

In search of a General Academic vocabulary: A corpus-driven study

Sylviane Granger & Magali Paquot
Université catholique de Louvain
Centre for English Corpus Linguistics

[DRAFT VERSION]

1. Introduction

Most studies of vocabulary in English for Academic Purposes (EAP) have emphasized the importance of a ‘sub-technical’ or ‘academic’ vocabulary alongside core words and technical terms in academic discourse (cf. Nation 2001: 187-216). A number of word lists have been compiled to meet the academic vocabulary needs of students in higher education settings (e.g. Xue and Nation 1984). The *Academic Word List* (AWL) (Coxhead 2000) is the most widely used today in language teaching, testing and materials development (e.g. Schmitt and Schmitt 2005). It consists of 570 word families which have wide range and reasonable frequency of occurrence in a 3,500,000 word corpus of academic texts (e.g. *approximate, capacity, link, presume, summary, widespread*, etc.). It is important to note that Coxhead’s list excludes the top 2,000 words of English included in West’s (1953) *General Service List* (GSL). Taken together, words in the GSL and the AWL plus discipline-specific items should approach the critical 95% coverage threshold necessary for reasonable reading comprehension (Nation 2001:197).

AWL word families consist of relatively formal vocabulary items that appear with a high frequency in a wide range of academic texts and are infrequent in non-academic texts (cf. Coxhead 2000). The term ‘academic vocabulary’, however, has also been used in a more restricted meaning to refer to words that “have in common a focus on research, analysis and evaluation – those activities which characterize academic work” (Martin 1976:92). Academic words “most probably occur because they allow academic writers to do the things that academic writers do. That is, they allow writers to refer to others’ work (*assume, establish, indicate, conclude, maintain*) and to work with data in academic ways (*analyse, assess, concept, definition, establish, categories, seek*)” (Nation 2001: 18).

Hyland and Tse (2007: 238), however, question the widely held assumption that “a single inventory can represent the vocabulary of academic discourse and so be valuable to all students irrespective of their field of study”. They show that the coverage of AWL items in a corpus of 3.3 million words from a range of academic disciplines is not evenly distributed. Of the 570 AWL families, 534 (94%) have irregular distributions across three field-specific sub-corpora (engineering, sciences and social sciences) with, in many cases, a majority of the occurrences located in just one sub-corpus. Hyland and Tse further argue that “all disciplines shape words for their own uses” (ibid: 240) as demonstrated by their clear preferences for particular meanings and collocations. In addition, an investigation of a set of potential homographs in the AWL reveals a considerable amount of semantic variation across disciplines. Science and engineering students, for example, are very unlikely to come across the noun *volume* in the meaning of “a book or journal series” unless they are reading book reviews. The noun *strategy* has different preferred associations across disciplines: it often appears in the multi-word unit *marketing strategy* in business, *learning strategy* in applied linguistics and *coping strategy* in sociology. The authors thus conclude that “By considering context, cotext, and use, *academic vocabulary* becomes a chimera” (ibid 250).

These findings pose a tremendous challenge to the growing number of students who enrol in interdisciplinary programmes and to ELT practitioners who are regularly faced with mixed groups of students, most notably in international EAP programmes. Two decades ago, in an article on second language teaching for academic achievement, Saville-Troike insisted that “vocabulary knowledge is the single most important area of second language competence when learning content through that language is the dependent variable” (1984:199). EAP courses need to ensure that sufficient attention is given to vocabulary development (cf. Sutarsyah et al 1994: 37). That being the case, if academic vocabulary is a chimera, one can wonder which words EAP tutors should teach to mixed groups of students. The main objective of our study is to investigate whether it is possible to identify a common core that can be taught in a general EAP context.

