recognise them by shape or by morphological structure providing a route to meaning, which does not rely on

phonological coding. Learners without this high literacy and who are tied to phonological decoding may develop more balanced lexicons with orthographic and phonological word knowledge more equal in size as suggested in Milton and Hopkins (2006) and Milton and Riordan (2006). However, the price to be paid for this, perhaps through the slowness of the reading process and the extra burden on memory, is that the lexicon tends to grow more slowly, limiting communicativeness in the written domain.

The research summarised above appears to support theories such as Ellis's Lexical Learning Hypothesis. Vocabulary development, however measured, appears to mesh very closely with other features of language such as grammatical development, and also with overall language ability. Developing learners' vocabulary knowledge appears to be an integral feature of developing their language performance generally. The link has not been established in a strongly causal sense and while it is not yet clear that the vocabulary knowledge is driving the other aspects of language development, vocabulary certainly appears to develop in size and depth alongside every other aspect of language. This very strongly supports the idea, as in the lexical approach (Lewis & Hill, 1997), that vocabulary should be built more explicitly into the development of any good language curriculum. This could be in the form of indicating particular words to be learned, as in the most frequent words in any language, but it might imply the introduction of size as a metric into curricula as a means of setting appropriate targets and monitoring progress without dictating the content of learning directly.

Even though this may seem quite commonsensical, we have evidence from the UK that details of vocabulary can be systematically downplayed from formal curricula in line with methodological approaches such as the Communicative Approach. Curriculum descriptions for B1 level foreign language exams in UK (e.g. Edexcel, 2003, for French) routinely contain only minimal core vocabularies of around 1,000 words, levels of vocabulary which are incompatible with performance attainment at B1 level observed elsewhere in Europe (Milton & Alexiou, 2009). We also have evidence that the teaching of foreign language vocabulary following these curricula rarely extends beyond 1,000 words at B1 level (Milton, 2006; 2008; David 2008). In other countries (as indicated in Milton & Alexiou, 2009) CEFR levels have an expectation of rather greater vocabulary knowledge than in the UK and since it is highly unlikely that learners can be as communicative with 1,000 words at B1 level as with the 2,000 or more words required for this level elsewhere in Europe, there is a clear mismatch in the applications of the CEFR level which vocabulary size estimates can demonstrate.

## References

- Adolphs, S. & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425-438.
- Albrechtsen, D., Haastrup, K., & Henriksen, B. (2008). *Vocabulary and writing in a first and second language: Processes and development*. Basingstoke: Palgrave Macmillan.
- Alderson, J. C. (1984). Reading in a foreign language: A reading or a language problem? In J.C. Alderson & A.H. Urquhart (Eds.), *Reading in a Foreign Language* (pp. 1-24). London: Longman.
- Anderson, R. C. & Freebody, P. (1981). Vocabulary Knowledge. In J. T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews* (pp. 77-117). Newmark: International Reading Association.
- Brumfit, C. (1984). *Communicative Methodology in Language Teaching*. Cambridge: Cambridge University Press.
- Chomsky, N. (1995). *The Minimalist Program*. (Current Studies in Linguistics, 28.) Cambridge, MA: MIT Press.
- Cook, V. (1998). Review of Skehan, P. (1998) *A Cognitive Approach to Learning Language*. Oxford: Oxford University Press. Accessed at <http://homepage.ntlworld.com/vivian.c/Writings/Reviews/SkehanRev.htm> on 16 Feb 07.
- Council of Europe (2001). *Common Framework of Reference for Languages*. Cambridge: Cambridge University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). Editors' introduction: conventions, terminology and an overview of the book. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 1-32). Cambridge: Cambridge University Press.
- David, A. (2008). Vocabulary breadth in French L2 learners. *Language Learning Journal*, 36(2), 167-180.
- Edexcel (2003). Edexcel GCSE in French (1226) First examination 2003 June 2003. Accessed at <http://www.edexcel.com/quals/gcse/gcse-leg/lang/french/Pages/default.aspx> on 08.03.2011.
- Ellis, N. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description Acquisition and Pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.
- Ellis, R. (1994). Factors in the incidental acquisition of vocabulary from oral input: a review essay. *Applied Language Learning*, 5(1), 1-32.
- Gyllstad, H. (2007). *Testing English Collocations - Developing Receptive Tests for Use with Advanced Swedish Learners*. Lund University: Media-Tryck.
- Häcker, M. (2008). Eleven pets and twenty ways to express one's opinion: the vocabulary learners of German acquire at English secondary schools. *Language Learning Journal*, 36(2), 215-226.

- Harris, V. & Snow, D. (2004). *Classic Pathfinder: Doing it for themselves: focus on learning strategies and vocabulary building*. London: CILT.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan.
- Laufer, B. & Nation, P. (1999). A productive-size test of controlled productive ability. *Language Testing*, 16(1), 33-51.
- Lewis, M. & Hill, J. (1997). *The Lexical Approach; the state of ELT and the way forward*. Boston, Mass: Heinle.
- Lightbown, P. & Spada, N. (2006). *How Languages are Learned (3<sup>rd</sup> Ed)*. Oxford: Oxford University Press.
- Littlewood, W. (1983). *Communicative Language Teaching*. Cambridge: Cambridge University Press.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35-53). Cambridge: Cambridge University Press.
- Meara, P. & Milton, J. (2003). *X\_Lex, The Swansea Levels Test*. Newbury: Express.
- Meara, P. & Wolter, B. (2004). V\_Links, beyond vocabulary depth. *Angles on the English Speaking World*, 4, 85-96.
- Milton, J. (2006). Language Lite: Learning French vocabulary in school. *Journal of French Language Studies* 16(2), 187-205.
- Milton, J. (2008). French vocabulary breadth among learners in the British school and university system: comparing knowledge over time. *Journal of French Language Studies*, 18(3), 333-348.
- Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Vedder, I. Bartning, & M. Martin (Eds.), *Communicative proficiency and linguistic development: intersections between SLA and language testing research* (pp. 211-232). Second Language Acquisition and Testing in Europe Monograph Series 1.
- Milton, J. & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: do vocabulary tests underestimate the knowledge of some learners. *The Canadian Modern Language Review*, 63(1), 127-147.
- Milton, J. & Riordan, O. (2006). Level and script effects in the phonological and orthographic vocabulary size of Arabic and Farsi speakers. In P. Davidson, C. Coombe, D. Lloyd, & D. Palfreyman (Eds.), *Teaching and Learning Vocabulary in Another Language* (pp. 122-133). UAE: TESOL Arabia.
- Milton, J. & Alexiou, T. (2009). Vocabulary size and the Common European Framework of Reference for Languages. In B. Richards, H.M. Daller, D.D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary Studies in First and Second Language Acquisition* (pp. 194-211). Basingstoke: Palgrave Macmillan.

- Milton J., Wade, J. & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torrealblanca-López (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 83-98). Bristol: Multilingual Matters.
- Mitchell, R. & Myles, F. (2004). *Second Language Learning Theories*. London: Hodder Arnold.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- O'Dell, F. (1997). Incorporating vocabulary into the syllabus. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp. 258-278). Cambridge: Cambridge University Press.
- Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, 56(2), 282-307.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, N. (2008). Review article: instructed second language vocabulary learning. *Language Teaching Research* 12(3), 329-363.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18 (1), 55-88.
- Schoonen, R. (2010). The development of lexical proficiency knowledge and skill. Paper presented at the Copenhagen Symposium on *Approaches to the Lexicon*, Copenhagen Business School on 8-10 December 2010. Accessed at <https://conference.cbs.dk/index.php/lexicon/lexicon/schedConf/presentations> on 03.03.2011.
- Schoonen, R., Van Gelderen, A., Stoel, R., Hulstijn, J., & De Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary-school students. *Language Learning*, 61(1), 31-79.
- Segalowitz, N. & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 371-388). Oxford: Oxford University Press.
- Snellings, P., Van Gelderen, A., & De Glopper, K. (2004). The effect of enhanced lexical retrieval on L2 writing. *Applied Psycholinguistics*, 25(2), 175-200.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152.
- Suárez, A. & Meara, P. (1989). The effects of irregular orthography on the processing of words in a foreign language. *Reading in a Foreign Language*, 6(1), 349-356.
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics* 22(2), 217-234.
- Wesche, M. & Paribakht, T. A. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13-40.
- Wilkins, D. A. (1972). *Linguistics in Language Teaching*. London: Arnold.
- Wolter, B. (2005). *V\_Links: A New Approach to Assessing Depth of Word Knowledge*. PhD Dissertation, University of Wales Swansea.
- Zimmerman, K. J. (2004). *The role of Vocabulary Size in Assessing Second Language Proficiency*. MA dissertation, Brigham Young University.

# *FREQUENCY 2.0: Incorporating homofoms and multiword units in pedagogical frequency lists*

Thomas Cobb

Université du Québec à Montréal

The importance of frequency as a principle for organizing language learning, while long promoted in principle (Palmer, 1941; West, 1953), has recently become feasible in practice with three new developments: theoretical support from acquisition theorists (Ellis, 2002); the assembly of truly enormous, representative and accessible language corpora (Davies, 2011; Leech, Rayson & Wilson, 2001); and the extraction of pedagogically relevant lexical information (Nation, 2006) and grammatical information (Biber et al., 1999) from them. Since about 1990, this frequency information has regularly been deployed in the development of language courses and learning resources, particularly lexical resources such as dictionaries and tutorial computer programs for learning vocabulary. Now, however, at least in the area of lexis, the frequency approach must face two consequences of its own success: larger corpora and stronger tools of analysis have revealed not just useful ranked lists of word forms, but (1) the extent of homonymy and homography hidden within them, and (2) the extent of multiword units with meanings independent of their component words. The present paper makes the case for including both types of information in pedagogically oriented frequency lists. It shows firstly why this should be done, then reviews some new research that is making it possible, and finally develops and pilot-tests a way of doing it. The underlying theme is that the technologies that raised the problems of homofoms and multiword units can also be used to solve them.

## **1. Introduction**

Applying corpus insights to language learning is slow work with roughly one or two interesting advances per decade. In terms of lexis and frequency: Tim John's corpus and concordance package MicroConcord became available in 1986, enabling language practitioners to build concordances and calculate word frequencies in their own texts and compare these to more general word frequencies in the small corpora bundled with the program. In the 1990's, Heatley and Nation's (1994) Vocabprofile, a computational deployment of West's (1953) General Service List (GSL) integrated with a series of academic lists, allowed

practitioners to perform MicroConcord's two functions together: analyzing texts in terms of the frequency of their individual words both in a particular text and in the English language as a whole. The 2000's have been largely devoted to exploiting the 100-million word British National Corpus (BNC; Aston & Burnard, 1998) and the frequency lists derived from it (Leech et al., 2001). Some important exploitations have been the pedagogical adaptation of these lists (Nation, unpublished), and then their incorporation in a vocabulary test (Beglar & Nation, 2007), deployment in a Vocabprofile update (Nation, 2006), use in a variety of research enterprises (discussed below), and dissemination to researchers, teachers and learners on the World Wide Web (partly via the *Compleat Lexical Tutor* Website, or *Lextutor*, [www.lextutor.ca](http://www.lextutor.ca)). A likely near-term development will be the incorporation of US English into the scheme from the COCA, or Corpus of Contemporary American English (Davies & Gardner, 2010).

A key element in the pedagogical adaptation of the BNC lists is the expansion of the grouping unit from the lemma (headword and inflections) to the word family (lemma and transparent derivations; Bauer & Nation, 1993). For example, the lemma for the noun *cup* would be *cup* and *cups*, but the family would be these plus the derived verb *to cup* (one's hands), which involves a changed part of speech but not a change in the basic meaning. The development of the family concept is based on learning principles rather than linguistics or computational principles: a learner who understands *cup* will have no problem understanding *cup your hands*.

The appeal of pedagogically oriented lexical frequency information in the language teaching industry appears to be large, an impression that can find quantitative support in Lextutor's user statistics. Since coming on line in 2005, Lextutor's user base has doubled every year and currently generates more than 10,000 concordances, frequency lists, or lexical profiles daily. Lextutor's most utilized resource is Web Vocabprofile, an online adaptation of both Heatley and Nation's original Vocabprofile (1994) and Laufer and Nation's (1995) Lexical Frequency Profiler (LFP), which categorizes every word of any text in terms of both family membership as well as the overall rank of the family in either the GSL or the BNC, calculating a profile by percentage. For example, five of the six words in this sentence, *The cat sat on the mat*, are very frequent (from the BNC's first 1,000 word families by frequency), but one, *mat*, is less frequent (from the fourth 1,000). One can thus state that the text comprises 83% first thousand items, and go on to predict that this text could probably be handled by an intermediate learner who could be predicted to know five of its six words leaving just one to work out from context or look up.

Teachers and learners use this type of analysis to determine and modify the difficulty level of texts. Frequency profiling thus connects the rough-and-ready



instructional design end of language learning with the frequency-based learning principles of acquisition researchers like Ellis and Larsen-Freeman (e.g., 2009) at the other. Vocabprofile analysis is fairly simple in both concept and function, and has received empirical validation in both English (Laufer & Nation, 1995; Morris & Cobb, 2004) and French (Ovtcharov, Cobb & Halter, 2006; Lindqvist, 2010) and is a mainstay in the ongoing text coverage and comprehension research (Nation, 2006; Schmitt, Jiang & Grabe, 2011; van Zeeland & Schmitt, in press).

Taking Vocabprofile as an example of how frequency information is being used in the language learning field, we can continue with a finer grained account of the slow but steady evolution roughed out above. As already mentioned, the original frequency list at the heart of Vocabprofiling (West's, 1953, two thousand-item General Service List) has now been replaced by the BNC list (Leech et al., 2001) as adapted and divided by Nation (unpublished) into 14 family thousand-lists. The increase in the number of lists from two to 14 allows much finer grained profiles of texts, clearer distinctions between texts, and a substantial reduction in the percentage of words that cannot be categorized. Other developments in the concept and software are mainly modifications suggested by practitioners, including colour coding of frequency zones, automated treatment of proper nouns, and the sequential re-analysis of evolving text modifications (Cobb, 2010). However, these and related developments have not involved a rethinking of the basic idea, which is to match text words to static frequency information straight out of a computer program whose knowledge of language is limited to counting up the items between empty spaces and judging where they are the same or different to each other and to words in a database.

While it has been possible to do a good deal of frequency work using this simple definition of *word*, the definition was based on two assumptions known to be incorrect but believed to pose relatively minor problems. It was assumed that homoforms (an umbrella term for homonyms, like *river banks* and *money banks*, and homographs, like *to read* and *well read*) could be provisionally ignored. It was also assumed that multiword units (MWUs, phrases with meanings independent of their individual words, like *up to you* and *a lot*) could be overlooked, at least for a while. But larger corpora and growing familiarity with their contents has now revealed the extent of the homoforms and MWUs that lie hidden in between-the-spaces frequency lists. That is, many single words are really two words, and many phrases are really single words. These arguably merit separate entries in a pedagogical frequency list, as well as revamped frequency ratings and pedagogical emphases. It may be that *a\_lot* (of anything) should be taught without reference to *a lot* (to build a house on), and *banks* (for money) should be introduced to beginners and *banks* (of rivers) reserved for later, rather than mixing everything together, as happens at pres-

ent and is tacitly supported by existing Vocabprofile software. Without accounting for this information within and beyond the space-defined word form, existing frequency profiles are almost certainly inaccurate to some unknown degree. Or to put it another way, frequency profiling could be even more useful than it is now. Fortunately, much of the human and computational spade work has already been done to achieve this.

## 2. FREQUENCY 2.0: Why it is needed

West's hand-made General Service List (1953) of 2,000 high-value lexical items for English teaching made careful distinctions not only between homofoms, which are clearly different words (*money banks* and *river banks*), but also between main senses of words (*cloud banks* and *river banks*). The limitations of this list are that it is small (2,000 word families), intuitive (with only rudimentary frequency information), narrowly pedagogical (no vulgarities allowed), and largely inapplicable to text creation or modification except through handwork with small texts. These shortcomings have now been more than compensated for by lists based not only on huge corpora like the BNC, but also by the systematic inclusion of *range* (the distribution of items across the BNC's 100 subdivisions) as a second consideration in their construction. And yet it is ironic that in the newer lists, the old distinctions have temporarily been lost between both word senses and homofoms. Distinguishing word senses may not be crucial to such an enterprise, if, as Beretta, Fiorentino and Poeppel (2005) argue, these are normally computed in real time from a single entry in the mental lexicon. Nation (e.g., 2001) has long argued for a pedagogy focusing on the "monosemic" concept underlying the polysemes. Nonetheless, homofoms do pose a problem.

The BNC frequency list produced by Leech et al. (2001), while lemmatized for part of speech, does not distinguish between different words that are merely linked by a common word form. A trip to the Web version of the BNC (at <http://bncweb.lancs.ac.uk/>) reveals that the program is able to output lemmas (related morphologies of the same word form) but not distinguish homofoms. Nor does the newer list by Davies and Gardner (2010) drawing on the even larger Corpus of Contemporary American English (COCA, 425 million words, see Figure 1).

The combined meanings of *bank* shown in Fig. 1 place the word-form at rank 701 in the frequency list, hence in the first 1,000 words by frequency. But this placement is almost certainly an artifact of lumping the two *banks* together, as shown by the collocates *account*, *loan*, and *river* in line 3. *Bank*<sub>1</sub> and *bank*<sub>2</sub> are clearly distinct words linked mainly by a resemblance of form (and possibly a common etymology that few language users would be aware of). The reason

**Figure 1.** Homoform lumping in Davies & Gardner (2010)

<p><b>701 bank n</b>  <i>adj</i> central, national, federal, large, commercial,  international, reserve, foreign, major, outer  <i>noun</i> world., .account, loan, river, money, invest &lt;  credit, development, fund, saving <i>verb</i> rob., buy,  .lend, own, .fail, finance, charge, issue, borrow,  operate  52366   0.93</p>
--

for failure to distinguish between the two *banks* is, of course, clear. The amount of textual information that is summarized in a small compilation like Figure 1 is vast (the figure 52,366 at the bottom refers to the number of instances of *bank* in the COCA corpus), such that there is no easy way to insert human judgment into the process. A human investigation of the context for each of these entries, followed by a count-up, is presumably the only way to tell the different *banks* apart, and this is an arduous task.

However, with some quick and dirty human-computer cooperation based on random sampling, this prising apart can be done for many practical purposes. For example, here is a mini-experiment for the word-form *bank* based on the 50 random non-lemmatized samples offered for free by the BNC website at <http://www.natcorp.ox.ac.uk/>. Entering a search for *bank* reveals that the BNC contains 17,603 lemmatized instances of this item (all noun forms combined). Then, eyeballing and counting up the separate meanings from the available 50 random concordance lines over 10 runs, we find a remarkably consistent 43 to 50 lines of money or blood *bank* and only 5 to 7 of river or cloud *bank*. Thus a rough 86% to 96% of the 17,603 uses pertain to money *bank*, or minimally 15,138 occurrences, so it is probably safe in its first-1,000 position (see Figure 1 for BNC cut-offs). But *river bank* is instead a medium frequency item (7 uses in 50, or 14% of the BNC's 17,603 total occurrences amounts to 2,465 occurrences, placing it near the end of the third 1,000 by frequency).

The recent large-corpus based lists also fail to distinguish between MWUs that are compositional, like *a+lot* (to build a house on), and ones that are non-compositional, like *a\_lot* (of money), in the sense that the individual words do not add up to the accepted meaning of the unit (as suggested in the notation of an underscore rather than a plus sign). But once again the corpora make it possible to do so. Passing large corpora through computer programs identifies a wealth of information about all the ways that words co-occur in more than random sequences and the extent to which they do so (Sinclair, 1991). In Figure 1, we see

COCA's main collocates of *bank*, with bullet signs indicating whether each falls consistently before or after the key word (*world•* = World Bank, *•account* = bank account). What the computer output does not show is that not all collocates are created equal. In some, the node word and collocate retain their independence (*an international bank*), while in others they do not (*World Bank, Left Bank, West Bank*). Degree of connectedness can to some extent be predicted by frequency of found versus predicted co-occurrence in such measures as mutual information or log-likelihood, as calculated by programs like BNC-Web (which gives *international bank* a mutual information (MI) value of 3.04 and *West Bank* a value of 5.82 or almost double).

In two BNC-based studies, both again involving computational analysis with human follow-up, Shin and Nation (2007) and Martinez and Schmitt (2012) identified unexpectedly large numbers of recurring word strings in the highest frequency zone of the language. Shin and Nation's co-occurrences (*you know, I think, a bit*) were for the most part compositional items which, if incorporated into the existing frequency scheme, would count as first 2,000 items. There was no proposal actually to incorporate these items into standard frequency lists, but merely to argue for their importance to language learners. Martinez and Schmitt's focus, on the other hand, was specifically on high-frequency co-occurrences that they judged to be non-compositional, or idiomatic, i.e. which have, in specific environments, independent meanings and hence deserve their own places within standard frequency lists. Using a methodology to be described below, these researchers identified 505 such MWUs in the first five thousand-lists of the BNC (or just over 10%), distributed over these lists in the manner shown in Table 1.

**Table 1.** Distribution of Martinez and Schmitt's MWUs by 1000-group

Number of MWUs	Zone (by 1000)	Proportion of zone (%)
32	1k	3.2
75	2k	7.5
127	3k	12.7
156	4k	15.6
97	5k	9.7

Incorporating homoform and MWU information into frequency lists could cause quite extensive changes in their composition. If a word form like *arm*, a first thousand item, were found to be about equally implicated in weaponry and anatomy, it is doubtful that either of these would remain a first 1,000 item: one or both might be bumped down to second thousand or beyond. If Martinez and

Schmitt’s 505 MWUs were given their rightful places and added to the current frequency lists, then quite a number of existing items would be displaced from zone to zone (which are arbitrary divisions in any case). The result would be a set of lists something like the one imagined in Table 2.

**Table 2.** The type of frequency list needed

1000 List	3000 List
bank_1	bank_2
of_course	course
something	something_of_a

Incorporating these two kinds of information would also have strong effects on the deployment of frequency information in the profiling of novel texts. Profiling would no longer be a simple matter of matching a word in a text to its family headword and thence to its counterpart in a frequency list. Rather, the profiler would have to interpret both homoforms and MWUs in context, in order to determine which meaning of a homoform was applicable (*bank\_1* or *bank\_2*), and in the case of MWUs whether a particular string was compositional or non-compositional (‘look *at all* the bugs’, or ‘I don’t like bugs *at all*’). It is this incorporation of context that is the qualitative transformation implied in the term Frequency 2.0.

### 3. The feasibility of reworked frequency lists

Frequency profiling up to present has been based on single word forms. It has relied on matching stable word frequencies to equivalent word forms in a given text. The modification proposed here involves not only extensive modification of the lists, but also a real-time contextual analysis of each potential homoform or MWU to determine its true identity in a particular text. These are dealt with in turn.

#### 3.1. Multiwords

Whether for homoforms or MWUs, the first task is to identify the item involved, assign it to a category (‘money *bank*’ or ‘river *bank*’; ‘a lot of money’ or ‘build on a *lot*’), calculate the frequency of each in a large corpus, and give each a place in the standardized vocabulary lists used by course developers, test writers, and computer programs like Vocabprofile. A methodology for doing this work is under development in a new crop of student research projects in vocabulary.

**Table 3.** The highest frequency MWUs from Martinez and Schmitt (2012)

Integrated List Rank	MWU	Frequency (per 100 million)	Example
107	<b>HAVE TO</b>	83092	I exercise because I <b>have to</b> .
165	<b>THERE IS/ARE</b>	59833	<b>There are</b> some problems.
415	<b>SUCH AS</b>	30857	We have questions, <b>such as</b> how it happened.
463	<b>GOING TO (FUTURE)</b>	28259	I'm <b>going to</b> think about it.
483	<b>OF COURSE</b>	26966	He said he'd come <b>of course</b> .
489	<b>A FEW</b>	26451	After <b>a few</b> drinks, she started to dance.
518	<b>AT LEAST</b>	25034	Well, you could email me <b>at least</b> .
551	<b>SUCH A(N)</b>	23894	She had <b>such a</b> strange sense of humor.
556	<b>I MEAN</b>	23616	It's fine, but, <b>I mean</b> , is it worth the price?
598	<b>A LOT</b>	22332	They go camping <b>a lot</b> in the summer.
631	<b>RATHER THAN</b>	21085	Children, <b>rather than</b> adults, tend to learn quickly.
635	<b>SO THAT</b>	20966	Park it <b>so that</b> the wheels are curbed.
655	<b>A LITTLE</b>	20296	I like to work out <b>a little</b> before dinner.
674	<b>A BIT (OF)</b>	19618	There was <b>a bit</b> of drama today at the office.
717	<b>AS WELL AS</b>	18041	She jogs <b>as well as</b> swims.
803	<b>IN FACT</b>	15983	The researchers tried several approaches, <b>in fact</b> .
807	<b>BE LIKELY TO</b>	15854	To be honest, I'm <b>likely to</b> forget.
825	<b>GO ON</b>	15610	He <b>went on</b> for a while before stopping for lunch.
845	<b>IS TO</b>	15232	Obama <b>is to</b> address the media this afternoon.
854	<b>A NUMBER OF</b>	15090	<b>A number of</b> concerns were raised.
879	<b>AT ALL</b>	14650	Do you have any kids <b>at all</b> ?
888	<b>AS IF</b>	14470	They walked together <b>as if</b> no time had passed.
892	<b>USED TO (PAST)</b>	14411	It <b>used to</b> snow much more often.
894	<b>WAS TO</b>	14366	The message <b>was to</b> be transmitted worldwide.
908	<b>NOT ONLY</b>	14110	<b>Not only</b> was it cheap, it was delicious.
913	<b>THOSE WHO</b>	13951	He would defend <b>those who</b> had no voice.
934	<b>DEAL WITH</b>	13634	The police had several issues to <b>deal with</b> .
939	<b>LEAD TO ('CAUSE')</b>	13555	Excessive smoking can <b>lead to</b> heart disease.
951	<b>SORT OF</b>	13361	It's <b>sort of</b> why I'm here.
974	<b>THE FOLLOWING</b>	12963	He made <b>the following</b> remarks.
984	<b>IN ORDER TO</b>	12762	We shared a room <b>in order to</b> reduce costs
988	<b>HAVE GOT (+NP)</b>	12734	I don't know what he <b>has got</b> planned.

