

Part I
Vocabulary and terminology
in academic writing

Christine S. Sing
Vienna University of Economics and Business, Austria

Writing for specific purposes: Developing business students' ability to 'technicalize'

Abstract

English: This article examines business students' ability to technicalize in an L2 English writing task. Building up technicality in discourse is a key component of writing competence. Despite the importance of technicality for discipline-specific writing, so far little attention has been paid to identifying the usage patterns characteristic of this type of writing. The aim of this study is to investigate how undergraduate writers technicalize in elaborating on technical terms by means of defining, exemplifying and explaining. Drawing on a self-compiled specialized corpus, the study adopts a mixed-methods approach of computation and interpretation. It was found that technicalizing is a two-stage process, which consists of naming a term and subsequently embedding it in taxonomic relationships. The resulting chains of reference are taken to be indicative of field-specific uses in writing. The findings have important implications for developing business students' writing skills in view of the conceptual challenges they meet in current specific-purpose instruction.

German: Gegenstand des vorliegenden Beitrags ist die Art und Weise, wie Studierende beim fachspezifischen Schreiben in ihrer L2 Englisch Technizität (*technicality*) herstellen. Hierzu werden diejenigen sprachlichen Verfahren ermittelt, die sie nutzen, um Fachtermini durch Definitionen, Beispiele und Erklärungen in den Text einzubetten. Die Fähigkeit, Technizität herzustellen, wird dabei als wesentliche Komponente der Kompetenz zum fachsprachlichen bzw. disziplinspezifischen Schreiben verstanden. Als Datengrundlage dient ein spezialisiertes Korpus von studentischen Texten aus vier Bereichen der internationalen Betriebswirtschaftslehre, die auf Englisch als L2 verfasst wurden. In einem *Mixed-Methods*-Ansatz werden korpuslinguistische mit interpretatorischen Verfahren kombiniert. Es zeigt sich, dass das Herstellen von Technizität als zweistufiger Prozess beschrieben werden kann, in dem ein Fachausdruck zunächst benannt und dann in eine taxonomische Beziehung eingebettet wird, wodurch Referenzketten entstehen, die für die untersuchten Texte charakteristisch sind. Aus den gewonnenen Erkenntnissen werden Schlussfolgerungen für die Didaktik der Schreibkompetenzförderung in wirtschaftswissenschaftlichen Lehr-/Lernkontexten gezogen.

1 Writing for specific purposes

English for Specific Purposes (ESP) is traditionally conceived of as a “materials- and teaching-led movement” (Dudley-Evans/St John 1998: 19), catering to specific student needs in specific contexts. ESP is thus “an approach to language teaching in which all decisions as to content and method are based on the learner’s reason for learning” (Hutchinson/Waters 1987: 19). Recent developments in the field, however, have heightened the need for a fresh perspective on ESP: “What once looked to many like a straightforwardly needs-oriented, a- or pan-theoretical [...] approach, now, like the constantly changing learning targets it addresses, is itself becoming harder and harder to capture in anything like a single stop-action frame” (Belcher 2006: 134).

The internationalization of higher education (HE) marks a sea change in the role of English in scholarship and instruction. English-medium instruction has become the rule rather than the exception in tertiary education across Europe. Given the centrality of writing in HE institutions (Hyland 2013), there is an increasing concern for students’ needs, lacks and wants in writing instruction. Indeed, it appears to be the case that students are lacking essential competencies when it comes to university writing and thus need to be acquainted with its primary purposes. This raises the question what differentiates university writing in English for General Academic Purposes (EGAP) from that in English for Specific Academic Purposes (ESAP) settings.

Arguably, the transition from secondary to tertiary education represents a “cultural shift” (Gee 1996: 155) for most students as they meet with different literacy expectations in their new environment. What is more, the novice writers face linguistic challenges in that they use English as an L2 in their writing, which means that they are struggling with both general and academic English at the tertiary level. Particularly, students are lacking in knowledge of the relevant target genres and are grappling with the highly formalized and conventionalized written academic style.

In ESP settings, the writing task has regularly been of central importance, most notably in business, academic and professional domains (Tardy 2012: 6266). And yet, in specific-purpose writing instruction, these two challenges – cultural and linguistic – are compounded by what Peters et al. (2014: 744 ff.) refer to as “conceptual challenges”. The authors report

several pedagogical issues in ESP contexts, arising from the fact that learning disciplinary knowledge and learning the specialist language are not carefully orchestrated. As a result, ESP students are not only linguistically challenged by having to overcome difficulties in both general academic English and the specialist language of their discipline; they also face conceptual challenges in that ESP programmes “require students to develop a more abstract understanding of concepts within the discipline, in order to be able to apply their knowledge effectively” (Peters et al. 2014: 755). This implies that ESP students may be familiar with field-specific concepts while lacking the language resources necessary to construe disciplinary knowledge in their writing.

It is this conceptual challenge that will be addressed in the present analysis. It will be argued that the students in the ESP setting of a business school fail to be empowered by the potentially effective cross-fertilization of subject knowledge and disciplinary language. Building on Tribble’s (2002) contextual-analysis framework and Flowerdew’s (2004) parameters for specialized corpora, it will be shown that, in order to be able to engage in disciplinary discourse, student writers need to develop the ability to “technicalize” (Ravelli 2004: 104), i.e., to grow aware of the linguistic resources highlighting that a given word or concept is embedded in a body of knowledge.

Drawing on a self-compiled specialized corpus, the corpus of Academic Business English (ABE), this account sets out to examine three major modes of constructing technicality in business student writing, namely defining, explaining and exemplifying. Prior to presenting the main findings, I will first outline the theoretical framework of Systemic Functional Linguistics (SFL), in which the notions ‘technicality’ and ‘technicalize’ originated. The in-depth contextual analysis will conclude with a discussion of how specific-purpose writing instruction could address these conceptual challenges.

2 Theoretical background

The notions ‘technicality’ and ‘technicalizing’ are deeply embedded in the SFL framework, which contends that language users represent experience through language by means of various genres and resources. Martin (1991: 308 f.) proposes a cross-classification of genres (report, explanation, exposition) and discourse functions (describing, explaining). In this

approach, the complexity of written texts can be described on the basis of two functional dimensions, technicality and abstraction. The rationale behind this classification is that, while both sciences and humanities rely on these genres and functions, there are clear discipline-specific preferences as to which features predominate. A case in point is the genre of report which, in science discourses, fulfils a taxonomizing function, while reports in the humanities tend to generalize. Not only does this highlight the instrumental character of discourses, it also illustrates the process of knowledge construction in different disciplines. Different fields name, order and classify similar phenomena differently.

This begs the question as to the linguistic resources that are instrumental in construing disciplinary knowledge in writing. In order to achieve representation, language typically fulfils three functions: 1. creating technical vocabulary, 2. classifying the experiential world, and 3. explaining the experiential world. Ravelli (2004: 104) argues that, in order to perform these functions, “writers must be able to give names to things, and to connect these names to each other, in order to theorize about the world around them”. She refers to these functional modes as technicalizing and rationalizing, respectively. Here the emphasis will be placed on technicalizing or theorizing, as it may also be referred to. It will be argued that, once capable of technicalizing, student writers demonstrate a deeper understanding of taxonomies, i.e., of how terms or concepts are to be placed in an ordered system.

Technicalizing, i.e., the process of building up technicality in writing, involves two stages: 1. naming the phenomenon and 2. making the name technical. The latter is aided by ‘discourse cues’ such as textual signals or macrostructures. Woodward-Kron (2008: 238 f.) identified several linguistic devices that may be used to flag technicality in discourse, examples being definitions and taxonomic relationships. The reliance on discourse cues, however, presupposes the existence of an engaging writer-reader relationship. Since the English language represents a “writer-responsible culture” (Dahl 2004), strategies such as “lexical familiarization” are particularly rewarding. The latter can be defined as “a contextual aid, intentionally and explicitly provided by the author when writing for a specific readership. The writer’s intention is to help his reader by providing him with sufficient

familiarity with the new word, as employed in its context, so that the reader can continue reading with understanding” (Bramki/Williams 1984: 170).

Taking the writer and reader community as a point of departure, Hyland (2010) conceives of academic writing as an interactive communicative practice. Of particularly strong appeal is Hyland’s (2005: 37) interpersonal model of metadiscourse, i.e., “the cover term for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community”. More specifically, he distinguishes between two dimensions of interaction, namely interactional and interactive metadiscourse. While the former is primarily concerned with reader engagement, it is the category of interactive metadiscourse that includes the linguistic resources instrumental in technicalizing. According to Hyland (2005: 37), interactive metadiscourse is grounded in the writer’s awareness of an audience. For this reason, writers rely on resources such as transitions, frame markers and code glosses. The latter, which “supply additional information, by rephrasing, explaining or elaborating what has been said, to ensure the reader is able to recover the writer’s intended meaning” (Hyland 2005: 52), are vital for building up technicality in discourse. Table 1 presents the code glosses that may be used to signal technicality in discourse.

Table 1: Technicalizing functions of code glosses

| <i>Discourse function</i> | <i>Linguistic resources</i> | <i>Metadiscourse category</i> |
|-------------------------------------------------|---------------------------------------------------------------------------------------------|-------------------------------|
| <i>Defining Exemplifying Explaining</i> | <i>called, defined as, known as such as, for example, e.g. that is, which means</i> | Code glosses |

In what follows, each of these functions will be subject to an in-depth analysis, involving both a quantitative analysis of their frequencies of occurrence and a more qualitative study of their discourse patterning. As will be shown, the student writers technicalize in that they establish so-called “chains of reference”, “i.e. sequences of noun phrases all referring to the same thing. This constitutes an important aspect of textual cohesion, which makes a text more than just a series of sentences” (Biber et al. 1999: 234 f.). The following corpus example showcases how such textual macrostructures are established:

- [1] I will **name** three of these companies, **explain** what kind of products they sell and the materials used. [ABE_Business]

The student clearly feels the need to first refer to the phenomenon, followed by embedding it in a given technical framework. Before this will be discussed, the data and context of analysis will be described.

3 Data and method

3.1 Corpus and context description

Recent developments in the field of corpus-based analyses of student writing have seen the compilation of several corpora of (under-)graduate writing in English, for example, the Michigan Corpus of Upper-level Student Papers (MICUSP) and the corpus of British Academic Written English (BAWE). While both provide a rich source of data, comprising several genres and disciplines, they need to be carefully differentiated from learner corpora on the one hand and Language for Specific Purposes (LSP) corpora such as the ABE corpus on the other. The former are clearly indebted to Second Language Acquisition (SLA) research, where the emphasis is placed on interlanguage features such as a limited lexical repertoire, semantic misuse or overuse of connective devices (Paquot 2010). The latter, by contrast, are referred to as specialized corpora, which “are corpora designed for the purpose of creating a sample of specialized language either by collecting texts of similar content [...] or of similar text-type or genre” (Gavioli 2005: 7). These tend to be small and localized collections of data, which means that scholars, like the author of this present study, are frequently forced to assemble their own corpora (Tognini-Bonelli 2008: 35). The ABE corpus, which makes up the database of the present study, is such a self-compiled, specialized corpus. As shown in Table 2, with a size of approximately one million tokens, the ABE corpus greatly exceeds the usual 250,000 word mark of specialized corpora. In actual practice, specialized corpora typically range between around 60,000 words and 250,000 words. Handford (2010: 258) thus concludes that “[i]n terms of actual size, a specialized corpus can be defined as large [...] if it contains a million words”.

Table 2: ABE corpus data

| <i>Subcorpora</i> | <i>Number of papers</i> | <i>Type/token ratio (STTR)</i> | <i>Number of running words (tokens)</i> |
|-------------------|-------------------------|--------------------------------|-----------------------------------------|
| <i>Business</i> | 103 | 41.22 | 236,917 |
| <i>Economics</i> | 104 | 40.76 | 273,455 |
| <i>Finance</i> | 104 | 39.60 | 253,203 |
| <i>Marketing</i> | 102 | 40.64 | 251,156 |
| <i>ABE Total</i> | 413 | 40.55 | 1,014,731 |

Given that a corpus is “not simply a collection of texts” (Biber/Conrad/Reppen 1998: 246), there should be a principled stand on corpus compilation. The compilation of the ABE corpus was guided by a clear set of design criteria. These are summarized in Table 3 below, drawing on Flowerdew’s (2004: 21) set of parameters for defining specialized corpora on the one hand and Tribble’s (2002: 133) contextual-analysis framework on the other.