In this paper, we report the preliminary results of the investigation of a new ESP corpus, the *Varieties of English for Specific Purposes dAtabase* (VESPA). We focus on lexical verbs, which are characterized by a high rate of EAP-specific collocates and preferences in terms of aspect, tense and voice and have been shown to create a minefield of difficulty for learners (e.g. Hinkel 2002, Swales and Feak 2004, Nesselhauf 2005). The investigation is part of a wider study into EAP vocabulary (De Cock et al 2007; Gilquin et al 2007; Paquot 2007a, 2007b, 2008) and is a follow-up to a previous study that compares the use of lexical verbs in academic writing by native writers and intermediate/advanced EFL learners (Granger and Paquot in press).

In section 2 we describe the corpora used to investigate the use of EAP verbs by expert and learner writers. In section 3 we identify the lexical verbs that are prominent in three fields (business, linguistics and medicine) and zoom in on the meaning and preferred phraseological patterns of one particular verb, viz. *analyze*, with a view to identifying the similarities and differences in use it displays in the three fields. In section 4 we give the results of the analysis of the verb *analyze* in a learner corpus containing term papers in linguistics and focus on the role played by learner corpora in the debate between advocates of a general EAP approach and scholars like Hyland who promote specificity. Section 5 contains concluding remarks.

2. Data

The corpus used in this study is a corpus that is currently under development at the Centre for English Corpus Linguistics (Université catholique de Louvain, Belgium), viz. the *Varieties of English for Specific Purposes dAtabase* (VESPA), which aims to represent a wide range of disciplines, genres and writer expertise in academic settings. The collection of expert academic discourse currently totals 1,701,351 words and contains 229 scientific articles from top journals in three domains: business, medicine and linguistics (cf. Table 1).

Disciplines	Number of texts	Number of words
Business (VESPA-BUS)	58	536,971
Linguistics (VESPA-LING)	58	519,109
Medicine (VESPA-MED)	113	645,271
TOTAL	229	1,701,351

Table 1: The VESPA expert corpus

The learner data consists of 57 linguistics term papers written by high-intermediate to advanced EFL students at the same university. The corpus currently totals 163,702 words (cf. Table 2). The project is still very much in its infancy and our aim is to collect essays written by L2 writers from a wide range of language backgrounds in different disciplines.

Discipline: Linguistics	Number of texts	Number of words
Morphology	23	69,702
Syntax	10	25,686
Corpus Linguistics	24	68,314
VESPA-EFL-LING	57	163,702

Table 2: The VESPA learner corpus

All sub-corpora were lemmatized and part-of-speech tagged with the *Constituent Likelihood Automatic Word-tagging system (CLAWS) C7* (cf. Garside and Smith 1997). The tagset includes six different tags for lexical verbs: VV0 (base form, e.g. *drink, work*), VVD (past tense, e.g. *drank, worked*), VVG (-ing participle, e.g. *drinking, working*), VVI (infinitive, e.g. *drink, work*), VVN (past participle, e.g. *drunk, worked*), VVZ (-s form, e.g. *drinks, works*). We applied a Perl program to the CLAWS output to create corpora containing lemmas + simplified POS-tags. POS-tags were automatically simplified to match the level of specificity of lemmas, i.e. the six tags available for lexical verbs (VV0, VVD, VVG, VVI, VVN, VVZ) were conflated into a single VV tag.

We made use of *WordSmith Tools 4* (Scott 2004) to create lists of lemmas + POS-tags for each sub-corpus. In this study, we analyze all lemmas that were assigned a verb tag.

3. Lexical verbs in academic writing

In order to identify the lexical verbs that are prominent in business, linguistics and medicine, we submitted the three sub-corpora to a keyword analysis. This technique has been used in a variety of fields to extract distinctive words or keywords, e.g. words typical of speech vs. writing (Leech et al. 2001), business English words (Nelson 2000) and terminological items typical of specific sub-disciplines of English for Information Science and Technology (Curado Fuentes 2001). Keywords are words that “occur with unusual frequency in a given text” (Scott 1997:236). The adjective ‘unusual’ does not refer to overall high frequency but to a high degree of distinctiveness in comparison with a reference corpus. Each VESPA sub-corpus was compared with a “strongly contrasting reference corpus” (Tribble 2001: 396) of fiction writing, i.e. the 1,000,000 word fiction sub-corpus of the Baby British National Corpus, on the grounds that typical EAP words would be particularly under-represented in this literary genre.