The largest investigation into non-compositional MWUs to date was performed by Ron Martinez and his PhD supervisor Norbert Schmitt (Martinez & Schmitt, 2012). These researchers set Scott's text analysis program WordSmith Tools 6.0 the task of generating a list of all the recurring 4, 3, and 2-word strings, or n-grams, in the 100-million word BNC, a computer run of just under four days. Lemmas rather than word forms or families were used for this stage of the analysis, so that for example all forms of a verb are included in the analysis (*have to* as well as *had to*) as is occasionally but not consistently marked in Table 3 (in the form of *is/are* and *alan*). From this massive output, those items with fewer than 787 occurrences were eliminated (787 is the cut-off for inclusion in the first 5,000 headwords of the existing BNC-based Vocabprofile scheme, the number 5,000 being chosen for pedagogical relevance as the words most language learners are likely to be concerned with). The surviving 15,000 items were then hand-sorted in a double randomization procedure. For each candidate MWU, WordSmith was asked to generate two random 100-word listings, which were then hand sorted into compositional vs. non-compositional meanings of the MWU. For example, in the case of the phrase *at first*, this process yielded 16 compositional uses like 'attack *at first* light' in a single iteration of this process and also 16 in the other. Non-compositional uses such as '*at first* I wasn't sure' were more frequent; there were 84 non-compositionals in one round and 85 in the other. In cases such as this, where there was a discrepancy, the lower of the two numbers was used. The original raw frequency per 100 million was then multiplied by (in this case) .84 to produce the frequency for the non-compositional meaning of the phrase (for *at first*,  $5177 \times .84 = 4275$ , placing it in the third thousand-list according to the cut-offs shown in Table 5). Following this method, instances of the non-compositional *at all* extrapolated to 14,650 occurrences, and thus it was placed at position 879 in the full BNC list, in other words in the first 1000 group (Table 2). In total, 505 MWUs were thus determined and situated throughout the first five lists. The 35 provisional first thousand level items are shown in Table 3, with BNC frequency and computed list rank.

It is almost certain that these rankings are not final. Some of the examples chosen suggest uncertainty in the groupings (such as the last item in Table 3 – the NP is present only with a transformation). But more broadly, compositionality, as Martinez and Schmitt propose, is a cline or continuum, such that different researchers could have selected different non-compositional units from the computer's offering. Research by Grant and Nation (2006), working with a different idea of compositionality, would suggest a less extensive list than the one proposed by Martinez and Schmitt. They feel that most of the proposed non-compositional MWUs are merely metaphorical extensions of the compositional (if *a lot* with a house on it is a large space, and *a lot* of money is a large

**Table 4.** Eighteen homographs where most common meaning < 90% of 500 concordance lines

<b>MISS</b>	fail to get or have	50.00%	title	50.00%				
<b>YARD</b>	land	56.60%	36 inches	43.40%				
<b>NET</b>	web	59.36%	total	40.64%				
<b>REST</b>	remainder	62.20%	recuperate	37.80%				
<b>RING</b>	(to produce the) sound of a bell	67.47%	circle	32.53%				
<b>WAKE</b>	become awake	75.80%	a track left behind	23.00%	vigil	1.20%		
<b>SPELL</b>	letter-by-letter/ incantation	75.95%	interval of time	24.05%				
<b>LIKE</b>	resembling	76.20%	opposite of dislike	23.80%				
<b>RIGHT</b>	not left	77.40%	legal rights	22.60%				
<b>POOL</b>	water	78.62%	combine resources	21.38%				
<b>LEAVE</b>	part from	78.96%	direction	17.03%	permission	0.80%	leaf	3.0%
<b>BAND</b>	group of people	79.00%	ring	21.00%				
<b>FIRM</b>	business	80.12%	strong/solid	19.88%				
<b>SET</b>	to place/to be firm	80.40%	a collection	19.60%				
<b>ARM</b>	body part	83.00%	weapon	17.00%				
<b>DEAL</b>	an amount	84.00%	to distribute	16.00%				
<b>HOST</b>	of a party	85.28%	multitude	13.91%	consecrated wafer	0.81%		
<b>WEAVE</b>	interlace threads	87.80%	move from side to side	12.20%				



amount of money, then there is a clear similarity between the two, such that they can be seen as members of a single ‘monoseme’). Thus the exact MWUs eventually to be integrated into standard frequency schemes remain to be determined. Nonetheless it seems likely that at least some of Martinez and Schmitt’s selections are not very controversial (*at all*, *as well as* from the first 1,000 list, and *as far as* and *as long as* from the second, clearly have both compositional and non-compositional meanings). It also seems clear that Martinez and Schmitt’s basic methodology for determining such items, a large-scale crunching of matched corpus samples followed by a principled selection by humans and the calculation of a frequency rating, is likely to prove the best means of working toward a standard set of MWUs. Following that, the question will be how to deploy this information in live Vocabprofiles of novel texts, and this is a question that can be tackled while the exact target items are not yet settled.

### 3.2. Homoforms

The work on homoforms was performed by Kevin Parent in the context of doctoral work with Nation. Parent took West’s GSL list of 2,000 high frequency items as a starting point, on the grounds that most homoforms are found in the highest frequency zones and also that these would be of greatest pedagogical relevance. Wang and Nation (2004) had already shown that there were only a handful of such items (about 10) in the 570-word Academic Word List (AWL; Coxhead, 2000; a compendium of third to sixth thousand level items). In the GSL, Parent identified 75 items with two or more headwords in the *Shorter Oxford English Dictionary* (SOED), a dictionary which marks homoforms explicitly with separate headwords. For each of these 75 items, he generated 500 random concordance lines from the BNC, and hand-sorted them according to the SOED’s headwords. He found that for 54 of the 75 items, the commonest meaning accounted for 90% or more of the 500 lines (surprisingly *bank* itself falls into this category, along with *bear* and *bit*; the others can be seen in Table 1 in the Appendix). Some of the remaining items whose homoformy is less skewed are shown in Table 4. Thus, we see in the first row that half of the uses of *miss* pertained to loss, or failing to have or to get something, while the other half occurred in titles (such as *Miss Marple*).

Some points about Table 4 are in order. First, the items are not lemmatized, or divided into parts of speech (POS), but are simple counts of word forms. This is because while the different meanings of a homoform sometimes correspond to a difference in POS (to *like* somebody vs. look *like* somebody), sometimes they do not (‘I broke my *arms*’ vs. ‘I left the *arms* outside the house’). In the absence of knowing which of these two types of homoform is predominant

in English, Parent's decision was to begin the analysis with word forms. Second, Parent's analysis was confined to true homoforms. This meant that he did not include words with plausible etymological relationships (gold *bar* and drink at a *bar*) and words that while undifferentiated in writing are nonetheless differentiated in speech ('*close* [shut] the door' and '*close* [near] to dawn'). The analysis is now being expanded to include all effective homoforms, roughly 100 items in the highest frequency zones. Third, as shown in Table 4, Parent's list was also confined to cases where the least important meaning of a homoform set was greater than 10% in the BNC. It has often been argued that there is no point in handling items where one meaning is vastly predominant (e.g., Wang & Nation, 2004) since the labour to do so would be great and the differences minor. However, once a methodology for assigning differential frequencies is developed, it is arguably feasible to deal with a larger number of homographs and take less frequently used members into account. For example, as already mentioned the 10% criterion leaves 'river *bank*' lumped with 'money *bank*', which intuitively seems an inaccuracy, and one that can easily be avoided once this analysis and technology is in place. A useful target is probably all the homoforms in the first 5,000 word families where the less frequent member or members account for more than 5% of cases.

Following the calculation of proportions from the 500-word samples, each item would be tagged (possibly as *miss\_1* and *miss\_2*) and assigned by extrapolation its two (or sometimes more) new places in the frequency lists. The evenly divided *miss* is currently a first-1,000 item, with 19,010 lemmatized occurrences in the BNC (raw information available from BNC-Web, <http://bncweb.lancs.ac.uk/>). But if half of these (about 9,505) are apportioned to each meaning of *miss*, then neither *miss\_1* nor *miss\_2* belongs in this first 1,000 category. As the first row of Table 5 shows, only lemmas occurring 12,696 times or more in the BNC qualify as first 1,000 items. Rather, both would feature in the second 1,000 zone (between 4,858 and 12,638 occurrences). In cases where a meaning distinction corresponds to a POS distinction, as with *miss*, then the POS-tagged BNC could provide even more precise information (in this case that the verb is 10,348 occurrences and the noun 8,662, both still in the second 1,000). Counts could be refined and cut-offs change as the proposed amendments are made and items shifted up and down the scale. List building would ideally be left to an expert in developing and applying inclusion criteria, with Paul Nation as the obvious candidate since he has already developed a principled method of balancing frequency and range, spoken and written data, and corpus as well as pedagogical validity, into the existing BNC lists.

**Table 5.** BNC's first five 1000-list cut-offs by token count (for lemmas)

K1	>12639
K2	4858 - 12638
K3	2430 - 4857
K4	1478 - 2429
K5	980 - 1477

Source: R. Martinez (2009)

Table 6 gives a sense of what this new arrangement would look like. Parent's proportions have been multiplied against BNC frequency sums and sorted according to Martinez' cut-offs in order to give a provisional look at the thousand-level re-assignments that could flow from Parent's data in Table 3. The thousand (or *k*) levels in the first column on the left are the current composite k-levels from the BNC; those in the third and subsequent columns are provisional new k-levels for the independent meanings of the homoform. (These are even *highly* provisional since they merely result from multiplying Parent's percentages from 500 lines against BNC word-form totals from 100 million words). The goal in presenting this data at this point is merely to give a flavour of the changes being proposed. Also of interest may be any compatibility issues arising from combining data from several analyses.

Note that the original 1,000-level ratings as presented in Table 6 may not be identical to those in Nation's current fourteen 1,000 lists in all cases (*spell* is shown as 2k in Table 6, but in Vocabprofile output it is 1k). That is because Nation's first two 1,000 levels (1k and 2k) are derived from the spoken part of the BNC corpus (10 million words, or 10 percent of the full corpus), in order to ensure for pedagogical reasons that words like *hello* will appear in the first 1,000 word families. All ratings in Table 6 are based on information from the unmodified BNC, in an attempt to employ a common scale to think about moving items between levels.

Table 6 shows provisional list assignments for the 18 items of Parent's analysis that would be most likely to affect frequency ratings, in that the less dominant meaning is nonetheless substantial (between 10% and 50%). As is shown, only seven items (the top six plus *pool*) would require shifting the dominant member to a lower frequency zone (e.g., from first thousand to second). Similarly, in the remainder of the homoforms identified by Parent, the reanalysis proposed here will most often leave the dominant member of a homoform at its existing level. (The remainder of Parent's analysis is shown in Table 1 in the Appendix [further analysis under way, January, 2013]). So is this reanalysis worth the trouble?

Table 6. Provisional adjustments to frequency ratings for homographs

<b>MISS</b> 19,010 (currently) 1k	fail to get or have	50.00% 9,505 (provisionally) 2k	title	50.00% 9,505 (provisionally) 2k
<b>YARD</b> 6,627 2k	land	56.60% 3,751 3k	36 inches	43.40% 2,876 3k
<b>NET</b> 7,578 2k	web	59.36% 4,494 3k	total	40.64% 3,076 3k
<b>REST</b> 18,368 1k	remainder	62.20% 11,425 2k	recuperate	37.80% 6,943 2k
<b>RING</b> 12,114 2k	sound of a bell	67.47% 4322 3k	circle	32.53% 2161 3k
<b>WAKE</b> 4,981 2k	become awake	75.80% 3,776 3k	a track left behind	23.00% 1,146 5k
<b>SPELL</b> 3,806 3k	letter-by-letter/ incantation	75.95% 2,889 3k	interval of time	24.05% 913 6k
<b>LIKE</b> 155,813 1k	resembling	76.20% 118,729 1k	opposite of dislike	23.80% 37,083 1k
<b>RIGHT</b> 103,410 1k	not left	77.40% 80,039 1k	legal rights	22.60% 23,370 1k
				vigil
				1.20% 60 >14k

&gt;&gt;&gt;

>>>									
<b>POOL</b>	water	78.62%	combine	21.38%					
<b>5,818</b>		<b>4,573</b>	resources	1,244					
<b>2k</b>		<b>3k</b>		<b>5k</b>					
<b>LEAVE</b>	part from	78.96%	direction	17.03%	permission	0.80%	Tree leaves	3.01%	
<b>63,807</b>		50,343		10,847		510		191	
<b>1k</b>		<b>1k</b>		<b>K2</b>		<b>8k</b>		<b>13k</b>	
<b>BAND</b>	group of people	79.00%	ring	21.00%					
<b>9,005</b>		7114		1891					
<b>2k</b>		<b>2k</b>		<b>4k</b>					
<b>FIRM</b>	business	80.12%	strong/solid	19.88%					
<b>19,890</b>		15,912		3,938					
<b>1k</b>		<b>1k</b>		<b>3k</b>					
<b>SET</b>	to place/ to be firm	80.40%	a collection	19.60%					
<b>53,544</b>		42,835		10,495					
<b>1K</b>		<b>1k</b>		<b>2K</b>					
<b>ARM</b>	body part	83.00%	weapon	17.00%					
<b>20,051</b>		16,725		3,426					
<b>1K</b>		<b>1K</b>		<b>3k</b>					
<b>DEAL</b>	an amount	84.00%	to distribute	16.00%					
<b>28,065</b>		23,575		4,490					
<b>1k</b>		<b>1k</b>		<b>3K</b>					
<b>HOST</b>	of a party	85.28%	multitude	13.91%	consecrated	0	wafer	34.6	
<b>4,327</b>		3,678		601					
<b>3K</b>		<b>3K</b>		<b>7K</b>				<b>&gt;14K</b>	
<b>WEAVE</b>	interlace threads	87.80%	move from side to side	12.20%					
<b>1,213</b>		1,065		148					
<b>5K</b>		<b>5k</b>		<b>&gt;14K</b>					

Bumping the minor member down a zone could yield rather different text profiles from those at present. If teachers are looking for texts at a particular level, say one matched to their learners as a means of building fluency, or ahead of their learners to build intensive reading skills, then just a few items (*band\_2* or *host\_2*) can push a short text above or below the 95% (Laufer, 1989) or 98% known-word comprehension threshold (Nation, 2006). Given the air time given in the recent research literature to the 95 vs. 98% difference as a factor in comprehension (Schmitt et al., 2011), small differences are clearly important. Similarly when Vocabprofiles are used to assess the lexical richness of student writing (Laufer & Nation, 1995) or speech (Ovtcharov et al., 2006; Lindqvist, 2010), a small number of lower frequency items can make a large difference to the lexical richness scores of short texts.

To summarize, the resources, methodologies, and motivation for a significant upgrade of the Frequency 1.0 scheme are largely in place. These include a methodology for identifying the main homoforms and MWUs for the pedagogically relevant zones of the BNC, a means of assigning them frequency ratings, and a first application of this methodology. There is clearly much more to do in this phase of the project, yet even when this is accomplished there will still be the matter of deploying this information in the real-time profiling of particular texts.

#### 4. Deployment of new lists in profiles of novel texts

A theme in this chapter is that the pedagogical application of a relatively simple frequency analysis of a large corpus has now necessitated a more sophisticated frequency analysis. The presence and then the extent of multiword units was first noticed and eventually tallied over the 2,000s, and now there is really no choice but to incorporate this information into the analysis. Similarly homoforms: the difference between ‘the *rest* of the day’ and ‘a *rest* for a day’ may seem a fairly minor phenomenon in a 1-million word corpus, where many minor partners in homograph pairs probably did not feature at all owing to the flukes of a small sample, but in the BNC’s 100-million there is no denying its importance. A second theme in this paper, however, is that while large corpora pose new problems, they also contain within them the solutions to these problems, as will be shown in the plan for deploying updated frequency information.

The goal is to reconfigure Vocabprofiling computer programs so that each *rest* or *bank* is tagged and assigned its own frequency level. In this way, two texts, like “Pound a stake into the bank to hold the dog” and “Stake out the bank for a hold up with a dog from the pound,” would be assigned quite different profiles. In considering how software can be programmed to make such distinctions, it is useful to ask how humans distinguish *bank*<sub>1</sub> from *bank*<sub>2</sub> and *at\_all*

from *at + all*. Clearly, they do it through an implicit analysis of the linguistic and situational context of the utterance, something a computer program cannot fully do at present, or maybe ever. However, a large part of a homoform's context is its particular lexical associates, which a computer program can easily identify.

The lexical associates in question are the frequent collocations that, while occurring with most words, are not so bound together that they form MWUs. In other words, these are collocates that maintain their independent or compositional meanings, as for example *fast* often collocates with *car*, and yet *fast car* is not normally viewed as a unit. In Davies and Gardner's list above (Fig. 1), the top noun collocations for 'money *bank*' are *account* and *loan*, and while no collocates are offered for 'river *bank*', these could include *grassy*, *steep*, *fishing*, or *Thames*. The discovery that large corpora have made available is, first, the great extent of these collocations, but second the fact that they are largely non-overlapping in character, at least in the case of homoforms and MWUs. We do not have *steep* money banks or *accounts* at river banks. We buy, look at, or covet *a lot* on which to build a house, but for this we need to pay or borrow *quite a lot* or *a whole lot* of money. Stubbs (2009) and Hoey (2005) both argue for systematic collocation as the means by which the mind distinguishes both polysemes and homoforms (Stubbs, p. 19, suggests this "can be done automatically" but with no reference to a running example). A test of this assertion begins with obtaining an adequate listing of collocations for a sample collection of homoforms and MWUs. A preliminary set of collocations for such a sample is explored in the next section by way of illustration.

## 5. A database of collocates

A listing of collocates for any single-word lemma can be generated at Sharp-Europe's BNC-based *Just-The-Word* online collocational database (at <http://www.just-the-word.com/>). The database supplies all collocates for an entered item if there are five or more instances of the item in the corpus; it looks within a span of five words on either side. Thus for Parent's collection of 178 homoforms, a collection of collocates down to a frequency of 10 was straightforward to produce. These collocations are, of course, not counted according to which meaning of a homoform they refer to (*between bank*, for example, is simply presented as a collocation having a frequency of 42), so once again the computer analysis has to be followed by a human sorting. This sorting is under way, but will be tested here on the first 10 items of Table 4, those most likely to cause a change in frequency rating. Table 2 in the Appendix shows the entire collocation listings for the two meanings of *bank* as generated by *Just-The-Word*.

Figure 2. BNC-Web's first 15 collocates for *at all* sorted by Mutual Information

Collocation parameters:						
Information:		collocations		Statistics:		Mutual information
Collocation window span:		3 Left	- 3 Right	Basis:		whole BNC
Freq(node, collocate) at least:		50		Freq(collocate) at least:		50
Filter results by:		Specific collocate:		and/or tag:		no restrictions
				Submit changed parameters		Go!
There are 5592 different types in your collocation database for "[word="at"%c] [word="all"%c]". (Your query "[word="at"%c] [word="all"%c]" returned 16941 hits in 2930 different texts, thinned with method <i>random selection</i> to 5000 hits)						
No.	Word	Total No. in whole BNC	Expected collocate frequency	Observed collocate frequency	In No. of texts	Mutual information value
1	levels	12,047	2.733	172	144	5.9758
2	hardly	8,397	1.905	69	63	5.1788
3	none	8,979	2.037	62	60	4.9278
4	times	29,194	6.623	195	167	4.8799
5	nothing	32,216	7.308	184	161	4.654
6	costs	15,161	3.439	65	63	4.2402
7	anything	27,431	6.223	114	103	4.1953
8	no	226,707	51.430	666	518	3.6948
9	any	121,761	27.622	291	255	3.3971
10	not	451,291	102.378	848	617	3.0502
11	look	51,972	11.790	83	77	2.8155
12	n't	316,187	71.729	438	330	2.6103
13	if	253,205	57.441	219	202	1.9308
14	like	147,567	33.476	97	89	1.5348
15	did	135,699	30.784	82	74	1.4134

A listing of collocates for MWUs is unfortunately not so simple to obtain, since *Just The Word* as presently configured does not perform searches for strings longer than one word (e.g., does not offer the typical collocates for a two-word string like *at all*). Fortunately, however, BNC-Web does handle multi-words, outputting a collocate list tagged by frequency and mutual information value (the degree of connectedness between headword and collocate). A small selection of high frequency MWUs from Martinez and Schmitt's collection (Table 3) was chosen for which there seemed to be little doubt of the existence of both a compositional and non-compositional version (*at all*, *as well as*, and *a lot* from the first 1,000, and *as far as* and *as long as* from the second).

The working hypothesis here is that the members of both homofoms and MWUs can be distinguished by collocations, but there are nevertheless some differences between the two. One is that some MWUs do not have a compositional meaning at all, or else it is extremely unlikely, and hence there is no point performing the collocational part of the analysis. For instance, it is hard to think of a compositional way to use *in order to* or *by and large* ('Zebras thundered *by and large* vultures flew overhead?') so these can be tagged as MWUs and assigned their frequency rank without deliberation.

BNC-Web can generate lists of lemmatized collocates for the 505 MWUs in question, and provide both raw frequency and mutual information values for each one, which allows for trimming of the list to a manageable human task. The program's output for the most connected 15 collocates of *at all* (sorted by mutual information value) is shown for illustration in Figure 2. For *at all* as a



compositional phrase, the frequent collocates mostly involve words like *levels*, *times*, and *costs* (thus *at all* levels, etc.) and as a non-compositional phrase they largely involve negative quantifiers like *none*, *hardly*, and *nothing* (thus nothing *at all*, etc.) and this once again must be hand sorted. A compilation of the most frequent 50 collocates of *at all*, sorted into compositional and non-compositional lists that an updated Vocabprofile can use to do its sorting is shown in Table 3 in the Appendix.

From these diverse sources, a database of collocates for both homoforms and MWUs can be fashioned.

## 6. Program function

The goal is for a modified Vocabprofile program to be able to assign homoforms and MWUs to their correct identities through an analysis of the high frequency collocates in the context (in this case choosing a span of four words on either side, following Sinclair's, 1991, suggestion). The program's job is to go through a text, and for any word or phrase it recognizes as a potential MWU or homoform (from an existing list), inspect the context for items from the two collocate sets from its database, and use this information to categorize the item as, e.g., *bank\_1* or *bank\_2*, or as *at\_all* (non-compositional unit) or *at all* (compositional separate words).

This procedure is intended to simulate a much reduced version of what humans do when they encounter ambiguous words or phrases. Further human-like functions of the program include (1) a coherent information assumption and (2) a competition procedure for conflicting information. For the first, once for instance *bank* has shown itself to be *bank\_2* (river bank) in a particular text, then in the absence of further information the next occurrence is also assumed to be this same kind of *bank* on the grounds that it is uncommon for the two senses of a homograph to appear in the same text (money *banks* located on river *banks*?). Where this does happen, however, by the second assumption collocates are simply counted up on a competition basis (most collocates wins) in an elemental version of the "cue summation model" proposed by MacWhinney (1989, p. 200) for similar language choices. In future, this calculation could be refined by inclusion of strength-of-relationship information from a corpus, such as mutual information value.

The way this procedure would work in a Frequency 2.0 Vocabprofile is as follows: The user enters a text for analysis. The Familizer subroutine ([lextutor.ca/familizer](http://lextutor.ca/familizer)) translates every word form in the text into a family headword (e.g., every *had* is changed to *have*) based on Nation's (2006) pedagogical rendering of the BNC frequency list. The disambiguator routine (living in pro-

totype form at [lextutor.ca/concordancers/text\\_concord/](http://lextutor.ca/concordancers/text_concord/)) then reads through the text-as-families, first in three-word, then two-word n-grams (to pick up any *at all*-like items) and then in singles. Every n-gram and single is weighed against the program's stop list of potential homoforms. In the singles phase, for example, the program comes across the headword *miss*, finds the item to be in its stop list, and thus opens its collocational database for this item (an abbreviated version of this database, coded for reading by a PERL routine, is shown in Fig. 3). The program inspects the eight words surrounding *miss* in the text (four to the left, four to the right). If it finds *bare*, *boat*, or *bus*, it parses the word as the 'loss' type of miss, *miss\_1*. If it finds *girl*, *young*, *pretty*, or other similar titles like *mistress*, or a following word with a capital letter (miss Smith), it parses the word as *miss\_2*. If there are multiple occurrences of *miss* and the program finds collocates supporting both interpretations, the majority association wins. In the event of a tie or a lack of any match, any previous parsing is repeated, following the reasoning already mentioned. In the rare event (except at the very beginning of a text) of no collocate matches and no previous parsing, then the parsing assigned is *miss\_0*.

**Figure 3.** Database with collocates for two members of the homograph *miss*

MISS	+ miss missed unmissed misses missing
<b>loss</b>	(I   you   he   she   they   we) miss   aim   bad   bare   beat   boat   bus   (can   can_1   cannot   cannot) miss   chance   date   deadline   dreadful   fail   family   foot   forget   heart   hit   lack   lose   lot   mark   match   mile   moment   much   narrow   near   never   opportunity   plane   point   putt   race   really   target   terrible   thing   train   trick   tube   want to miss
<b>title</b>	daughter   dress   girl   hair   kiss   lady   little   marry   master   mister   mistress   mrs   niece   pretty   sister   spinster   universe   victorian   young   Miss ([*the])([A-Z][a-z]+[ ]*)   she

In the n-gram phase of the analysis, if an instance of *at all*, for example, is found, it is tested against the non-compositional collocates for this entry (Fig. 4), and if none is found in the environment, then the individual components are returned to the analysis as single words (where *at* and *all* will both be classed 1k items). The collocational criteria for the two meanings of *at all* are shown in Fig 4. The prepositional meaning is nearly always followed by *the*; the quantity meaning of *at all* is almost always preceded by a negating term like *never*, plus optional intervening other words (like '*never saw him at all*', which can be picked up by the regular expression  $[a-z^*]$ ).

**Figure 4.** Distinguishing collocates for a multi-word unit

AT_ALL	+
quantity	(no   not   nothing   none   any   never   any   anything   few) [a-z]* at_all   if [a-z]* at_all
preposition	at_all the

## 7. How well do collocates do their work? A Mini-Experiment

### 7.1. Research question

Can homoforms including MWUs with a compositional and non-compositional meaning be reliably distinguished by the collocational resources currently available?

