

Exploitation and assessment of a Business English corpus through language learning tasks

Alejandro Curado Fuentes
University of Extremadura

1 Introduction

Several ESP (English for Specific Purposes) scholars use corpus data for practical purposes. The findings are diverse, and contribute empirical evidence to establish categories of word use and collocations (eg Luzón Marco 2000), and to offer detailed lexical profiles in specific subject areas (eg Nelson 2000). The general purpose of the study reported on in this paper is to evaluate the validity of corpus-driven information for ESP courses. However, I consider the results gathered in my lexical analysis as preliminary and makeshift, subject to change if the learning situation so demands. A Business English corpus of about one million words has been used, containing six main thematic divisions and eight different academic genres. I believe that the specific contents of the corpus should be consistently revised for effective integration in the ESP setting, and that this revising process needs insight gained from academic task application. In addition, the development of ESP activities ought to take place in an authentic communicative academic context, which greatly influences special learning conditions (Starfield 2001).

In this case, the academic context is the College of Business Science at the University of Extremadura, Spain, where my colleagues and I work as ESP instructors in a team effort to match real language situations with academic tasks. During the last ten years, ESP teaching has dramatically evolved in our institution from a grammar-based syllabus presentation to a subject-specific learning approach. Corpus techniques, constantly enhanced for linguistic analysis, have influenced this evolution, as the application of such instruments fosters direct involvement with content and language. My doctoral dissertation (Curado Fuentes 2000) focuses on applying Corpus Linguistics to the ESP

learning situation. Genre and register variation in Business Science and Technology can be explored and analyzed with great accuracy for teaching purposes. Such lexical diversity can be found, for instance, in e-mail writing, featuring lexical items and phrases that can either resemble conversational input or provide a dimension that is closer to academic writing, in agreement with Gains (1999). Corpus data enables the realization of important language functions, which 'are widespread and consistent in usage, and most importantly, meaningful' (Guest 1998: 31).

Corpus studies also focus on special terminology, such as academic and technical vocabulary. For example, James and Purchase (1996) describe key Economics and Business Science vocabulary at the university level, which was selected according to the frequency and dispersion of subject-based lexical items in textbooks read at the Hong Kong University of Science and Technology during the early 90s. Dealing with Engineering English, Farrell (1991) insists on the need to carry out lexical analyses for subject area description in ESP. And Pedersen (1995), underlining the importance of technical collocation in nominal compounds, discusses the significance of technical and semi-technical combinations for register analysis: words acquire fixed positions in compounds and phrases, and thus form characteristic language bonds in scientific-technical registers (eg *heat resistance*, *superior heat resistance*, *provide superior heat resistance to*).

The purpose of my study is to follow a similar type of lexical analysis based on written corpora. In contrast to the focus of the scholars cited above, however, my main concern is not a classification of corpus data for register description. Rather, my main concern is the conjunction of topic and task in the design and arrangement of Business English corpus material: the texts have been considered within the area of Business Science and Technology according to common core subject criteria (a topic-centered approach), whereas the learning situation has provided the testing ground for the confirmation of corpus analysis results (a task-based focus).

My corpus analysis is directed towards word behavior acquisition, which implies awareness of special learning needs as regards the use of lexical chunks. Thus, these linguistic units are assessed in terms of how well learners perform specific tasks, such as delivering technical reports. Learners' communicative weaknesses are determined by accounting for contextual variables in the learning situation. Hence, working with a representative corpus involves a dual quest: identifying the sample of texts that satisfies topic interests according to external criteria (ie university and professional curricula) and adjusting to learning factors such as types of skills to be emphasized in the ESP context. As a result,

three stages have been devised in this paper: 1. specification of the methodology to satisfy language and content requirements, 2. lexical analysis to fit learning interests, and 3. evaluation of findings in the target learning situation.