The comparisons were made using the compare list option of *WordSmith Tools* (Scott 2004) which produces a list of all the words that present a statistically significant difference in frequency between two corpora. The three resulting lists of key verbs were then compared to identify (1) verbs that are key in the three domains (business, linguistics and medicine) and are thus potential candidates for inclusion in a list of EAP verbs; and (2) verbs that are exclusive to one domain. In this study, we focus on the 106 verbs that are key in the three domains. Table 3 shows that these shared key verbs account for c. 40% of the verbs that are key in each domain (36.6% in business, 36.7% in linguistics and 40% in medicine).

Discipline	Number of key verb lemmas	Number of shared key verb lemmas	% of shared key verb lemmas
Business	290	106	36.6%
Linguistics	289	106	36.7%
Medicine	267	106	40%

Table 3: Percentage of shared key verb lemmas across disciplines

As appears from Table 4, the 106 shared key verbs largely consist of verbs that are typically used to serve organisational or rhetorical functions prominent in academic writing: reviewing the literature, describing research, exemplifying, reporting and quoting, expressing cause and effect, describing tables and figures, contrasting and summarising.

account, achieve, affect, analyze, apply, assess, assign, associate, base, calculate, characterize, classify, compare, comprise, compute, conduct, consider, consist, construct, contribute, correlate, correspond, define, demonstrate, denote, depict, derive, describe, determine, develop, differ, differentiate, document, eliminate, enhance, establish, estimate, evaluate, examine, exclude, exhibit, explore, facilitate, favor, focus, form, generate, highlight, hypothesize, identify, illustrate, improve, include, increase, indicate, influence, initiate, interact, investigate, involve, lack, limit, link, maintain, manifest, measure, minimize, model, modify, note, obtain, occur, participate, perform, predict, present, process, produce, promote, provide, range, reduce, refer, reflect, relate, report, represent, require, result, reveal, review, score, select, show, signal, study, suggest, summarize, support, target, test, underlie, use, utilize, vary, yield

Table 4: Shared key verbs in business, linguistics and medicine

It is reasonable to argue that these 106 key verbs should be taught in general EAP courses. However, as pointed out by Hyland and Tse (2007: 245-246), although words are used in a wide range of disciplines, “all disciplines adapt words to their own ends, displaying considerable creativity in both shaping words and combining them with others to convey specific, theory-laden meanings associated with disciplinary models and concepts”. They show that the verb *analyze*, which belongs to our list of 106 shared key words, tends to refer to “methods of determining the constituent parts or composition of a substance” in engineering, while in the social sciences it often simply means “considering something carefully” (ibid: 244). This difference in meaning is reflected in our data, as illustrated by example (1) from the medical corpus and examples (2) and (3) from business and linguistics.

- (1) Cells were **analyzed** on a FACSCalibur apparatus (BD Biosciences). (VESPA-MED)
- (2) Boardroom debates, decisions, behaviours and relationships were **analysed** with a view to identifying key relevant themes and developing associated categories. (VESPA-BUS)
- (3) The interactions **analysed** in this paper are drawn from the database of the Wellington Language in the Workplace Project (LWP) which includes a wide range of New Zealand workplaces, ranging from government departments, commercial companies, and small businesses, to factories, and a hospital ward. (VESPA-LING)

In spite of this semantic difference we would like to argue that it is possible to teach the verb *analyze* in an interdisciplinary EAP context. Our objective here is not to downplay differences

in meaning but to try to reconcile research findings and the reality of EAP teaching practice. Like Wang and Nation (2004:310), we consider that “learners should be encouraged to look for the central concept behind a variety of uses”. As regards the verb *analyze*, this central concept can be defined as ‘to examine data using specific methods or tools in order to make sense of it’, with ‘data’ and ‘methods or tools’ varying across fields. It is only by invoking more general definitions of this type that EAP tutors will help L2 learners deal with the various uses of verbs that they may come across even within a single discipline. In linguistics, for example, *analyze* is also often used in the sense of carrying a statistical analysis (example 4), submitting data to a computer-aided analysis (example 5) and distinguishing the constituents of a word, phrase or sentence (example 6).