### 7.2. Context

It is frequently claimed that there are few true synonyms in a language owing to differences in contexts of use and especially the distinct collocations that different senses of words typically enter into (Sinclair, 1991). This claim should be even more applicable to forms which are not just synonyms but have no related meaning whatever. However, to date many examples but few proofs are offered for this claim, which therefore remains intuitive. The proof of the claim would be if the collocations that appear to distinguish the meanings of a homoform in a particular corpus could predict the same distinctions in a novel text or corpus.

### 7.3. Procedure

The BNC was mined for all collocations with a frequency > 10 for the first three items from Parent’s selection in Table 6 (*miss*, *yard*, and *net*) and two selections from Martinez and Schmitt’s selection in Table 3 (*a lot* and *at all*) in the manner of the information in Table 2 in the Appendix for *bank*. For each item, roughly 200 collocations, with some variability in the number, were hand sorted into those corresponding to each meaning, which in the case of *miss* was tagged as *miss\_1* or *miss\_2*. The collocations were coded in the PERL scripting language to match text strings within ten words on either side of each test item, including strings with an unpredicted intervening word (*miss train* would also match *missed their train*). Novel contexts for the five items were obtained by searching a corpus of simplified stories for texts containing both meanings of each of the homoforms. For example, Wilde’s *The Picture of Dorian Gray* (Oxford Bookworms Series; 10,500 running words; 1,000 headwords) bears three instances of *miss* with both parsings represented. All instances were extracted as concordance lines

of roughly 30 words (80 characters on either side of the keyword). These concordance lines served as a greatly truncated ‘text’ that would test the program’s ability to use context information to disambiguate the homographs. The next step was to feed this test text into a computer program that accesses the collocational database. The program breaks a text (in this case, the set of concordance lines with homographs) into family headwords, identifies the current search term, and looks for pattern matches in its collocation set. Each time it makes a match it records the fact and awards a point to the relevant meaning.

#### 7.4. Results

The collocational information is clearly able to distinguish the two meanings of the homograph *miss*. Figure 5 shows the Dorian Gray output for *miss*, followed by the record of the decision process.

**Figure 5.** “*miss*” in simplified *The Picture of Dorian Gray* - Bookworm Level 4

---

#### **Parsed concordance**

034.    omething to say to you.'That would be lovely. But wont you MISS\_1 your train?' said  
Dorian Gray, as he went up the step
035.    , You look like a prince. I must call you Prince Charming.' MISS\_2 Sibyl knows how to  
flatter you.'You dont understand
036.    g, Harry. I apologize to you both.' My dear Dorian, perhaps MISS\_2 Vane is ill,' said  
Hallward. We will come some other

#### **Program’s reasoning**

34.    2 0 miss\_1  
to you'That would be love But wont you MISS you train' say DORIAN Gray as he go up  
— miss 'you MISS'  
— miss 'train'
35.    0 1 miss\_2  
like a prince I must call you Prince Charming' MISS Sibyl know how to FLATTER you'You dont understand  
— miss 'MISS Sibyl' (CAP)
36.    0 1 miss\_2  
I apology to you both' My dear Dorian perhaps MISS Vane be ill' SAY Hallward We will come some  
— miss 'MISS Vane' (CAP)
- 

The program’s reasoning as shown in the output is thus: Before starting, the algorithm reduces all words to familized headwords (e.g., *go* not *went* in line 34). To parse the instance at concordance line 34, a pronoun subject (*I|you|he*, etc) before the keyword, and the presence of the high frequency collocate *train* anywhere in the string, give a score of 2-0 for *miss\_1* (loss). The challenge point in

this and the many other runs of this experiment is where the meaning of the homoform changes. This happens in line 35, where there is no match suggesting miss\_1 (loss), and one piece of evidence for miss\_2 (title), namely *miss* followed by a word with a capital letter, giving a score of 0-1 and a verdict of miss\_2. In line 36, a capital letter is once again the decider, now backed up by the coherent information assumption. A score of 0-0 would have led to a continuation of the previous parsing and that would have been correct.

Similarly, the Bookworms version of Conan Doyle’s *Tales of Mystery and Imagination* was found to bear both meanings of *at all*, and once again the collocations were able to distinguish these (Fig. 6), largely through discovering various quantifiers like *few*, *none*, *any* and *if* for the non-compositionals and a following *the* for the compositional (these are underlined in the concordance output for emphasis).

**Figure 6.** “*at all*” in simplified *Tales of Mystery & Imagination* – Bookworm Level 3

- 
- 020. sons of the richest families of England. There was nothing at\_all\_1 to stop me now. I spent my money wildly, and passed
  - 021. n and the strange fears I had felt. If I thought about them at\_all\_1, I used to laugh at myself. My life at Eton lasted f
  - 022. htening, and few people were brave enough to enter the room at\_all\_1. In this room, against the farthest wall, stood a hu
  - 023. nd held it there for many minutes. There was no life in him at\_all\_1. Now his eye would not trouble me again. Perhaps you
  - 024. lantern was closed\_2, and so no light came out of it, none at\_all\_1. Then slowly, very slowly, I put my head inside the
  - 025. d it. I started walking around the streets at night looking at\_all\_2 the cats, to see if I can\_1 find another one like Pl
- 

In the five test cases, all significantly longer than the ones shown here, the collocation database was able to correctly identify the relevant meaning of the single word or multiword homoform in at least 95% of cases. Accuracy can be increased by expanding the size of the database (Fig. 4 is far from an exhaustive list of at all the collocates Web-BNC offers for *at all*), but at the expense of slowing the program down and making it less useful for practitioners.

### 7.5. Discussion

There is thus evidence that collocations can indeed simulate the function of human judgment in this task and hence that the full database of collocates for the high frequency homoforms and MWUs is worth building.

Further, it should be noted that the task set to the computer program in

the mini-experiment is unrealistically difficult. As mentioned, few natural/normal/real texts contain both meanings of a homoform in as close proximity as in the special texts used here to test the program, which were chosen precisely for the presence of both meanings of the homoform. In a natural text, one meaning is normally established and then the algorithm's default procedure ("use previous") almost invariably leads to a correct assignment – and the success rate over the many trials performed by the author is more like 98%.

## 8. Conclusion

The pieces of Frequency 2.0 are at hand and, although hailing from quite disparate quarters, merely require assembly. The most frequent and most pedagogically relevant homoforms have been identified, separated, and assigned initial frequency ratings, and a methodology is in place to move the analysis down the scale to the vast number of homoform items in English where the minor member represents fewer than 5% of occurrences. Refinements there will certainly be, and the question of what makes an MWU non-compositional will need further thinking, but the methodology is likely to be something similar to the one proposed here. Further, while the first round of this work had to be accomplished by humans, prizing apart the *banks* and *at all's* by inspecting samplings of concordance lines, for subsequent rounds a means is available to automate this task using a computer program in conjunction with a collocational database such that sampling should not be necessary: within a year or two, the collocational database should be completed for both the Parent and Martinez items, or principled sub-sets thereof, and it should be possible to assemble the pieces and create a complete set of trial lists, incorporating both types of homoforms, as hypothesized in Table 2.

When that happens, an important task will be to establish new cut-offs – that is, new frequency counts. The alert reader will have noticed that in several of the analyses above, the original word-form cut-offs were used for proposed new frequency assignments, whereas in fact, every re-assignment will shift all the cut-offs. For example, if the first thousand list is defined as every BNC lemma represented by more than 12,369 occurrences (Table 5), and the non-compositional meaning of *a lot* is found to have more occurrences than this, then it should be included as a first thousand item – and the current last item will be bumped to the second thousand list.

Also on the to-do list will be to establish a coding format for the different meanings of homographs (*bank\_1* and *bank\_2*, or *bank\_money* and *bank\_river?* and *at\_all* for non-compositional MWUs but plain *at* and *all* for compositional?); to settle on the exact list of MWUs to include; to settle on the percentage of main-meaning occurrences (90% or 95%) that makes handling separate

meanings worth program time; and to decide whether to limit the single word analysis to the first five thousand-word families or to proceed further. Benefits to be realized will be more accurate Vocabprofiling (extent to be determined), greater credibility for this methodology within the scientific community, and more effective language instruction.

## References

- Aston, G., & Burnard, L. (1998). *The BNC handbook: exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253-279.
- Beglar, D., & Nation, P. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: an MEG study. *Cognitive Brain Research*, 24, 57-65.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Pearson Education.
- Cobb, T. (2010). Learning about language and learners from computer programs. *Reading in a Foreign Language*, 22(1), 181-200.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Davies, M., & Gardner, D. (2010). *Frequency dictionary of contemporary American English: Word sketches, collocates, and thematic lists*. New York: Routledge.
- Davies, M. (2011). Word frequency data from the Corpus of Contemporary American English (COCA). [Downloaded from <http://www.wordfrequency.info> on 2012-07-02.]
- Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24(02), 143-188.
- Ellis, N. C., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, 59, 90-125.
- Grant, L., & Nation, P. (2006). How many idioms are there in English? *International Journal of Applied Linguistics*, 151, 1-14.
- Heatley, A., & Nation, P. (1994). *Range*. Victoria University of Wellington, NZ. [Computer program, available with updates at <http://www.vuw.ac.nz/lals/>].
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Taylor and Francis.
- Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, 14(2), 151-162.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Clevedon, UK: Multilingual Matters.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written

- production. *Applied Linguistics*, 16, 307-322.
- Leech, G., Rayson, P., & Wilson, W. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.
- Lindqvist, C. (2010). La richesse lexicale dans la production orale de l'apprenant avancé de français. *La Revue canadienne des langues vivantes*, 66(3), 393-420.
- Martinez, R. (2009). *The development of a corpus-informed list of formulaic expressions and its applications to language assessment and test validity*. PhD thesis, University of Nottingham.
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299-320.
- MacWhinney, B. (1989). Competition and lexical categorization. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), *Linguistic categorization* (pp. 195-242). Amsterdam: Benjamins.
- Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of TESL student performance. *System*, 32(1), 75-87.
- Nation, P. (2001). *Learning vocabulary in another language*. London: Cambridge.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, P. (Unpublished). *The frequency ordered 1,000 word family lists based on the British National Corpus*.
- Ovtcharov, V., Cobb, T., & Halter, R. (2006). La richesse lexicale des productions orales: mesure fiable du niveau de compétence langagière. *Revue Canadienne des Langues Vivantes*, 63(1), 107-125.
- Oxford Bookworms Library*. London: Oxford University Press.
- Palmer, H. E. (1941). *A grammar of English words: One thousand English words and their pronunciation, together with information concerning the several meanings of each word, its inflections and derivatives, and the collocations and phrases into which it enters*. London: Longman.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95(1), 26-43.
- Sharp, Europe. Just The Word: Collocational Database. [Website <http://www.just-the-word.com/>, accessed 20 November 2011.]
- Shin, D., & Nation, P. (2007). Beyond single words: The most frequent collocations in Spoken English. *ELT Journal*, 62(4), 339-348.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. London: Oxford University Press.
- Stubbs, M. (2009). Technology and phraseology: With notes on the history of corpus linguistics. In U. Romer & R. Schulze (Eds.), *Exploring the lexis-grammar interface* (pp. 15-32). Amsterdam: Benjamins.
- van Zeeland, H. and Schmitt, N. (in press). Lexical coverage and L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*.
- Wang, K., & Nation, P. (2004). Word meaning in academic English: Homography in the Academic Word List. *Applied Linguistics*, 25, 291-314.
- West, M. (1953). *A general service list of English words*. London: Longman.



APPENDIX

Table 1. Full list of Parent’s GSL homographs

air	coast	faint	late	page	rest	step
arm	company	fall	lay	pan	right	stick
article	concentrate	fast	lead	park	ring	still
ball	contract	fence	leave	passage	rock	stir
band	count	figure	left	patient	roll	store
bank	country	file	letter	pen	row	stone
bar	course	fine	lie	pick	scale	strike
bear	court	fire	light	plant	school	swallow
bear	cross	firm	like	pole	season	table
belt	crush	flat	line	policy	second	tend
bill	cry	fold	live	pool	sense	train
bit	culture	foot	lock	port	sentence	trip
boil	cure	formal	lot	pot	set	type
boot	curl	forward	love	pound	shoot	wake
bowl	current	game	match	present	shoot	watch
box	date	general	mean	present	shower	weave
bridge	deal	go	metre	press	slip	well
brush	degree	habit	might	pretty	sock	whip
camp	die	hand	minute	private	sound	wind
can	down	hide	miss	produce	spell	wound
case	drag	host	mouse	pupil	spirit	yard
cell	draw	how	nature	race	spring	
charge	drive	just	net	rail	staff	
chest	duty	kind	nut	rank	stage	
close	ear	knot	order	realize	state	
club	egg	last	organ	repair	steep	
	even					

Table 2. Collocates for two banks, from Just-The-Word database, frequency >10, span=5 word-forms either side, hand-sorted into independent meanings

Money banks

world bank	714	development bank	86	director of bank	51
central bank	690	bank on	84	bank announce	50
bank account	422	bank balance	78	bank credit	50
bank holiday	409	swiss bank	76	bank provide	49
bank manager	298	bank rate	74	private bank	49
national bank	272	major bank	73	money in bank	49
commercial bank	226	bank lend	71	clearing bank	48
european bank	215	state bank	67	international bank	48
merchant bank	201	bank clerk	64	president of bank	48
royal bank	191	bank and company	62	bank offer	47
bank loan	189	British bank	61	bank statement	47
investment bank	165	american bank	57	french bank	45
between bank	142	bank and institution	57	bank official	45
go to bank	117	borrow from bank	55	leave bank	44
midland bank	113	include bank	55	german bank	43
big bank	104	branch of bank	55	reserve bank	43
governor of bank	97	bank or building society	55	clearing bank	40
bank deposit	95	bank hold	53	creditor bank	40
foreign bank	91	bank note	53	bank strip	40
bank and building society	90	japanese bank	52	bank lending	39
large bank	87	data bank	51	bank agree	38

>>>

&gt;&gt;&gt;

bank pay	38	bank seek	22	accept by bank	14
chairman of bank	38	irish bank	22	deposit in bank	14
work in bank	37	issuing bank	22	make by bank	14
join bank	37	bank interest	22	set up bank	14
bank buy	37	head of bank	22	offer by bank	14
leading bank	37	group of bank	22	owe to bank	14
bank governor	37	Western bank	21	shanghai bank	14
break bank	36	role of bank	21	write to bank	14
bank lending	36	clear bank	20	bank step	14
overseas bank	35	enable bank	20	retail bank	14
bank charge	35	close bank	20	jeff bank	14
bank debt	35	bank operate	20	bank employee	14
allow bank	34	bank raid	20	bank finance	14
have in bank	33	line bank	19	bank funding	14
rob bank	33	sponsor by bank	19	bank customer	14
issue by bank	33	bank charge	19	bank estimate	14
bank issue	33	bank require	19	consortium of bank	14
bank sell	32	trust bank	19	building society and bank	14
bank able	32	bank borrowing	19	bank and government	14
land bank	32	bank corporation	19	receive from bank	13
bank branch	32	bank vault	19	draw on bank	13
loan from bank	32	subsidiary of bank	19	sell to bank	13
way to bank	32	establishment of bank	19	co-op bank	13
northern bank	31	take to bank	18	deposit with bank	13
be bank	30	bank create	18	bank to bank	13
bottle bank	30	asian bank	18	get in bank	12
street bank	30	account with bank	18	hold by bank	12
bank robbery	30	Government and bank	18	pay to bank	12
bank base rate	30	eastern bank	17	take by bank	12
memory bank	29	piggy bank	17	bank assistant	12
put in bank	28	state-owned bank	17	bank guarantee	12
bank cut	28	city bank	17	bank creditor	12
bank staff	28	bank card	17	Balance at bank	12
manager of bank	28	debt to bank	17	currency and bank	12
force bank	26	oblige bank	16	Building society or bank	12
provide by bank	26	approach bank	16	bank and credit	12
Independent bank	26	bank publish	16	bank or company	12
bank report	26	bank deal	16	deposit with bank	11
pay into bank	25	bank overdraft	16	bank grant	11
street bank	25	agreement with bank	16	bank intervene	11
union bank	25	name of bank	16	failed bank	11
bank robber	25	available from bank	16	gene bank	11
account at bank	25	bank and house	16	bank post	11
customer of bank	25	bank up	16	bank operating	11
fund and bank	25	own by bank	15	bank interest rate	11
bank and fund	25	work for bank	15	chair of bank	11
regional bank	24	persuade bank	15	money from bank	11
bank act	22	bank president	15	company and bank	11
bank refuse	22	bank show	15		

**River banks**

west bank	240	steep bank	45	left bank	28
river bank	210	opposite bank	42	east bank	27
along bank	194	west bank	42	left bank	26
south bank	166	top of bank	42	stand on bank	15
far bank	94	grassy bank	41	occupied bank	14
its banks	85	north bank	41	shingle bank	12
down bank	73	sit on bank	30	situate on bank	11
up bank	53	swain bank	30	walk along bank	11
south bank	48	burst bank	28		

**Table 3.** Collocates for *at all* (57 idiomatic or non-compositional, 11 compositional) selected from the BNCWeb’s most frequent and most connected 100 (by log-likelihood of co-occurrence) as the basis for database entry (Fig. 6)

**Non-Compositional**

(anything) at all wrong	(no) interest at all	at all — (phrase end)
(didn't) notice at all	(no) problem at all	at all' (phrase end)
(didn't) seem at all	(no) reason at all	at all possible
(didn't) sleep at all	(no) sense at all	at all! (sentence end)
(doesn't) bother (me) at all	(no) sound at all	at all. (sentence end)
(doesn't) exist at all	(no) trouble at all	at all? (sentence end)
(doesn't) look at all	(not) aimed at all	did (not) at all
(don't care) at all about	(not) at all actually	hardly at all
(don't care) at all except	(not) at all clear	if at all
(don't care) at all really	(not) at all easy	mention at all
(don't see it) at all	(not) at all sure	never (did it) at all
(don't) like at all	(not) at all surprised	no ... at all
(don't) mind at all	(not) changed at all	nobody at all
(don't) remember at all	(not) doubt (it) at all	none at all
(don't) see at all	(not) pleased at all	not at all
(no) good at all	(not) worried at all	nothing at all
(no) harm at all	any at all	n't ... at all
(no) help at all	anything at all	scarcely at all
(no) idea at all	anywhere at all	without (any) at all

**Compositional**

avoided at all (costs)	at all sites	at all events
avoid at all (costs)	at All Saints	at all costs
at all times	at all levels	at all ages
at all stages	at all hours	



# *A new approach to measuring lexical sophistication in L2 oral production*

Christina Lindqvist\*, Anna Gudmundson\*\* and Camilla Bardel\*\*  
\*Uppsala University, \*\*Stockholm University

The aims of this chapter are a) to give a comprehensive description of a new tool for lexical profiling by reporting how it was developed, and b) to indicate possible areas of use and future developments of the tool. The tool has been used for measuring the lexical sophistication of Swedish learners of French and Italian. The different steps of development have partly been presented in previous studies (Bardel & Lindqvist, 2011; Bardel, Gudmundson & Lindqvist, 2012; Lindqvist, Bardel & Gudmundson, 2011) but are complemented here through a detailed account of the tool, in order to enable replication and use of the method with other languages. The outline of this chapter is as follows: first, as a background, we provide a survey of methods designed to measure lexical richness in L2 production. Then we discuss the inherent differences between written and spoken language and what these differences may imply when lexical richness is measured. Next, we present a new method for analyzing L2 learners' lexical profiles in oral production data, giving a detailed technical description of the creation of the tool. We then discuss pros and cons with frequency-based measures in general and present our solutions to some of the problems brought up. Finally, we suggest some potential areas of use and discuss some possible improvements of the method.

## **1. Background: a survey of methods designed to measure lexical richness in L2 production**

In the study of L2 vocabulary, lexical richness can be seen as an umbrella term, covering four different dimensions: *lexical density*, *lexical diversity*, *lexical sophistication* and *proportion of errors* among the words used by an L2 learner (Read, 2000, pp. 200-201). Lexical density can be measured as the proportion of semantically full words (or lexical words) as opposed to function words. Lexical diversity, or variation, can be measured by the simple type/token ratio (TTR). The TTR is a calculation of the number of types divided by the number of tokens in a text. The basic problem with TTR is its sensitivity to text length, as is well known. As explained by McCarthy and Jarvis (2007, p. 460), "the more words (tokens) a text has, the less likely it is that new words (types) will occur". If a text is so long that certain words start to be repeated, high-frequency words

will be repeated more often as compared to low-frequency words, and this tendency will increase the longer the text is. Several measures have been proposed in order to solve the problem with text length. One example is the index of Guiraud (Guiraud, 1954), which is a type/token based measure that is supposed to be independent of text length. The index of Guiraud results from dividing the number of types by the square root of the number of tokens. For a long text, this procedure will result in a higher lexical richness than what would have been obtained with a simple TTR. However, according to Daller, Van Hout and Treffers-Daller (2003, p. 200) neither TTR nor the index of Guiraud are valid measures of lexical richness at later stages of L2 acquisition. A development of the Guiraud index is the advanced Guiraud, which takes in frequency as a factor (Daller et al., 2003). Furthermore, Malvern, Richards, Chipere and Durán (2004) have suggested the D measure, which is freely available in CHILDES. D models the falling TTR curve by calculating TTRs for samples of different text lengths, ranging from samples of 35 words to samples of 50 words, which are taken randomly from the text. However, in their critical evaluation of D, McCarthy and Jarvis (2007) conclude that even though the D measure was the most reliable of those investigated, it still retains a certain degree of sensitivity to text length.

Lexical sophistication is defined as the percentage of sophisticated or advanced words in a text. There are, however, different definitions of sophisticated/advanced vocabulary. Low-frequency words, for instance, are generally considered to be advanced and sophisticated (Laufer & Nation, 1995; Vermeer, 2004). It has even been suggested that words are learned in rough order of frequency (Cobb & Horst, 2004; Vermeer, 2004). The difficulty of words, as measured by their frequency, should therefore be taken into account when measuring the lexical richness of L2 learners. A method which relies on the raw frequency of words in the target language is the *Lexical Frequency Profile*, LFP (Laufer & Nation, 1995). The LFP measures the proportion of high-frequency words vs. the proportion of low-frequency words in a written text. All the words are divided into different categories, which have been established on the basis of frequency bands based on written language corpora (Laufer & Nation, 1995). *Vocabprofile* is a program that executes this categorization according to the following frequency bands: the 1000 most frequent word families, the next 1000 most frequent word families, and the Academic Wordlist, which contains the 570 most frequent word families drawn from academic texts (Coxhead, 2000, see also [www.lextutor.ca/vocabprofile](http://www.lextutor.ca/vocabprofile)). The words that do not appear in any of these categories end up in the 'not-in-the-lists' category.<sup>1</sup>

---

1 There is also an updated version of Vocabprofile for English (but not for French), which distinguishes 20 different frequency bands.

Laufer and Nation (1995) have shown that the LFP measure is able to distinguish between different proficiency levels. The English version of LFP was validated by Laufer and Nation and there is also a French version, with the program *Vocabprofil*, also based on written data, which has been validated in a study of the oral production of advanced French L2 learners by Ovtcharov, Cobb and Halter (2006). It is interesting to note that Ovtcharov et al. actually used oral learner data and ran those against frequency bands based on written data. Still, they found significant differences between learners at different proficiency levels.

## 2. Lexical sophistication in written vs. spoken language

Even though Ovtcharov et al. (2006) were able to validate the French version of LFP using learners' oral production data, the appropriateness of comparing learners' *spoken* language with written data bases can be questioned. Lindqvist (2010) used the French version, *Vocabprofil*, comparing two groups at different proficiency levels.<sup>2</sup> In contrast to Ovtcharov et al. (2006), she found no significant differences between the two learner groups. She also conducted a qualitative analysis of the words classified in the not-in-the-lists category, and found that many words typical in oral French were classified in this category, such as *ben* ('well'), *ouais* ('yeah'), *rigolo* ('fun'), *prof* (short for 'teacher'), *sympa* ('nice'), although these are frequent in everyday speech. Lindqvist suggested that frequency lists based on L1 oral data should be used when investigating L2 learners' oral production. This has also been pointed out by Tidball and Treffers-Daller (2008, p. 311), who call for an oral version of the *Vocabprofil* program, so that oral data can be compared to an oral data base, which would better reflect the informants' lexical profile. For instance, the words *ben* and *ouais* are discourse markers that are often found in spoken language, but not in written production (McCarthy, 1998; Tidball & Treffers-Daller, 2008), so even if they are produced often by a learner a comparison to a written data base would give the impression that the learner uses rare words, and the conclusion that the learner in question has an advanced vocabulary might be wrong. According to McCarthy (1998, p. 122), frequency lists based on spoken language differ from those based on written sources. Generally, the differences between spoken and written language are considerable (see e.g. Linell, 2005, p. 28), something that must

---

2 The levels of proficiency of the learners were established on the basis of a morpho-syntactic analysis (cf. Barting & Schlyter, 2004).

have consequences at the lexical level of language. Considering this, there is a clear risk of running into validity problems when comparing spoken language to written corpora.

### **3. A new method for analyzing learners' lexical profiles in oral production data: the Lexical Oral Production Profile (LOPP)**

Considering the background described above, and in order to avoid not only a written language bias (cf. Linell, 2005), but also methodological problems of validity, we set out to create a new tool for analyzing lexical sophistication in French and Italian L2, within the on-going project *Aspects of the advanced L2 learner's lexicon*.<sup>3</sup> We developed a lexical profiler explicitly for the analysis of spoken language. In order to create frequency bands based on spoken target language data, we used the *Corpaix* corpus for French and the *C-Oral-Rom* and *LIP* corpora for Italian.<sup>4</sup> We also developed a program that runs learner data against the frequency bands. In the following, we will describe the process of creating the tool.

#### *3.1. SQL: a tool for manipulating data bases*

SQL stands for *Structured Query Language* and is a declarative programming language initially developed at IBM with the purpose of manipulating big data bases. Work with data bases emerged in the 1960s due to cheaper storage and computing power (Wilton & Colby, 2005, p. 7), and the first scientific article discussing SQL was published in 1970 by the IBM researcher Codd (1970). SQL is now standardized by both the International Standards Organization (ISO) and by the American National Standards Institute (ANSI) (Jones et al., 2005, p. 2).