2 Specification of methodology

A corpus-based lexical analysis of specific texts should concentrate on the readings required or recommended by subject instructors (eg Breland et al 1994; James and Purchase 1996; Lozano Palacios 2000). In my institution, academic readings and technical knowledge are intrinsically related, as reading competence generally leads to a good level of content knowledge, and this degree of expertise, in turn, fosters effective reading skills. A key exponent of academic ability is, in this sense, familiarity with the various genres and text types handled in class; in fact, different genres must be coped with by learners in order for them to widen their knowledge and enhance their linguistic competence (Conrad 1996: 302).

Three main learning levels are identified with relation to academic genres (see eg Huddleston 1971): first, an advanced text represented by writings of 'high-brow' genres, eg research papers; second, a 'low-brow' genre which points to a less complex type of discourse, such as introductory textbooks; and third, a middle position held by descriptive genres, eg reports of products and sales. The three co-exist at my university, and yet, only the first two clearly serve academic purposes, whereas the third type tends to be more common in professional situations.

In the ESP setting, additional written and oral genres are acknowledged for Business English. For example, conceived at a high (research) academic level, conference papers are becoming widely used by advanced students. This is due to the increase of travels abroad, which provide students with the challenge of presenting a technical paper (eg a final project report). Another new genre is the electronic discussion, gaining greater recognition and acceptance among academic peers, and often supporting interesting and innovative ideas concerning the specific subject area. Electronic discussions are also treated as an advanced genre in my corpus, mainly due to their degree of specification and expertise in the subject area. Other texts include mid-brow sources, such as the aforementioned technical reports, and pieces of news presenting specific issues on Business technology. Both texts exemplify an intermediate level of difficulty, a discourse considered less complex than research writing, but usually more demanding than introductory textbooks.

The integration of the genres in my corpus therefore reflects, as a primary concern, the learning stages of the ESP teaching context. Coming to terms with different genres is equivalent to acquiring a certain competence and academic achievement in polytechnical courses. In this progression towards discourse command, understanding stylistic diversity is essential for the effective management of register, which is mainly identified through the *field*, *mode* and *tenor* of discourse (Halliday and Hasan 1985: 38). In fact, the more technically prepared the recipient of the information, the more effectively s/he can operate with these discourse factors to distinguish concepts, formats and writer/reader relationships in the texts. All learning of discourse implies that the subject matter in the corpus should vary according to different genre levels. This correlation chiefly depends on the year of studies when the subjects are taught. For instance, General Business, a first year course, is introductory, presented in textbook chapters and book reviews; these texts are low-brow genres.

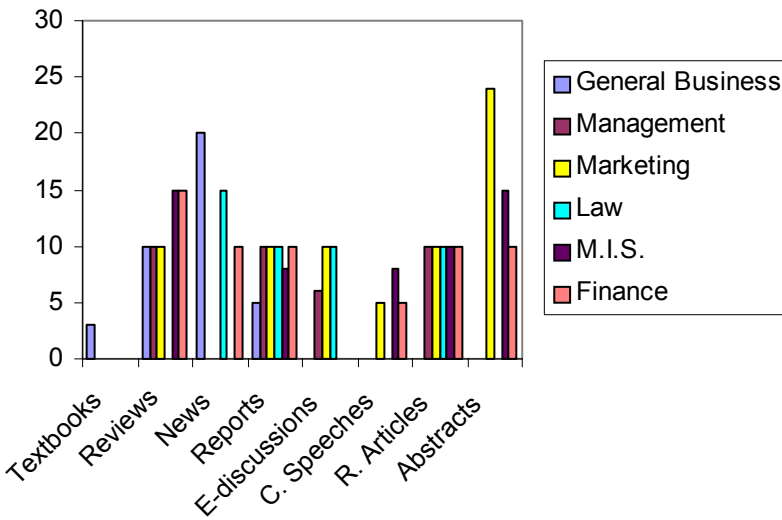


Figure 1: Distribution of genres in relation to subject areas

As Figure 1 shows, six main academic areas have been selected. These areas share a major theme in Business Science and Technology: the relationship

between Business Science and Information Technology. In this respect, the corpus is topic-oriented.¹

Three textbooks have been included in General Business, represented by the left-most column. These textbooks deal with the concept of and developments in electronic commerce, a chief issue in Business Technology. Some mid-brow genres have also been included in General Business: for example, 25 short pieces of news sharing a common theme, and the description of computer companies operating worldwide, published in recent issues of *The New York Times*.