Table 3: Description of data and context

| DATA DESCRIPTION | |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------------|
| <i>Specific purpose for compilation</i> | To investigate business student writing in a localized setting |
| <i>Size</i> | Large specialized corpus (ABE) of approx. 1 million running words |
| <i>Language used</i> | Written academic English, used as L2 |
| CONTEXTUAL DESCRIPTION | |
| <i>Genre family</i> | Assessed university student writing |
| <i>Communicative purpose</i> | To demonstrate / be accredited for proficiency in academic English and knowledge of business concepts |
| <i>Text type</i> | 413 simple assignments, seminar papers |
| <i>Subject matter</i> | Broad thematic range of business topics |
| <i>Setting</i> | ESP setting of a business school; committed to a wide-angled approach to the teaching of business English |
| <i>Participants</i> | Advanced undergraduate students of international business administration |

In what follows, each of these features will be described, detailing the specifics of the ABE corpus and the institutional context in which it originated. First and foremost, “[h]aving a clearly articulated question is an essential step in corpus construction since it will guide the design of the corpus” (Reppen 2010: 31). And yet, the compilation of specialized corpora tends to be prompted by more specific purposes. The most obvious reason for building “DIY corpora” (McEnery/Xiao/Tono 2006: 21) lies in the fact that none of the existing corpora include a sample of the genre investigated. Another main driving force behind self-compiled corpora are methodological considerations. Even if corpora are available, not all data lend themselves to conducting contextualized analyses. Studying business student writing with a view to enhancing these students’ development of writing skills necessitates the retrievability of contextual information, ranging from details about the institutional setting and the curriculum in operation to localized academic practices. Thus, methodologically, small self-compiled corpora have the edge over larger corpora whenever detailed, fully contextualized analyses are intended (Nelson 2010: 55). Yet another reason for compiling the ABE consisted in creating a specialized corpus that would contain the type of language characteristic of this particular ESP setting.

As mentioned earlier, the ABE corpus can be considered a large specialized corpus. The question of corpus size is closely intertwined with the issue of representativeness. Unlike general corpora, for which representativeness may be achieved through size, this “aim [...] is rendered impossible through the need to target a disciplinary or thematic specialty” (Williams 2002: 45). Instead, “the representativeness of specialized corpora is usually measured by reference to external selection criteria (i.e. *by/for whom* the text is produced, what is its subject matter), which are regarded as somewhat subjective” (Flowerdew 2004: 18; highlighting in the original). In order to mitigate the effects of subjectivity, the collection of data needs to meet clear sampling criteria. Accordingly, texts had to meet the following external criteria for inclusion in the ABE: All texts had to be complete rather than samples. They had to be single-authored and of equal length (i.e., not exceeding the word limit of 2,400 words). They had to be organized in terms of the macrostructure laid down in the institute’s style sheet. Their thematic focus was required to be on either business, economics, finance or marketing, which corresponded to the subject matter dealt with in the

seminars. These external criteria clearly limit the range of texts eligible for inclusion, meaning that “[b]y restricting the scope of the corpus, energy can be directed towards assembling a detailed collection of texts that fully represent the kind of academic language one is likely to encounter” (Meyer 2002: 36).

However, the conditions of data collection are highly unpredictable, particularly when student work is involved. Alsop & Nesi (2009: 73) report similar sampling problems when compiling the BAWE corpus. As they did not quite know what to expect from the assignments – neither in terms of volume nor nature – they used a matrix of disciplinary groupings on four different levels of study. Following Alsop & Nesi, the ABE corpus texts were sampled on the basis of a matrix of four disciplinary groupings in order to plan the structure of the corpus, i.e., the sub-fields of business, economics, finance and marketing served to organize corpus holdings. Accordingly, texts were cross-classified making up matrix cells, which were based on these sub-fields on the one hand and on the external criteria mentioned above on the other. In order to achieve balance, texts were then added to the corpus until all cells were filled in equal numbers.

Prior to moving on to the description of context, a passing reference to the language used in the corpus texts is in order. The specialized language represented in the ABE corpus is “[t]he academic business English required by students on courses in disciplines such as business, finance, accounting and banking [which] has more in common with the study of other EAP [English for Academic Purposes] disciplines” (Dudley-Evans/St John 1998: 53). The student writing included in the ABE is therefore strongly informed by the institutional context. The close interweaving of texts and the institutional setting in which they originated warrants the distinction between text type, genre and communicative purpose, all of which relate to the categorization of written texts. The distinction between genre and text type, in particular, is a moot point. Broadly, the distinction is grounded in whether external or internal criteria are used for grouping texts together. As originally laid down by Biber (1988), genre “is defined as a category assigned on the basis of external criteria such as intended audience, purpose, and activity type, that is, it refers to a conventional, culturally recognized grouping of texts based on properties other than lexical or grammatical (co-)occurrence features” (Lee 2001: 38). The category text type, by con-

trast, characterizes texts in terms of their textual properties and linguistic patterning. However, as pointed out by Paltridge (1996), the distinction is anything but watertight and is best seen as complementary. More precisely, genres can be realized by different text types, that is, “more than one genre may share the same type. [...] Equally, a single genre [...] may be associated with more than one text type” (Paltridge 1996: 239).

In modelling the ‘seminar paper’ (SP) as a specific type of written text, the present account subscribes to Biber’s (1988) corpus-linguistic classification. Accordingly, the SP is viewed as a specific, institutionalized form of assignment writing, which a) has a particular organization, b) deals with a specialized topic and c) realizes an identifiable set of social and communicative functions. The SP is deeply rooted in German-speaking HE (Kruse 2006). It is an argued text similar to an exposition, frequently taking on the form of research articles (RA) in miniature (Ehlich 2003). In many disciplines, particularly in the social sciences and humanities, it continues to be the standard type of assessed writing. In the context of Austrian HE, the seminar paper has been studied extensively (Gruber 2006). A case in point is Gruber’s (2004: 50) identification of the SP’s prototypical structure: introduction, problem definition, descriptive and argumentative discussion of the problem(s), case description and conclusion.

Its generic properties are further determined by a detailed set of instructions, according to which the paper is to present an ‘academic argument’ rather than make recommendations for (case-)specific situations. The strict formal criteria laid down for the SP typically relate to principles of layout and text organization, citation practices and bibliographical conventions. The SP is typically a simple assignment, i.e., the number of texts equals the number of writing tasks; conversely, the number of texts equals the number of authors. Thematically, SPs are subject to sub-field specific variation. The internal criteria used to determine the SP as an abstract text type were topic-based. In order to ensure that the teaching and learning context is adequately reflected in the corpus data, the corpus was subdivided into four thematic subcorpora, each relating to the topical focus of seminar teaching: business studies, economics, finance and marketing. Regarding the social and communicative functions of SPs, they can be argued to realize a particular genre, namely assessed student writing. In fact, university student writing has been shown to constitute a genre family consisting of various

subgenres (Nesi/Gardner 2012). In combining several of these subgenres, the SP fulfils what is perhaps the primary function of assessed university writing: accreditation (Nesi/Gardner 2012: 23).

The setting in which the texts were collected and the participants can be specified as follows: The institutional context at hand is a fully-fledged academic unit, which, despite its full-departmental status, does not offer bona fide degree programmes. It instead caters to business programmes, such as the four-year diploma programme of International Business Administration (IBA). Owing to the content-based curriculum, the focus clearly is on specialist language (Business English – BE), leaving little scope for developing general English language skills. Similarly, there are few opportunities of familiarizing students with academic English, which means that the students will not have received any formal writing instruction prior to producing the assignment. In the light of this curricular framework, it is perhaps not surprising that a wide-angled approach (Basturkmen 2010: 143) to the teaching of BE is adopted, implying that English is taught through a variety of topics in several business fields. The integration of subject-specific knowledge with language learning is thus hinged on a terminology-driven approach to BE. The participants are usually in their third or fourth year of the diploma programme when producing the SP. They are referred to as advanced students in the sense that they will have spent at least three years in HE, in other words, ‘advanced’ does not describe their proficiency in English. As there are no initial placement tests nor any official language testing throughout the programme, no assessed information as to their actual level of English can be provided. Neither are there any detailed metadata available on individual student authors. The vast majority are however L1 users of (Austrian) German. At the same time, there is a large intake of international students, accounting for approximately 10% of the entire student population.

3.2 Method

Discipline-specific writing has mainly been studied from either a corpus-linguistic, variationist perspective or using ethnographic research methods. Corpus-based approaches aim to retrieve lexico-grammatical patterns in order to identify field-specific uses of language. These variationist accounts

tend to focus on variables such as novice vs. expert practices or English L1 vs. English L2 uses in discipline-specific writing. There is a substantial body of research which has examined disciplinary variation in academic discourse, including citation practices (Samraj 2013; Charles 2006), stance (Chang/Schleppegrell 2011; Biber 2006) and shell noun uses (Flowerdew 2006; Aktas/Cortes 2008). However, approaches of this kind carry with them various well-known limitations. First, it seems to be the case that, due to the contrastive approach taken, variationist studies conceive of disciplinary forms of writing as uniform practices. The inherent normativity then leads to an analytical focus “on identifying academic conventions [...] and on (or with a view to) exploring how students might be taught to become proficient or ‘expert’” (Lillis/Scott 2007: 13). Second, expert writing is frequently conceived of as constant and stable, ignoring the fact that lexical complexity in the writing of experts is itself subject to considerable variation over time (Trinh 2011). Thus, once both a timescale and a time window are taken into account, expert practices no longer serve as the reliable yardstick they are made out to be. Yet another shortcoming of this approach consists in the decontextualized lexico-grammatical patterns identified; more often than not, the analysis does not include the level of discourse, i.e., language beyond the sentence level.

Ethnographically-oriented studies seek to address the social conditions of the production and interpretation of academic texts. Accordingly, Gnutzmann & Rabe (2014: 33) observe “a certain neglect of the wider context of academic writing and publishing, since many conventions of disciplinary cultures are unwritten and have to be reconstructed through other methods”. Using interpretative methods, the authors propose an interview-based study of the language demands and attitudes of German researchers using English as an L2. The ultimate aim of such an account consists in homing in on precisely those contextual factors amiss in much corpus-based research.

The present account proposes a ‘third way’ of analysing ESP writing, intending to reconcile these two approaches. It combines the empirical rigour of corpus-based studies with the interpretative methods of ethnographical research. The study at hand thus provides a fully contextualized analysis in drawing on the ABE as a localized collection of data. Specialized corpora permit the analyst to study patterns of language use in the contexts in which they originated (Handford 2010: 258 f.). More precisely, with

specialized corpora, “the analyst is probably also the compiler and does have familiarity with the wider socio-cultural context in which the text was created [...]. The compiler-cum-analyst can therefore act as [...] specialist informant to shed light on the corpus data” (Flowerdew 2005: 329). Not least the growing complexity of ESP settings has highlighted that corpus data are invariably situated. The importance of studying corpus texts in their contextual environment results from the tight interweaving of text and context: “To make sense of the way particular texts and particular grammatical patterns occur, and why they occur in a particular order, it is essential to consider what is distinctive about texts within their particular institutional context” (Veel 1997: 162). The author of the present study is such an insider. This field knowledge will be used when discussing the findings of the corpus study below.

In what essentially constitutes a mixed-methods approach involving computation and interpretation, the corpus data were subject to a bottom-up, inductive analysis, followed by the study of discourse patterns, which was more qualitative in nature. Using *Wordsmith Tools 6* (Scott 2012), a wordlist analysis was first conducted for all four subcorpora. This served to retrieve the most frequently used code glosses across the ABE, as exemplified in Table 1 above. These instances were then concordanced to identify their most frequent collocates. In a second stage, the discourse patterns in which these code glosses are embedded were examined using *MAXQDA*. For this purpose, the chains of reference were coded as definitions, exemplifications and explanations respectively. It should be noted that the classification of metadiscourse devices is problematic in itself, since these describe a continuum of functional resources rather than mutually exclusive categories. Therefore, metadiscourse is notoriously difficult to quantify, which is why the emphasis was placed on exploring the usage patterns of these discourse phenomena in the data at hand, paying particular attention to intra-disciplinary variation.

4 Results and discussion

The analysis set out to examine the linguistic resources used by business students to build up technicality in their writing. Technicalizing was introduced as a two-stage process: 1. referring to/naming the phenomenon concerned

and 2. making it technical by means of elaboration. Due to the interactive nature of academic writing, the first step of the analysis involved retrieving all relevant code glosses, assuming that they highlight instances of technical language use. The emphasis was placed on glosses that mark defining, exemplifying and explaining. As can be seen in Fig. 1, which shows the main findings for this first set of analyses, the three metadiscoursal devices are used fairly differently across the subcorpora.