- (4) Adjective evaluation scores were **analyzed** in a 2 (verb evaluation: positive or negative) x 2 (verb class: experiencer or action) x 2 (thematic role: agent/stimulus or patient/experiencer) ANOVA. (VESPA-LING)
- (5) Gaudio then proceeded to **analyze** the speech signals of the eight speakers, in an attempt to identify any prosodic differentiation between those who self-identified as gay and those who self-identified as straight that may have served as a salient cue to the listener population. (VESPA-LING)
- (6) According to these analyses, NRCs are syntactically part of the clause that appears to contain them—typically they are **analysed** as adjoined to the antecedent nominal, cf. figure 2 below (e.g. Jackendoff 1977, Perzanowski 1980, Kempson 2003, Arnold 2004). (VESPA-LING)

We do not question the fact that different disciplines may show preferences for particular meanings and phraseological patterns (cf. Hyland 2008) but we argue that besides these discipline-specific features, there remains a teachable common core. It is true that the nouns *mice*, *cell lysates*, *protein levels* and *thymocytes* only appear as significant collocates of the verb *analyze* in the medical corpus (cf. example 7) while the nouns *language choice*, *language forms* and *lexical items* are exclusive to linguistics (cf. example 8).

- (7) **Cell lysates were analyzed for** TGF-β1 mRNA and protein, and secreted cytokine was measured by ELISA. (VESPA-MED)
- (8) Metalinguistic awareness refers to the ability to identify, **analyze**, and manipulate **language forms**. (VESPA-LING)

However, the verb *analyze* was found to co-occur with the nouns *study*, *data* and *effect*, the adverb *further* and the complex preposition *in terms of* in all three domains (examples under 9 and 10).

- (9) (a) Further, we **analysed** NFIB’s archival **data** for the 5,957 firms in our sampling frame, which included firm size and industry membership, and found a significant difference only for size; on average, responding firms (73 employees) were larger than the average NFIB member firm. (VESPA-BUS)
- (b) The **data analyzed** in Payne’s study were collected when the children were between 8 and 20 (Payne 1976:110, 112, Fig. 3.2). (VESPA-LING)
- (c) The **data** were collected and **analyzed** by the Data Coordinating Center of the American Lung Association Asthma Clinical Research Centers, which vouches for the data analysis. (VESPA-MED)
- (10) (a) **To analyze the effects of** the hypothesized individual, household and neighborhood factors, along with the effects of segregation processes, a logistic

regression analysis was completed predicting the log odds of being self-employed rather than a wage or salary worker. (VESP-BUS)

(b) **The study analyzes the effects of** orthographic-phonological and morphological awareness in both L1 Hebrew and L2 English and their effects on single-word reading in (L2) English. (VESPA-LING)

(c) **We first analyzed the effects of** UDCA and dexamethasone on the expression of AE2 mRNA isoforms in human liver cells from both cholangiocyte and hepatocyte lineages. (VESPA-MED)

The dominance of the passive is much more marked in medicine (63% of all occurrences of the verb *analyze*) than in the other two varieties, but it still accounts for c. 30% of the occurrences of *analyze* in business and linguistics.

(11) (a) Some of the ways in which the context affects strategies, and knowledge acquisition in particular, **have been analysed** in the literature since L&S published their study. (VESPA-BUS)

(b) The sequences **were then analyzed** in detail, using conversation analysis as applied to the study of institutional interaction (Drew and Heritage, 1992; Hutchby and Wooffit, 2003 [1998]). (VESPA-LING)

(c) 97 serum samples from (A) convalescing patients with invasive diseases and (B) 40 healthy adults **were analyzed**; data are expressed as ELISA units calculated from absorbance at 405 nm. (VESPA-MED)

In the next section, we examine EFL learners' use of lexical verbs in term papers in morphology, syntax and corpus linguistics and compare our findings with results from the expert corpora.