High-brow genres are favored in advanced courses of M.I.S. (*Management Information Systems*), taught in the fourth and fifth years of Business Science and Economics. As Figure 1 shows, eight conference speeches, ten research articles, and 14 abstracts have been selected. Lectures dealing with economics statistics and evaluation in the e-business world, for instance, constitute some of the conference topics. The discourse of academic research thus prevails in these sources, demanding special content knowledge from the audience.

In addition, some reports and reviews have been included in M.I.S. This integration responds to both external (institutional) and internal (ESP) criteria. Regarding the former, study programs and course readings demand the growing use of Business reports and reviews after the second year of studies. In the latter case, as an ESP instructor, I deem the combination of low- and high-brow genres as productive, in agreement with Conrad (1996). This blending of genres should be maintained at all learning stages. In introductory General Business texts, for instance, not only textbooks but also a moderate number of news articles and reports are read.

Study programs and bibliographies serve as pivotal reference for the selection of sources, as mentioned before. Some additional guidance is provided by subject teachers, whose advice usually points to the essential literature available in the topic-driven area of Business Technology. My colleagues have identified textbooks and articles as main readings in the corpus. The aid of references on the internet is also relevant; eg *WWW resources in Economics* offers a great deal of information on reports, conference papers, abstracts, and electronic discussions. Finally, business technology reviews and news can be selected daily in electronic publications such as *The New York Times*, *Finance Review*, *ACME*, *The Economist*, *Newsweek*, etc.

My corpus contains a total of 1,010,435 words, which have been distributed across all six subject sub-corpora (academic areas) mentioned. The greatest margin is 77,646 words (between Finance and Law). Figure 2 displays the organization of the data according to three categories: 1. tokens or total number of running words, 2. types or distinctive items, and 3. standardized ratio of tokens

and types, ie the average number of distinctive items per 1,000 running words. These ratios vary from 35.47 words as the lowest score (for Law) to 38.43 items as the highest (Marketing). This means that Law texts present fewer distinctive items, whereas Marketing has a more varied set of words despite its low number of tokens (second lowest in the corpus).

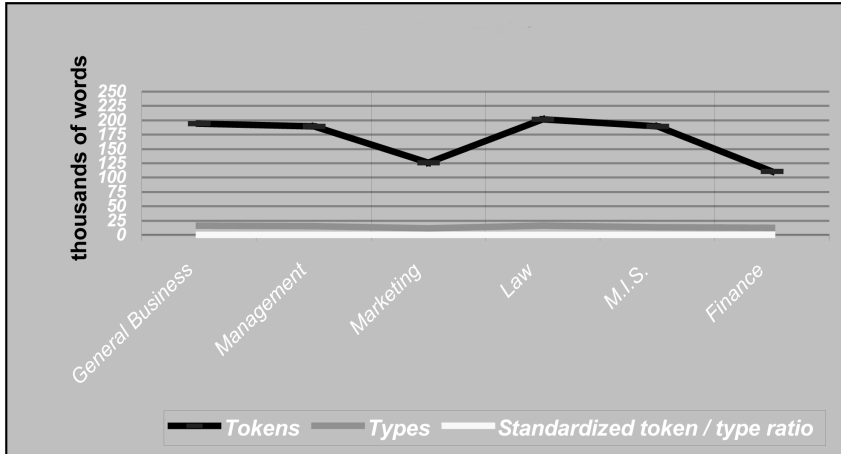


Figure 2: Corpus size

For my analysis of the data, I have selected the top two categories in the graph (Figure 2). These are the sets of Law texts (201,986 items) and General Business (193,786 tokens). I have chosen these two sub-corpora because of their representativeness in yielding a clear picture of lexical variation in the ESP context. Allowing for different discourse types to take place, this diversity of linguistic input provides suitable mixed conditions for task development from thematic and stylistic viewpoints.