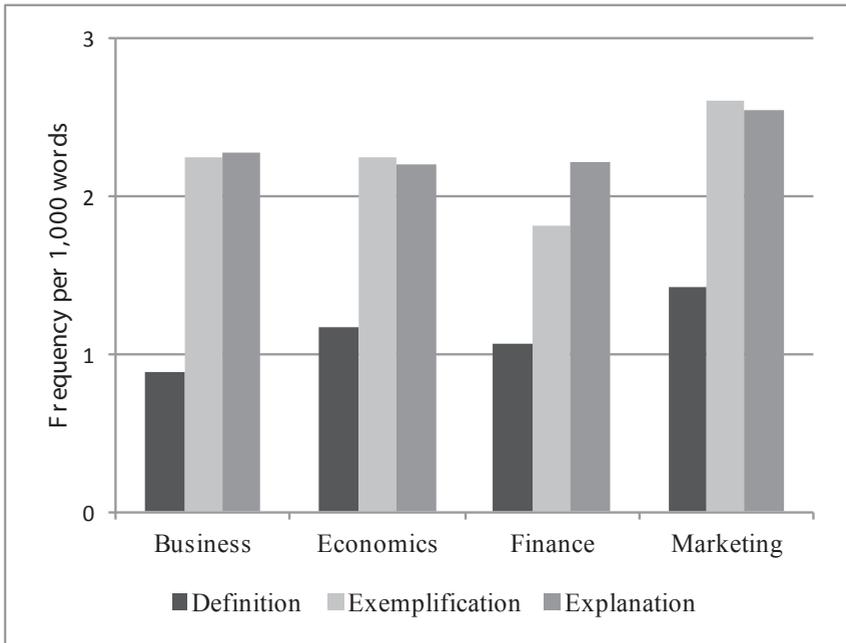


Fig. 1: Distribution of code glosses across the ABE subcorpora

Generally, it is evident that the construction of technicality chiefly occurs by means of exemplification and explanation. Explicit glosses used to signal acts of defining are clearly less frequent than those relating to exemplification and explanation. Referring more specifically to individual subcorpora and zooming in on the more subtle differences between them, several interesting findings were produced. First and foremost, the in-corpus variation is remarkable, given that all four ABE subcorpora are made up of the same

text type. While definitions clearly play a subordinate role in building up technicality in these corpus texts, the distribution of exemplifications and explanations in the corpus data varies greatly. Focusing on frequency of occurrence alone, it could be argued that the results for the business and economics subcorpora are fairly similar. However, when taking into account the relations between the most pervasive code glosses exemplification and explanation, the distribution for business and finance on the one hand, and economics and marketing on the other is fairly similar. In the latter, the student writers use slightly more exemplifications than explanations while the situation is reversed for the business and finance subcorpora. This can be accounted for by the differences in composition structure: While both business and finance papers tend to be organized as case studies and thus follow the problem-solution pattern, papers in economics and marketing tend to be argued essays, both descriptive and argumentative in nature, implying that the student writers rely on exemplifications to illustrate the points they make rather than explaining case specifics in detail. Even though these facts and figures of corpus analysis highlight important differences between the ABE subcorpora, they represent only a small part of a much bigger picture. Therefore, the second part of the analysis involved identifying the usage patterns and textual macrostructures in which these code glosses are embedded.

4.1 Technicalizing by means of defining

Due to their knowledge-telling function, definitions expand on a given term that has been ‘named’ in the preceding context. When labelling such a discourse act, the student authors whose writing is documented in the ABE tend to resort to the following linguistic resources: *called*, *the term X*, *defined as*, *define* and *known as*, which makes them the top five most frequently used glosses of this category across the ABE subcorpora. The analysis revealed several usage patterns which will now be discussed (the subcorpora from which the examples are taken are indicated in square brackets):

- [2] AUTHOR and AUTHOR (1995: 17) **define** CC [= Corporate Citizenship] as the engagement of companies, using corporate resources in order to solve [ABE_Business]
- [3] type of tourism called events tourism. AUTHOR **defines** events tourism as “the planning, development [ABE_Marketing]

As shown, the most obvious way of introducing a new term is by what Ädel & Garretson (2006: 273) refer to as “attribution”, the procedures by which “writers [...] frequently attribute statements and acts to other researchers”. The student writers, too, make use of this standard procedure by means of paraphrasing or quoting verbatim from a source, which means that the verb *to define* is used as a reporting verb. In both instances, however, the term is not defined on first mention; instead it has already been used in the preceding context, which testifies to the existence of “chains of reference”. Accordingly, the term *Corporate Citizenship* has already been introduced as it is defined in abbreviated form. Example [3] provides an introductory sentence in which reference is made to the term *event tourism*, which is subsequently defined in a more technical way.

Chains of reference are also operative in yet another pervasive pattern identified in business student writing. As exemplified below, the reference to a given term is established by the section heading that serves as the antecedent and immediately precedes the defining act:

- [4] Joint Ventures [= section heading] A joint venture can be defined as a company which is owned corporately by two or more parties [ABE_Business]
- [5] Definitions [= section heading] The Role of Information [= section heading] Information can be defined as “data with attributes of relevance and purpose [ABE_Economics]

Technicality is thus gradually built up in first naming the phenomenon in question, followed by according it a technical status. As indeed highlighted by the corpus examples above, technicalizing clearly has a taxonomizing function in that the term is positioned in an ordered system. The students’ taxonomic organization of their writing topic creates such a system, whose structure becomes manifest in the tables of contents of their papers. The fact that the student writers seem to be preoccupied with taxonomies rather than developing ideas is further supported by the relatively low lexical density (see STTR scores in Table 2 above), suggesting that their writing has a low informational value and that topic development occurs on the basis of few new language items. In their study of technicality in geography textbooks, Wignell, Martin & Eggins (1993: 157 ff.) found that chapter or (sub-)section headings and the boldening of terms correspond to “superordination taxonomy” in contrast to “composition taxonomy”. The former is

primarily concerned with classifying a given term while the latter provides a list of its elements. For example, the superordination taxonomy of the technical term *climate* would include different types of climate, e.g., tropical, moderate and dry, while the composition taxonomy would include a list of descriptive elements such as temperature, pressure systems or atmospheric moisture (Wignell/Martin/Eggins 1993: 159).

It can be argued that the above examples are indicative of a strong concern for superordination rather than composition taxonomy. The findings also suggest that students are concerned with composition taxonomy as illustrated by the following examples:

- [6] cross-cultural cooperation in a positive manner. The first dimension is called “universalism versus particularism [ABE_Economics]
- [7] change process is called “managing the transition”. The third phase is called “sustaining momentum”. [ABE_Finance]

Taxonomizing presupposes that students are capable of identifying technical vocabulary in the sources used. In so doing, they make use of the composite nature of technical language and seek to decompose it by foregrounding component parts such as *step*, *dimension*, *phase*, *stage* or *mode*. This has the effect of an unfolding discourse, which can be achieved in two directions: from naming to signalling or from signalling to naming.

The following examples show a reversal of the process of technicality construction:

- [8] long a put option. This strategy is called protective put and limits the maximum loss to the premium [ABE_Finance]
- [9] expatriates normally return back home, this process is called repatriation. Cost of foreign assignments [ABE_Business]

Here the description of the phenomenon in question occurs first, followed by the introduction of the terms, namely *protective put* and *repatriation*. In this pattern, the chains of references are established by means of so-called ‘signalling nouns’ or ‘shell nouns’ (Sing 2013). “A signalling noun is potentially any abstract noun which is unspecific out of context, but specific in context” (Flowerdew 2006: 348). Nouns such as *attitude*, *difficulty* and *process* are pervasive in academic discourse, performing important cohesive functions within or between sentences.

The preceding discussion has clearly shown that these student writers, despite imperfections of language and style, demonstrate the ability to technicalize. However, they do not seem to conceive of themselves as ‘knowledge-makers’.

- [10] In the third part of my paper I will try to define the term hooliganism and give some basic explanations for the phenomenon hooliganism. Finally, in the last part [ABE_Business]
- [11] This section will **define** the term new media and then shortly discuss the change in language culture. [ABE_Marketing]
- [12] The first part basically outlines the aim of the agreement and **defines** the term ‘trade in services’. The second part gives an overview [ABE_Economics]

In contrast to the uses of *to define* as a reporting verb above, these uses highlight that defining proper is restricted to attribution. Although the student writers use the first person pronoun *I*, they do not perform the act of defining. Instead the main emphasis is placed on the paper’s organization, as evidenced by the co-occurrence with so-called frame markers such as *first, then, finally* and *in the last part*, which “can be used to sequence parts of the text or to internally order an argument” (Hyland 2005: 51). This means that, in this particular pattern, the verb serves as a genuine metatextual gloss. Alternatively, the students make use of non-human agents (*section, part*), which also frame, rather than perform, the defining act.

One possible explanation for this pattern is students’ awareness of the genre, i.e., assessed university writing, and its main social purpose, accreditation. The student writers thus show compliance and do what is expected of them. Given that the business students will not have received any writing instruction prior to producing the assignment, they come equipped with lay beliefs about what constitutes good academic writing. The fact that they are still in the process of internalizing their knowledge is shown in the following set of examples:

- [13] Before examining the topic in detail, it is necessary to define the reoccurring terms ‘endorsement marketing’ and ‘celebrity endorser’. AUTHOR (2009: 126) states that endorsement [ABE_Marketing]
- [14] First it is important to define the term that is probably the most important in the whole paper, ethnic minority. Since there is no general [ABE_Marketing]

These examples highlight that the patterning is clearly learning-related. It is evident that these business students are anxious to define key terminology in their writing. At the same time, they are overwhelmed by the task, given the abundance of definitions for relevant concepts. They find it difficult to tolerate this ambiguity, seeking to reduce the complexity to the one-and-only correct definition of a given phenomenon.

- [15] confused and are used in a much diversified sense (AUTHOR 2004: 3). For this reason, it is important to **define** e-commerce in comparison to e-business and illustrate the differences. [ABE_Marketing]
- [16] paper is dealing with cultural aspects of the COUNTRY capital, it seems to be a good approach to **define** more narrowly the use of the word culture. Culture is certainly one of the two or three most complicated words [ABE_Business]

All in all, the business students at hand tend to be familiar with different modes of elaborating on technical terms, thereby establishing chains of reference to build up technicality. The students are also clearly alert to expectations in terms of genre conventions. With a view to complying with the requirements of the writing task, the student writers have internalized that, in academic writing, they are to define key words prior to discussing them in more detail. In the sections that follow, the findings for the modes of technicalizing by means of exemplification and explanation will be presented.

4.2 Technicalizing by means of exemplifying

Technicality may also be constructed when writers elaborate on a term by means of exemplification. The most pervasive metadiscourse devices across the ABE subcorpora are: *such as*, *for example*, *include*, *e.g.* and *for instance*. Crucially, there is considerable variation with regard to the concepts that are exemplified in discourse. As shown in Table 4, the above-mentioned metadiscourse resources co-occur with both field-specific (e.g., *companies*, *products*, *investors*) and 'signalling noun' (e.g., *factors*, *issues*, *problem*) L1 collocates (listed in terms of their frequency, with a minimum occurrence of 4):

Table 4: L1 collocates across subcorpora (minimum frequency of 4 occurrences)

| Corpora | L1 collocates |
|----------------------|----------------------------------------------------------------------------------------------------------------------|
| <i>ABE Business</i> | issues, values, activities, factors, problem, countries |
| <i>ABE Economics</i> | countries, factors, companies, issues, problems, products, services |
| <i>ABE Finance</i> | instruments, investors, areas, failures, products, sectors, banks, countries, factors, institutions, events |
| <i>ABE Marketing</i> | events, attributes, countries, technologies, activities, areas, communication, industries, media, problems, services |

As shown below, exemplifying proper serves to elaborate field-specific or procedural vocabulary. The following examples illustrate this point:

- [17] Patents require inventions. In general, the main characteristics of inventions such as novelty, inventiveness, [ABE_Economics]
- [18] Equity capital markets deal with stocks and derivative instruments, such as forward contracts, futures, and options (AUTHOR 2011: 14). [ABE_Finance]

If field-specific vocabulary is exemplified, the term will most likely have been introduced in the preceding context, which may also involve attribution. What is more, exemplification may also be used to instantiate taxonomies. Following on from the discussion in Section 4.1, the effect of superordination taxonomy on subsection headings is also noticeable here.

- [19] Recruitment [= section heading] Recruitment and selection is more than just hiring people. Pre-recruitment activities include writing a job description as well as developing [ABE_Business]
- [20] Introduction [= section heading] There are many forms of business communication, such as marketing or image communication, crisis communication [ABE_Marketing]

Interestingly, the effect of composition taxonomy on how students elaborate a given term by means of exemplification seems to be weaker than when they technicalize by means of defining resources. More importantly still, exemplification appears to encourage students to opt for chains of references in which ideas are loosely strung together, enforcing a linear, associative structuring rather than building taxonomies.

- [21] There exist different types of tour operators- such as 'niche tour operators', who specialize in certain destinations as **for example** a region [ABE_Marketing]
- [22] (compound options) and options on swaps (swaptions). Compound options, **for instance**, appear in four basic forms: call on a call, call on a put, put on a call and put on a put. [ABE_Finance]

The virtual absence of hierarchical structures in exemplifying chains of reference may also point to another task-related requirement. The SP is to show students' ability to reason. Witness the following corpus examples:

- [23] sponsors, suppliers and employees. Secondary stakeholders are, **for example**, the host community and media. In addition, the external environment [ABE_Marketing]
- [24] The terms have also been used to invoke a variety of associated concepts. **For instance**, exploration has been linked to radical innovation, market expansion [ABE_Finance]

Of course, the student writers cannot be assumed to construct a discursive argument in the examples above. Contrasting different types of stakeholders on the basis of exemplification is conceptually not very challenging. And yet, while still remaining descriptive on the surface, the students show initial signs of truly elaborating rather than simply naming a given term.