SQL is a data base management system allowing one to access and manipulate data bases. A data base could be described as a set of one or more tables organized in a systematic way or as "one or more large structured sets of persist-

---

3 This study is part of the research program High-Level Proficiency in Second Language Use, funded by the Bank of Sweden Tercentenary Foundation (grant M2005-0459).

4 An inherent problem with spoken language corpora is the relative limitations that the oral language mode implies, in terms of technical adjustments needed, transcription etc. As a consequence, these corpora are rather small, in comparison to, for example, the BNC.



ent data, usually associated with software to update and query the data” (The Free On-line Dictionary of Computing: <http://foldoc.org/database>). When working with sets of associated tables, i.e. retrieving, organizing, joining, counting and comparing table contents, work is very much facilitated if a query language such as SQL can be used.

### *3.2. Construction of the French and Italian frequency bands*

The French frequency bands are based on the oral corpus Corpaix, compiled at the Université de Provence (Campione, Véronis, & Deulofeu, 2005). The corpus consists of about 1 million tokens based on different tasks such as interviews, conversations and meetings on different topics such as personal memories, travel, politics and professional experiences. A token-frequency list, based on Corpaix, has been created and published online at <http://sites.univ-provence.fr/veronis/data/freq-oral.txt> by Jean Véronis and that list was used when creating the French frequency bands discussed in the present study.<sup>5</sup> All tokens in the list were lemmatized with the software TreeTagger (Schmid, 1994, 1995) and later run through the software WordSmith (Scott, 2004) to calculate the frequency of each lemma. Hence, the final result consists of a lemma-frequency list composed of 2746 different lemmas.<sup>6</sup>

In regard to the Italian frequency bands, they were based on the already lemmatized versions of two different oral corpora: the LIP (Lessico di frequenza dell’italiano parlato) (De Mauro, Mancini, Vedovelli, & Voghera, 1993), which is freely available at the site BADIP (Schneider, 2008) and the C-Oral-Rom corpus (Cresti & Moneglia, 2005). The LIP corpus is based on several types of oral production: face-to-face conversations, telephone conversations, non-free dialogical interactions, monologues and radio and TV programs. C-Oral-Rom is based on both formal and informal speech, face-to-face conversations, telephone conversations and broadcasting. The social context of data collection is both private, within the family, and public, for example political speech and debate. A Perl programming language script was run on the XML versions of the two corpora in order to create a lemma-

---

5 Only tokens that appear ten times or more in the Corpaix corpus were added to the list created by Véronis.

6 This number has been corrected compared to earlier studies (Bardel, Gudmundson, & Lindqvist, 2012; Lindqvist, Bardel, & Gudmundson, 2011) in which the number of lemmas was estimated to 2766, due to a technical error. This small difference does not have any effect on the division of the lemmas into the frequency bands.

frequency list based on both LIP and C-Oral-Rom. The final result consists of a lemma-frequency list composed of 19962 different lemmas based on a total of 789070 tokens.

When creating the French and Italian frequency bands it was decided to use the lemma as counting unit instead of the word family, for the following reasons (for a more detailed discussion, see Lindqvist et al., 2011). A word family can include both derivations and inflected forms of a headword, which implies that the word family might include quite a high number of forms. For example, an Italian regular verb has six different forms in present tense: *canto*, *canti*, *canta*, *cantiamo*, *cantate*, *cantano* (from inf. *cantare*). This marking of person is compounded with marking of tense, aspect and modality (e.g. past tense of subjunctive 1<sup>st</sup> person plural: *cantassimo*). Hence, Italian has a very rich verb morphology. Furthermore a word family can also include nouns, adjectives, etc, whose relationships with the base are not always very transparent, such as *canzone* (song), *cantante* (singer) and, possibly, *cantautore* (a compound of *cantante* and *autore*, singer/songwriter). The fact that a learner uses one particular form does not necessarily mean that he or she has knowledge of all the related forms in the word family. This claim is particularly relevant in our research, which concerns oral production. It is plausible that the learner knows several word forms that are simply not used in one particular recorded session, which makes it impossible to draw any conclusions regarding how many forms related to a specific word family are actually known. Using the lemma as counting unit is an option that reduces the number of forms attached to a headword, even though this does not solve the problem completely. In conclusion, the French and Italian frequency bands described in this paper are different from the ones elaborated by Laufer and Nation (1995) and Cobb and Horst (2004), which are based on word families.

2746 lemmas from the French lemma-frequency list and 3127 lemmas from the Italian lemma-frequency list were divided into three frequency bands consisting of about 1000 lemmas each. Hence, band 1 includes the

**Table 1.** The French frequency bands

Band	Lemma range	Lemmas (n)	Tokens (n)	Relative token frequency (%)
1	1-986	986	896347	95.93
2	987-1939	953	28003	3.00
3	1940-2746	807	10034	1.07
Total		2746	934384	100

most frequent 1000 lemmas, band 2 the 2<sup>nd</sup> 1000 most frequent lemmas and band 3 the 3<sup>rd</sup> 1000 most frequent lemmas. The lemmas not appearing in any of these three bands are categorized as off-list lemmas, i.e. those not belonging to the most frequent 3000 lemmas in Italian or French. Table 1 shows the frequency distribution of the French frequency bands and table 2 the frequency distribution of the Italian frequency bands.

**Table 2.** The Italian frequency bands

Band	Lemma range	Lemmas (n)	Tokens (n)	Relative token frequency (%)
1	1-1019	1019	676098	91.82
2	1019-2047	1028	39726	5.39
3	2048-3127	1080	20526	2.79
Total		3127	736350	100

The tokens included in the French frequency bands (1-3) cover 93.44% of the total number of tokens included in the Corpaix corpus, and the tokens included in the Italian frequency bands (1-3) cover 93.32% of the total number of tokens included in the Italian corpus, i.e. the combination of LIP and C-Oral-Rom. As can be seen from the tables above, the number of lemmas included in the Italian frequency bands is slightly higher than that of the French bands. It can also be noted that the number of lemmas included in each band within each language varies between 807 and 986 for French and between 1019 and 1080 for Italian. The reason for this is that the line between two frequency bands must be drawn where two lemmas differ in frequency; for example, in the French list, all lemmas from rank 971 to 986 occur 50 times in the corpus, while the lemma ranked as number 987, *journal* (newspaper) occurs 49 times. *Journal* could not be included in the first frequency band since it would have been necessary to include all other lemmas that occur 49 times as well. The number of lemmas included in each band could therefore not be established and decided beforehand. The aim, however, was to distribute them as evenly as possible. It can be noted that more than 90% of all tokens that appear in the two corpora belong to band 1 and that only a small percentage belong to bands 2 and 3. The French and Italian frequency bands were imported into an SQL data base.

### 3.3. The lexical oral production profiler (LOPP): running analysis

French and Italian learner production can be compared to the frequency bands to measure the proportion of lemmas that fall within each frequency band. In order to do that, all data has to be lemmatized and information about lemma

frequency must be added. Other information, such as name of informant/name of recording, the language status (i.e. whether it's an L1 or an L2 speaker), and the linguistic level, i.e. proficiency level, can also be included.<sup>7</sup> Figure 1 shows part of an input file.

**Figure 1.** Part of a French input file

<b>informant name</b>	<b>language status</b>	<b>lemma</b>	<b>lemma freq</b>	<b>linguistic level</b>
lda4int	L2	le	74	6
lda4int	L2	être	71	6
lda4int	L2	on	70	6
lda4int	L2	avoir	58	6
lda4int	L2	je	58	6
lda4int	L2	de	51	6
lda4int	L2	un	44	6
lda4int	L2	que	41	6
lda4int	L2	oui	36	6
...	...	...	...	...

The following SQL query can be used to compare French learner data to the French frequency bands (named 'corpaixband').

```
(1)
SELECT
  i.InformantName,
  i.LinguisticLevel,
  sum(LemmaFreq) as "number of lemmas",
  sum(case when band = 1 then freq else 0 end) as "band 1",
  sum(case when band = 2 then freq else 0 end) as "band 2",
  sum(case when band = 3 then freq else 0 end) as "band 3",
  sum(case when band is null then freq else 0 end) as "offlist"
FROM FrenchInputFile i
left outer join corpaixband b on i.lemma = b.lemma
group by InformantName
order by LinguisticLevel
```

In example (1) above, the content of the field/column 'LemmaFreq' from the table 'FrenchInputFile' is compared to that of 'corpaixband', creating an output file with information about the number of lemmas in the 'FrenchInputFile' belonging to band 1, band 2, band 3 and offlist. The result is grouped and ordered by 'InformantName' and 'LinguisticLevel' as shown in the figure below.

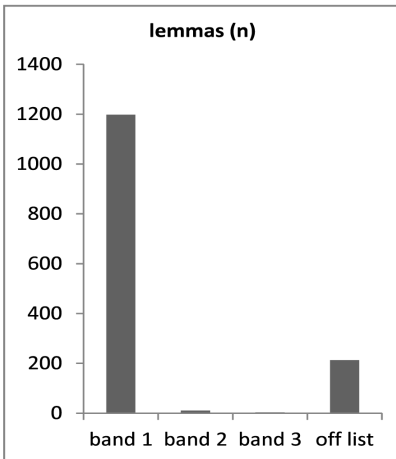
<sup>7</sup> Proficiency level was operationalized as a 1-6 scale based on Bartning & Schlyter's (2004) framework, where 6 corresponds to a very advanced level.

**Figure 2.** Part of a French output file

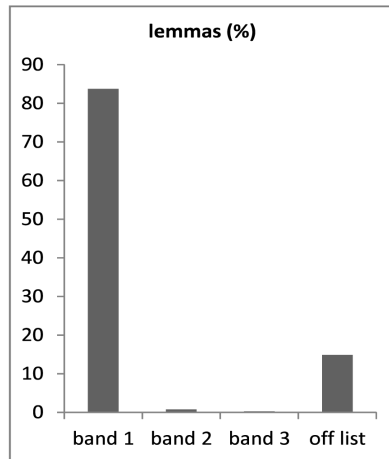
informant name	linguistic level	number of lemmas	band1	band2	band3	offlist
Yvonne1int	4	1841	1573	14	7	247
Christina1int	4	1430	1190	15	14	211
Christina4int	4	1384	1159	30	10	185
Eva1int	4	1669	1385	21	9	254
Eva4int	4	1430	1199	12	5	214
Malena1int	4	2068	1737	16	6	309
Mona1int	4	1249	988	14	3	244
Pernilla1int	4	1126	976	14	8	128
Pernilla4int	4	1171	963	11	3	194
Yvonne4int	4	2284	1904	23	6	351
Ida4int	6	1488	1224	29	9	226
Kerstin1int	6	1764	1457	22	4	281

The output shown in figure 2 can easily be exported to an Excel spreadsheet where the number of lemmas can be converted into proportions. The following figures illustrate the lexical frequency profile, in terms of number and proportions of lemmas, for Eva4int.

**Figure 3.** Lexical richness output: number of lemmas in Eva4int



**Figure 4.** Lexical richness output: proportion of lemmas in Eva4int



Another useful query provides information about the informant's name, the lemma, the frequency of the lemma, the linguistic level of the informant, and the band to which the lemma belongs. The query is shown in example (2) and it returns an output file represented in figure 5.

```
(2)
select
  i.InformantName,
  i.lemma,
  i.LemmaFreq,
  i.LinguisticLevel,
  b.band
from FrenchInputFile i
left outer join corpaixband b on i.lemma = b.lemma
```

**Figure 5.** Part of a French output file

<b>informant name</b>	<b>linguistic level</b>	<b>lemma</b>	<b>lemma freq</b>	<b>band</b>
lda4int	6	le	74	1
lda4int	6	être	71	1
lda4int	6	penser	5	1
lda4int	6	valise	4	offlist
lda4int	6	aimer	3	1
lda4int	6	arabe	3	1
lda4int	6	avion	3	2
lda4int	6	vouloir	3	1
lda4int	6	aéroport	2	offlist
lda4int	6	alors	2	1
lda4int	6	beaucoup	2	1
lda4int	6	bon	2	1
lda4int	6	brancher	2	2
lda4int	6	cinquante	2	1
lda4int	6	comprendre	2	1
lda4int	6	contenter	2	2
lda4int	6	devant	2	1
lda4int	6	inscrire	1	3
lda4int	6	intéressant	1	1
lda4int	6	irréel	1	offlist
...	...	...	...	...

As can be seen from the output file in figure 5, the last column indicates the band to which the lemma belongs. This is useful information when single lemmas have to be studied and analyzed.

#### **4. Pros and cons with frequency-based measures**

Two important advantages with the lexical frequency profiling analysis are that it is able to distinguish between proficiency levels in oral production and that

this measure of lexical richness seems to correlate with the other measures of proficiency used in our earlier studies. However, there are also some important drawbacks with this kind of measure in general. Some of them will be discussed at the end of this paper. There are also problems related to the frequency criterion *per se*. The method relies exclusively on (low-) frequency as a criterion of high level proficiency (or difficulty for the learner). Other factors that may have an impact on learnability (and lexical richness) are cognateness and the role of teaching materials (cf. Horst & Collins, 2006; Milton, 2007). Horst and Collins showed that the use of cognates decreased with higher proficiency, suggesting that cognates (although of low frequency) are not indicative of an advanced vocabulary, in the sense of LFP. As for the role of teaching materials, Milton has pointed out that words that are introduced early, covering certain thematic fields, like travelling or eating out, are learned early, even though they are not used in everyday speech by native speakers, and these words are erroneously classified when regarded as advanced vocabulary. These issues were explored in Bardel and Lindqvist (2011), which led to certain modifications of the LOPP method. These modifications are described in the following section.

#### 4.1. LOPPa: further elaborations of LOPP

Bardel and Lindqvist (2011) investigated the role of cognates and thematic vocabulary in two learners of French and two learners of Italian at different proficiency levels, focusing on the use of low-frequency words. They found that among the low-frequency words produced by the learners there were many cognates and thematic words related to teaching materials, i.e. words, although infrequent, that could be considered rather easy for a Swedish learner of French or Italian. The authors therefore suggested an elaboration of the LOPP tool in order to measure lexical richness in a way that takes not only the proportion of words belonging to a certain frequency band into account, but also the cognate-factor and the thematic word-factor. A new tool, LOPPa, was therefore created. While the old tool, henceforth LOPP<sub>f</sub>, splits the learner data into three frequency bands, LOPPa classifies each word in the learner data as either basic or advanced.<sup>8</sup> The basic vocabulary is composed of a combination of high frequency words, basic cognates and basic thematic words, while the advanced vocabulary is composed of low-frequency words, advanced cognates and advanced thematic words. In order to operationalize the concept of basic cognates and basic thematic words vs. advanced cognates and advanced thematic words, teachers' judgements were used (cf. Tidball & Treffers-Daller, 2008). A full description

---

<sup>8</sup> *a* stands for advanced and *f* for frequency.

of the methodology used to carry out the teachers' judgement test can be found in Bardel et al. (2012).

In order to evaluate the LOPPa tool, data from a previous study carried out with the LOPPf tool (Lindqvist et al., 2011) were re-analyzed with the LOPPa tool (Bardel et al., 2012). It was found that the distinction between basic and advanced words resulted in a higher intra-group homogeneity compared to the purely frequency based perspective. Thus, by taking cognateness and the notion of thematic words into consideration, the lexical richness measure improved, an improvement that was shown by an increased effect size as expressed by  $\eta^2$ .

## 5. Potential areas of use of the method

On the basis of our research we can claim that there are two main advantages with lexical frequency profiling analyses: (1) They are able to distinguish between proficiency levels in oral production. This has been shown both for the method relying only on frequency (Lindqvist et al., 2011) and for the elaborated version of the method, which takes cognates and thematic vocabulary into account (Bardel et al., 2012). (2) LOPPa provides results that seem to correlate with other measures of proficiency used in our earlier studies (mainly measures of morpho-syntactic development).

Another advantage that we would like to point out is that it is possible to conduct both quantitative and qualitative analyses using LOPPa, as opposed to using formulas of lexical richness, e.g. D or TTR. The procedure of LOPPa is to first provide a quantitative result, i.e. the division of the lemmas into bands. In a second phase, it is possible to make an in-depth analysis of the words actually used, by looking at the lists provided by the program. This is possible for a whole data set as well as for individual learners. By making such a thorough analysis it is also possible to continuously improve the method by analyzing the words that appear in the off-list for instance. It is plausible that new cognates and words belonging to thematic vocabulary will appear in the off-list when new data is used in the program. We also believe that the method could be used for pedagogical purposes, for example in order to assess learners' lexical richness in oral production. Teachers could use the basic/advanced word list as a point of reference in vocabulary teaching. The method is also suitable for self-assessment, if learners are given the possibility to analyze their own production within a specific course component at higher levels of education.

It has to be admitted that there are some limitations to the method at this stage of our research. One of the limitations concerns the fact that it is oriented towards learners with Swedish as their L1 and French or Italian as their L2 (and also taking into account that English is an additional second language for all



learners). This certainly limits the number of potential users. However, given the detailed description of the elaboration of the method provided in this paper, there are good possibilities to adapt it for use with other languages. Another limitation is that the method is most suitable for oral data. As we have discussed elsewhere, it is preferable to compare learner data to the same type of data in the target language, as word frequency may differ between oral and written language.

There are also some important drawbacks with this kind of measure of lexical richness in general. One is that it only taps formal aspects of word knowledge. Deep knowledge of vocabulary is not accounted for, e.g. use of words with multiple meanings or use of multi-word units (cf. Nation, 2006; Cobb, this volume). Furthermore, another aspect that remains ignored is non-targetlike use of target language forms. Possible solutions to these problems will be discussed in the following section.

## 6. Possible improvements of LOPPa

There are several aspects that must be learned in order to achieve complete knowledge of a word: form (spoken and written, i.e. pronunciation and spelling), word structure (morphology), syntactic pattern of the word in a phrase and sentence, meaning (referential – including multiplicity of meaning and metaphorical extensions of meaning; affective – the connotation of the word; pragmatic – the suitability of the word in a particular situation), lexical relations of the word with other words (e.g. synonymy, antonymy, hyponymy) and collocations. All these aspects can be more or less well known. The more advanced a learner, the more aspects of a word are likely to be known, and the more developed are the different aspects, for example, more meanings of a homograph are known, more synonyms, more collocations and idiomatic expressions are mastered (Laufer, 1997, p.141).

Qualitative knowledge about the single word is sometimes referred to as *depth*. In his attempt to pinpoint what researchers have in mind when investigating depth of knowledge, Read (2004) distinguishes three approaches to vocabulary learning in the literature, *comprehensive word knowledge*, *precision of meaning* and *network knowledge*. According to the first approach, depth covers different types of knowledge of a word, like those indicated by Laufer (1997, p. 141), all of which, if they are fulfilled, can be called *comprehensive word knowledge*. With *precision of meaning*, Read (2004, p. 211) refers to “the difference between having a limited, vague idea of what a word means and having much more elaborated and specific knowledge of its meaning”. It seems problematic to establish a criterion for precise knowledge. Typically, the criterion is that of the adult native speaker. However, as Read (2004, p. 213) points out, “knowl-

edge of specialized, low-frequency vocabulary reflects in the first instance a person's level and field of education but also their social background, occupation, personal interests and so on". Depth can also be understood as *network knowledge*, i.e. the incorporation of a word into the network surrounding it in the mental lexicon. Word knowledge is sometimes thought of as a network, and words as interconnected nodes. The nodes are interconnected in different dimensions, thematically, phonologically, morphologically, conceptually etc. (Vermeer, 2001, p. 218; Meara, 2009; Gyllstad, this volume).

Two aspects of deep knowledge that are crucial parts of complete word knowledge concern the multiple meaning of polysemic words or homographs and the meaning of multi-word units. Knowing several meanings of a single word form is a kind of deep knowledge that is referred to as *range of meaning* in addition to precision of meaning (see above) by Read (2000, p. 92). The role of context is essential for the interpretation of the meaning of words, and this becomes obvious when dealing with words with multiple meanings and with multi-word units. In lexical frequency profiling, these two aspects become problematic, since the profilers normally do not take context into account. A disadvantage with frequency-based measures such as LFP or LOPPa is that they do not account for the frequency of each meaning attached to a word form (see also Nation, 2006, p. 66). A homograph like French *louer* will always be categorized in the same frequency band independently of the meaning attached to it (*rent* or *praise*), even though the different meanings of the word may not be equally frequent (see Cobb, this volume). It has been suggested that more advanced learners know more meanings of a word than less advanced learners (cf. Bensoussan & Laufer, 1984). It would therefore be a great advantage if lexical profilers could be adapted in order to account for the frequency of the meaning of the word used in a particular context. In that way, the measure would be sensitive to the possible variation of frequency of different meanings of words in the learners' input.

Another qualitative aspect of word knowledge is the knowledge and ability to use multi-word units. A multi-word unit can be defined as a particular combination of words that generates one meaning (see Henriksen, this volume, for an overview of different definitions). One approach to multi-word units is that of Wray (2002), according to whom such combinations of words seem to be retrieved as a whole unit from memory (Wray, 2002, p. 9). This usage of particular word combinations cannot be measured in the LFP, nor in LOPPa, because the programs use graphic criteria to define a word. This means that expressions in French like *tout le monde* (everybody) or *tout à fait* (exactly) will be regarded as three separate words and not as one unit that generates one meaning. Moreover, the words contained in a multi-word unit may belong to different frequency bands. As for *tout à fait*, *tout* and *à* belong to Band 1, while

*fait* is an off-list word. Treating these words separately means that the number of words categorized as highly frequent will rise, although this may not correspond to the frequency of the whole expression in the target language input. In order to account for the frequency of multi-word units, we would have to find a way to integrate them in the frequency lists. It is encouraging to see that work in this direction has started for English (Cobb, this volume; Martinez & Schmitt, 2012). However, considering our approach in the LOPPa framework, we find it pertinent to include multi-word units that are cognates (Wolter & Gyllstad, 2011) and thematic in a basic and an advanced vocabulary.

How could this be accomplished within the LOPPa framework? Every multi-word unit present in the corpus to be analyzed must be tagged as a unit in order to make it appear as a unit and not as several different words. This would lead to a non-match with the baseline corpora, if they are not tagged in exactly the same way, and consequently the multi-word units would end up in the off-list among the low-frequent advanced words. If the aim is to get a picture of the role of frequency for vocabulary learning, as in the LFP, one must make them appear in the frequency bands they actually belong to, and in order to do this the actual frequency of the multi-word units must be looked up in the corpora used as baseline data. Of course, the same goes for the multiple meanings of words. Words occurring in the baseline corpora must be sorted into frequency bands on the basis of the meaning they have in context.

Another important aspect, which is not accounted for in lexical profiling analyses, is the use of words that do not exist in the TL. In fact, non target-like word forms and non target-like use of words (although correct at the formal level) represent an important aspect of vocabulary knowledge. Our main focus thus far has been on the vocabulary use by relatively advanced learners, but earlier research has shown that cross-linguistic influence occurs more frequently at the earlier stages of development (Lindqvist, 2009; Williams & Hammarberg, 2009 [1998]). It is important to integrate this aspect when analyzing the lexical profile of learners. Moreover, as noted above, Read (2000) considers that the proportion of errors is one aspect of lexical richness.

Non target-like use can be instances of code-switching, lexical inventions or other deviant forms of words in the TL (Bardel & Lindqvist, 2007; Dewaele, 1998; Williams & Hammarberg, 2009 [1998]). Vocabprofile gives the instruction to remove code-switches and other deviant forms, and this was also done in the Laufer and Nation (1995) study. We followed this methodology in the LOPPf/a analyses. The main reason for that is that if they had been kept, words belonging to another language than the TL would end up in the off-list, thus adding to the proportion of advanced words. However, in our view, code-switches are also part of the learner's vocabulary, and have something to say about the level of vocabulary proficiency. Moreover, the fact that a learner uses

a correct TL word form does not automatically imply that it is appropriate in the context. However, since lexical profiling methods are not sensitive to context, this type of deviance will not be captured. An example of a word (in this case a multi-word unit) from one of the learners in the present study is the expression *tout le monde* (everybody), which is used in the sense of *le monde entier* (the whole world). The non target-like use of the expression cannot be captured without a closer look at the context.

## 7. Conclusions

As we have shown, several efforts have been made within the project *Aspects of the advanced L2 learner's lexicon*, to create and improve a tool for lexical profiling of Swedish L2 learners' oral production of French and Italian. In a number of steps we have improved our original method LOPP, but there are still many things to develop further. On top of the ideas put forward in this chapter, given that the method is now only available to the research group, an important step forward would be to make the method and the data accessible to other users by providing a user-friendly interface.

## References

- 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), 269-290.
- Bardel, C. & Lindqvist, C. (2007). The role of proficiency and psychotypology in cross-linguistic influence. A study of a multilingual learner of Italian L3. In M. Chini, P. Desideri, M.E. Favilla & G. Pallotti (Eds.), *Atti del XI congresso internazionale dell'Associazione italiana di linguistica applicata. Napoli 9-10 febbraio 2006* (pp. 123-145). Perugia: Guerra.
- Bardel, C. & Lindqvist, C. (2011). Developing a lexical profiler for spoken French and Italian L2: The role of frequency, cognates and thematic vocabulary. In L. Roberts, G. Pallotti, & C. Bettoni (Eds.), *EUROSLA yearbook 11* (pp. 75-93). Amsterdam: Benjamins.
- Bartning, I. & Schlyter, S. (2004). Itinéraires acquisitionnels et stades de développement en français L2. *Journal of French Language Studies*, 14(3), 281-289.
- Bensoussan, M. & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, 7(1), 15-32.
- Campione, E., Véronis, J., & Deulofeu, J. (2005). The French corpus. In E. Cresti, & M. Moneglia (Eds.), *C-ORAL-ROM: Integrated reference corpora for spoken romance languages* (pp. 111-133). Amsterdam: Benjamins.