The analysis of the lexical data is described in section 3. This examination is given as a preparation stage where language divergence is checked; however, this variation must be tested further by contrasting feedback from task application (section 4).

3 Lexical analysis

Lexical activities in ESP are based on the observation of word behavior according to specific context. This is done by heeding Key Word In Context as a central measuring device for small corpora (KWIC) (cf Tribble 1997; Scott 1999a). My goal is to determine language variation by describing the how's and why's of lexical distinction. Contextual factors, such as the presence of specialized subject matter, significantly influence this language change. In addition to this thematic aspect, three other elements have been found to be influential. Such are grammatical functions, core senses, and specific genre traits. The four features contribute to shaping the lexical profile which learners aim to acquire in ESP learning situations. In this respect, command of common core language, grammatical properties, subject matter, and genre features becomes the chief learning target in the communicative approach.

3.1 Common coreness

As a primary step, the data should be organized in the form of word lists. Two chief frequency indexes have been found in the corpus, provided by *WordSmith Tools 3.0* (Scott 1999b): first, the Detailed Consistency List (DCL) orders items according to frequency and dispersion in as many texts as needed; second, the overall corpus list displays data in terms of frequency alone.

The DCL has been used to highlight content words. These are differentiated from grammatical items so that lexical collocations can be distinguished from grammatical phrases. The concept of lexical collocation is defined as a main unit of meaning, from which contextual feedback is obtained. Content words, in addition, function as primary matter shaping academic and technical language. In the case of the DCL, as items are ordered not only according to frequency, but also to distribution, the resulting nouns, verbs, adjectives and adverbs are considered semi-technical, 'words which are not specific to a subject specialty but which occur regularly' (Kennedy and Bolitho 1984: 57). Their appearance is also that of academic lexis: words measured across various academic genres, and 'high-frequency words that are broadly applicable in university-level courses' (Burgmeier et al 1991: viii).

The corpus includes eight genres. Table 1 lists the 41 most frequent semi-technical / academic words in these genres. These 41 elements have been selected from the first 100 words of the DCL; the remaining 59 are grammatical / discourse words, excluded from the DCL.

Table 1: List of academic words in the DCL (among the top 100 items)