4.3 Technicalizing by means of explaining

The corpus analysis revealed that the top five most frequently used code glosses in this category are *means that, i.e., explain, namely* and *refers to*. By and large, the usage patterns retrieved are fairly consistent with those previously identified. In general, it seems that students perceive of the repertoire of discourse cues available to them as alternative modes of construing technicality, while being largely unaware of the more subtle differences between them. In what follows, I will first briefly comment on the established patterns and subsequently pinpoint the causes for variation between them.

In explaining a given term, students also establish chains of reference in which the phenomenon in question is first named and only subsequently explained more accurately, frequently involving attribution (see examples [25] and [26]).

- [25] reduce their carbon footprint, i.e., "the total greenhouse gas emissions caused directly and indirectly [ABE_Business]

- [26] The term mass tourism was born. Mass tourism is characterized by “short-term travel of non-residents [ABE_Economics]

On the level of taxonomies, the effects of both superordination ([27] and [28]) and composition taxonomy ([29] and [30]) are comparable to those identified for elaboration.

- [27] Methods and techniques of traffickers [= section heading] Ways of recruitment [= subsection heading] Trafficking is characterized by different recruitment methods which [ABE_Business]
- [28] Marketing ethics and consumer behavior [= section heading] Marketing ethics refers to the moral principles behind the operation [ABE_Marketing]
- [29] possible identity shifts. One group of the returned expatriates is classified as “identity shifters” which can be explained through their exposure [ABE_Economics]
- [30] consumer manipulation in at least four domains. The first field refers to food products, where, for instance, manufacturers [ABE_Business]

The two modes of expressing technicality are also consistent with regard to the following usage pattern:

- [31] I shall briefly describe the idea of fair trade. Then I shall **explain** and discuss the impact of the social, economic [ABE_Business]
- [32] CSR measures at COMPANY. In a first step, I shall **explain** what drives companies to engage in CSR [ABE_Business]

Just like the verb *to define*, *to explain* is not used as a reporting verb either. Instead, it tends to function as a gloss in the strict sense of a metatextual comment (witness the co-occurrence with frame markers). Once more, the business students are certainly aware of genre conventions and assessment criteria, which, as mentioned in Section 3.1, include presenting an academic argument on the basis of reasoning.

- [33] characteristic is the oligopoly market situation which **means that** the market is dominated by a few companies that fight over market [ABE_Economics]
- [34] developing countries and has not changed after independence. Furthermore, these countries are characterized by top-down management with “authoritarian and paternalistic [ABE_Business]

As these examples indicate, students have their ways of responding to the requirements laid down for SP writing. They demonstrate their understand-

ing of the concept ‘oligopoly market situation’ by providing an explanation using their own words rather than relying on attribution. It remains to be seen, however, whether or not they do so as a result of the nature of the SP “in which writers want to show their instructor what they have read (and hence what they know), but what they think is of minor importance for their paper” (Gruber 2004: 54). While this compliance may well be linked to the SP’s main purpose of accreditation, it may also point to the absence of more assertive writer identities (Sing 2015). Alternatively, both factors could be related to the institutional setting and viewed as the outcome of localized forms of learning. Prior to addressing this important point, I will briefly wrap up the discussion of the results.

Centring on the notion of technicality and relating it to Hyland’s (2005) interpersonal model of academic writing, the present study conceived of technicalizing as a two-stage process, involving the naming of the phenomenon concerned as well as integrating it into a field of knowledge. The corpus analysis revealed that student writers use a number of metadiscourse devices, so-called code glosses, in order to flag the technical nature of their writing. While it is understood that writers technicalize in a variety of ways and using a range of linguistic resources other than the ones examined here, the corpus analysis succeeded in retrieving highly relevant usage patterns that may be fed back into teaching. Cases in point are the different uses and functions of the chains of reference established by the student writers when elaborating, or trying to elaborate, on new terminology. The addition *trying to* is especially noteworthy as the preceding discussion may have suggested that the usage patterns identified are taken at face value. Whether or not the student writers intend to define, explain or exemplify when using the respective metadiscourse devices is a moot point. Owing to the explorative character of the present study, the issue of congruence is an important corollary of the analysis. Further empirical work, possibly involving focus group interviews with the students concerned, is required to establish if the discourse action and the discourse label used to denote the action can, with some justification, be argued to converge.

5 Conclusions: Meeting conceptual challenges

At the outset, the article invoked the challenges of contemporary ESP settings in view of the major changes in HE across Europe. Many of the challenges – linguistic, cultural and conceptual – have regularly dominated specific-purpose instruction. This ongoing process can be accounted for by a widening of the gap between ESP research and practice. To date, teaching and materials development are surprisingly little informed by state-of-the-art research, even in those cases which would lend themselves to direct teaching applications.

What is more, needs analysis retains a strong foothold in more traditionally inclined ESP settings, assuming that once the area and nature of students' difficulties are identified, they can be more easily assisted in mastering the language-based skills or tasks required. In view of the growing complexity of HE, ESP programmes are, however, frequently forced to cater to vastly different literacy demands, spanning the workplace, academia and the corporate world. What is needed is a stronger commitment to the relevance of context in ESP, including an acknowledgement of the situatedness of learning.

Using the field knowledge of an ethnographic researcher, I will now turn to discussing several pedagogical implications for ESP writing instruction. The inconsistency between language and content typical of ESP settings becomes manifest in the writing task as the distinction between 'carrier content' and 'real content', which is central to ESP. In specific-purpose instruction, teaching activities are essentially context-based, implying that content knowledge is used to teach and practise specific language. In the ESP setting at hand, the business topics of the SP are not the aim of the task; rather, these topics constitute the carrier content for which written academic English is the real content, i.e., the explicit teaching target. This is, however, not the only reason why the SP is a genuine showcase for addressing conceptual challenges in ESP writing instruction.

What is more, students seem to struggle with the writing task itself. Given the absence of any formal writing instruction prior to producing the assignment, students come equipped with rather common (mis-)conceptions about academic writing, regarding it as a monolithic entity. Setting the SP as a simple assignment further misconstrues writing as a product. Instead,

it would be useful to foreground the processual character of writing, for example, in setting up a compound assignment, in which individual subgenres are trained separately. In this manner, the different resources that can be used to technicalize in writing could be singled out, paying particular attention to structural issues such as composition taxonomy.

It is on the level of categorization rather than language that students are in need of assistance. Despite imperfections in grammar and style, students' actual conceptual problem lies in taxomizing. As the findings suggest, students rely heavily on the structure already provided to them in the sources they use. Only few student writers have managed to reorganize the material covered in an original way. In view of the patterns identified, the large majority of students seem to address this requirement through retaining the superordination taxonomy. Bold as it may seem, such an assertion is fully supported by the contextual information available to the compiler-cum-analyst of the ABE corpus. On the basis of this inside knowledge, such a student response is hardly surprising. On the contrary, it is a conditioned reflex acquired in a learning environment that rewards learning strategies such as rote learning, which is then simply transferred to writing. The business students at hand emulate what they consider expert models and seek to apply these to their own writing.

This begs the question as to what type of expertise these students are aspiring to. Being susceptible to a technocratic view of the experiential world, it seems likely that these expert genres are realized by text types that are dramatically different from the SP. Thus the discrepancy between the educational genre, the training text used for writing practice and the professional genres used in the workplace could not be more apparent. If assignment writing requires students to "fictionalize" (Pohl 2009), this is particularly valid for this ESP setting. Using the SP with its cultural baggage as a training text only serves to alienate students from the professional target genres. With a view to developing their writing skills as business students, these professional genres need to be firmly anchored in the relevant curricula. Although policy makers pay lip service to the importance of language learning, formal writing instruction in particular has been increasingly marginalized in curricular development. It seems to be the case that specific-purpose instruction is clearly seen as training rather than an integral part of language education.

References

- Ädel, Annelie / Garretson, Gregory, 2006: "Citation practices across the disciplines: The case of proficient student writing." In: Pérez-Llantada, Auría (ed.): *Academic and Professional Communication in the 21st Century: Genres, Rhetoric and the Construction of Disciplinary Knowledge*. Zaragoza: Prensas Universitarias de Zaragoza, 271–280.
- Aktas, Rahime Nur / Cortes, Viviana, 2008: "Shell nouns as cohesive devices in published and ESL student writing." *Journal of English for Academic Purposes* 7(1), 3–14.
- Alsop, Sian / Nesi, Hilary, 2009: "Issues in the development of the British Academic Written English (BAWE) corpus." *Corpora* 4(1), 71–83.
- Basturkmen, Helen, 2010: *Developing Courses in English for Specific Purposes*. Houndmills, Basingstoke: Palgrave Macmillan.
- Belcher, Diane D., 2006: "English for specific purposes: Teaching to perceived needs and imagined futures in worlds of work, study, and everyday life." *Tesol Quarterly* 40(1), 133–156.
- Biber, Douglas / Conrad, Susan / Reppen, Randi, 1998: *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, Douglas / Johansson, Stig / Leech, Geoffrey / Conrad, Susan / Finegan, Edward, 1999: *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber, Douglas, 1988: *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas, 2006: "Stance in spoken and written university registers." *Journal of English for Academic Purposes* 5(2), 97–116.
- Bramki, Doudja / Williams, Ray, 1984: "Lexical familiarisation in economics text, and its pedagogic implications in reading comprehension." *Reading in a Foreign Language* 2(1), 169–181.
- Chang, Peichin / Schleppegrell, Mary, 2011: "Taking an effective authorial stance in academic writing: Making the linguistic resources explicit for L2 writers in the social sciences." *Journal of English for Academic Purposes* 10(3), 140–151.

- Charles, Maggie, 2006: "Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines." *English for Specific Purposes* 25(3), 310–331.
- Dahl, Trine, 2004: "Textual metadiscourse in research articles: a marker of national culture or of academic discipline?" *Journal of Pragmatics* 36(10), 1807–1825.
- Dudley-Evans, Tony / St John, Marie Jo, 1998: *Developments in ESP. A Multi-Disciplinary Approach*. Cambridge: Cambridge University Press.
- Ehlich, Konrad, 2003: „Universitäre Textarten, universitäre Struktur.“ In: Ehlich, Konrad / Steets, Angelika (eds): *Wissenschaftlich schreiben – lehren und lernen*. Berlin: De Gruyter, 13–29.
- Flowerdew, John, 2006: "Use of signalling nouns in a learner corpus." *International Journal of Corpus Linguistics* 11(3), 345–362.
- Flowerdew, Lynne, 2004: "The argument for using English specialized corpora to understand academic and professional language." In: Connor, Ulla / Upton, Thomas A. (eds): *Discourse in the Professions. Perspectives from Corpus Linguistics*. Amsterdam: Benjamins, 11–37.
- Flowerdew, Lynne, 2005: "An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies." *English for Specific Purposes* 24(3), 321–332.
- Gavioli, Laura, 2005: *Exploring Corpora for ESP Learning*. Amsterdam: Benjamins.
- Gee, James Paul, 1996: *Social Linguistics and Literacies. Ideology in Discourses*. London: Taylor & Francis.
- Gnutzmann, Claus / Rabe, Frank, 2014: "'Theoretical subtleties' or 'text modules'? German researchers' language demands and attitudes across disciplinary cultures." *Journal of English for Academic Purposes* 13(0), 31–40.
- Gruber, Helmut, 2004: "Scholar or consultant? Author-roles of student writers in German business writing." In: Ravelli, Louise J. / Ellis, Robert (eds): *Analyzing Academic Writing. Contextualized Frameworks*. London: Continuum, 45–65.
- Gruber, Helmut, 2006: *Genre, Habitus und wissenschaftliches Schreiben. Eine empirische Untersuchung studentischer Texte*. Lit: Wien.

- Handford, Michael, 2010: "What can a corpus tell us about specialist genres?" In: O'Keeffe / McCarthy (2010). 255–269.
- Hutchinson, Tom / Waters, Alan, 1987: *English for Specific Purposes*. Cambridge: Cambridge University Press.
- Hyland, Ken, 2005: *Metadiscourse. Exploring Interaction in Writing*. London: Continuum.
- Hyland, Ken, 2010: "Metadiscourse: Mapping interactions in academic writing." *Nordic Journal of English Studies* 9(2), 125–143.
- Hyland, Ken, 2013: "ESP and writing." In: Paltridge, Brian / Starfield, Sue (eds): *The Handbook of English for Specific Purposes*. Malden, Mass: Wiley-Blackwell, 95–113.
- Kruse, Otto, 2006: "The origins of writing in the disciplines. Traditions of seminar writing and the Humboldtian ideal of the research university." *Written Communication* 23(3), 331–352.
- Lee, David, 2001: "Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle." *Language Learning & Technology* 5(3), 37–72.
- Lillis, Theresa / Scott, Mary, 2007: "Defining academic literacies research: Issues of epistemology, ideology and strategy." *Journal of Applied Linguistics* 4(1), 5–32.
- Martin, James R., 1991: "Nominalization in science and humanities: Distilling knowledge and scaffolding text." In: Ventola, Eija (ed.): *Functional and Systemic Linguistics. Approaches and Uses*. Berlin: Mouton de Gruyter, 307–339.
- McEnery, Tony / Xiao, Richard / Tono, Yukio, 2006: *Corpus-Based Language Studies. An Advanced Resource Book*. London: Routledge.
- Meyer, Charles F., 2002: *English Corpus Linguistics*. Cambridge: Cambridge University Press.
- Nelson, Mike, 2010: "Building a written corpus." In: O'Keeffe/McCarthy (2010): 53–65.
- Nesi, Hilary / Gardner, Sheena, 2012: *Genres across the Disciplines. Student Writing in Higher Education*. Cambridge: Cambridge University Press.
- O'Keeffe, Anne / McCarthy, Michael (eds), 2010: *The Routledge Handbook of Corpus Linguistics*. London: Routledge.