- Cobb, T. & Horst, M. (2004). Is there room for an academic wordlist in French? In P. Boogards, & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15-38). Amsterdam: Benjamins.
- Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377-387.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Cresti, E. & Moneglia, M. (2005). *C-ORAL-ROM: Integrated reference corpora for spoken romance languages*. Amsterdam: Benjamins.
- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197-222.
- De Mauro, T., Mancini, F., Vedovelli, M., & Voghera, M. (1993). *Lessico di frequenza dell'italiano parlato* (1st ed.). Milano: Etaslibri.
- Dewaele, J. (1998). Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics*, 19(4), 471-490.
- Guiraud, P. (1954). *Les caractéristiques statistiques du vocabulaire*. Paris: Presses Universitaires de France.
- Horst, M. & Collins, L. (2006). From *faible* to strong: How does their vocabulary grow? *Canadian Modern Language Review*, 63(1), 83-106.
- Jones, A., Stephens, R., Plew, R. R., Garrett, B., & Kriegel, A. (2005). *SQL functions programmer's reference (programmer to programmer)*. Indianapolis: Wiley Pub.
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. N. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20-34). Cambridge: Cambridge University Press.
- Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Lindqvist, C. (2009). The use of the L1 and the L2 in French L3: Examining cross-linguistic lexemes in multilingual learners' oral production. *International Journal of Multilingualism*, 6(3), 281-297.
- Lindqvist, C. (2010). La richesse lexicale dans la production orale de l'apprenant avancé de français. *Canadian Modern Language Review*, 66(3), 393-420.
- Lindqvist, C., Bardel, C., & Gudmundson, A. (2011). Lexical richness in the advanced learner's oral production of French and Italian L2. *IRAL*, 49(3), 221-240.
- Linell, P. (2005). *The written language bias in linguistics*. London: Routledge.
- Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.
- Martinez, R. & Schmitt, N. (2012). A phrasal expression list. *Applied Linguistics*, 33(3), 299-320.
- McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- McCarthy, P. M. & Jarvis, S. (2007). *Vocd: A theoretical and empirical evaluation*. *Language Testing*, 24(4), 459-488.

- Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: Benjamins.
- Milton, J. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 47-58). Cambridge: Cambridge University Press.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review* 63(1), 59-82.
- Ovtcharov, V., Cobb, T., & Halter, R. (2006). La richesse lexicale des productions orales: Mesure fiable du niveau de compétence langagière. *The Canadian Modern Language Review*, 61(1), 107-125.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146-161.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *International Conference on New Methods in Language Processing*, Manchester, UK.
- Schmid, H. (1995). Improvements in part-of-speech tagging with an application to German. *Proceedings of the ACL SIGDAT-Workshop*, Dublin, Ireland. 1-9.
- Schneider, S. (2008). *BADIP*. Retrieved 10/10, 2008, from <http://languageserver.uni-graz.at/badip/badip/home.php>
- Scott, M. (2004). *WordSmith tools version 4*. Oxford: Oxford University Press.
- Tidball, F., & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgment can tell us about basic and advanced words. *French Language Studies*, 18(3), 299-313.
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(2), 217-234.
- Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Boogards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 173-189). Amsterdam: Benjamins.
- Williams, S. & Hammarberg, B. (2009 [1998]). Language switches in L3 production: Implications for a polyglot speaking model. In B. Hammarberg (Ed.), *Third language acquisition* (pp. 28-73). Edinburgh: Edinburgh University Press.
- Wilton, P. & Colby, J. W. (2005). *Beginning SQL*. Indianapolis: Wiley.
- Wolter, B. & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, 32(4), 430-449.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

# *Lexical properties in the writing of foreign language learners over eight years of study: single words and collocations*

Tami Levitzky-Aviad and Batia Laufer  
University of Haifa, Israel

Lexical proficiency has been defined and researched in terms of lexical knowledge, use and fluency. Different studies have shown that use of vocabulary in a foreign language (or L2) develops more slowly than vocabulary knowledge, either passive or active. However, many studies of free production compared learners of two or three proficiency levels and examined single words, not multi-word units, even though the latter are characteristic of idiomatic language, and should be considered a component of lexical use.

The data for the present study was collected as part of the on-going compilation of an Israeli learner corpus of written English. The data was analyzed to examine progress in vocabulary use over 8 years of learning, starting with students at the end of elementary school (grade 6) and ending with English majors at the university. The passages were compared on lexical richness – the proportion of frequent to non-frequent vocabulary, on lexical variation – type token ratio, and on the number of collocations. A total of 290 essays (200 words each) were analyzed using the VocabProfile, a software program that calculates the percentage of a text's words at different frequency levels and provides the text's type-token ratio. Significant increases in the use of infrequent vocabulary and collocations were found only with the university students. A significant increase in lexical variation was found at the end of high school. The lack of substantial progress during school years, on the one hand, and the significant progress during the one year at university, on the other hand, corroborate previous research. 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.

## **1. Vocabulary and writing in a foreign language**

The goal of the present study is to examine the development of several 'active' lexical dimensions across eight years of learning English. More specifically, the study aims at investigating developments in active vocabulary knowledge and in three dimensions of vocabulary use: variation, richness and the use of collocation.

tions. Vocabulary is a clear indicator of how well foreign language (FL) learners can communicate (Lewis, 1997; Widdowson, 1989). Effective vocabulary use in writing has been found to have a positive influence on measures of the quality of writing and on one's general language level (e.g. Lee, 2003; Llach & Gallego, 2009; Morris & Cobb, 2004). Also, language learners themselves mention vocabulary as a crucial aspect in writing (Leki & Carson, 1994; Polio & Glew, 1996). It is therefore not surprising that research interest in the importance of vocabulary for writing in a foreign language is growing.

To understand the relationship between vocabulary and writing, we will first explain several key terms in lexical research: lexical knowledge vs. lexical use; depth, breadth and strength of knowledge; passive and active vocabulary knowledge; recall and recognition; lexical variation and lexical richness; and collocations. We will then refer to available research on vocabulary and writing, first for single words, then for collocations.

Vocabulary acquisition can be discussed in terms of both 'lexical knowledge' and 'lexical use'. Lexical knowledge is the information about the word that learners have stored in their mental lexicons, while lexical use is the manifestation of this knowledge in real-time production (Laufer, 2005; Laufer & Goldstein, 2004). This distinction implies that lexical knowledge in a foreign language is typically more advanced than lexical use, because not all words stored in learners' mental lexicons are necessarily activated and used in free writing (Laufer, 1991).

Vocabulary knowledge can be assessed qualitatively, in terms of 'depth' of knowledge, and quantitatively in terms of 'breadth' of knowledge and 'strength' of knowledge. Depth of knowledge refers to the degree of acquaintance with the various form and meaning components of a given lexical entry (e.g. its morphological structure, its grammatical or lexical patterns, and its relations with other lexical items) (Richards, 1976). Breadth of knowledge refers to vocabulary size, i.e. the quantity of lexical entries stored in one's mental lexicon. In measuring vocabulary size, a word is considered 'known' when the correct meaning is associated with the correct word form. However, form-meaning associations can take different forms, reflecting different parameters according to which strength of knowledge is assessed (Laufer, Elder, Hill, & Congdon, 2004; Laufer & Goldstein, 2004). These parameters have been defined along the active-passive and recall-recognition distinctions of meaning-form relationships. More details on how the distinctions were operationalized are provided in the 'Measurement tools' section). The first distinction implies that there is a difference in knowledge between people who can retrieve the FL word form in order to convey a certain meaning ('active' knowledge) and those who cannot do this, but can retrieve the meaning once the FL word is presented to them ('passive' knowledge). The second



distinction implies that there is a difference between those who can recall the form or the meaning of a word and those who cannot do this, but can recognize the form or meaning in a set of options. Four modalities of strength of knowledge thus emerge from these distinctions: active recall, passive recall, active recognition and passive recognition. Of these, active recall is the hardest to achieve, and therefore represents the strongest degree of knowledge, followed by passive recall, active recognition and passive recognition, respectively (Laufer & Goldstein, 2004). In sum, strength of knowledge is a combination of four aspects of knowledge of meaning that constitute a hierarchy of difficulty: passive recognition (easiest), active recognition, passive recall, and active recall (hardest).

Lexical 'variation' and lexical 'richness' are two quantitative measures of vocabulary use. Variation, (or 'diversity'), is a measure of the number of different words (types) used, or, more specifically, the type-token ratio (TTR). 'Richness', on the other hand, is the proportion of low-frequency words in a piece of writing (Laufer, 1994; Laufer & Nation, 1995).

Phraseological analyses suggest that at least one-third to one-half of language is composed of multi-word units (MWU) (Erman & Warren, 2000; Hill, 2000). They are retrieved faster than individual lexical items, indicating perhaps that certain phrases are stored and retrieved as a whole (Erman, 2007; Schmitt, Grandage, & Adolphs, 2004; Wray, 2002). There also seems to be a processing advantage for formulaic sequences, at least in reading (Underwood, Schmitt & Galphin, 2004). Therefore, a good knowledge of formulaic language is advantageous for language learners and users.

Though there are several kinds of MWUs, we focused on the knowledge and use of lexical collocations (henceforth, 'collocations') as it was shown to be one possible indicator of native-like competence (Howarth, 1998; Hill, 2000). We have adopted Nesselhauf's (2003) definition of collocations as word combinations in which one of the words (the 'base' or headword) retains its independent meaning, while the meaning of the other word, (the 'collocate') is restricted to the specific context and can only be used with some semantically related headwords (though not even with all of them). The combinations chosen for investigation in the present research were thus only MWUs which were found compatible with this definition. These included examples such as 'make a decision' or 'heavy rain', but not combinations such as 'eat breakfast' or 'play ball'.

Active vocabulary has been found to be (i) smaller in size, (ii) develop more slowly (Laufer, 1998; Laufer & Goldstein, 2004; Nemati, 2010) and (iii) decay faster (Schneider, Healy, & Bourne, 2002) than passive vocabulary. Accordingly, as mentioned earlier, the most advanced degree of knowledge has been found to be active recall, followed by passive recall, active recognition and passive recognition, respectively (Laufer & Goldstein, 2004). Test results on progress in for-

eign language vocabulary use in writing have shown statistically significant improvements in richness in groups of university English majors (Laufer, 1991; Leńko-Szymańska, 2002), but not so much in school students or in university students who are not English majors. Results of correlation tests between active vocabulary size and use are not as consistent. Some studies have found significant correlations between active vocabulary size and richness and/or variation (e.g. Laufer & Nation, 1995) while others have not (e.g. Laufer, 1998; Lemmouh, 2010).

Knowledge of collocations by FL learners has been found to increase as learning progressed (e.g. Gitsaki, 1999), but not always to the same extent as knowledge of general vocabulary (e.g. Bahns & Eldaw, 1993). In fact, research on the use of collocations by FL learners has demonstrated that even advanced learners have considerable difficulties in producing collocations (Nesselhauf, 2003) and tend to use free combinations where collocations could be used (Nesselhauf, 2005).

## 2. The study

### 2.1. *Research questions and hypothesis*

Our research questions were the following:

- 1) What developments occur in the following dimensions of lexical proficiency during the years of formal English learning?
  - a. the size and strength of active vocabulary knowledge of English words
  - b. the lexical richness of learners' written samples
  - c. the lexical variation of learners' written samples
  - d. the use of collocations in learners' written samples
- 2) Is there a correlation between the improvements in each of the lexical dimensions over the years?

The basic hypothesis underlying the present research was that improvement would occur in all investigated parameters of vocabulary knowledge and use, throughout the eight years of EFL learning. With regard to collocations, this assumption was made despite the limited use of collocations by foreign language learners found in previous research (e.g. Nesselhauf, 2003, 2005). While acknowledging these findings, it was nevertheless assumed that, at least in some of our data that included participants majoring in English, some improvement would occur in this respect too.

## *2.2. Sources of data*

The primary source of data for the present study consists of 290 passages written by learners of English in Israel during the beginning phases of the compilation of the Israeli Learner Corpus of Written English (ILcoWE, Waldman & Levitzky-Aviad, in preparation). This part of the corpus includes 215 passages written by school-aged students in seven consecutive grades (6-12) and 75 passages written by first year university English majors. The topics of essays varied as learners at very different proficiency levels cannot be expected to write on identical topics. The younger students (grades 5-7/8) were mainly asked to write essays of a narrative or descriptive nature in response to prompts such as 'Describe a family event that you attended' or 'Tell the story of what is going on in the drawing...'. The older students (grades 9-12/university majors) were asked to write descriptive and argumentative essays such as 'What would you do if you got a huge sum of money for your birthday? Explain your choices', 'Which is the most important meal of the day, and why?' or 'Argue for and against the use of computers in the classroom'.

Due to time limitations and the need to keep students' personal information confidential, a longitudinal corpus collection, and accordingly, a longitudinal study were impossible. The study is thus cross-sectional, examining the writing of different students at different years of learning.

The second source of data consists of the results of a bilingual test of active vocabulary knowledge, including both active recall and active recognition. Of the students who provided the corpus samples mentioned above, 101 were also administered a test of active vocabulary knowledge (see section 2.3.1). This sample included students at the end of elementary school (in grade 6), at the end of junior high (in grade 9), at the end of high school (in grade 12) and at the beginning of the first year in the English department.

## *2.3. Measurement tools*

### *2.3.1. Measuring active knowledge*

To measure active knowledge in the present research, we designed an active vocabulary test, modelling it upon two well established and validated vocabulary tests. The items were selected from the monolingual versions of the Vocabulary Size Test (VST, Nation & Beglar, 2007). The number of items was also the same as in VST. The methodology of testing, on the other hand, was modelled on the Computer Adaptive Test of Size and Strength (CATSS, Laufer, 2007).

The monolingual version of the VST tests words sampled from the 7,000 most frequent word families in English, based on the British National Corpus. The list can be divided into seven frequency levels (k1-k7), each comprising

1,000 words (Nation, 2006). In the VST, each of these levels is represented by a sample of 20 words. Hence, VST tests peoples' knowledge of a total of 140 items which represent the above mentioned 7,000 word families. As part of the VST, test-takers show their understanding of each English word tested by choosing the correct option from four options of synonyms and definitions of the word.

Though based on the VST, the test used for the current study was a bilingual test. Since the groups which were compared included beginners and low level learners, a bilingual test was considered more appropriate than a monolingual test. Additionally, while the VST tests passive knowledge, or, more specifically, passive recognition (since learners choose the correct paraphrase of the target item), the test designed for the purpose of the present research tested *active* knowledge.

The other test upon which our test was modelled is the CATSS. The specific feature of CATSS, in addition to testing words at different frequency levels, is that it tests the four modalities of strength of knowledge from strongest to weakest (see section 1): active recall, passive recall, active recognition and passive recognition. The test proceeds as follows: In the first modality (active recall), a prompt appears on screen, which is the L1 translation of the target word. The first letter of the target English word is also provided and the test-taker needs to use this letter and type the English equivalent. Words known in this modality are not tested again in subsequent modalities. Representing the hardest, hence strongest degree of knowledge, each correct answer accounts for 1 point of the final CATSS score. In the second modality (passive recall), the English target word appears on screen for the test-taker to translate into the L1. Words known in this modality are not tested again. Each correct answer accounts for 0.75 points of the final CATSS score. In the third modality (active recognition), the test-taker needs to choose the correct English equivalent for the L1 word out of four English options. Words known at this modality are not tested again. Each correct answer accounts for 0.5 points of the final CATSS score. In the last modality (passive recognition) the test-taker needs to choose the correct L1 equivalent for the English target word out of four L1 options. Representing the 'weakest' degree of knowledge, a correct answer at this modality receives 0.25 points of the final CATSS score. Words not known in any of the four modalities receive zero points in the final score. The items tested proceed from frequent to less frequent. Hence, the final CATSS score has been claimed to represent both size and strength of knowledge as it takes into account not only the number of words test-takers know, but also the 'way' in which these words are known (Laufer et al., 2004; Laufer & Goldstein, 2004).

Modelled upon CATSS, the test designed for the present study also takes into account different strength modalities, yet with several modifications. While

CATSS tests both passive and active knowledge, the test in this study tests only active knowledge (hereafter referred to as ACATSS). Another feature distinguishing ACATSS from the original CATSS is that the Hebrew (L1) prompt words in the ACATSS do not appear in isolation, but rather in between two asterisks *within* a Hebrew sentence. The decision to present the word within a sentence was made so as to avoid ambiguity in cases of polysemy of the Hebrew words. Such an approach also follows the model used in the VST.

In the ACATSS, the learners' task is to provide the English equivalent of the word in asterisks. To do so, the test includes three cycles: two for testing active recall and one for testing active recognition.

First, the target item is tested for active recall without any cues, to mirror a real life situation of independent writing. This is demonstrated in the following example, where the target word is 'lake' and the Hebrew sentence means: *This \*lake\* is nice*. The instructions for the test were given in both English and Hebrew so that young learners could also clearly understand what they were expected to do.

Example: cycle 1

Translate the words in \*asterisks\* into English:

ה. אגם\* הזה נחמד.

A word known in this cycle is not tested again. If it is not known, it is tested again in the second cycle. Here too active recall is tested, but now with the first letter of the English word provided. Whereas in cycle 1 learners may provide a non-target word which nevertheless fits the context, the first letter in cycle 2 limits word choice, trying to direct the learners to elicit the target word.

Example: cycle 2

Translate the words in \*asterisks\* into English

(use the first letter of the English word as provided for you):

ה. אגם\* הזה נחמד. 1

Based on the assumption that words known in active recall would also be known in active recognition (Laufer et al., 2004; Laufer & Goldstein, 2004), only words which were not known in either one of the active recall stages are tested again for active recognition. In this third cycle, learners are presented with four English words of which they are asked to choose the correct equivalent for the Hebrew word in asterisks. The distracters in the recognition stage were sampled from the same frequency level as the English target word to eliminate the effect that word frequency might have on the choice of the response.

Example: cycle 3

Circle the correct translation for each of the words in \*asterisks\*:

הַגַּגְמָה \*אֶגֶם\* הַיָּם הַיָּבֵשׁ

a. tale

b. rhythm

c. lake

d. lawn

Once all 20 words at one frequency level are tested, the test moves on to the next frequency level. A word scores 1 point if known in the first cycle (active recall with no cue), 2/3 if known in the second cycle (active recall with a cue), 1/3 in the third cycle (active recognition) and 0 for lack of any knowledge. The total score for each frequency level is calculated by adding up the scores learners receive for the 20 words. The total scores of all seven frequency levels are then summed up to provide one total ACATSS score. As in the VST, since the 140 words tested in the ACATSS represent a vocabulary size of 7,000 word families, the total ACATSS score can be multiplied by 50 to provide an indication of active vocabulary size as affected by the strength modalities tested.

### 2.3.2. *Measuring use - VocabProfile*

The sampled written passages were analyzed with the experimental BNC-20 version of the Web-VocabProfile (WebVP) program on the Lextutor website (<http://www.lex Tutor.ca> Cobb, n.d.). The WebVP is an adapted version of Heatley and Nation's Range program (1994). Both the Range and the WebVP programs match a text with frequency lists and show the relative proportion of words used from different frequency levels. The relative proportion is called LFP (Lexical Frequency Profile). The program also calculates the type-token ratio (TTR) of an essay. The profiles created with these programs present the proportions of k1, k2 and Academic Word List (AWL Coxhead, 2000) words in a text. The experimental BNC-20 version, on the other hand, presents the proportion of words in a text which are taken from the revised 20 frequency levels based on the British National Corpus (Nation, 2006; Cobb, 2007). In this sense, it seems to provide a more detailed and fine-grained profile of the learners' writing. Additionally, as with the use of the ACATSS for active vocabulary knowledge, the experimental BNC-20 WebVP might be more sensitive than earlier versions to developments in vocabulary profiles between different learning stages.

To use the VocabProfile, various steps had to be taken regarding the corpus data that were used. The profile has been shown to be less stable with essays shorter than 200 words. Such essays were therefore excluded from the present research. Furthermore, in light of the sensitivity of the TTR to composition length (e.g. Kucera & Francis, 1967; Linnarud, 1986), only 200 words of each passage were sampled, even if more words were written.

Three scores were obtained with the VocabProfile. Following the distinction between the first 2000 words (k1-k2) as the most frequent words and the beyond-2000 levels (k3-k20) as the low frequency words (Nation & Kyongho, 1995), we first added up the percentages of k3-k20 to obtain the general percentage of the low frequency vocabulary in the passages. The score obtained was thus considered an indication of how 'rich' the piece of writing was. However, since some of the learners whose essays were sampled for the research were at the very early stages of EFL learning, we also separated the percentages of the 1st and the 2nd 1000 words. Additionally, the TTR obtained with the VocabProfile program was taken to be an indication of variation.

### *2.3.3. Testing the use of collocations*

No measurement tool was employed for testing the use of collocations in the written samples. These were manually counted. Once a word combination was identified as a possible collocation, a further step was taken to check whether these specific combinations were used in native-speakers' language. To this end, three sources based on native-speakers corpora were consulted: the Longman Exams Coach (Summers, Mayor, & Elston, 2006), the Oxford Collocations Dictionary (McIntosh, Francis, & Poole, 2009) and the word frequency list of American English (Davis & Gardner, 2010). If the expression appeared in at least one of these sources, it was considered a collocation.

As we performed the manual check, three things became apparent. First, as in Hsu (2007), the collocations were mostly verb-noun, or adjective-noun collocations. Therefore, only the use of these grammatical combinations was examined. Secondly, the number of collocations in each of the 200-word samples seemed quite small (see table 4.1), and, in many cases, they were the same ones used more than once (in accordance with Nesselhauf, 2005). Counting the total number of such collocations, then, (with many of them repeatedly used), did not seem to be of great value in checking for lexical growth over the years. Hence, following a similar procedure to that used by Zhang (1993) and Hsu (2007), each specific collocation was counted only once even if it was used repeatedly (in much the same way as the counting of 'word types' with single words).

### *2.4. Data Analysis*

When we applied the three procedures outlined in section 2.3, five scores were obtained, each representing one dimension of active lexical proficiency. The total ACATSS scores were used as a measure of active knowledge size and strength. The proportions of k2 words and k3-k20 words in the written samples as calculated by the VocabProfile were used as two measures of lexical richness in writing. The type-token ratio as calculated by the VocabProfile was used as a meas-

ure of lexical variation in writing. Finally, the total number of different verb-noun and adjective-noun collocations was used to examine their prevalence in the written samples.

Four sets of one-way ANOVAs and post-hoc tests were used to compare learners at different points of learning on each of the four dimensions of lexical proficiency: size and strength of active vocabulary knowledge, richness, variation and the use of collocations.

Pearson correlations were then used to test whether the improvements in each of the lexical dimensions over the years correlate with each other.

## 2.5. Results

Our first research question addressed the developments in each of the dimensions of lexical proficiency. Tables 1.1 – 4.2 show the results for each dimension. As noted in section 2.2, the written data analyzed in the present study consisted of the 290 passages written by school-aged students in grades six to twelve and by first year university English majors. However, the ACATSS results were only obtained for 101 of these students. Thus, tables 1.1 and 1.2, showing the results for active knowledge, refer only to students in grades 6, 9 and 12 and the university students at the beginning of their first year in university. Tables 2-4 then show the results for the different measures of vocabulary use in the written passages for all the school grades tested and for the university students at the beginning and at the end of their first year.

### 2.5.1. RQ 1a: What developments occur in the size and strength of active vocabulary knowledge of English words during the years of formal English learning?

Table 1 presents the means of the raw scores for each of the English learning stages tested by the ACATSS. Table 2 shows the significance of differences between the different pairs of learning stages. As noted in section 2.2, only 101 of the 290 students were tested with the ACATSS. Accordingly, the results in tables 1 and 2 only refer to these students. Table 1 shows that the mean ACATSS scores increase at each learning stage; table 2 shows that the differences between all pairs of stages are statistically significant.

**Table 1.** Raw ACATSS scores (out of a maximum of 140) (n=101 learners)

Learning Stage	N	Min	Max	Mean	SD
Grade 6 (end of element. school)	15	9	21	15	4
Grade 9 (end of junior-high)	27	27	50	37	5
Grade 12 (end of high school)	29	30	62	46	9
Eng. Majors- beginning	30	39	74	57	10



**Table 2.** Differences in mean ACATSS scores between learning stages

Learning Stage	Grade 6	Grade 9	Grade 12
Grade 9 (end of junior-high)	22**		
Grade 12 (end of high school)	32**	10**	
Eng. Majors- beginning	42**	20**	10**

\*\*p&lt;0.01

### 2.5.2. RQ 1b: What developments occur in the lexical richness of learners' written samples during the years of formal English learning?

Table 3 presents the mean proportions of k3-k20 words in the written samples. Table 4 shows the significance of differences in these proportions between all of the different pairs of learning stages. Table 5 presents the mean proportions of k2 words in the written samples. Table 6 shows the significance of differences in these proportions between all of the different pairs of learning stages.

Table 3 shows a general increase across the learning stages represented by school/university years in the mean proportion of k3-k20 words in the written samples, despite some slight decreases between some of the learning stages (e.g., grade 9 - 3.84%, grade 10 - 3.65%). However, as shown in table 4, in school years all these changes appear to be statistically insignificant. In other words, in the six years between the end of elementary school (grade 6) and the end of high-school there are no statistically significant increases in the use of low frequency words of k3-k20. Statistically significant improvements occur between each of the school grades 6-12 and the English majors at the end of their 1st year in the English department and between each of the school grades 6-10 and the English majors at the beginning of their first year. Another significant improvement occurs in the one year of English studies at the English departments in the college or university.

**Table 3.** Mean proportions (in %) of k3-k20 words in the written samples (n=290 learners)

Learning Stage	N	Min (%)	Max (%)	Mean (%)	SD
Grade 6	15	1.5	5.45	3.24	1.20
Grade 7	21	.99	5.37	2.85	1.11
Grade 8	35	1	6.40	3.28	1.54
Grade 9	30	.98	6.86	3.84	1.62
Grade 10	39	0	8.16	3.65	1.78
Grade 11	36	.51	7.92	4.04	1.80
Grade 12	39	.50	8.54	4.17	1.78
Eng. Majors- beginning	36	1.49	12.75	5.48	2.74
Eng. Majors-end of 1st year	39	.50	16.58	7.75	3.37

**Table 4.** Differences in k3-k20 proportions between stages of learning

<b>Learning Stage</b>	<b>Grade 6</b>	<b>Grade 7</b>	<b>Grade 8</b>	<b>Grade 9</b>	<b>Grade 10</b>	<b>Grade 11</b>	<b>Grade 12</b>	<b>Eng. Majors-beginning</b>
Grade 7	.39							
Grade 8	.04	.43						
Grade 9	.60	.99	.56					
Grade 10	.41	.80	.37	.19				
Grade 11	.80	1.19	.76	.20	.39			
Grade 12	.93	1.32	.90	.33	.52	.13		
Eng. Majors-beginning	2.24*	2.63**	2.21**	1.64*	1.83**	1.44	1.31	
Eng. Majors-end of 1st year	4.51**	4.91**	4.48**	3.91**	4.10**	3.71**	3.58**	2.27**

\* $p < 0.05$  \*\* $p < 0.01$

Table 5 shows a general increase in the use of k2 words. Table 6 shows that significant increases in the use of these words occur already during school years between each of the grades 6-10 and grade 12. Statistically significant improvements also occur between each of the school grades 6-10 and the two university stages.