4. The added value of learner corpora

EFL learners have been shown to underuse a whole range of lexical verbs that fulfil important rhetorical or organisational functions in academic writing (cf. Granger and Paquot forthcoming, Paquot 2007b). Closer examination of the top 50 underused verb lemmas in VESPA-EFL-LING shows that 52% of the underused verbs belong to our list of the 106 shared key verbs in business, linguistics and medicine (cf. Table 4). Examples include *achieve, contribute, demonstrate, establish, examine, identify, provide, report, suggest* and *support*. On the other hand, the verb *analyze* is significantly overused in the learner corpus.

As Lorenz (1999:72) has demonstrated, overuse is often accompanied by patterns of non-native usage. French-speaking students make repeated use of the verb *analyze* in active structures introduced by a first person pronoun (38% of all occurrences of the verb) as illustrated in examples 12 and 13. This lexico-grammatical pattern is significantly overused when compared to our business, linguistics and medicine corpora of expert writing. By contrast, French-speaking students significantly underuse passive structures in their linguistics term papers (16% vs. 30.4% in linguistics articles written by professional writers). They also make use of a whole range of infelicitous lexico-grammatical patterns. Examples 14 and 15 illustrate French learners' use of the first person plural imperative *let us/let's/let me* to organise discourse, a rhetorical strategy that does not conform to English academic writing conventions but rather to French academic style and which can therefore be ascribed to L1 transfer of rhetorical conventions (cf. Paquot 2008, see also Neff van Aertselaer 2006).

- (12) **I will analyse** the syntactic structure of the sitcom dialogs in comparison to a corpus of compiled authentic conversations (the Baby BNC corpus).
- (13) **We also analysed** some properties of the root and asserted that it often was a substantive and that it was shortened in more than 20% of the cases.
- (14) **Let's analyse** the Greek prefix anti- and not forget that Brinton includes it among negative prefixes whereas Lightner and Quirk et al. consider it as a prefix of opposition.
- (15) **Let me thus analyse** the 3ADJ entries more closely.

As illustrated by examples (16) to (20), the verb *analyze* is often used in clumsy sequences in VESPA-EFL-LING. In (16), for example, it co-occurs with the adverb *deeply* instead of *carefully* or *thoroughly*. In (20), its subject is *table* rather than the expected animate noun. This inappropriate use would be avoided if learners had a proper understanding of the meaning of the verb *analyze*, i.e. 'examine data using specific methods or tools in order to make sense of it'.

- (16) **It is now interesting to analyse these results more deeply** by looking at the way the different speakers use these three verbs: what are the frequencies of the that-clauses?
- (17) Wordsmith Tools Controller **enables me to analyse** these data as accurately as possible.
- (18) **This paper tries to analyse** some processes of derivation used by Panofsky, by focusing on adjectivalizers and prefixes combined with proper nouns.
- (19) In this case, the grammatical properties of the item **can be guessed by analyzing** its stress placement or its context.
- (20) **The table below analyses** this phenomenon.

These findings show that the population of EFL learners under scrutiny in this study has difficulty using the verb *analyze* in its core meaning and typical lexico-grammatical patterns. By contrast, they seem to master a whole range of discipline-specific collocations such as *analyse a phrase* and *analyse the (grammatical) structure of sth* (cf. examples 21 to 26).

- (21) **A noun phrase can be analyzed** according to both its form and its grammatical function.
- (22) As Biber (2002) points out, linguistics can **analyze both the structure and the use of language** but the emphasis has always been put more on the first one than on the second one.
- (23) Jespersen **analyses** after-, off-, out-, over-, under- and up- **as particles**.
- (24) **The prefixes we are analysing** are very often used in nicknames.
- (25) The aim of this study is **to analyze** Clinton's and Bush's **speeches**.
- (26) Some terms and concepts that could be useful to **analyse our corpus** will also be defined and illustrated in this section.