- Paltridge, Brian, 1996: "Genre, text type, and the language learning classroom." *ELT Journal* 50(3), 237–243.
- Paquot, Magali, 2010: *Academic Vocabulary in Learner Writing. From Extraction to Analysis*. London: Continuum.
- Peters, Pam / Smith, Adam / Middledorp, Jenny / Karpin, Anne / Sin, Samantha / Kilgore, Alan, 2014: "Learning essential terms and concepts in Statistics and Accounting." *Higher Education Research & Development* 33(4), 742–756.
- Pohl, Thorsten, 2009: *Die studentische Hausarbeit. Rekonstruktion ihrer ideen- und institutionsgeschichtlichen Entstehung*. Heidelberg: Synchron.
- Ravelli, Louise J., 2004: "Signalling the organization of written texts: hyper-Themes in management and history essays." In: Ravelli, Louise J. / Ellis, Robert (eds): *Analyzing Academic Writing. Contextualized Frameworks*. London: Continuum, 104–130.
- Reppen, Randi, 2010: "Building a corpus. What are the key considerations?" In: O'Keeffe / McCarthy (2010): 31–38.
- Samraj, Betty, 2013: "Form and function of citations in discussion sections of master's theses and research articles." *Journal of English for Academic Purposes* 12(4), 299–310.
- Scott, Mike, 2012: *WordSmith Tools. Version 6*. Liverpool: Lexical Analysis Software.
- Sing, Christine S., 2013: "Shell noun patterns in student writing in English for specific academic purposes (ESAP)." In: Granger, Sylviane / Gilquin, Gaëtanelle / Meunier, Fanny (eds): *Twenty Years of Learner Corpus Research. Looking back, Moving ahead*. Louvain-la-Neuve: Presses universitaires de Louvain, 411–422.
- Sing, Christine S., 2015: "'My own summary is definitely less scientific but more easy to grasp': ESP writing and disciplinary identity." *Fremdsprachen Lehren und Lernen* 44(1), 82–96.
- Tardy, Christine M., 2012: "Writing and Language for Specific Purposes." In: Chapelle, Carol A. (ed.): *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell, 6266–6274.
- Tognini-Bonelli, Elena, 2008: "Corpora and LSP. Issues and implications." In: Torsello, Carol Taylor / Ackerley, Katherine / Castello, Erik (eds): *Corpora for University Language Teachers*. Bern: Peter Lang, 31–47.

- Tribble, Christopher, 2002: "Corpora and corpus analysis: new windows on academic writing." In: Flowerdew, John (ed.): *Academic Discourse*. Harlow: Longman, 131–149.
- Trinh, Thi Giang Thanh, 2011: *An Adult's Language Variability and Development*. (MA Thesis). Rijkuniversiteit Groningen: Groningen.
- Veel, Robert, 1997: "Learning how to mean – scientifically speaking: apprenticeship into scientific discourse in the secondary school." In: Christie, Frances / Martin, James R. (eds): *Genre and Institutions. Social Processes in the Workplace and School*. London: Cassell, 161–195.
- Wignell, P. / Martin, James R. / Eggins, S., 1993: "The discourse of geography. Ordering and explaining the experiential world." In: Halliday, Michael A. K. / Martin, James R. (eds): *Writing Science. Literacy and Discursive Power*. London: Falmer Press, 151–184.
- Williams, Geoffrey, 2002: "In search of representativity in specialised corpora: categorisation through collocation." *International Journal of Corpus Linguistics* 7(1), 43–64.
- Woodward-Kron, Robyn, 2008: "More than just jargon – the nature and role of specialist language in learning disciplinary knowledge." *Journal of English for Academic Purposes* 7(4), 234–249.

Hans Malmström¹, Diane Pecorari², Magnus Gustafsson¹
¹Chalmers University of Technology, Gothenburg, Sweden;
²Linnaeus University, Växjö, Sweden

Coverage and development of academic vocabulary in assessment texts in English Medium Instruction¹

Abstract

English: This paper is centred in the context of English Medium Instruction (EMI) and is primarily concerned with advanced students' productive knowledge of English academic vocabulary, widely regarded as a crucial dimension of successful academic communication. The study problematizes the claim that EMI is beneficial for students' development of academic vocabulary knowledge. The investigative context is a technical university in Sweden where all degree programmes at graduate level use English as the medium of instruction. The corpus data include texts (n=80, approx. 720,000 words) produced by Master of Science students in their first and second year of study, written by home and international students. The study, using the Academic Vocabulary List (Gardner/Davis 2014), sets out to answer three research questions relating to knowledge and development of academic vocabulary in EMI: 1. What is the lexical coverage of advanced (master's) level student writing, i.e., what proportion of words in students' texts is academic? 2. Are home students and international students (all of whom have English as a foreign language) comparable in terms of their productive academic vocabulary knowledge? 3. Does students' productive knowledge of academic words appear to develop during their studies? The results of the investigation can be summarized as follows: In the corpus as a whole, academic vocabulary items account for approximately 20% of all tokens. This figure is considerably higher than that found in many earlier studies. There are no significant differences between home and international students in any of the measures of vocabulary used (pertaining to lexical sophistication and diversity). Finally, the findings regarding lexical development across years of study are somewhat mixed; however, the overall picture presented by the various measures is one of significant but very modest gains in some areas and none in others. These

1 This research was supported by the Swedish Research Council (grant number: VR2013-2373).

findings call into question the actual effectiveness of EMI for academic vocabulary development. The overall contribution of the paper is an important step towards a more comprehensive understanding of what expectations we may reasonably have of the development of English language competency in EMI.

German: Die Fähigkeit, wissenschaftlichen Wortschatz aktiv zu gebrauchen, gilt als eine entscheidende Komponente erfolgreicher Wissenschaftskommunikation. Dabei herrscht die Annahme vor, dass English Medium Instruction (EMI) einen positiven Einfluss auf die Entwicklung des wissenschaftlichen Wortschatzes Studierender habe. Im vorliegenden Beitrag werden die Ergebnisse einer Studie vorgestellt, die diese Annahme kritisch beleuchtet. Die Studie wurde an einer Technischen Universität in Schweden durchgeführt, an der alle Masterstudiengänge auf Englisch unterrichtet werden. Das Korpus umfasst Texte (n=80, ca. 720.000 Wörter), die von schwedischen und internationalen Master-of-Science-Studierenden in ihrem ersten und zweiten Studienjahr verfasst wurden. Die Studie, für die die Academic Vocabulary List (Gardner/Davis 2014) genutzt wurde, geht drei Forschungsfragen nach, in deren Mittelpunkt der Umfang und die Entwicklung wissenschaftlichen Wortschatzes in EMI-Kontexten stehen: 1. In welchem Umfang beherrschen fortgeschrittene Masterstudierende den wissenschaftlichen Wortschatz, d. h., welcher Anteil der Wörter in den studentischen Texten ist wissenschaftlich? 2. Sind schwedische und internationale Studierende mit Englisch als L2 vergleichbar in ihrer Kompetenz, wissenschaftlichen Wortschatz aktiv zu gebrauchen? 3. Entwickelt sich dieser Wortschatz während des Studiums erkennbar weiter? Die Ergebnisse der Untersuchung lassen sich wie folgt zusammenfassen: Im Korpus hat wissenschaftliches Vokabular einen Anteil von ungefähr 20 % aller Tokens und damit einen deutlich höheren Anteil, als in vielen früheren Studien nachgewiesen werden konnte. Im Gebrauch wissenschaftlichen Wortschatzes (sowohl hinsichtlich dessen lexikalischen Anspruchs als auch dessen Differenziertheit) lassen sich zwischen schwedischen und internationalen Studierenden keine signifikanten Unterschiede feststellen. Die Befunde zur Entwicklung des Wortschatzes während des Untersuchungszeitraums sind ambivalent; insgesamt lässt sich jedoch ein moderater Zugewinn in einigen Bereichen feststellen, wohingegen in anderen Bereichen kein Fortschritt zu verzeichnen ist. Diese Ergebnisse geben Anlass zu berechtigten Zweifeln an der Annahme, dass sich mit EMI Wortschatz effizient erweitern lasse. Die Studie leistet somit einen wichtigen Beitrag zur realistischeren Einschätzung der Möglichkeiten, die EMI für die Förderung der englischen Sprachkompetenz bietet.

1 Introduction

The broader context of this study of academic vocabulary knowledge is English Medium Instruction (EMI). For the purposes of this paper, this may be defined as the deliberate use of English (typically as a result of an official educational policy) to engage students communicatively in academic study, i.e., by asking students whose first language is not English to read, write, speak and listen in English rather than using their first language (cf. Coleman 2006; Dearden 2014). While in some contexts a distinction between EMI and Content and Language Integrated Learning (CLIL) is becoming increasingly difficult to maintain, EMI and CLIL should not be considered the same thing. The difference might be one of degree but the defining factor distinguishing the two is the extent to which students receive deliberate language education (CLIL) as opposed to mere immersion (EMI) (Marsh 2005; Lasagabaster 2008; Gustafsson et al. 2011; Gustafsson/Jacobs 2013).

EMI is a “rapidly growing global phenomenon” (Dearden 2014: 2) and a number of different “drivers of the Englishization” (Coleman 2006: 4) of higher education exist. In this respect, there is a widespread assumption articulated to differing extents by the various stakeholders that students enhance their academic as well as general English competency as a result of studying in EMI contexts. However, this is only an assumption (like many other assumptions about EMI reviewed by Dearden), and it has yet to be confirmed by empirical research. Dearden (2014: 2) highlights the fact that “we are quite some way from a ‘global’ understanding” of EMI and notes that there is an “urgent need for a research-driven approach [...] which measures the complex processes involved in EMI”, for example, the conditions for “the acquisition of English proficiency”. In other words, research that confirms, refutes, or at least problematizes the claim that EMI is beneficial for students’ development of English language skills is called for.

In this paper, we are concerned with a single but crucial dimension of English proficiency development in EMI contexts, namely academic vocabulary knowledge, and, more specifically, students’ productive knowledge of academic words as reflected in their writing. Our starting point is the widely held claim (see, e.g., Stæhr 2008) that there is a correlation between knowing many words, i.e., having good vocabulary knowledge, and overall communicative competence. Milton (2010: 212) notes that “vocabulary

knowledge is key to both comprehension and communicative ability”, and Laufer & Nation (1999: 34) talk of the “enabling” function of vocabulary vis-à-vis other dimensions of communication. In academic discourse, the same kind of correlation obtains between academic words and general academic literacy (see, e.g., Corson 1997; Coxhead 2000; Milton 2010). These correlations are supported by research indicating that understanding virtually all the input words is fundamental to comprehension in any kind of communicative situation (Coady/Huckin 1997; Schmitt 2000; Nation 2001; Bogaards/Laufer 2004). It has been suggested that in excess of 95% of the running words must be understood for “adequate comprehension” to be possible in connection with reading and listening, and for optimal comprehension as much as 98% of the words should be known (Nation 2001; 2006). It seems evident that, if a large vocabulary is needed for reading comprehension, it must be of at least equal importance for the productive assessment tasks in an EMI environment.

In this respect, it is reasonable to ask what might actually be expected of students in EMI in terms of academic vocabulary knowledge. Numerous methods of measuring vocabulary knowledge exist, including self-assessment scales and definition tasks among other measures of receptive knowledge. However, in most EMI settings, students need not only to *understand* English in lectures, textbooks, etc., but also to *produce* it in assessment tasks, so their productive vocabulary knowledge is of interest. Our first research question is therefore:

1. What is the lexical coverage of academic vocabulary in student writing, i.e., what proportion of words in students’ texts is academic?