**Table 5.** Mean proportions (in %) of k2 words in the written samples (n=290 learners)

<b>Learning Stage</b>	<b>N</b>	<b>Min (%)</b>	<b>Max (%)</b>	<b>Mean (%)</b>	<b>SD</b>
Grade 6	15	2.5	7.35	4.55	1.40
Grade 7	21	1.46	8.37	4.63	2.06
Grade 8	35	1.95	8.29	5.13	1.83
Grade 9	30	0	10.26	4.82	2.64
Grade 10	39	1.46	9.80	5.34	2.88
Grade 11	36	.50	11.50	5.79	2.99
Grade 12	39	1.99	12.56	7.25	2.58
Eng. Majors- beginning	36	2.49	13.93	7.27	3.18
Eng. Majors-end of 1st year	39	2.42	18.65	7.37	3.22

**Table 6.** Differences in k2 proportions

Learning Stage	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	Eng. Majors-beginning
Grade 7	.08							
Grade 8	.58	.50						
Grade 9	.28	.19	.30					
Grade 10	.79	.71	.21	.51				
Grade 11	1.24	1.16	.66	.96	.45			
Grade 12	2.70*	2.62**	2.12*	2.43**	1.91*	1.46		
Eng. Majors-beginning	2.72*	2.64**	2.14*	2.44**	1.93*	1.48	.02	
Eng. Majors-end of 1st year	2.83*	2.74**	2.25**	2.55**	2.03*	1.59	.12	.11

\* $p < 0.05$  \*\* $p < 0.01$

### 2.5.3. RQ 1c: What developments occur in the lexical variation in learners' written samples during the years of formal English learning?

Table 7 presents the mean type-token ratio reflecting lexical variation, i.e., the percentage of different words in the text. Table 3.2 shows the significance of differences between all the different pairs of EFL learning stages in regard to the type-token ratios.

**Table 7.** Type-Token ratios (in %) of the written samples (n=290 learners)

Learning Stage	N	Min (%)	Max (%)	Mean (%)	SD
Grade 6	15	41	57.29	50.98	4.79
Grade 7	21	43.41	58.25	49.77	3.83
Grade 8	35	43.37	60.50	53.09	3.97
Grade 9	30	41.09	60.10	53.05	4.63
Grade 10	39	42.36	60.50	52.95	3.78
Grade 11	36	43.07	59.41	52.56	4.19
Grade 12	39	46.83	64.71	56.78	4.06
Eng. Majors- beginning	36	46.77	63.96	56.83	3.93
Eng. Majors-end of 1st year	39	48.74	66.50	56.77	4.32

Table 7 shows a general increase in the type-token ratios in the writing samples, despite some slight decreases which occasionally occur (e.g., grade 6 – 50.98%, grade 7 – 49.78%). The only statistically significant differences, however (table 8) are between each of the grades 6-11 and grade 12 and between each of the grades 6-11 and each of the university stages.

**Table 8.** Differences in the Type-Token ratios

Learning Stage	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	Eng. Majors-beginning
Grade 7	-1.21							
Grade 8	2.11	3.32						
Grade 9	2.07	3.27	.04					
Grade 10	1.97	3.18	.13	.09				
Grade 11	1.58	2.78	.53	.49	.40			
Grade 12	5.80**	7.01**	3.69**	3.73**	3.82**	4.22**		
Eng. Majors-beginning	5.85**	7.06**	3.75**	3.79**	3.88**	4.28**	.05	
Eng. Majors-end of 1st year	5.79**	7**	3.69**	3.73**	3.82**	4.22**	0	.06

\* $p < 0.05$  \*\* $p < 0.01$

#### 2.5.4. RQ 1d: What developments occur in the use of collocations in the learners' written samples during the years of formal English learning?

Table 9 presents the raw means of different (not repeated) verb-noun and adjective-noun collocations found in the learners' written samples of 200 tokens each. Table 10 shows the significance of differences between all the different pairs of EFL learning stages in regard to the use of these collocations.

Table 9 shows a general increase in the use of collocations, despite some decreases which occur occasionally (e.g., grade 10 – 0.72, grade 11 – 0.42). However, table 10 demonstrates that the only statistically significant differences are between each of the school grades (6-12) and the English majors at the end of their first year and between each of the grades 6-9 and 11 and the English majors at the beginning of the first year.

**Table 9.** Raw means of different collocations in the 200-word samples (n=290 learners)

Learning Stage	N	Min (raw)	Max (raw)	Mean (raw)	SD
Grade 6	15	0	1	0.13	0.35
Grade 7	21	0	2	0.38	0.59
Grade 8	35	0	2	0.23	0.55
Grade 9	30	0	2	0.37	0.61
Grade 10	39	0	5	0.72	1.15
Grade 11	36	0	2	0.42	0.60
Grade 12	39	0	4	0.72	0.94
Eng. Majors- beginning	36	0	7	1.31	1.65
Eng. Majors-end of 1st year	39	0	5	1.56	1.57

**Table 10.** Significance of differences between the raw means of collocations

Learning Stage	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	Eng. Majors-beginning
Grade 7	.25	Grade 8	.10	.15				
Grade 9	.23	.01	.14					
Grade 10	.58	.34	.49	.35				
Grade 11	.28	.04	.19	.05	.30			
Grade 12	.58	.34	.49	.35	.00	.30		
Eng. Majors-beginning	1.17*	.92*	1.08**	.94*	.59	.89*	.59	
Eng. Majors-end of 1st year	1.43**	1.18**	1.34**	1.20**	.85*	1.15**	.85*	.26

\* $p < 0.05$  \*\* $p < 0.01$

Table 11 shows the results of Pearson product moment correlations between the developments, that is, the mean differences of the various lexical dimensions over the years. Correlations with the ACATSS were conducted only for the 101 students who took this test. All other correlations were conducted for all 290 students.

The table 11 shows that the improvements in almost all lexical dimensions over the years correlate significantly with each other. Lack of significant correlation was found only between the results of the progress on the ACATSS and the progress in the use of collocations.

**Table 11.** Correlations between the mean differences of the various lexical dimensions

	Active knowledge size & Strength (ACATSS) (N=101)	Variation (TTR) (N=290)	Richness #1 (k3-k20) (N=290)	Richness #2 (k2) (N=290)
Variation (TTR) (N=101)	.380**			
Richness #1 (k3-k20) (N=290)	.207**	.297**		
Richness #2 (k2) (N=290)	.298**	.348**	.316**	
Use of collocations (N=290)	.149	.326**	.222**	.201**

\*\* $p < 0.01$

### 3. Discussion

The main focus of this study was the similarities and differences in the developmental patterns of several dimensions of L2 lexical proficiency over eight years of study. We will therefore discuss the progress found for each dimension and compare the development of vocabulary knowledge with that of vocabulary use.

Continuous statistically significant improvements were found in active knowledge as reflected in the ACATSS scores across all stages of English learning (see tables 1 and 2). And yet, these significant improvements should also be considered vis-à-vis what they mean in terms of active vocabulary size and its growth, and, even more so, in terms of the manifestation of this knowledge in vocabulary *use*.

An increase in the size of knowledge suggests that there is an increase in the amount of low-frequency words learners know. We can therefore expect that at least those learners who have demonstrated a relatively high command of the language and are accepted to the English department would also possess knowledge of more lower-frequency words than would the general population of school-aged students for whom English is not the major area of study. When multiplying the mean ACATSS score of the first year English majors (see table 1) by 50 to reach the more general estimate of their active vocabulary size (see section 2.3.1), the figure reached is 2850 (57x50). Hence, despite the statistically significant increase in active vocabulary size from the 12th grade to the beginning of the 1st year in the English department (see table 1.2), even the advanced students in the latter group know fewer than 1000 words beyond the 2000 most frequent words in English.

Furthermore, although these figures represent the development in active *knowledge*, they do not necessarily reflect a similar vocabulary growth in free writing. With regards to free writing, the results show a gradual, and sometimes statistically significant, progress in the three dimensions of vocabulary use we tested: richness, variation and the use of collocations. However, while active knowledge demonstrated a *continuous* significant increase throughout the years, our findings, similar to previous ones (Laufer, 1991; Laufer & Nation, 1995; Laufer & Paribakht, 1998; Lemmouh, 2010; Leńko-Szymańska, 2002; Muncie, 2002) indicate that six or more years must pass before students' ability to put this knowledge into use also significantly improves. More specifically, a statistically significant improvement in lexical variation was evident only at the end of high-school (see table 8), whereas statistically significant improvements in the use of the k3-k20 low-frequency words were completely lacking during school years and occur only during the one year of university (see table 4). Lack of significant progress is also evident in the use of collocations, not only during school years, but also during the one

year of university (see table 10). These results corroborate previous findings (Laufer & Waldman, 2011; Nesselhauf, 2003; Pawley & Syder, 1983) and provide a clear indication of the specific difficulty involved in incorporating collocations into the writing of even advanced learners. Laufer and Waldman (2011) explained this difficulty in terms of semantic transparency of collocations and their difference from L1. As many collocations are easily understood, they go unnoticed in the input, and as a collocate in an L2 collocation is often different from L1, learners cannot rely on their L1 and on the knowledge of the individual words in L2.

The lack of statistically significant improvements in students during the six earlier school years, as well as the lack of significant progress in the use of collocations even during the one advanced year at university, are even more puzzling given that richness and variation in vocabulary use can improve even over the course of a single year at university. Since not all school students eventually become English majors, some of them may never again study English in a formal setting. It is hard to accept, then, that what school students end up with is only an active vocabulary size of just over 2000 word families ( $46 \times 50 = 2300$ ), and, perhaps, a higher ability to vary the vocabulary they are able to use, without similar increases in the numbers of lower-frequency words or collocations they use.

A few possible explanations can be provided to account for the discrepancies between vocabulary knowledge and use and for the lack of significant progress in vocabulary use during earlier school years. One possible assumption which could have been made is that the nature of vocabulary learning may be such that active knowledge and use are separate traits of lexical proficiency, which develop in totally different ways. However, the moderate correlations we found between vocabulary knowledge and use (see table 11), similar to previous studies (Laufer & Nation, 1995; Leńko-Szymańska, 2002), point to a different interpretation of the results. These correlations indicate that, despite the discrepancies between vocabulary knowledge and use, an increase in learners' active vocabulary knowledge may be moderately reflected in their *use* of richer vocabulary. Also, the statistically significant increase in the use of k3-k20 words during the one year at university suggests that rapid progress in vocabulary use is possible. Hence, taken together, the significant correlations found between active vocabulary knowledge and use and the progress in the use of low-frequency words over the one year of university suggest that the lack of statistically significant growth we found in lexical use could be changed.

Therefore, another explanation for the lack of significant progress in vocabulary use during earlier school years could be the lack of sufficient language training and practice during these years, which could result from learners' writing strategies, the teaching methods applied and/or the time of expo-

sure to English during school years. Coming up with a word to express a certain idea in writing requires learners to know more features of that word than they need when they are asked to provide the word in some controlled setting. However, due to factors such as the rarity of low frequency words, the arbitrary nature of collocations or various incongruencies between L1 and L2 collocations, learners may experience uncertainties regarding the use of such lexical items and may thus simply refrain from using them (Fan, 2009; Hill, 2000; Laufer, 1998; Laufer & Waldman, 2011; Nesselhauf, 2003). Instead, they may resort to using high frequency single words which convey the same, or at least similar, ideas. This strategy is reinforced by teachers who believe that for communication to be effective, foreign language learners' ability to express their ideas using any appropriate vocabulary is satisfactory in many cases. Unfortunately, such a claim, especially when made by teachers, downplays the need for sufficient practice of non-basic vocabulary (Laufer, 2005; Nemati, 2010; Milton, this volume) and, consequently, perpetuates stagnation of vocabulary in free expression. This lack of progress is not something that any education system should welcome.

To achieve progress, specific and realistic goals need to be set, and effective teaching methods need to be implemented. Such teaching methods should involve acknowledging the importance of encouraging FL learners' use of low-frequency vocabulary and collocations in their writing. Previous studies have shown the effectiveness of Form-Focused Instruction (FFI) in activating learners' lexical knowledge and putting some of it to use (Laufer, 2005; Laufer, 2010; Laufer & Girsai, 2008; Lee, 2003; Nesselhauf, 2003; Webb, 2005; Xiao & McEnery, 2006). Such an approach advocates explicit vocabulary instruction, either as part of more general communication tasks (Focus on Form-FonF) or as a goal in itself (Focus on Forms – FonFs). A longitudinal systematic syllabus of FFI which gradually introduces low-frequency words and collocations and encourages their use could be a possible solution for enhancing the knowledge and use of such items at all stages of L2 learning.

Future research could compare the development of EFL vocabulary use in writing in different educational systems, in different classes or in different controlled experimental conditions. Such comparisons might be useful to show the effectiveness of different pedagogical approaches for the development of L2 vocabulary use over the years.



## References

- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21(1), 101-114.
- Cobb, T. (n.d.). *Web Vocabprofile: An adaptation of Heatley & Nation's (1994) Range*. Computer program. Available on-line at <http://www.lextutor.ca/vp/>
- Cobb, T. (2007). *The revised frequency lists of k8-k14*. Available on-line at [http://www.lextutor.ca/vp/bnc/cobb\\_6](http://www.lextutor.ca/vp/bnc/cobb_6)
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2): 213-238.
- Davis, M. & Gardner, D. (2010). *Word Frequency List of American English*. Available on-line at [www.wordfrequency.com](http://www.wordfrequency.com)
- Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics* 12(1), 25-53.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse*, 20(1), 29-62.
- Fan, M. (2009). An exploratory study of collocational use by ESL students-A task based approach. *System*, 37(1), 110-123.
- Gitsaki, C. (1999). *Second language lexical acquisition: A study of the development of collocational knowledge*. San Francisco, CA: International Scholars Publications.
- Heatley, A. & Nation, P. (1994). *Range*. Victoria University of Wellington, NZ. Computer program. Available on-line at <http://www.vuw.ac.nz/lals/>
- Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching Collocation: Further Development in the Lexical Approach* (pp. 47-70). Hove: Language Teaching Publications.
- Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161-186). Oxford: Clarendon Press.
- Hsu, J. (2007). Lexical collocations and their relation to the online writing of Taiwanese college English majors and non-English majors. *Electronic Journal of Foreign Language Teaching*, 4(2), 192-209.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press Providence, RI.
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, 75(4), 440-448.
- Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal*, 25 (2), 21-33.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2), 255-271.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5(1), 223-250.
- Laufer, B. (2007). *CATSS: The Computer Adaptive Test of Size and Strength*. Computer program. Available on-line at <http://hcc.haifa.ac.il/~blaufer/>

- Laufer, B. (2010). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, 24(1), 29-49.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48 (3), 365-391.
- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202-226.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Journal of Learning Language*, 54(3), 399-436.
- Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29, 694-716.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647-672.
- Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System*, 31(4), 537-561.
- Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *Tesol Quarterly*, 28(1), 81-101.
- Lemmouh, Z. (2010). *The Relationship among Vocabulary Knowledge, Academic Achievement and the Lexical Richness in Writing in Swedish University Students of English*. Ph.D. Dissertation, Department of English, Stockholm University.
- Leńko-Szymańska, A. (2002). How to trace the growth in learners' active vocabulary? A corpus based study. *Teaching and Learning by Doing Corpus Analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora*. Graz (pp. 19-24).
- Lewis, M. (1997). Pedagogical implications of the lexical approach. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 255-270). Cambridge: Cambridge University Press.
- Linnarud, M. (1986). Lexis in composition: A performance analysis of Swedish learners' written English. *Dissertation Abstracts International. C: European Abstracts*, 47 (4), 812.
- Llach, M. P. A., & Gallego, M. T. (2009). Examining the relationship between receptive vocabulary size and written skills of primary school learners. *ATLANTIS*, 31, 129-147.
- McIntosh, C., Francis, B., & Poole, R. (Eds.) (2009). *The Oxford Collocations Dictionary*. Oxford: Oxford University Press.
- Morris, L., & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of teaching English as a second language trainees. *System*, 32(1), 75-87.
- Muncie, J. (2002). Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System*, 30(2), 225-235.
- Nation, I.S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, 63(1), 59-82.

- Nation, I.S.P., & Kyongho, H. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23(1), 35-41.
- Nation, I.S.P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7): 9-13.
- Nemati, A. (2010). Active and passive vocabulary knowledge: The effect of years of instruction. *The Asian EFL Journal Quarterly* 12(1), 30-46.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: Benjamins.
- Pawley, A., & Syder, F. H. (1983). Natural selection in syntax: Notes on adaptive variation and change in vernacular and literary grammar. *Journal of Pragmatics*, 7(5), 551-579.
- Polio, C., & Glew, M. (1996). ESL writing assessment prompts: How students choose\* 1. *Journal of Second Language Writing*, 5(1), 35-49.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77-89.
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In Schmitt, N. (ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 127-151). Amsterdam: Benjamins.
- Schneider, V. I., Healy, A. F., & Bourne L. E. Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, 46(2), 419-440.
- Summers, D., Mayor, M., & Elston, J. (Eds.), (2006). *The Longman Exams Coach*. Essex: Pearson-Longman.
- Underwood, G., Schmitt, N., & Galphin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In Schmitt, N. (ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp 153-172). Amsterdam: Benjamins.
- Waldman, T. & Levitzky-Aviad, T. (in preparation). The Israeli Learner Corpus of Written English (ILcoWE).
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(01), 33-52.
- Widdowson, H. G. (1989). Knowledge of language and ability for use. *Applied Linguistics*, 10(2), 128-137.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27(1), 103.
- Zhang, X. (1993). *English collocations and their effect on the writing of native and non-native college freshmen*. PhD. Dissertation, Indiana University of Pennsylvania.



# *Automatic extraction of L2 criterial lexico-grammatical features across pseudo-longitudinal learner corpora: using edit distance and variability-based neighbour clustering*

Yukio Tono

Tokyo University of Foreign Studies

The aim of this study is to offer a generic technique of extracting lexico-grammatical features that serve as criteria for distinguishing one CEFR level from the others in pseudo-longitudinal learner corpora. Semi-automatic error tagging for surface error taxonomy was performed on a written corpus of 10,038 Japanese EFL learners by comparing the original essays against the proofread ones, by using edit distance and automatic POS tagging. The output was further processed using multivariate statistics called correspondence analysis and variability-based neighbour clustering to examine whether those automatically assigned errors could serve as criterial features. The results show that this new approach of error annotation and clustering is useful to identify criterial features for lower levels that are not provided by the English Profile Programme and suggest an alternative classification of features for all CEFR levels.

## **1. Introduction**

In SLA, it is becoming increasingly popular to use techniques and resources developed in the field of corpus linguistics and natural language processing. The use of learner corpora, systematically sampled collections of learner speech or writing in a machine-readable format, is rapidly gaining ground among ELT materials developers, practitioners and SLA researchers (Granger, 1998; Granger, Hung, & Petch-Tyson, 2002). Behind all of this, there is a growing awareness that frequency of items to be acquired in input plays an important role in L1 and L2 acquisition processes (Gries & Divjak, 2012). According to Goldberg (1995, 2006), the Saussurian concept of a symbolic unit, that is a form-meaning pair, is assumed to cover not only the level of words, but also applies to constructions at all levels of semantic linguistic representation from morphemes and words to increasingly complex syntactic configurations. This symbolic unit is acquired through the exposure to the target language in context. I would argue that with the advent of corpus linguistics and natural language processing, SLA researchers should once again

Table 1. Possible criterial feature types

Type of feature	Descriptions	Examples (based on Hawkins & Buttery 2010)
Positive linguistic properties of the L2 levels	Correct properties of English that are required at a certain L2 level and that generally persist at all higher levels. E.g. property P acquired at B2 may differentiate [B2, C1 and C2] from [A1, A2 and B1] and will be criterial for the former.	The ditransitive NP-V-NP-NP structure ( <i>she asked him his name</i> ) appears at B1, and is thus criterial for [B1, B2, C1, C2]. The object control structure, NP-V-NP-AdjP ( <i>he painted the car red</i> ) is criterial for [B2, C1, C2].
Negative grammatical properties of the L2 levels	Incorrect properties or errors that occur at a certain level or levels, and with a characteristic frequency. Both the presence versus absence of the errors, and the characteristic frequency of error can be criterial for the given level or levels. E.g. error property P with a characteristic frequency F may be criterial for [B1 and B2].	Errors involving incorrect morphology for determiners, as in Derivation of Determiners (abbreviated DD) <i>She name was Anna</i> (instead of <i>Her name ...</i> ), show significant differences in error frequencies that decline from B1 > B2 > C1 > C2.
Positive usage distributions for correct L2 properties	Positive usage distributions for a correct property of L2 that match the distribution of native speaking (i.e. L1) users of the L2. The positive usage distribution may be acquired at a certain level and will generally persist at all higher levels and be criterial for the relevant levels.	The distribution of relative clauses formed on indirect object/oblique positions (e.g. <i>the professor that I gave the book to</i> ) to relativizations on other clausal positions (subjects and direct objects) appears to approximate that of native speakers at the C levels, but not at earlier levels. Hence this is a positive usage distribution that is criterial for [C1, C2]
Negative usage distributions for correct L2 properties	Negative usage distributions for a correct property of L2 that do not match the distribution of native speaking (i.e. L1) users of the L2. The negative usage distribution may occur at a certain level or levels with a characteristic frequency F and be criterial for the relevant level(s).	The distribution of relative clauses formed on indirect object/oblique positions is the negative usage distribution, criterial for B2 and below.

focus on descriptive aspects of IL processes, in addition to already available introspective and experimental methods. By identifying the use/misuse of language features and their relative frequencies at different developmental stages in more detail, one can take into account frequency effects in language acquisition and learning.

To this end, a very unique project called the English Profile Programme (EPP) has started. It is sponsored by the Council of Europe and is maintained by the research team including Cambridge ESOL Examinations, Cambridge RCEAL, and University of Bedfordshire. The aim of the EPP is to create a 'profile' or set of Reference Level Descriptions (RLDs) for English linked to the Common European Framework of Reference (CEFR). The EPP website (<http://www.englishprofile.org/>) states that reference level descriptions

will provide detailed information about the language that learners can be expected to demonstrate at each CEFR level (A1 & A2: basic user; B1 & B2: independent user; C1 & C2: proficient user), offering a clear benchmark for progress that will inform curricula development as well as the development of courses and test material to support learners, teachers and other professionals in the teaching of English as a foreign language.

What is unique in the EPP is its corpus-based method of finding 'criterial features' from learner corpora sampled from the subjects at different CEFR levels. Salamoura and Saville (2009) defined a 'criterial feature' as follows (Salamoura & Saville, 2009, p. 34).

A 'criterial feature' is one whose use varies according to the level achieved and thus can serve as a basis for the estimation of a language learner's proficiency level. So far the various EP research strands have identified the following kinds of linguistic feature whose use or non-use, accuracy of use or frequency of use may be criterial: lexical/semantic, morpho-syntactic/syntactic, functional, notional, discourse, and pragmatic.

Hawkins and Buttery (2010), for example, have identified four types of feature that may be criterial for distinguishing one CEFR level from the others. Table 1 shows the classifications.

The English Profile (EP) researchers have done preliminary studies with regard to the criterial features, using the Cambridge Learner Corpus (CLC) (Williams, 2007; Parodi, 2008; Hendriks, 2008; Filipovic, 2009; Hawkins & Buttery, 2010). The CLC currently comprises approximately 50 million words of written learner data, roughly half of which is coded for errors. It has also been parsed using the Robust Accurate Statistical Parser (RASP) (Briscoe, Carroll & Watson, 2006). Salamoura and Saville (2009) state that the CLC mainly covers A2 level and above, which is the reason why the EP researchers started to build

a new corpus called the Cambridge English Profile Corpus (CEPC), mainly focusing on lower-proficiency level students' writing and speech.

Considering the sheer size of the CLC with error annotations and the CEFR as a framework, this EP programme seems to create a new research paradigm in learner corpus research. Those who are interested in using learner corpora in SLA research can relate their findings to the EP researchers' findings in terms of criterial features. Those who are involved in syllabus/materials design will find the RLDs for English very informative once those items are actually identified. Test developers will make full use of the results of the EP research for improving their test design and contents.

Some may argue that this whole approach is affected by the 'comparative fallacy' (Bley-Vroman, 1983). Bley-Vroman warned that L2 speakers' interlanguage systems should be seen as independent of their L1s and target languages and should thus be studied in their own right. This implies discarding the notion of 'target-like' performance. Most learner-corpus-based IL studies rely on the comparison between L2 learners and their mother tongues or target-like performance by native speakers of the target languages. In my opinion, this again depends on research purposes. If one wishes to describe interim states of IL systems, independent of both L1s and target languages, Bley-Vroman's position makes perfect sense. However, as Kasper (1997) said, SLA researchers have legitimate and important interests in assessing learners' IL knowledge and actions not just as achievements in their own right but also measured against some kind of standard (ibid: 310). From pedagogical and assessment viewpoints, there is nothing wrong with setting native speakers' well-formed sentences as a goal, because that is the language taught in the classroom. Therefore, L2 profiling research is worth the effort, as long as we properly understand its aims.

One of the issues of identifying criterial features is deciding how to extract errors from learner data and judge whether they serve as criterial features or not. The CLC is manually tagged for errors, but it would be quite difficult to extract learner errors from generic learner data without error annotations. There are two main purposes of this paper; to propose a new approach of annotating errors semi-automatically by comparing the original learner data against the proofread data, by using edit distance and automatic POS tagging, and to judge whether or not those errors can serve as criterial features by employing multivariate statistics called correspondence analysis and variability-based neighbour clustering. This is especially useful because it provides a set of criterial features for lower levels that are not provided by CLC, in order to identify a set of features for Japanese learners of English in specific L2 contexts, to suggest an alternative classification of features for all CEFR levels, and to offer a generic technique of extracting criterial features from any learner corpora.