5. Conclusion

Needs analysis lies at the heart of EAP course design and teaching (e.g. Jordan 1997). Current corpus-based studies have emphasized the specificity of different academic disciplines and genres. As a result, Hyland (2002: 394) has argued that "effective teaching in the universities involves taking specificity seriously. It means that we must go as far as we can". EAP tutors, however, are left wondering how they can possibly bring specificity to classes which are

“more often composed of students from different disciplines and/or language backgrounds with different purposes for taking the class” (Huckin 2003: 6) and what they should teach to law students, for example, who also take courses in economics, history, sociology, psychology, etc. With the emergence of a wide range of interdisciplinary curricula, the problem is likely to become even more acute in the future, not only for students but also for their teachers as “it seldom happens, especially in mixed classes, that the LSP teacher has the disciplinary knowledge needed to provide reliably accurate instruction in technical varieties of language” (Huckin 2003: 8).

Faced with this difficulty, we advocate a ‘happy medium’ approach which, while concurring with Hyland and Tse’s (2007: 247) rejection of approaches of EAP as “an undifferentiated, unitary mass”, also subscribes to Eldridge’s (2008: 111) claim that “though one function of research is to unravel what distinguishes different fields and genres, another function is to find similarities and generalities that will facilitate instruction in an imperfect world”. Our analysis of the verb *analyze* shows that it is possible to identify both the common core features of an academic word and its discipline-specific characteristics in terms of meaning, lexico-grammar, phraseological patterns, etc. This dual approach allows teachers in mixed EAP contexts to both highlight the field-specific collocates which will not cause difficulties to students as they bring their discipline practices to the classroom, and place emphasis on cross-disciplinary collocates which “are not likely to be glossed by the content teacher” (Flowerdew 1993: 236) (see also Mudraya 2006). To be maximally useful, this approach should not take the GSL list “as any kind of ‘given’ in the compilation of more specialized wordlists” (Hancioğlu et al 2008: 466). As shown by Paquot (2007 a & b), a wide range of high frequency words are used to organise academic texts and structure their content along logico-semantic relations.

The field of EAP clearly needs more studies of the type presented in this paper to assess the generalizability of the approach. Ideally, those studies should also make use of learner corpora to pinpoint EFL learners’ difficulties with academic vocabulary. Only then will we be able to assess whether academic vocabulary is a chimera or a highly useful theoretical construct.

References

- Coxhead A. (2000). A new Academic Word List. *TESOL Quarterly* 34 (2): 213--238.
- Curado Fuentes A. (2001) Lexical behaviour in academic and technical corpora: implications for ESP development. *Language Learning and Technology* 5(3):106--129.
- De Cock, S., Gilquin, G., Granger, S., Lefer, M-A., Paquot, M., and Ricketts, S. (2007) Improve your writing skills. In M. Rundell (editor in chief) *Macmillan English Dictionary for Advanced Learners* (second edition) IW1--IW50. Oxford: Macmillan Education.
- Eldridge, J. (2008) “No, there isn’t an ‘academic vocabulary,’ but ...” A reader responds to K. Hyland and P. Tse’s “Is there an ‘academic vocabulary?’” *TESOL Quarterly* 42(1): 109--113.
- Flowerdew, J. (1993) Concordancing as a tool in course design. *System* 21(2): 231--244.
- Garside R. and N. Smith (1997) A Hybrid Grammatical Tagger: CLAWS4. In Garside R., G. Leech and A. McEnery (eds) (1997) *Corpus annotation: linguistic information from computer text corpora*. New York: Addison Wesley Longman, 102--121.