A second perspective we want to explore concerning students’ productive knowledge of academic words relates to the Englishization of higher education as a result of globalization/student mobility. One effect of globalization and the concomitant proliferation of EMI has been a rise in EMI outside of the traditionally English-speaking world, in the “Expanding Circle” (Kachru 1992).² In Sweden, where this study is set, as in many other

2 Kachru (1992) identifies three English “Circles”. The Inner Circle is represented by countries like the United Kingdom, USA, Canada, Australia, and New Zealand, i.e., countries where English is the primary language and the native language of most people. In Outer Circle countries (such as Singapore, India,

Expanding Circle countries, a result has been that significant numbers of ‘international’ students (students domiciled outside Sweden) are now studying alongside ‘home’ students (domiciled in Sweden with an almost 100% Swedish language background and English as their first foreign language at school), creating an international mix and a multilingual learning environment in what used to be a linguistically relatively homogenous teaching/learning environment. The (English) language demands placed on international students and home students in the EMI classroom are naturally the same, but our overall knowledge of how this more diverse group of students actually performs linguistically vis-à-vis home students is very limited.

It is easy to problematize international students and find isolated and categorical statements that speak in general negative terms about international students’ shortcomings with regard to English proficiency. For example, in response to a survey about attitudes towards English in higher education administered to university teachers in Sweden (reported in Pecorari et al. 2011), one teacher offered the following comment:

“English texts, especially academic English texts that we use, are demanding for students. The same is true for writing in English. The problem is especially pronounced for our foreign students who are particularly challenged to write acceptable English.”

However, with very few exceptions (see, e.g., Jochems et al. 1996), the research available concerning the overall academic performance and linguistic ability of international students in relation to home students is restricted to Inner Circle countries (see, e.g., Warwick 2006; Carroll/Ryan 2005; Morrison et al. 2005), meaning that our knowledge of how international students compare to home students in Expanding- and Outer Circle countries is almost non-existent.

A first step towards a more comprehensive understanding of the relationship between home students and international students as regards their

Pakistan, Bangladesh, Nigeria, and Kenya) English is not the native language but firmly established as the *lingua franca* in most areas of society and typically has the status of “official language”. Finally, in countries in the Expanding Circle (the Nordic countries are a case in point), although often widely used for the purpose of international communication (e.g., in much business communication), English is not an official language or the language used in government.

English proficiency is to establish whether home students and international students in EMI environments have equally sized academic English vocabularies, clearly a pertinent investigation given the centrality of academic words to comprehension. This leads us to pose the second of our three research questions:

2. Are home students and international students (all L2 users of English) comparable in terms of their productive vocabulary knowledge?

Finally, this paper directly addresses the supposition that EMI is conducive to developing students' English language skills, focusing on vocabulary gains. There is support in the literature that incidental exposure to English vocabulary in study contexts does lead to positive lexical gains over time (Huckin/Coady 1997; Laufer/Hulstijn 2001); however, the vast majority of this research has focused on learners' receptive knowledge of vocabulary and there is a dearth of research concerned with students' productive lexical knowledge (cf. Durrant 2014). In addition, the bulk of the literature on vocabulary acquisition, and indeed second-language acquisition generally, has focused on learners who are much less advanced than those who are in a position to undertake study at university through the medium of their L2.

Laufer (1994), the most widely cited study on productive vocabulary development available, looked at "changes in the productive lexicon of advanced second language learners' writing over a period of one academic year" (1994: 21) using a construct she calls "lexical quality". Focusing on writing compositions produced in a controlled environment, and by drawing on two basic types of analytical measures, a frequency profile and a Type-Token Ratio, Laufer found no significant lexical gains with regard to general high-frequency words, but there were significant gains for words from the University Word List (Xue/Nation 1984), i.e., Laufer's measure used for academic words, and for words of lower frequency (words beyond the 2,000 most common). With respect to lexical variation, as measured by the Type-Token Ratio, no significant longitudinal gains were recorded for any type of lexis.

Additional longitudinal perspectives on the development of productive lexical knowledge have been provided by more recent research from Australia. This research comes to a different conclusion. Knoch et al. (2015) investigated to what extent the writing of English L2 students developed

positively over three years. Development was measured by looking at a set of discourse-analytic measures, among which was lexical complexity, operationalized with reference to the proportion of words from the Academic Word List (Coxhead 2000), lexical sophistication and lexical richness. For all three measures of lexical complexity, the differences between the first and the second writing collection point fell well short of statistical significance, suggesting that there is little support for the notion that studying in an EMI context has a significantly positive effect on students' productive knowledge of academic lexis.

The setting provided by Laufer's study as well as the study reported by Knoch et al. is in many ways different from EMI contexts in Europe outside Great Britain and Ireland and elsewhere today. It is noteworthy, for example, that the participants in Laufer's investigation were all language learners and therefore possibly 'primed' to attend to linguistic matters like vocabulary, unlike the vast majority of EMI students enrolled in subject courses or degree programmes where there is little or no attention devoted to language *per se*. In addition, it seems unfair to compare the EMI situation of English L2 students studying in English Inner Circle countries with the situation in Expanding- or Outer Circle countries; the complete immersion in an Inner Circle environment presumably affords many more opportunities for engagement with English vocabulary (academic or otherwise). The issue of lexical development in the EMI context facing a great number of students in Expanding- or Outer Circle countries must therefore be investigated independently of such research in Inner Circle countries. Thus, the third research question that this study asks is:

3. Does students' productive knowledge of academic words appear to develop during their studies in an EMI context?

2 Data collection and methods

This section describes the context of the investigation, the data collection and the analytic procedures adopted.

2.1 Study context

Our study is set at a prestigious technical university in Sweden where policy stipulates that all degree programmes at master's level use English as the medium of instruction. The university has approximately 11,000 students, 2,400 of whom are enrolled in one of 41 master's programmes. While we have no reliable record of the first language of the students, the majority of the home students are Swedish L1 speakers and, as far as we have been able to ascertain, none of the international students' whose data were included were domiciled in an English L1 country.

2.2 Data

The data for this study is a small student text corpus made up of 80 texts in English (totalling just over 720,000 running words) written by Master of Science (MSc) students from four different disciplines (applied physics, chemical engineering, chemistry, and mechanical engineering) and from the first and second year of study at the master's level. A total of 30 texts, comprising approximately 115,000 running words, were primarily technical- or mini-project reports written as part of students' course work at some point during the first year. The 50 second-year texts, comprising approximately 605,000 words, were full-length master's theses written during the last term of a two-year study programme. At this university, master's theses are generally reports of project work. Thus, although there is greater variation in the first-year corpus in terms of the assignment set, the two sub-corpora can be regarded as broadly similar in terms of text type.

Because virtually all written course work at this university is done in groups of two or more students, we were unable to obtain first-year and second-year texts from the same author or team of authors. Therefore, 'development' of academic vocabulary refers to change across levels of study rather than change in individual students. In all cases, all authors were either Swedish or international; texts with a mixed authorship with regard to national origin were excluded from the sample in order to enable the comparison between Swedish and international MSc students regarding their productive knowledge of English academic vocabulary.

2.3 Analytical procedure

To address the three research questions, we used several measurements, following Milton (2009) in distinguishing between two basic constructs: lexical sophistication and lexical diversity. The former refers to the extent to which more or less common words are used: in Milton's example, the difference between *the cat sat on the mat* and *the feline reposed on the antique Persian rug* (2009: 131). The latter refers to the extent to which the same or different words are used.

One measure of lexical sophistication is the presence of academic vocabulary. Over the years, several descriptions and compilations of academic vocabulary have been developed (see Gardner/Davies 2014; and Charles/Pecorari 2016: 109 ff. for a discussion). Until recently, the most widely used list, for both teaching and research, has been Coxhead's (2000) Academic Word List (AWL). As a result of this wide use, a number of limitations of the AWL have been identified,³ and these limitations have provided the impetus for the newer Academic Vocabulary List (Gardner/Davies 2014); this list is adopted as a basis for what counts as academic vocabulary in the present study.⁴

The AVL was developed from a 120-million-word academic sub-corpus (featuring texts with a heavy emphasis on journal articles from across nine different academic disciplines) taken from the 425-million-word Corpus of Contemporary American English (COCA 2015). Rather than using an existing word list to exclude general vocabulary (the way Coxhead used the General Service List by West 1953), a criterion for relative frequencies was developed, such that words were considered to be part of an academic core if they occurred in the academic corpus with a frequency 50% greater than in the general, non-academic reference corpus (the non-academic portion of COCA). Words needed to be represented at or above a threshold

3 It is not within the scope of this article to criticize the AWL, but Hyland & Tse (2007) and Gardner & Davies (2014) both offer a comprehensive account of the perceived problems with the AWL.

4 The pedagogic utility of lists of general academic vocabulary is widely accepted and Gardner & Davies (2014: 2) note several areas in which such lists are purposeful (see also Schmitt/Schmitt 2014). It should be stressed, however, that pedagogic utility is not a central concern in this study (though see Section 6 where various didactic implications are discussed).

frequency in seven of the nine subject areas, and both a criterion for dispersion and relative frequencies were implemented to exclude words which have a particular affinity with one or a few subject areas (thus allowing for a distinction to be made between core/general academic vocabulary and subject-specific/technical vocabulary). As a result of this process, the AVL consists of 3,015 words (lemmas) that occur across a wide range of academic disciplinary areas more frequently than they do in general discourse. Table 1 includes examples of words from the AVL.

Table 1: Most and least frequent words (lemmas) in the Academic Vocabulary List

| AVL words 1–10 | | AVL words 3006–3015 | |
|----------------|--------------|---------------------|-----------------|
| study n. | however adv. | unusable adj. | imprimatur n. |
| group n. | research n. | unpalatable. adj. | coherently adv. |
| system n. | level n. | causally adv. | component n. |
| social adj. | result n. | prioritization n. | tangential adj. |
| provide v. | include v. | overemphasis n. | relevancy n. |

Because the AVL contains words that are more common in academic than general discourse, they are in that sense ‘advanced’ vocabulary. Two measures of lexical sophistication used in the present study were based on the AVL: the proportion of coverage afforded by the AVL, and the number of types from the AVL. However, AVL items vary greatly in frequency, and so not all are equally ‘advanced’. For example, the most frequent word on the AVL is *study*, and it also is among the first 1,000 general words by frequency. We thus distinguish between the first 500 words on the AVL and the rest of the list.

The second construct we were interested in was lexical diversity. One of the oldest (Johnson 1939; Mann 1944) and most common measures of lexical diversity is the Type-Token Ratio (TTR). However, the TTR has limitations (Malvern/Richards 2013; Vermeer 2000), including the fact that it is sensitive to text length (Holmes 1994; Baker 2006). While there are no entirely unproblematic measures of lexical diversity, the Guiraud Index (Guiraud 1954) and the Advanced Guiraud (Daller et al. 2003) compensate for the TTR’s sensitivity to length and perform more reliably, and were thus adopted here. The former is calculated by dividing the number of types in a text by the square

root of the number of tokens, and the latter uses the same calculation after very common words have been excluded (in this case, the first thousand most frequent words in the BNC and COCA corpora, as provided by Paul Nation's Range files (Range 2015)). Because the advanced measure eliminates the most frequent types, there is a basis for considering it to be an indicator of lexical sophistication as well as diversity, as Daller & Xue (2009) do. Because there is no established baseline for these measures in texts of the type analysed here, they are primarily of value in this study in the two between-group comparisons.

Once the data had been collected, the texts were cleaned⁵, converted into text files and processed using AntWordProfiler (Anthony 2015) to determine the frequency of AVL words. In vocabulary profiling (see, e.g., Laufer 1994; Nation 2006) the number of words (tokens) in the texts is counted and the words' distribution relative to pre-established lists is calculated. In this case we used two lists: one list consisting of the 500 most frequent types in the AVL (called AVL 500 here), and one comprising the remaining less frequent lemmas from the AVL (AVL 501+).

Two additional procedures were needed to enable a comparison of the present findings with the Gardner & Davies (2014) study (to the best of our knowledge, the only study based on the AVL to date). Unlike the COCA academic sub-corpus used in that study, our corpus is untagged, meaning that it does not distinguish between words like *study*, n., which is on the AVL, and *study*, v., which is not. To estimate the effect of this difference, a manual search was done among the first 300 words of the AVL for candidates for overcounting (such as *study*, v.). A second procedural issue is that the profiling tools used in the study may have idiosyncrasies which cause them to perform somewhat differently. To estimate the extent of this effect, samples of each corpus text were submitted individually to the lexical profiling tool on Mark Davies' Word and Phrase website (Davies 2015) which analyses the first 1,000 words from each text.

5 The following features were removed from the texts: extensive visual information in the form of tables and figures (table and figure captions were left in); all equations/formulae and/or parts thereof, unless some element featured as a syntactic constituent in which case it was treated as technical vocabulary; finally, all tables of contents, reference sections and acknowledgement sections were also removed.

Finally, where relevant, SPSS was used to test the significance of between-group differences. Because a random distribution could not be assumed, a non-parametric test was appropriate (Turner 2014). The independent samples Kruskal-Wallis test was used, and differences were considered significant when $p < .05$.

3 Results

The research questions guiding this investigation related to 1) academic vocabulary coverage; 2) comparisons between home and international students; and 3) comparisons between first-year and second-year texts.