## 2. Method

### 2.1. The JEFLL Corpus and its parallel version

The JEFLL Corpus is a corpus of 10,038 Japanese students' written compositions in English, totalling 669,281 running words (available online at <http://scn02.corpora.jp/~jefll04dev/>). The subjects were sampled across six school years (from Year 7 to 12 in terms of the U.S. school system). In Japan, English is generally introduced in Year 7 for the first time, so JEFLL consists of samples from beginning to lower-intermediate levels. The students were asked to write a short in-class essay in English in 20 minutes without the help of a dictionary. Essay topics were also controlled; there were six different topics in total (3 argumentative and 3 narrative/descriptive). The corpus can be queried on the basis of learner profile information such as school year, school type, and school level, as well as task variables (e.g. topics).

Using the JEFLL Corpus, my research team conducted a series of studies for identifying features characterising different stages of acquisition. Table 2 summarises the results.

**Table 2.** Previous studies using the JEFLL Corpus

Language features	References	Main findings
Morpheme orders	Tono (1998)	<ul style="list-style-type: none"> <li>• Article errors are persistent and the development of accurate article use is much slower than reported in previous research.</li> <li>• Possessive <i>-s</i> is easier than the universal order proposed in previous research.</li> </ul>
N-gram <sup>1</sup> analysis	Tono (2000, 2009)	<ul style="list-style-type: none"> <li>• The early stages are characterized by trigrams associated with V.</li> </ul>
Verb subcategorization	Tono (2004)	<ul style="list-style-type: none"> <li>• Subcategorization errors are influenced by inherent verb semantics and are not affected so much by input from the textbooks.</li> <li>• Overuse/underuse phenomena are related to textbook input.</li> </ul>
Verb & noun errors	Abe (2003, 2004, 2005) Abe & Tono (2005)	<ul style="list-style-type: none"> <li>• Verb errors are more frequent at lower proficiency levels.</li> <li>• Noun errors occur more frequently at higher levels.</li> </ul>
NP complexity	Kaneko (2004, 2006); Miura (2008)	<ul style="list-style-type: none"> <li>• Internal structures of NP are closely related to developmental stages.</li> <li>• Clause modifiers (relative clauses and <i>that</i>-clauses) are associated with the most advanced level.</li> </ul>

1 N-gram is a contiguous sequence of n items from a given sequence of text. In corpus linguistics, items in question can be words, parts-of-speech, or combinations of those. An n-gram of size 3 is called 'trigram.'

One of the methodological problems is the difficulty in error annotations. Some studies (Tono, 1998; 2004; Abe, 2003; 2004) examined errors in the JEFLL Corpus, but only smaller sets of texts, approximately 10,000 words for each subset, were used for manual error tagging. It is very time-consuming to tag the entire corpus for all types of errors, so we focused on certain grammatical errors only and performed so-called ‘problem-oriented’ tagging for errors. Currently, there are not very many fully error-tagged corpora available. The Cambridge Learner Corpus may be the only exception but again the corpus sampling tends to be skewed toward intermediate to advanced learners of English and unfortunately it is for in-house use only.

Instead of manually annotating every error in the files, a proofread version of the JEFLL Corpus was prepared. For this, one educated adult native speaker, who worked as an English instructor at a university in Tokyo, was hired to read through and correct errors in all the essays in the JEFLL Corpus. A single person did the job, because previous experiences show that annotation by a single person was more consistent than several people working together, although sufficient training was needed. A one-month training session was conducted, in which the proofreader was asked to correct several essays at different levels. The proofreader then discussed with the researcher the way errors were identified and corrected. Only local sentence-level lexico-grammatical errors were corrected. No corrections were made beyond sentence levels, such as coherence, connectivity, or the use of discourse markers across sentence or paragraph levels, for these error corrections usually involve a change in converting sentence orders or putting two sentences into one or vice versa. The sentence alignments in the essays were maintained strictly. One of the difficulties of proofreading the data in the JEFLL Corpus is that the compositions contain Japanese words or phrases. In the composition tasks, the use of Japanese was allowed especially for learners at the very beginning-level. Therefore, a proofreader competent in Japanese was chosen in order to produce corrected versions of the corpus.

## 2.2. *Edit distance*

A metric called an edit distance was employed. The edit distance between two strings of characters is the number of operations required to transform one of them into the other. There are several different ways to define an edit distance (for instance, Hamming distance, longest common subsequence, Levenshtein distance). Usually, an edit distance produces the actual number (e.g. the dis-

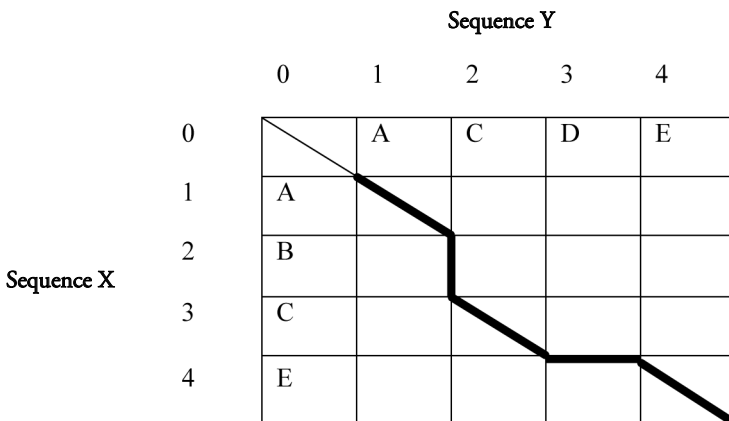
---

2 Differences between the two words are positions No. 2, 3, 5 and 7 in the letter sequence of “sitting”. Thus the distance is 4.

tance is 4, between “seaten” and “sitting”<sup>2</sup>), showing the amount of difference between the two sequences, but in the present study, I used this heuristic for identifying the same and different parts in the aligned sentences. My colleague, Hajime Mochizuki, helped to implement the program into the programming language Ruby, and the algorithm he used was basically the same as the so-called Levenshtein distance (Levenshtein, 1966). A commonly-used bottom-up dynamic programming algorithm for computing the Levenshtein distance involves the use of an  $(n + 1) \times (m + 1)$  matrix, where  $n$  and  $m$  are the lengths of the two strings. Figure 1 illustrates the matrix. The two sequences can be aligned in three possible ways, as (1) shows.

- (1) a. Two elements are identified as the same and aligned to each other (“\” path in the matrix)
- b. X is aligned to a gap (“|” path)
- c. Y is aligned to a gap (“-” path)

Suppose X has a sequence “ABCE” and Y has “ACDE,” the thick black line in Figure 1 indicates the optimal path for alignments. There is possibly more than one path from the starting point (0,0) to the end point (4,4). A Dynamic Programming (DP) algorithm checks all available paths from the start to the end and calculates each cost to identify the optimal path.



**Figure 1.** Dynamic Programming matrix

In our case, two aligned sequences correspond to two sentences, and the parts in the sequences (A to E in Figure 1) are actual words in the sentences. Figure 2 shows in matrix form how this algorithm checks the two aligned sentences, an original sentence (vertical) and its corrected counterpart (horizontal).



best tagged alignment results with the highest total of individual scores as an optimal alignment. The three error types are identified automatically based on the alignment results, and then tagged for each error type: <msf> for misformation, <add> for addition, and <oms> for omission. Correction candidates are specified in the case of misformation tags, as in <msf crr= "correct answer">. The output of the program is shown in (6):

(6) I eat <add>a</add> bread and <msf crr=fried>flied</msf> <oms>eggs</oms> every morning.

If the alignments are accurate, chances are that surface strategy taxonomy errors can be extracted fairly accurately and automatically.

### 2.3. Procedure

Using the heuristics described in 2.2., the parallel (i.e. original and proofread) version of the entire JEFLL Corpus was processed for the Levenshtein distance and then automatically tagged for three types of surface strategy taxonomy error: omission, addition and misformation. The output of the program was checked manually, and problematical cases of word order errors were identified and corrected. In order to capture an overall tendency of extracted errors, all the tagged surface strategy taxonomy errors were processed for part-of-speech (POS) information, using an automatic POS tagger. This made it possible to analyse extracted errors in terms of their parts of speech. At this level, the error annotation in the corpora is only related to the surface strategy taxonomy errors and their POS information. I am fully aware of the limitations of dealing with errors using the surface taxonomy and POS only. It needs further analysis in terms of linguistic classification, e.g. agreement errors, tense errors, verb subcategorization errors, among others. Furthermore, a POS tagger developed for analysing native speakers' data may not be entirely suitable for interlanguage data. But I have the following justifications for my approach. First, the main purpose of this chapter is to propose a method of annotating errors semi-automatically in learner language and not to propose comprehensive criterial features from learner data. Using the approach described in this paper, researchers can work on their learner data and make further analysis of each error type they are interested in. Second, the overview of POS-related errors based on the surface strategy taxonomy still provides a very interesting summary regarding the state of ILs at each stage and helps to generate new hypotheses related to different aspects of acquisition. For instance, omission errors of determiners are quite frequent across all the stages of acquisition in the JEFLL Corpus, while the repertoire of nouns in lexicon will also increase as the level increases. This means that the use of articles improves for particular noun groups, but the knowledge

of the article system is not fully acquired as more lexical items are introduced in the lexicon. This kind of microscopic analysis can be done for each error type, but this should be dealt with elsewhere. Third, automatic annotation described in this paper can be used to annotate large samples of learner corpora, which is cost-effective, and helps to conduct profiling research such as EPP to provide a bird's eye view of how learner performance will change from one stage to another.

The frequency distributions of the above error types in terms of POSs were obtained across the school years. Multivariate statistics were used in order to capture complex relationships between school years and different error types. Correspondence analysis was used first to obtain biplots between major error types and school years, which was supplemented by clustering techniques called "variability-based neighbour clustering (VNC)" (Gries & Stoll, 2008). Both are techniques of data reduction and summarisation. Correspondence analysis is a descriptive/exploratory technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns. The results provide information which is similar in nature to that produced by Factor Analysis techniques, and they allow one to explore the structure of categorical variables included in the table. Graphical representations of two variables mapped onto the two extracted dimensions are especially useful in order to see relative proximity of the items in each variable. VNC differs from standard approaches because it only clusters neighbouring data points, thus preserving the data points' temporal sequence. This is important because the order of school years needs to be taken into account as we cluster linguistic features characterising each level.

### 3. Results

#### 3.1. *The performance of edit distance*

The results of the Levenshtein distance show that this technique seems to work well. The precision and recall<sup>3</sup> rates for omission errors were 98.25% and 100% respectively (F measure is 0.9911 at  $\alpha = 0.5$ ). For the addition errors, the precision rate was 96.83% and the recall was 100% (F=0.9839). Only misformation errors were found to be less accurate. The number of incorrectly analysed items

---

3 Precision is defined as a measure of the proportion of selected items that the system got right:  $\text{precision} = (\text{true positive}) / ((\text{true positive}) + (\text{false positive}))$ . Recall is defined as the proportion of the target items that the system selected:  $\text{recall} = (\text{true positive}) / ((\text{true positive}) + (\text{false negative}))$  (Manning & Schütze 1999: 268).

was 179 out of 641 (precision = 72.07%), which shows that alignment of misformation was very difficult in comparison to the other two error types. Consequently, F measure was also low ( $F = 0.8373$ ). The sample output is shown in (7), where no error was found in the analysed sentence:

```
(7) <result>
<sentence id= "ns">
Today I ate bread and milk
</sentence>
<sentence id= "st">
Today I ate bread and milk
</sentence>
<trial no= "01a">
Today I ate bread and milk
</trial>
</result>
```

The first sentence labelled “ns” is the one proofread by a native speaker. The second sentence labelled “st” is the student’s original sentence and the third one is the output of comparing the pair (“ns” and “st”). If there is no error in the sentence, the output is the same as the two sentences above.

The sentences in (8) show the case in which the sentence pair (“ns” and “st”) has several differences. In the first output labelled “trial No. 01a”, differences between the pair were identified in terms of omission, addition and misformation (tagged <oms>, <add>, and <msf> respectively) along with suggested corrections shown in the attribute “crr=”. The edit distance program works in such a way that the first trial was retained as long as there was no overlapping word found in the identified error items. If there was any overlapping word, for example, “breakfast” in the output “01a”, additional analysis was made to re-classify the two overlapped words into a single case of transposition from one position to another in a sentence. Thus, in the output “02”, the word “breakfast” is tagged as <trs\_add> for the first one and <trs\_oms> for the second one, showing that these two words both belong to the same misordering error.

```
(8) <result>
<sentence id= "ns">
I like breakfast but I don't eat rice and miso soup for breakfast
</sentence>
<sentence id= "st">
I like breakfast but I don't eat in breakfast rise and misosoup
</sentence>
```

```

<trial no= "01a">
I like breakfast but I don't eat <add>in</add> <add>breakfast</add> <msf crr=
"rice">rise</msf> and <oms>miso</oms> <msf crr= "soup">misosoup</msf>
<oms>for</oms> <oms>breakfast</oms>
</trial>
<trial no= "02">
I like breakfast but I don't eat <add>in</add> <trs_add crr=
"breakfast">breakfast</trs_add> <msf crr= "rice">rise</msf> and <oms>miso</oms>
<msf crr= "soup">misosoup</msf> <oms>for</oms> <trs_oms crr=
"breakfast">breakfast</trs_oms>
</trial>

```

This technique of dealing with transpositions is quite similar to Damerau-Levenshtein distance, but the algorithm used here is a partial implementation of the formula, developed by Hajime Mochizuki (Tono & Mochizuki, 2009).

### 3.2. Distributions of surface strategy taxonomy errors

Figure 3 shows overall distributions of four types of surface strategy taxonomy errors (addition, omission, misformation and misordering). In terms of the number of error tags, misformation errors were found to be most frequent ( $n = 67,176$ ), followed by omission errors ( $n = 49,077$ )<sup>4</sup>, addition errors ( $n = 16,156$ ) and misordering errors<sup>5</sup> ( $n = 2,082$ ). Table 3 shows the breakdown of four types of errors across school years and parts of speech. This time, the frequencies are normalised per 10,000 words for comparison across different subcorpora. Overall, noun and verb errors are very frequent, followed by determiner errors. This has to be interpreted with caution because the total number of occurrences of nouns and verbs is usually greater than the other parts of speech. In this study, normalization was done for corpus size, but not for POS categories, so it is difficult to say exactly the error frequencies for nouns and verbs are greater than those of the other parts of speech. A relative measure will be needed in the future study to tease these possibilities out. Interestingly, the number of noun misformation errors ( $n = 594.8$ ) in Year 7 decreased dramatically through Year 7 to 9, and stayed the same across Year 10-12. One of the reasons is that Year 7 students overused Japanese words in the essays, which happened to be tagged as nouns since a POS tagger did not recognise Japanese words. There are also

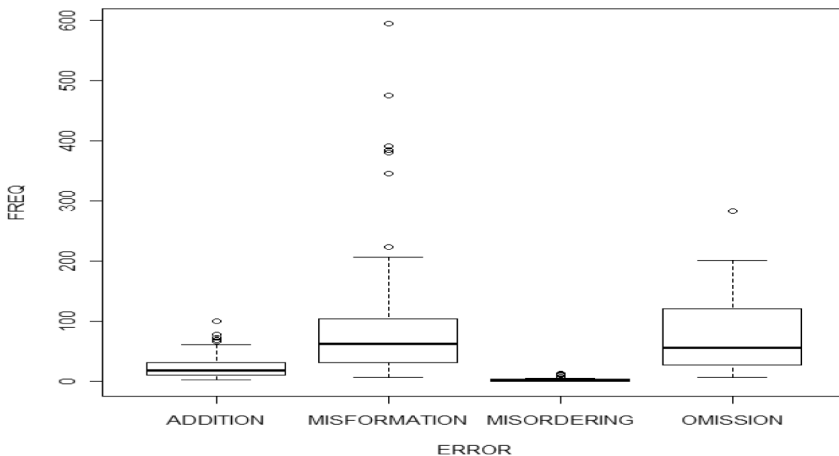
4 Please note, however, that this figure is based on the automatic extraction, whose precision is roughly 72%.

5 The number of misordering errors has to be interpreted carefully because this feature was added after the first evaluation was done for the other three types of errors and the accuracy rate was not checked against manually corrected data.



many misformation and omission errors on verbs. However, verbs behave differently from nouns in several respects. First, the number of verb misformation errors stays almost the same throughout the school years while noun misformation errors decrease in the first three years. This may be again related to the use of Japanese words in the compositions. Second, verb omissions are very high in year 7, they decrease considerably in Year 8 and after another slight decrease in Year 9 they tend to remain constant; noun omission errors seem to follow a U-shaped curve, with a high initial proportion gradually shrinking in Years 8 and 9, to then grow again in later years. Verbs are also different from nouns in the way addition errors occur. While the number of noun addition errors decreases constantly from Year 7 to 10, verb addition errors increase from Year 7 to 10. This is mainly due to the increasing overuse of “have” as an auxiliary besides its use as a lexical verb, as learners experiment with more complex grammatical constructions.

**Figure 3.** Distributions of surface strategy taxonomy errors



Determiner errors are especially frequent in the case of omissions. The frequencies of omission errors are five to six times higher than addition errors, which shows that Japanese-speaking learners of English tend to omit determiners rather than oversupply them. Error rates remain almost the same throughout the school years, which shows that determiner omission errors are quite persistent in nature. Prepositions are also problematical and they are frequently omitted. Interestingly, preposition omission errors have a typically U-shaped error curve, where the errors decrease for the first three years and then increase again in a later stage. Although the number is relatively smaller, addition errors of prepositions also increase steadily as the school year increases. Preposition errors



will become more frequent as learners learn more prepositions and try to use them to express more complex ideas in English.

It is noteworthy that errors observed with a frequency analysis based on the surface strategy taxonomy have some general characteristics, which may point to some general interlanguage developmental trends. First, omission errors are more common than additions. Naturally, L2 learners start with simplified structures, which lack required elements such as determiners, prepositions, verbs, and nouns to form well-formed sentences. As their proficiency levels go up, however, the ratio of addition errors to omission errors will become higher. This indicates that the more proficient L2 learners become, the more varieties of language they will use and they will thus take increasingly more risks in expressing themselves, which will lead to more errors. This is clearly shown in the increasing frequencies of errors related to verbs, adverbs, adjectives, prepositions, conjunctions and modals (see Table 3). This tendency is closely related to lexical choice errors with major content words and is known to have an inverted U-shaped curve (Hawkins & Buttery, 2010), which indicates that errors of this type will continue to increase as learners become proficient from the beginning to the intermediate levels and as the repertoire of language becomes wider and errors will decrease or disappear when they reach near-native proficiency levels. In JEFLL, because of the lower proficiency levels, most addition errors continue to grow in number or stay the same throughout the six years.

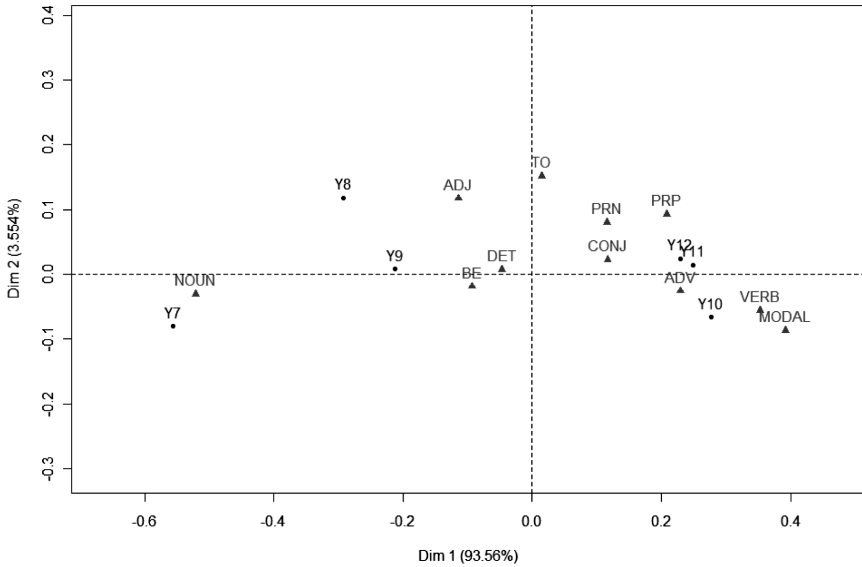
The statistics, however, have to be interpreted carefully in the case of misformation errors, given that the identification of misformation errors by edit distance has lower precision/recall scores in comparison to the other error types. There is also an influence of the use of Japanese words in the essays, which boosted the frequencies of noun errors, especially in Year 7.

### *3.3 Correspondence Analysis*

There are many ways to approach multifactorial data. The primary purpose of this study is to identify criterial features that distinguish one proficiency level from another. What is meant by criterial features here is a set of surface strategy errors classified according to parts of speech. Therefore, what needs to be done is to extract error categories that are salient enough to serve as criteria for distinguishing learners' proficiency levels. Hawkins and Buttery (2010) examined error frequencies across different CEFR levels by setting thresholds of error ratio to determine the significance of errors as criteria. Since the JEFLL Corpus was not categorised for CEFR levels, a different approach had to be taken. The simplest way to analyze contingency tables like Table 3 is the Chi-square test, but unfortunately, the Chi-square test does not provide a solution to the prob-

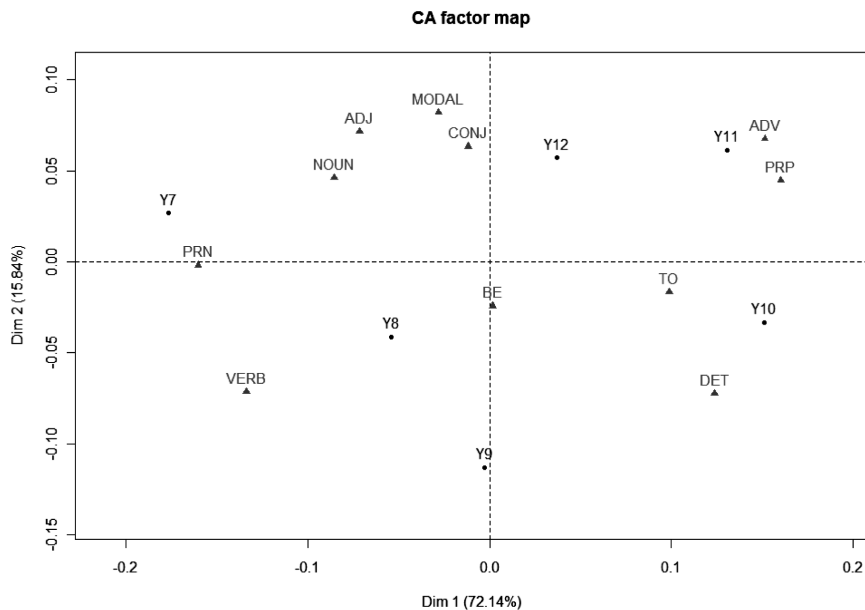
lem of identifying detailed relationships among column and row variables. Though it tests whether two variables are independent of each other, it does not allow us to characterize the school years in terms of the distribution of POS errors. Answers to the question are provided by correspondence analysis. Correspondence analysis is a statistical visualization method for picturing the associations between the levels of a two-way contingency table. In this case, the two variables were school years (row variables) and POS errors (column variables). This technique plots together in a bi-dimensional space groups of texts (Years 7-12) and features, thus representing graphically which features are more significant in identifying each group. Dimension scores were first calculated independently for the two variables, thus the distance between column or row variables is meaningful in independent row or column plots, which are not listed here. On the biplots like Figures 4 onwards, only the dimensions between row and column points are meaningful, because the elements for the two variables were plotted at the same time on the bi-dimensional space using a technique called symmetrical normalization. The simplest way to interpret the biplots is to draw a line on the plot through the origin (0,0) and the point corresponding to the POS error in question (NOUN, for instance). Perpendiculars to this line are dropped from each school year's position on the plot. Look at how close each POS error is on this line to the point, NOUN. One can see Y7 is the closest, Y8 and Y9 follow, and the other three (Y10, 11, and 12) are furthest. The relative positions between the school years and the POS errors show that NOUN is the most closely associated with Year 7 and VERB, MODAL, PRP, ADV tend to be related to more advanced levels (Years 10-12). DET, on the other hand, is positioned almost in the center (0,0), which means that DET is relatively the same in frequency across school years. An analysis was made independently for each of the four error types, due to the complexity of multiple correspondence analysis. Figure 4 shows the results of correspondence analysis for addition errors.

The horizontal axis (Dimension 1) explains 93.56% of the overall Chi-square value (or inertia), which means that we can interpret the results almost exclusively with regard to their positions on the first axis. Regarding the positions of the school year, Year 7 was placed on the leftmost edge, Year 8 and Year 9 were close together on the left side, much closer to the origin for the first axis, while Year 10, Year 11, and Year 12 appeared very closely together on the right side of the origin for the first axis. Therefore, it is fair to conclude that the first axis separates essays written by junior high school students from those by senior high school students, which means the first axis basically shows the differences in proficiency levels. Interestingly, all three groups in senior high school (Year 10-12) were very close in position, which indicates that as far as addition errors are concerned, the three groups were very similar. The same thing can be

**Figure 4.** Correspondence analysis (addition errors)

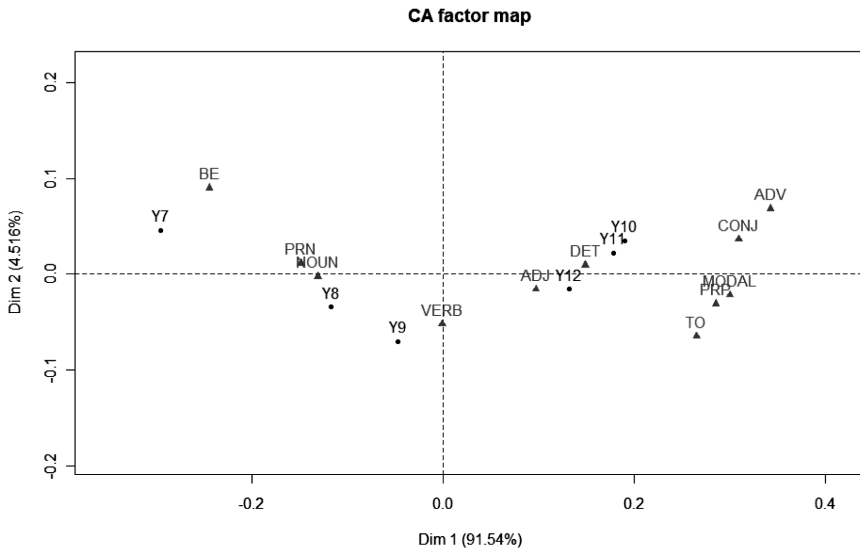
said about Year 8 and Year 9. Year 7 was apart from the other groups, showing that the group behaved very differently. The positions of POS errors in relation to the school years revealed interesting patterns. Noun errors (NOUN), for example, were close together with Year 7, far from the other error groups. As can be seen from Table 3, noun errors were very high in frequency for Year 7, mainly due to the fact that Year 7 students used Japanese words very often in the compositions, which were analysed as nouns by a POS tagger. Thus, high frequencies of noun errors involve the use of Japanese words in the passages. Another reason why noun errors were located far from the other groups is that their frequencies kept going down significantly from Year 7 to 9 until they became stable for higher levels. On the other hand, verb errors (VERB) and modal auxiliary errors (MODAL) showed opposite tendencies, with their frequencies continuing to increase toward Year 12. Figure 5 shows the results of correspondence analysis for omission errors.