- Gilquin, G., Granger, S. and Paquot, M. (2007) Learner corpora: the missing link in EAP pedagogy. In Thompson, P. (ed.) *Corpus-based EAP Pedagogy*. Special issue of *Journal of English for Academic Purposes* 6(4): 319--335.
- Granger, S. and Paquot, M. (in press) Lexical verbs in academic discourse: a corpus-driven study of learner use. In M. Charles, Hunston, S. & Pecorari, D. (eds.) *At the Interface of Corpus and Discourse: Analysing Academic Discourses*.
- Hancioğlu, N., Neufeld, S. and Eldridge, J. (2008) Through the looking glass and into the land of lexico-grammar. *English for Specific Purposes* 27(4): 459-479.
- Hinkel, E. (2002) *Second Language Writers' Text. Linguistic and Rhetorical Features*. Mahwah, New Jersey & London: Lawrence Erlbaum Associates.
- Huckin, T. (2003) Specificity in LSP. *Ibérica* 5: 3--17.
- Hyland, K. (2002) Specificity revisited: how far should we go now? *English for Specific Purposes* 21(4): 385--395.
- Hyland, K. (2008) As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27(1): 4--21.
- Hyland, K. and Tse, P. (2007) Is there an "Academic Vocabulary"? *TESOL Quarterly* 41(2): 235--253.
- Jordan, R. (1997) *English for Academic Purposes. A guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Leech G., P. Rayson and A. Wilson (2001) *Word frequencies in written and spoken English*. London: Longman.
- Lorenz G. (1999) Learning to cohere: causal links in native vs. non-native argumentative writing. In W. Bublitz, U. Lenk and E. Ventola (eds) *Coherence in Spoken and Written Discourse. How to create it and how to describe it*. Amsterdam and Philadelphia: Benjamins, 55-75.
- Martin A. (1976) Teaching Academic Vocabulary to Foreign Graduate Students. *TESOL Quarterly* 10(1): 91--97.
- Mudraya, O. (2006) Engineering English: A lexical frequency instructional model. *English for Specific Purposes* 25(2): 235--256.
- Nation P. (2001) *Learning Vocabulary in another Language*. Cambridge: Cambridge University Press.
- Neff Van Aertselaer, J. (2006) A rhetorical analysis approach to English for Academic Purposes. *Revista de Lingüística y Lenguas Aplicadas* 1: 63--72.
- Nelson M. (2000) *A Corpus-Based Study of Business English and Business English Teaching Materials*. Unpublished PhD Thesis. Manchester: University of Manchester.
- Nesselhauf, N. (2005) *Collocations in a learner corpus*. Amsterdam: Benjamins.
- Paquot, M. (2007a) Towards a productively-oriented academic word list. In J. Walinski, K. Kredens & S. Gozdz-Roszkowski (eds) *Corpora and ICT in Language Studies. PALC 2005*. Frankfurt am main: Peter Lang, 127--140.
- Paquot, M. (2007b) *EAP vocabulary in EFL learner writing: from extraction to analysis: A phraseology-oriented approach*. Unpublished PhD thesis. Université catholique de Louvain, Centre for English Corpus Linguistics.
- Paquot, M. (2008) Exemplification in learner writing: a cross-linguistic perspective. In S. Granger and F. Meunier (eds) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: Benjamins, 101--119.
- Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly* 18: 117--31.
- Schmitt N. and D. Schmitt (2005) *Focus on vocabulary: Mastering the Academic Word List*. London: Longman.
- Scott M. (1997) PC analysis of keywords and key keywords. *System* 25(2): 233--245.

- Scott M. (2004) *WordSmith Tools 4*. Oxford: Oxford University Press.
- Sutarsyah, C., Nation, P. and Kennedy, G. (1994) How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal* 25(2): 34--50.
- Swales, J. and Feak, C. (2004) *Academic Writing for Graduate Students. Essential Tasks and Skills*. Ann Arbor: The University of Michigan Press.
- Tribble C. (2001) Small corpora and teaching writing: towards a corpus-informed pedagogy of writing. In Ghadessy M., A. Henry and R. Roseberry (eds) *Small corpus studies and ELT: theory and practice*. Amsterdam and Philadelphia: Benjamins, 381--408.
- Wang Ming-Tzu, K. and Nation, P. (2004) Word Meaning in Academic English: Homography in the Academic Word List. *Applied Linguistics* 25(3): 291--314.
- West, M. (1953) *A General Service List of English Words*. London: Longman.
- Xue, G. and Nation, P. (1984) A University Word List. *Language Learning and Communication* 3(2): 215--229.