The results of the AVL profiling (Fig. 1), relating to the lexical sophistication of the texts, showed that 19.3% of the tokens in the corpus are academic words. This is a considerably higher proportion of academic vocabulary than previous studies have shown. The most relevant earlier study is Gardner & Davies (2014), who found that the AVL gave coverage of the academic sections of the COCA and BNC in the vicinity of 14%. To estimate the extent to which procedures may have contributed to these different results, two additional analyses were conducted.

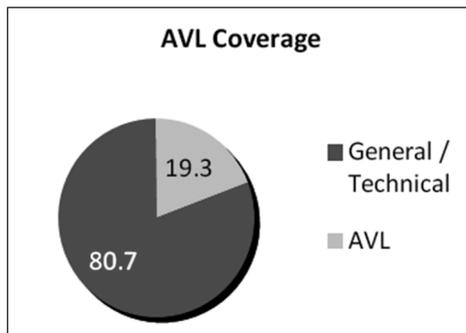


Fig. 1: Academic vocabulary coverage in the MSc writing corpus

To account for the effect of the untagged corpus, a manual search was done among the first 300 words of the AVL for candidates for overcounting (such as *study*, *v.*). A total of 2,028 such forms were identified, or approximately 1.4% of the total corpus size. The effect of the untagged corpus is therefore real but relatively modest. When the first 1,000 words of each text were

submitted to the Word and Phrase profiling tool (Davies 2015), an average of 23% of the tokens came from the AVL. It is therefore reasonable to conclude that academic words do in fact make up approximately 20% of our corpus and that procedural issues play a relatively small role. Other explanations for the difference between these findings and earlier studies are taken up in Section 4 below.

A further measure of lexical sophistication was the proportion of infrequent AVL words. In the COCA academic corpus, the 500 most frequent of the 3,015 AVL types (i.e., 17%) account for 74% of all of AVL tokens. The average number of tokens representing each type in the 1–500 list was 14 times greater than on the 501+ list (22,599 versus 1,480) (see Fig. 2). As Figure 2 shows, the figures for the present corpus are comparable: 70.5% of the AVL tokens come from the first 500 words of the AVL, with only 29.5% coming from the remainder of the list, and the average type on the 1–500 list had nearly 13 times as many tokens as the average type on the 501+ list.

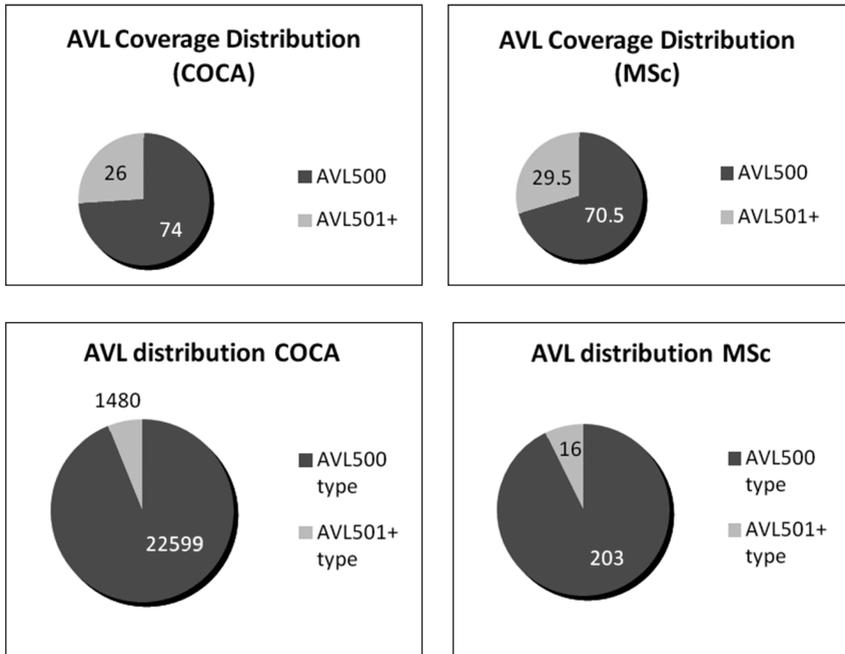


Fig. 2: Academic vocabulary diversity and sophistication in the COCA and MSc writing corpora

The second research question addressed the relative productive vocabulary of native Swedish and international students. Given that large numbers of international students are a relatively recent phenomenon in Swedish university classrooms, there is a need to understand whether the English skills of this new constituency permit them to participate in EMI on the same terms as their local counterparts. Measures for lexical sophistication and lexical diversity were therefore considered for the two groups separately.

As Table 2 shows, for all measures, the differences were small, and indeed none of them was statistically significant. This measure thus suggests, encouragingly, that this relatively new student group is able to take on EMI education on a level playing field with their Swedish peers, at least when assessed on the basis of the productive vocabulary knowledge.

Table 2: AVL distribution for home versus international students

| Measure | Overall | Home | International |
|------------------------------|---------|-------|---------------|
| AVL coverage (entire list) | 19.3% | 18.9% | 19.9% |
| AVL 500 coverage | 13.6% | 13.2% | 14.1% |
| AVL 501+ coverage | 5.74% | 5.7% | 5.8% |
| AVL Types per 100 words | 5.69 | 5.88 | 5.40 |
| AVL 500 types per 100 words | 3.60 | 3.76 | 3.37 |
| AVL 501+ types per 100 words | 2.08 | 2.12 | 2.03 |
| Guiraud Index | 14.2 | 14.01 | 14.48 |
| Advanced Guiraud | 8.68 | 8.33 | 9.20 |

Since language development is one of the reasons offered for implementing EMI, it would be reasonable to think that students' vocabulary – particularly academic vocabulary – develops during their course of study. The third research question was therefore whether the second-year texts showed greater lexical sophistication and variation than the first-year texts.

The measures of lexical sophistication failed to reflect gains between the two groups. As Table 3 shows, a small increase was found for the overall coverage afforded by the AVL, from 19.0% to 19.5%. On closer investigation, this increase is seen to be driven by increased usage of the 500 most frequent AVL items, from 12.9% to 13.9%, which was in turn offset by a

slight decrease in coverage from the remainder of the list. However, none of these changes were significant.

The normalized frequencies of types from the entire AVL, the most frequent 500 items and the less frequent items, all showed a significant decrease in the second year ($p=.000$). This (in combination with the change in the Guiraud Index and Advanced Guiraud, see below) is an indication that the greater diversity of lexis came either from general (i.e., non-academic) vocabulary in the less frequent range, and/or from technical terminology, rather than increased usage of general academic vocabulary.

In terms of variation, both the Guiraud Index and the Advanced Guiraud showed a modest but significant ($p=.000$) trend toward greater lexical variation in the year-two texts (see Table 3). Because it excludes the first thousand most commonly used words, the Advanced Guiraud also reflects lexical sophistication.

Table 3: AVL distribution in first- versus second-year texts

| Measure | Overall | Year 1 | Year 2 |
|------------------------------|---------|--------|--------|
| AVL coverage (entire list) | 19.3% | 19.0% | 19.5% |
| AVL 500 coverage | 13.6% | 12.9% | 13.9% |
| AVL 501+ coverage | 5.74% | 6.0% | 5.6% |
| AVL Types per 100 words | 5.69 | 7.40 | 4.73 |
| AVL 500 types per 100 words | 3.60 | 4.85 | 2.90 |
| AVL 501+ types per 100 words | 2.08 | 2.55 | 1.82 |
| Guiraud Index | 14.2 | 12.96 | 14.90 |
| Advanced Guiraud | 8.68 | 7.46 | 9.37 |

4 Discussion

Section 3 presented findings that were in some ways unexpected, and thus merit further discussion.

4.1 Academic vocabulary coverage

The finding that approximately 20% of this corpus consisted of academic words contrasts strikingly with the much lower figures in previous studies. For example, Coxhead (2000) and Hyland & Tse (2007) found that

the Academic Word List afforded about 10% coverage of their respective corpora. A 20%-figure was approached only in Chung & Nation's (2003) study concerning Applied Linguistics textbooks. As noted above, methodology accounts for only a minor part of the difference, and it is interesting to speculate as to what may account for the rest.

One likely explanation is the composition of the corpus; the proportion of academic vocabulary varies according to text type (Chung/Nation 2003; Li/Qian 2010) and academic discipline (Chung/Nation 2003; Coxhead 2000; Hyland/Tse 2007). However, no other study based on a fully comparable corpus exists. Engineering was one of the fields investigated by Hyland & Tse (2007) and Mudraya (2006), but the former corpus contained a mix of text types, while the latter consisted of textbooks and provided no overall academic vocabulary coverage figure. Thus, while it is probable that academic subject area and text type explain some of the difference between the present findings and earlier ones, it is not possible to ascertain the extent of their influence.

A second explanation lies in the use here of the Gardner & Davies (2014) AVL, while previous investigations have employed Coxhead's (2000) AWL. The AVL's ability to represent core academic vocabulary better than the AWL has been demonstrated empirically. Gardner & Davies (2014) profiled the academic sections of the BNC and COCA with both the AWL and the AVL, using word families to enable a comparison with the AWL, and found that the top AVL 570 word families provided nearly twice the coverage with respect to the AWL (13.8% versus 7.2% in COCA; 13.7% versus 6.9% in the BNC). In this light, the fact that the present study found approximately twice as much academic vocabulary as earlier studies is unsurprising; indeed an investigation of the present MSc writing corpus (with some modifications) found that the AWL provided just under 10% coverage (Gustafsson/Malmström 2013).

A further question is why AVL coverage is higher for the MSc writing corpus than the academic portions of COCA and the BNC. Here too, corpus composition undoubtedly plays a role. In addition, there is likely to be an effect due to an aspect of Gardner & Davies' (2014) methodology. While the AVL (unlike the AWL) is not based on word families, their figure of approximately 14% coverage comes from a case study which, in order to permit comparisons with the AWL, used part of the AVL grouped

into 570 word families. The findings of these studies are therefore not fully comparable, and because of the lack to date of studies using the AVL, further research is needed.

4.2 Home versus international students

There is a fairly commonplace belief on the part of many Swedish university teachers that international students have lower English proficiency than home students. This perception is frequently offered almost apologetically; there is a widespread perception that international students enrich the Swedish university classroom and that their presence is therefore desirable, but that achieving an international student presence requires the use of English as an academic *lingua franca*, and while this puts all participants at a disadvantage, those who have gone through the Swedish educational system, which emphasizes English, are better able to cope than most incoming mobile international students. It is not clear how to explain the disparity between this belief and the findings of the present study.

A possible explanation is that teacher perceptions are based less on reality and more on an awareness of differences. More specifically, the English used by Swedish university students is familiar to their teachers, and the non-standard transfer features that characterize it are unmarked, while those of students with other origins are more salient. Another possibility, which indeed is applicable to all of the findings reported here, is that students recruited to the prestigious university where this study was conducted are a relatively homogeneous, skilled group of English users. Were the study to be replicated at another institution, between-group differences might be identified. It is also possible that these groups may differ in English proficiency, but that the differences manifest themselves in other domains than productive academic vocabulary (i.e., in other domains of oral and written communication). Future research would be required to establish the extent to which any of these explanations is a factor.

4.3 Vocabulary development over time

One of the intended benefits of EMI is that it creates exposure to the language and can therefore result in incidental vocabulary acquisition. Academic vocabulary would appear to be a prime candidate for such ac-

quisition, since it is an area of language to which students can be expected to have greatest exposure in a university setting. It is therefore somewhat counterintuitive that the findings for academic vocabulary development were mixed.

One reason for this may be that even the least experienced writers in this study were highly proficient. By virtue of being deemed capable of doing postgraduate academic work through the medium of English, these students can be classed as advanced users of English, and this is additionally indicated by the fact that their texts were richly populated with academic vocabulary. As Hyltenstam & Abrahamsson (2012) note, research on very proficient L2 learners is in short supply compared with the voluminous body of second language acquisition research on learners at lower proficiency levels. However, it is reasonable to expect their learning to progress at a slower pace, simply because they have less ground to cover. In other words, there may be a phenomenon at play akin to a ceiling effect, according to which the year two texts did not show much greater lexical diversity and sophistication because the year one texts were already satisfactory in that regard.

Similarly, it may be thought that these students had relatively limited opportunities for vocabulary development. The EMI environment provides a context in which only incidental language acquisition can occur, rather than an EAP/TEFL environment where language development is the target of explicit instruction. As a result, opportunities for language learning are closely linked to *exposure* to the linguistic features that are candidates for learning. Less proficient learners have more opportunities for exposure to new forms than advanced learners, precisely because more of what they are exposed to is new. In the case of the high-register academic vocabulary that was the focus of the present investigation, the opportunities for exposure to the infrequent words decrease logarithmically, not arithmetically, once the first bands of very frequent words have been learned.

5 Conclusions

This article has reported an investigation into the academic vocabulary knowledge of students in an EMI setting. Students' knowledge of academic vocabulary is important in this context because it is essential both

for adequate comprehension of academic texts and for producing register-appropriate assessment work. As a consequence of the fact that study in the EMI environment places demands on students' receptive and productive academic vocabulary knowledge, it is an aspect of linguistic proficiency which could reasonably be expected to develop over the course of their studies. A measure of students' productive academic vocabulary is therefore a useful indicator (though by no means the only one) of two important factors: students' preparedness for academic study, and their development in English.