The overall picture here is different from addition errors. The relationship between the two variables (POS omission errors X school year) summarised in the biplots in Figure 5 can be interpreted by looking at Table 3 again. The students' groups were not plotted in the order of the school years. Rather, Year 12 was placed toward the centre, and Year 10 and Year 11 were on the rightmost end. This is partly due to the fact that error frequencies reported in Table 3 suddenly increased in Year 10 after a gradual decrease from Year 7 to 9. It seems that omission errors did not simply decrease as the school year went up. In

**Figure 5.** Correspondence analysis (omission errors)

many cases, omission errors decreased in frequency from Year 7 to 9, rose again in Year 10 and either stayed the same toward Year 12 or fluctuated through the three years in senior high, which explains why the points for these years do not follow a straight line from left to right in the biplot. Also there were two different groups of POS errors, divided by the origin of the axis. Those placed on the left side of the origin for the first axis (PRN, NOUN, VERB, and ADJ) all shared the same tendency that their frequencies in Year 7 were much higher, compared to the other errors (ADV, PRP, DET, and TO), whose frequencies were not very high in Year 7 and gradually became higher in Year 10 - 12. The former group consists of parts of speech that are primary components of constructions and open class in nature (except for PRN) whereas the latter group belongs to closed class and their primary functions are connecting components in a sentence. This shows that learners at the beginning stage of acquisition fail to supply major elements such as verbs or nouns, but these omission errors tend to decrease as they progress. On the other hand, they will have more errors on function words such as prepositions, determiners, infinitives, and adverbs, which help to modify principal elements in a sentence to make it more complex.

Figure 6 illustrates the way misformation errors occurred and their relationship with school years.

**Figure 6.** Correspondence analysis (misformation errors)

For misformation errors, Dimension 1 explains 91.5% of the inertia, thus this horizontal axis tells us most of the relationship between error types by POS and the school years. As is shown in Figure 6, the school years were basically plotted in the order of the progression of the grades, but again the senior high school groups (Year 10 to 12) appeared close together in almost the same area, which shows that error patterns in the upper-grade groups were quite similar. A striking difference was found in two groups of POS errors. By examining frequencies in Table 3 to interpret the plot, the group plotted on the left side of the origin for the first axis (BE, PRN, NOUN) all had the tendency to be very high in frequencies in Year 7, gradually decrease to Year 9, and then stay at the lower level throughout Year 10 to 12. On the other hand, the group plotted on the right side of the origin for the first axis (ADV, CONJ, MODAL, PRP, TO) all showed the similar tendency that the error frequencies increased constantly toward Year 12. The other POS errors (VERB, ADJ, DET) showed almost the same error frequencies throughout the six years. Misformation errors showed a tendency similar to addition errors in the sense that the growth of learners' vocabulary and their repertoire, as they move from the beginning to the lower-intermediate stages of learning, will lead to taking more risks to use newly learned items, thus resulting in more errors. This also has something to do with

the syntactic elaboration of sentences, which is shown in the errors of closed system such as CONJ, MODAL, PRP and TO.

### *3.4. Refining the analysis by using neighbour clustering*

Even though correspondence analysis shows a graphical image of the relationship between the variables in terms of distances, it does not give us any information about how items in the variables can be clustered meaningfully. Cluster analysis is usually a common technique for classification tasks, but it has a serious problem in the sense that standard cluster analysis cannot take into account the 'time factor'. The present data is pseudo-longitudinal in nature, and it is desirable to find meaningful clusters based on error frequencies, but at the same time sensitive to the order of data points along the time sequence.

Gries & Stoll (2009) dealt with these 'variability problems' of children's mean MLUs over time as 'developmental problems'. He rightly commented that "one cannot simply lump together all utterances with a particular MLU value because this procedure would be completely blind to the order of elements and the developmental implications this may have" (ibid: 222). This problem is similar to mine, and his solution was to employ 'variability-based neighbour clustering (VNC)'. VNC is a hierarchical cluster-analytic approach, which takes into account the temporal ordering of the data (Hilpert & Gries, 2009, p. 390). What VNC basically does is to access the first and the second time period (Year 7 and Year 8, for instance) and compute the similarity measures of their respective two values (using e.g. variation coefficients or summed standard deviations, depending on the nature of the data), then proceed to do the same for all successive pairs of values, the second and the third, the third and the fourth, etc. always storing the similarity measures. After that, VNC identifies the largest similarity score, which indicates the values that are most similar to each other and thus merit being merged into one group. After the first iteration, there are only five data points, the first two groups having been merged. This process will be repeated until only one data point is left.

Figure 7 shows the result of VNC for noun addition errors. The left panel of Fig. 7 plots the distance in summed SD as an analogue to scree plots in principal component analysis, where they are used as a guideline to determine how many factors should be included in a model. The plot indicates how many different stages could be identified within a developmental progression, as in our case, the series of school years. The plot shows substantial distances between the first three largest clusters, i.e. steep slopes between the first three points. After the third cluster, the curve levels off to the right and becomes nearly horizontal. This suggests a division into three separate developmental stages, each represented by a cluster. The dendrogram (right panel) illustrates what these clusters are.



Dendrograms are best read from the bottom, since they join together groups starting from those having the lowest distance. The distance is represented not in the horizontal but in the vertical axis, which means that a short vertical line represents closely associated points while a long one represents a greater distance between them. Cluster 1 distinguishes Year 7 from the rest. Cluster 2 ranges from Year 8 and Year 9, and cluster 3 ranges from Year 10 to Year 12.

**Figure 7.** VNC for noun addition errors (LEFT: scree plots; RIGHT: dendrogram)

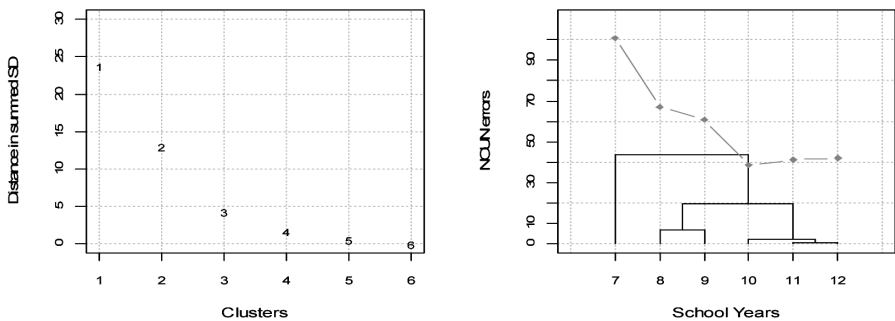
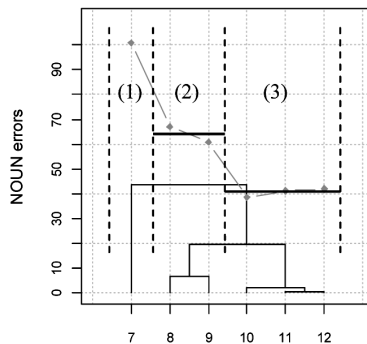


Figure 8 shows the three clusters by dividing them by vertical dotted lines. Horizontal lines under the numbers (2) and (3) indicate the mean frequencies that are observed in the data for the three clusters.

**Figure 8.** Three clusters in the dendrogram of noun addition errors



Dendrograms of VNC for addition and omission errors sub-classified by POS are reported in a separate file which can be accessed online at the URL [http://eurosla.org/monographs/EM02/tono\\_fig9-10.pdf](http://eurosla.org/monographs/EM02/tono_fig9-10.pdf). Misformation and misordering errors were not examined because of lower precision/recall scores.

The analysis revealed that some POS errors could not produce meaningful clusters. When the scree plots did not show any steep slope between the points,

the results were not very useful even though the dendrograms in Figures 9 and 10 made two clusters anyway, just for the sake of giving an idea of where the division could be made. Regarding the addition errors in Figure 7, only nouns, adverbs, verbs, modals and prepositions made two meaningful clusters. Except for noun addition errors, which produced three clusters due to the effects of the intensive use of Japanese in Year 7, the first cluster ranges from Year 7 to Year 9, and the second ranges from Year 10 to Year 12, thus clearly dividing the junior high group and the senior high group in terms of the error occurrence patterns. This confirms the findings observed in correspondence analysis in Figure 4, and without VNC it was difficult to state which POS errors actually contributed to the divisions.

The omission errors show slightly more complicated pictures. As was shown in Figure 5, there is a tendency for omission errors to decrease throughout Year 7 and Year 9, and increase again in Year 10 toward Year 12, which is due to the fact that learners took more risks to extend their repertoire of English at later stages, yielding more errors. Learners tended to master the use of basic lexis and grammar that they had learned at the early stage, but as they moved onto more advanced stages, they produced different types of omission errors. In terms of accuracy rates, this is a well-known inverted U-shaped developmental curve. Among the omission errors, only nouns, pronouns, and verbs seemed to show meaningful clusters. Interestingly, the two clusters are Year 7 and the rest in most cases. It is worth pointing out again in this connection the results of correspondence analysis. Those errors placed on the left side of the origin for the first axis (PRN, NOUN, VERB, and ADJ) in Figure 5 nearly correspond to the ones showing meaningful clusters in Figure 8, namely nouns, verbs, and pronouns. One should bear in mind that their frequencies in Year 7 were much higher, compared to the other errors (ADV, PRP, DET, and TO), whose frequencies were not very high in Year 7 and gradually became higher in Year 10 - 12. Therefore, the results of VNC suggest that three omission errors above all (noun, verb and pronoun) are useful in distinguishing Year 7 from the rest of the groups, while for the other POS errors the results are not conclusive.

#### **4. Discussion**

So far, I have proposed a new way of extracting errors from learner corpora and judging the status of those extracted errors as criterial features. Edit distance is a common metric to spot differences between two strings of characters. It is used intensively in other areas such as the analysis of DNA sequences. By extending its use to a comparison of learner production and target-like per-

formance, it is possible to identify surface strategy errors semi-automatically over a large amount of learner data. The present study also shows that data reduction techniques such as correspondence analysis are useful in summarising the data. However, correspondence analysis plots do not show exactly what meaningful clusters are. In order to solve this problem, a special clustering technique called variability-based neighbour clustering was introduced. The results of the combination of these two techniques revealed the contribution of addition/omission errors for particular POSs as criterial features of the developmental stages.

Table 4 summarises the results in terms of extracted criterial features to characterise Japanese EFL learners' acquisition stages.

**Table 4.** Extracted criterial features for the learning stages of Japanese EFL learners

Types	POS	Criterial for:	mean error freq. of errors
Addition	nouns	[Year 7] > [Year 8 - 9] > [Year 10 -12]	58.4
	adverbs	[Year 10 - 12] > [Year 7 - 9]	24.93
	verbs	[Year 10 - 12] > [Year 7 - 9]	48.81
	prepositions	[Year 10 - 12] > [Year 7 - 9]	23.62
	modals	[Year 10 - 12] > [Year 7 - 9]	11.65
Omission	nouns	[Year 7] > [Year 8] = [Year 10 -12] > [Year 9]	177.98
	verbs	[Year 7] > [Year 8 - 12]	120.62
	pronouns	[Year 7] > [Year 8 - 12]	111.73

Note: '>' means "occur more frequently than ...";

As shown in the column of mean error frequencies, the relative frequencies of omission errors are much higher than those of addition errors. However, a closer look into the categories of omission errors by POS reveals that omission errors are only useful for distinguishing the very beginning stage of learning from the rest, as shown in the third columns in Table 4. Overall, omission errors tend to decrease toward Year 9 and then jump up again in upper grades. Since the primary purpose of this paper is to present a heuristic to identify criterial features, I will not develop this point any further. More research into omission errors at a lexical level will be needed in order to describe in more detail what is happening in this U-shaped phenomenon.

Addition errors are more sensitive to level differences and thus work as criterial features distinguishing the lower level from the upper. It is noteworthy that in all cases but noun errors, addition errors are more frequent in the upper levels (Year 10-12). Adverbs, prepositions or modals are the elements that modify main constituents of a sentence. For instance, adverbs modify either verbs,

adjectives or other adverbial phrases. Prepositions usually modify nouns or verbs. Modals modify verbs to add epistemic or deontic meanings. As proficiency levels increase, learners have a wider repertoire of these lexical items and feel more confident in using basic lexis and grammar, which leads to a greater chance that they take risks to use new items to convey subtler meanings. Sometimes they fail to make the right word choices, and thus have more lexical choice errors, but in other cases they overuse and add unnecessary words to sentences, yielding non-target-like outcomes.

There are a few methodological issues related to this approach. One is the issue of “normalisation”. In this study, a parallel set of the original students’ essays and their proofread versions were used for edit distance. In order to produce parallel corpora, one native speaker instructor, who was trained for error corrections, worked on all of the 10,000 essays. It is a well-known fact (cf. Milton & Chowdhury, 1994) that a certain error in a sentence can be corrected (i.e. normalised) in more than one way. I am aware of such multiple interpretations of L2 learner errors and that there is also a system of multi-layered annotations, such as MMAX2 (Müller & Strube, 2006), so that one can annotate possible choices of normalisation in more than one way. In this study, however, I did not take that approach for two main reasons. First, native speakers’ correction possibilities could be almost infinite if we allow for multiple possibilities of normalization. If the native speaker wanted to extend their correction to stylistic or discourse elements, a number of different ways of correcting and refinement could be possible, and it would thus be almost impossible to incorporate those into the analysis, although the variation in native speakers’ judgments could be a valuable research object in its own right. The second reason is that even though there were some minor inconsistencies in normalisation patterns, corrections in more than 10,000 essays should cause some patterns of use/misuse to emerge, which help to explain the patterns of development over different school years. There is no error annotation system that can be said to be superior to others in and of itself. Error annotation adequacy is always relative to the research goals.

It would be pedagogically very significant to identify criterial features from learner corpora. If those performance features can work as ‘classifiers’ in the sense of text mining, it is possible to produce an automatic performance analysis system, in which the input by an L2 learner will undergo text analysis and his or her proficiency level will be determined by checking the existence of criterial features. In language testing, with such criterial features available, the assessment procedure of speech or writing can be facilitated by first automatically assessing the text based upon known criterial features and then by human intervention only on those aspects that need human judgements. What we need

is a formal procedure for extracting and identifying criterial features. This paper proposes a formal, methodological procedure for identifying criterial features in IL development. Using edit distance, possible error candidates are automatically extracted. Subcategorising those errors by POS can be done by automatic POS tagging. Variability-based neighbour clustering will make it possible to aggregate similar groups and cluster variables into meaningful stages of learning. This procedure can be applied to any kinds of learner corpora if they have parallel versions of the data set ready for edit distance. A word of caution is in order here. The approach presented in this paper is only applied to extracting surface strategy taxonomy errors. It will not deal with semantic errors such as tense/aspect morphology, for this kind of information is not revealed on the surface. Also this method is only applicable to “errors” as criterial features. It will not be used to extract well-formed language features as criteria. This should not be the limitation of this study, however, because well-formed linguistic features are usually much easier to extract, using ordinary corpus analysis tools such as concordancing or n-gram analysis over different sets of learner data. I hasten to add that VNC can also be used for analysing both errors and non-errors as long as frequency information is available regarding given linguistic features across different stages.

Some final notes are in order with respect to methodological issues. The detection of misformation errors could be improved. At the moment, the accuracy of misformation errors is sufficiently high with respect to one-to-one lexical mapping relation. If the mapping is between one to multiple words or vice versa, the accuracy rate suddenly drops. In order to solve this problem, ontological knowledge such as POS-labelled wordlists or something of the kind will be needed, which is more complex than simple surface character-level similarities. The results of multivariate analysis should also be further interpreted from both macroscopic and microscopic viewpoints. In macro views, my findings should be related to a much larger framework of criterial features and CEFR levels. If several dozen criterial features were identified, it would be necessary to re-classify those criterial features in terms of their relative importance. Also there are some cases in which a bundle of criterial features will work better than a single feature, thus some methods have to be proposed in order to figure out how to deal with such possibilities. I should admit that identifying criterial features is one thing, but constructing the overall framework is quite another. This whole process of identifying criterial features using learner corpora and constructing the overall theoretical framework based on those criterial features seems to me a very promising research strand, which definitely links learner corpus research to SLA and English language teaching and assessment in a meaningful way.

## References

- Abe, M. (2003). A corpus-based contrastive analysis of spoken and written learner corpora: the case of Japanese-speaking learners of English. In D. Archer, P. Rayson, A. Wilson, & T. McEnery (Eds.), *Proceedings of the Corpus Linguistics 2003 Conference (CL 2003)* (pp. 1-9). Lancaster University: University Centre for Computer Corpus Research on Language.
- Abe, M. (2004). A corpus-based analysis of interlanguage: errors and English proficiency level of Japanese learners of English. In Y. Tono (Ed.), *Handbook of An International Symposium on Learner Corpora in Asia* (pp. 28-32). Tokyo: Showa Women's University.
- Abe, M. (2005). A comparison of spoken and written learner corpora: analyzing developmental patterns of grammatical features in Japanese Learners of English. *The Proceedings of the NICT JLE Corpus Symposium* (pp. 72-75). Kyoto: National Institute of Communications Technology.
- Abe, M. & Tono, Y. (2005). Variations in L2 spoken and written English: investigating patterns of grammatical errors across proficiency levels. *Proceedings from the Corpus Linguistics Conference Series* ( Vol. 1, no.1) Retrieved from <http://www.corpus.bham.ac.uk/pclc/index.shtml>
- Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning*, 33, 1-17.
- Briscoe, E., Carroll, J., & Watson, R. (2006). *The second release of the RASP System*. Retrieved January 15, 2012, from <http://acl.ldc.upenn.edu/P/P06/P06-4020.pdf>
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language Two*. Oxford: Oxford University Press.
- Filipovic, L. (2009). *English Profile – Interim report*. Internal Cambridge ESOL report, April 2009.
- Goldberg, A. E. (1995). *Construction: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, A.E. (2006). *Constructions at Work: the nature of generalization in language*. Oxford: Oxford University Press.
- Granger, S. (Ed.). (1998). *Learner English on Computer*. London/New York: Addison Wesley Longman.
- Granger, S., Hung, J. & Petch-Tyson, S. (Eds.). (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins.
- Gries, S. Th. & Divjak, D. (2012). *Frequency Effects in Language Learning and Processing*. Berlin: Mouton de Gruyter.
- Gries, S. Th. & Stoll, S. (2009). Finding developmental groups in acquisition data: variability-based neighbor clustering. *Journal of Quantitative Linguistics* 16(3), 217-242.
- Hawkins, J. A. & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. *English Profile Journal*, 1(1), 1-23.

- Hendriks, H. (2008). Presenting the English Profile Programme: In search of criterial features. *Research Notes*, 33(3), 7-10.
- James, C. (1998). *Errors in Language Learning and Use: Exploring Error Analysis*. London: Longman.
- Kaneko, E. (2004). Development of noun phrases in the interlanguage of Japanese EFL learners. Poster session presented at the 6th Conference of the Japanese Society for Language Sciences (JLS 2004), Nagoya.
- Kaneko, E. (2006). Corpus-based research on the development of nominal modifiers in L2. Paper presented at the American Association of Applied Corpus Linguistics (AAACL), Flagstaff, Arizona.
- Kasper, G. (1997). "A" stands for acquisition: A response to Firth and Wagner. *Modern Language Journal*, 81(3), 307-312..
- Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707-710.
- Manning, C. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge MA: MIT Press.
- Milton, J.C.P. & Chowdhury, N. (1994). Tagging the interlanguage of Chinese learners of English. *Proceedings of the joint seminar on corpus linguistics and lexicology* (pp. 127-143). Hong Kong: Language Centre, HKUST.
- Miura, A. (2008). Kaiwa (NICT JLE) vs. Sakubun (JEFL) Corpus no hikaku to bunseki [A comparison of spoken and written corpora]. *English Corpus Studies*, 15, 135-148.
- Müller, C. & Strube, M. (2006). Multi-level annotation of linguistic data with MMAX2. In S. Braun, K.Kohn, & J. Mukherjee (Eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods* (pp. 197-214). Frankfurt: Peter Lang.
- Parodi, T. (2008). L2 morpho-syntax and learner strategies. Paper presented at the Cambridge Institute for Language Research Seminar, Cambridge, UK.
- Salamoura, A. & Saville, N. (2009). Criterial features of English across the CEFR levels: evidence from the English Profile Programme. *Research Notes*, 37, 34-40.
- Tono, Y. (1998). A computer learner corpus-based analysis of the acquisition order of English grammatical morphemes. In *TALC (Teaching and Language Corpora) '98 Proceedings* (pp. 183-187). Oxford: Seacourt Press.
- Tono, Y. (2000). A corpus-based analysis of interlanguage development: Analysing part-of-speech tag sequences of EFL learner corpora. In B. Lewandowska-Tomaszczyk, B., & J.P. Melia (Eds.), *PALC'99: Practical Applications in Language Corpora* (pp. 323-340). Frankfurt: Peter Lang.
- Tono, Y. (2004). Multiple comparisons of IL, L1 and TL corpora: the case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and Language Learners* (pp. 45-66). Amsterdam: Benjamins.
- Tono, Y. (2009). Variability and invariability in learner language: A corpus-based approach. In Y. Kawaguchi, M. Minegishi, & J. Durand (Eds.), *Corpus Analysis and Variation in Linguistics* (pp. 67-82). Amsterdam: Benjamins.

- Tono, Y. & Mochizuki, H. (2009). Toward automatic error identification in learner corpora: A DP matching approach. Paper presented at Corpus Linguistics 2009, Liverpool, UK.
- UCLES-RCEAL Funded Research Projects. Retrieved January 15, 2012, from [http://www.englishprofile.org/images/pdf/ucles\\_rceal\\_projects.pdf](http://www.englishprofile.org/images/pdf/ucles_rceal_projects.pdf).
- Williams, C. (2007). *A preliminary study into the verbal subcategorisation frame: Usage in the CLC*. Unpublished manuscript.



## *About the authors*

**Camilla Bardel** is professor of modern languages and language education at Stockholm University, Sweden. Her research mainly regards the learning of third languages, with special focus on cross-linguistic influences in vocabulary and syntax. She has published articles and co-edited works on the L2 and L3 learning of French, Italian, Swedish and other languages. She also has an interest in lexicographical issues.

**Tom Cobb** teaches and does research in applied linguistics at l'Université du Québec à Montréal. His main interests are lexical acquisition and computing in research and learning - and the crossovers between them. His website Lextutor ([www.lextutor.ca](http://www.lextutor.ca)) is devoted to making research tools and research-based instruction accessible to the applied linguistics community.

**Anna Gudmundson** has a PhD in Italian and does research in L2 and L3 acquisition at the department of language education at Stockholm University, Sweden. Her thesis concerns the acquisition of grammatical gender and number in Italian as a second language. She is currently engaged in research on lexical acquisition and cross-linguistic influences from previously acquired languages.

**Henrik Gyllstad** is a senior lecturer of English linguistics at Lund University, Sweden. His main research interests straddle the fields of second language acquisition and language testing. In particular, he is interested in bilingual lexical processing, L2 vocabulary learning and testing, and English phraseology. He is the co-editor of the international volume *Researching Collocations in Another Language – Multiple Interpretations* (2009), published by Palgrave Macmillan, and several articles in international journals.

**Birgit Henriksen** is associate professor (reader) at the University of Copenhagen, Denmark. Her main research interests are vocabulary acquisition and teaching and academic language use in English medium instruction. From 2008-2011 she was the director of the Centre for Internationalization and Parallel Language Use. She has been active in developing the field of foreign language acquisition in Denmark, e.g. through her teaching on a number of in-service courses for foreign language teachers at various levels of language education.

**Batia Laufer** is professor of applied linguistics at the University of Haifa, Israel. Her main contribution to the field of applied linguistics is her research on vocabulary acquisition in additional languages (vocabulary threshold for reading, the limitations of input-based learning, factors of word difficulty, quantitative assessment of vocabulary, task effect on learning, vocabulary attrition). Her additional research areas are lexicography, cross-linguistic influence, reading, and testing.

**Tami Levitzky-Aviad** is a PhD student at the department of English language and literature at the University of Haifa. She teaches courses in pedagogical grammar, introduction to linguistics, writing and English for academic purposes. Her research interests include foreign language acquisition (FLA), learner corpora and lexicography in the context of FLA.

**Christina Lindqvist** is a research fellow in romance languages, especially French linguistics, at the department of modern languages at Uppsala University. Her research interests include third language acquisition and cross-linguistic influence, vocabulary acquisition, particularly lexical richness and lexical profiling.

**James Milton** is professor of applied linguistics at Swansea University, UK. A long-term interest in measuring lexical breadth and establishing normative data for learning and progress has led to extensive publications including *Modelling and Assessing Vocabulary Knowledge* (CUP 2007 with Michael Daller and Jeanine Treffers-Daller) and *Measuring Second Language Vocabulary Acquisition* (Multilingual Matters 2009).

**Yukio Tono** is professor of corpus linguistics and English language teaching at Tokyo University of Foreign Studies. His main research interests include corpus-based second language acquisition, use of corpora in language teaching, L2 vocabulary acquisition and dictionary use. He co-authored volumes for Routledge and John Benjamins and he is also a member of the editorial board for the *International Journal of Lexicography* and *Corpora*.