To the extent that the findings presented in this paper speak to preparedness, they permit an optimistic interpretation: academic vocabulary items accounted for approximately 20% of all tokens, a rather higher figure than that found in many earlier studies. Although knowledge of academic vocabulary alone cannot be interpreted as evidence that students are equal to the challenges of study through the medium of English, a more cautious claim can be made: there is no reason to believe that this cohort of students lacks an adequate knowledge of academic vocabulary.

The high level of coverage also provides support for the principles underlying the construction of the AVL. By including items which occur in academic texts more frequently than in general ones, and by excluding items which occur disproportionately frequently in some disciplines only, the AVL is designed to give a better representation of general academic vocabulary than earlier lists, and the incidence of AVL items in the present corpus provides indirect evidence that the AVL behaves the way it was intended. While this does not resolve all of the problematic aspects of the notion of an academic core vocabulary (cf. Hyland/Tse 2007), it suggests that, in circumstances where an academic vocabulary list is necessary or desirable, for pedagogical or research purposes, the AVL is the list of choice.

Perhaps more significantly, this measure of productive academic vocabulary gives no support for the idea that international students and local Swedish students differ in their abilities in English. This is reassuring given the fact that the economic and policy imperatives in Swedish higher education (and reflected elsewhere in Europe) will for the foreseeable future lead to an increase in inward student mobility.

With regard to vocabulary development between the first and the second year, evidence was limited; there were modest gains by some measures

and none by others. This is a finding of relevance given the current rapid expansion of EMI, and the twin motivations behind it. EMI is expected to be both a tool to facilitate mobility in higher education and a vehicle for improved English language skills on the part of participants but this study of academic vocabulary knowledge provides little indication that the latter ambition is realized, at least in the context under investigation.

6 Pedagogical implications

In this volume, with its focus on the pedagogical aspects of assessment, the pedagogical implications of the findings merit exploration. However, the EMI environment is complex in its pedagogical objectives. One objective of EMI is simply to enable the teaching and learning of subject matter by using English as an academic *lingua franca*. In many EMI contexts, though, an additional objective is to provide a context which facilitates students' incidental acquisition of English. The pedagogical implications of students' vocabulary knowledge and development are different for these two different objectives.

In terms of content learning, these students appear to be well equipped with a productive knowledge of academic vocabulary sufficient to complete assessment tasks (and therefore by implication with a receptive vocabulary sufficient to read academic texts). This means that teachers (provided they have similar student profiles and communication genres) can concentrate on, for instance, promoting the critical reading of the disciplinary vocabulary. From a collaborative learning perspective, peer learning can enable the further exploration of the enhanced understanding of technical vocabulary.

In the scenarios where the EMI context involves an element of collaboration or contact between language lecturers and subject lecturers, the language lecturer might help the subject lecturer highlight the way in which academic vocabulary serves to carry the disciplinary argument. Such a shared focus would help students articulate the necessary disciplinary connections between argumentative components. A subject lecturer might contribute with useful insights for prompts, exercises, and classroom assessment techniques focused on exploring technical vocabulary.

With respect to language development, teachers may conclude that basic academic vocabulary knowledge can be taken as confirmed. They can there-

fore use this apparent communicative resource of academic vocabulary as a stepping-stone to explore the remaining dimension of written disciplinary communication. For example, they might have students extract and master the technical vocabulary in the texts they encounter via basic critical reading using genre and corpus analyses.

The productive knowledge of the frequent academic vocabulary items demonstrated here could also be a potential stepping-stone toward command of the less frequent AVL items. However, the evidence of this study is that development along those lines does not happen automatically, and indeed there is no reason to suppose it should, given that opportunities for exposure to infrequent vocabulary are limited. A key pedagogical implication of these findings is therefore that *incidental* acquisition is unlikely to be *accidental*, and that teachers who hope their students' academic vocabulary will develop during an EMI course should create opportunities for exposure to and practice of a broader range of academic lexis.

This study underscores a reality of many EMI settings. EMI is intended to be a *de facto* form of Content and Language Integrated Learning (CLIL) but while CLIL settings work actively both with content knowledge and with language development, in EMI the expectation is frequently that the preconditions for incidental language acquisition are put in place simply by dint of offering instruction in English. This study has provided evidence that those expectations are not entirely justified.

References

- Anthony, Laurence, 2015: *AntWordProfiler (Version 1.4.1)* [Computer Software]. Tokyo, Japan: Waseda University. <http://www.laurenceanthony.net/> (last accessed: 30 July, 2015).
- Baker, Paul, 2006: *Using Corpora in Discourse Analysis*. London: Continuum.
- Bogaards, Paul / Laufer, Batia (eds), 2004: *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (Vol. 10). Amsterdam: John Benjamins Publishing.
- Carroll, Judith / Ryan, Janette (eds), 2005: *Teaching International Students – Improving Learning for all*. Abingdon: Routledge.

- Charles, Maggie / Pecorari, Diane, 2016: *Introducing English for Academic Purposes*. London: Routledge.
- Chung, Teresa Mihwa, / Nation, Paul, 2003: "Technical vocabulary in specialised texts." *Reading in a Foreign Language* 15(2), 103–116.
- Huckin, Thomas / Coady, James, 1997: *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge: Cambridge University Press.
- COCA, 2015: *Corpus of Contemporary American English*. [Computer software]. <http://corpus.byu.edu/coca> (last accessed: 30 July, 2015).
- Coleman, James A., 2006: "English-medium teaching in European higher education." *Language Teaching* 39(1), 1–14.
- Corson, David, 1997: "The learning and use of academic English words." *Language Learning* 47(4), 671–718.
- Coxhead, Averil, 2000: "A new academic word list." *TESOL Quarterly* 34(2), 213–238.
- Daller, Helmut / van Hout, Roeland / Treffers-Daller, Jeanine, 2003: "Lexical richness in the spontaneous speech of bilinguals." *Applied Linguistics* 24(2), 197–222.
- Daller, Helmut / Xue, Huijuan, 2009: "Vocabulary knowledge and academic success: A study of Chinese students in UK higher education." In: Richards, Brian / Daller, H. Michael / Malvern, David D. / Meara, Paul / Milton, James / Treffers-Daller, Jeanine (eds): *Vocabulary Studies in First and Second Language Acquisition: The Interface Between Theory and Applications*. Basingstoke: Palgrave Macmillan, 179–193.
- Davies, Mark, 2015: *Word and Phrase*. [Computer software]. <http://www.wordandphrase.info/academic/> (last accessed: 30 July, 2015).
- Dearden, Julie, 2014: *English as a Medium of Instruction – a Growing Global Phenomenon*. (Research report). <http://www.education.ox.ac.uk/wordpress/wp-content/uploads/2014/09/EMI-a-Growing-Global-Phenomenon-new-cover.pdf> (last accessed: 30 July, 2015).
- Durrant, Phil, 2014: "Discipline and level specificity in university students' written vocabulary." *Applied Linguistics* 35(3), 328–356.
- Gardner, Dee / Davies, Mark, 2014: "A new academic vocabulary list." *Applied Linguistics* 35(3), 305–327.

- Guiraud, Pierre, 1954: *Les caracteres statistiques du vocabulaire*. Paris: Presses Universitaires de France.
- Gustafsson, Magnus / Eriksson, Andreas / Räisänen, Christine / Stenberg, Ann-Charlotte / Jacobs, Cecilia / Wright, Jenny / Wyrley-Birch, Bridget / Winberg, Chris, 2011: “Collaborating for content and language integrated learning: The situated character of faculty collaboration and student learning.” *Across the Disciplines* 8(3). <http://wac.colostate.edu/atd/clil/gustafssonetal.cfm> (last accessed: 30 July, 2015).
- Gustafsson, Magnus / Jacobs, Cecilia, 2013: “Student learning and ICLHE – frameworks and contexts.” *Journal of Academic Writing* 3(1). <http://e-learning.coventry.ac.uk/ojs/index.php/joaw/article/view/141> (last accessed: 30 July, 2015).
- Gustafsson, Magnus / Malmström, Hans, 2013: “Master level writing in engineering and productive vocabulary: What does measuring academic vocabulary levels tell us?” In: Johannesson, Nils-Lennart / Melchers, Gunnel / Björkman, Beyza (eds): *Of Butterflies and Birds, of Dialects and Genres: Essays in Honour of Philip Shaw*. Stockholm: Stockholm Studies in English 104, 123–140.
- Holmes, David I., 1994: “Authorship attribution.” *Computers and the Humanities* 28(2), 87–106.
- Hyland, Ken / Tse, Polly, 2007: “Is there an ‘academic vocabulary’?” *TESOL quarterly* 41(2), 235–253.
- Hyltenstam, Kenneth / Abrahamsson, Niclas, 2012: “Introduction.” *Studies in Second Language Acquisition* 34(2), 177–186.
- Jochems, Wim / Snippe, Joke / Smid, Harm Jan / Verweij, Agnes, 1996: “The academic progress of foreign students: study achievement and study behaviour.” *Higher Education* 31(3), 325–340.
- Johnson, Wendell, 1939: *Language and Speech Hygiene: An Application of General Semantics. Outline of a Course*. General semantics monographs 1. Lakeville, CT: Institute of General Semantics.
- Kachru, Braj, B., 1992: *The Other Tongue: English Across Cultures*. Champaign, IL: University of Illinois Press.
- Knoch, Ute / Rouhshad, Amir / Oon, Su Ping / Storch, Neomy, 2015: “What happens to ESL students’ writing after three years of study at an English medium university?” *Assessing Writing* 28, 39–52.

- Lasagabaster, David, 2008: "Foreign language competence in content and language integrated courses." *The Open Applied Linguistics Journal* 1(1), 30–41.
- Laufer, Batia, 1994: "The lexical profile of second language writing: does it change over time?" *RELC Journal* 25(2), 21–33.
- Laufer, Batia / Hulstijn, Jan, 2001: "Incidental vocabulary acquisition in a second language: The construct of task-induced involvement." *Applied Linguistics* 22(1), 1–26.
- Laufer, Batia, / Nation, Paul, 1999: "A vocabulary-size test of controlled productive ability." *Language Testing* 16(1), 33–51.
- Li, Yongyan / Qian, David, 2010: "Profiling the Academic Word List (AWL) in a financial corpus." *System* 38(3), 402–411.
- Malvern, David / Richards, Brian, 2013: "Measures of lexical richness." In: Chapelle, Carol A. (ed.): *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell Publishing Ltd, 3622–3627.
- Mann, Mary Bachmann, 1944: "The quantitative differentiation of samples of written language." *Psychological Monographs* 56(2), 39–74.
- Marsh, David, 2005: Project D3 – CLILMatrix – Central workshop report 6/2005 (Graz, 3–5 November 2005). European Centre for Modern Languages 6. http://archive.ecml.at/mtp2/CLILmatrix/pdf/wsrepD3E2005_6.pdf (last accessed: 30 July, 2015).
- Milton, James, 2009: *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Milton, James, 2010: "The development of vocabulary breadth across the CEFR levels." *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research*, 211–232.
- Morrison, Jo / Merrick, Beatrice / Higgs, Samantha / Le Métails, Joanna, 2005: "Researching the performance of international students in the UK." *Studies in Higher Education* 30(3), 327–337.
- Mudraya, Olga, 2006: "Engineering English: A lexical frequency instructional model." *English for Specific Purposes* 25(2), 235–256.
- Nation, Paul, 2001: *Learning Vocabulary in Another Language*. Cambridge, England: Cambridge University Press.

- Nation, Paul, 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.
- Pecorari, Diane / Shaw, Philip / Irvine, Aileen / Malmström, Hans, 2011: "English for Academic Purposes at Swedish universities: teachers objectives and practices." *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)* 22, 55–78.
- Range, 2015: [Computer software]. Available at: <http://www.victoria.ac.nz/lals/about/staff/paul-nation>.
- Schmitt, Norbert, 2000: *Vocabulary in Language Teaching*. Cambridge, England: Cambridge University Press.
- Stæhr, Lars Stenius, 2008: "Vocabulary size and the skills of listening, reading and writing." *Language Learning Journal* 36(2), 139–152.
- Turner, Jean L., 2014: *Using Statistics in Small-Scale Language Education Research: Focus on Non-Parametric Data*. New York: Routledge.
- Vermeer, Anne, 2000: "Coming to grips with lexical richness in spontaneous speech data." *Language Testing* 17(1), 65–83.
- Warwick, Philip, 2006: *International Students in the UK: How Can We Give Them a Better Experience?* (Working Paper). Department of Management Studies, University of York, York.
- West, Michael, 1953: *A General Service List of English Words*. London: Longman, Green & Co.
- Xue, Gua-yi / Nation, Paul, 1984: "A university word list." *Language Learning and Communication* 3(2), 215–29.