N	Word	Files	Total	Abst	Artic	Book	Conf	Disc	News	Repo	Rev
27	Data	8	1804	111	198	847	143	1	5	465	34
33	Model	7	1612	21	473	530	58	54	0	470	6
35	Market	8	1565	57	483	187	264	66	17	43	54
36	Information	8	1492	42	444	339	202	10	29	365	61
39	Management	8	1285	46	469	107	326	95	3	159	80
41	Financial	8	1268	8	341	336	265	92	2	212	12
44	Analysis	8	1158	21	176	657	67	8	4	203	22
45	Growth	8	1126	15	221	141	80	96	4	548	21
46	Internet	8	1106	1	430	131	261	2	27	143	111
48	Countries	8	1097	9	167	391	31	14	2	461	22
49	Example	7	1063	5	198	640	85	18	0	97	20
52	International	8	1021	45	344	235	50	11	6	263	67
55	Business	8	973	24	295	135	194	21	22	213	69
58	Economic	8	958	10	192	135	116	4	3	479	19
62	Case	8	918	25	182	368	72	28	12	222	9
63	Different	8	900	10	192	417	88	13	1	168	11
65	Based	8	854	22	301	256	78	12	17	146	22
66	Capital	8	853	2	147	119	225	114	1	232	3
68	Level	8	826	22	224	182	79	10	1	291	17
70	Change	8	810	1	453	53	38	79	6	169	11
72	Article	8	803	56	24	26	74	9	1	611	2
73	Design	8	797	2	127	589	12	2	8	52	5
75	Good	7	774	1	70	90	51	22	0	523	17
76	GIF	6	742	72	0	509	0	19	43	76	23
77	Innovation	8	738	19	81	11	10	1	1	601	14
78	Function	8	733	4	81	398	34	74	34	82	26
83	Control	7	703	5	244	201	46	9	0	185	13
84	High	8	703	8	238	140	79	10	12	179	37
85	Effects	7	687	4	115	340	33	3	0	186	6
86	General	8	672	3	77	352	30	12	12	179	7
87	Development	8	667	15	188	168	62	6	2	196	30
89	Firms	7	659	39	160	26	126	13	0	285	10
90	Large	8	655	26	181	175	48	16	2	197	10
93	Marketing	7	623	56	426	22	12	3	0	49	55
94	Country	8	619	11	73	160	5	15	5	321	29
95	Models	7	615	12	225	136	75	2	0	158	7
96	Group	8	615	7	209	238	35	8	12	86	20
97	Factors	8	614	10	71	413	34	3	1	75	7
98	Firm	8	611	15	320	24	105	9	4	130	4
99	Effect	8	604	14	82	210	41	5	1	245	6
100	Distribution	7	595	4	59	436	38	3	0	54	1

N = word position in DCL *Abst* = Abstracts / *Artic* = Research articles / *Book* = Textbooks / *Conf* = Conference speeches / *Disc* = Electronic discussions / *News* = Pieces of news / *Repo* = Technical reports / *Rev* = Book reviews

At least four genres must be recorded under the ‘Files’ column in order that the lexical item be considered semi-technical. In addition, under ‘Total’ (ie the overall number of repetitions of the word), a minimum of ten instances of the item must emerge. In this count, I have excluded grammatical words (eg *the, of, in, although*), discourse markers (eg *next, however, therefore*), and indexical verbs (eg *make, get, be, go, do, give*).

The resulting lexical set is thus made up exclusively by content words appearing frequently and widely enough in the domain of Business Science and Technology. Because of this recurrent use, such content lexis constitutes core linguistic material for the ESP setting: a main semi-technical list of subject matter words (cf Kennedy and Bolitho 1984; Farrell 1990; Nelson 2000). These items combine characteristically, leading to lexical coreness in the area of specialization. In other words, the fact that these elements are repeated consistently in the texts means that they are relevant as basic language for Business Technology. My interest lies in those meanings to be learned directly through recurrent exposition in the subject area. Such denotations are semi-technical, and academic due to the majority of common core meanings associated with them in the academic genres of Table 1.

Table 2 illustrates some of the verb, noun and adjective combinations in the corpus:

Table 2: Core semi-technical language

<i>Semi-technical combinations</i>	<i>Proportional ratios</i>
<i>Send + data</i>	18 %
<i>Capture + data</i>	12 %
<i>Receive + data</i>	7 %
<i>Store + data</i>	5 %
<i>View + data</i>	5 %
<i>Lack of data</i>	4 %
<i>Data files</i>	3 %
<i>Voice, data, and video</i>	2 %
<i>Data stream</i>	2 %
<i>Electronic data interchange</i>	2 %
<i>Our model of</i>	12 %
<i>Develop + model</i>	10 %
<i>Simulation model</i>	4 %
<i>Build + model</i>	4 %

<i>The perfect model for</i>	1 %
<i>Political economy model</i>	1 %
<i>The job market</i>	14 %
<i>Enter + market</i>	11 %
<i>Market summary</i>	10 %
<i>The stock market</i>	10 %
<i>Market technician</i>	6 %
<i>Fiber-to-desktop market</i>	2 %
<i>On-line market</i>	1 %

The percentages in Table 2 indicate the proportion of that particular expression or word combination in relation to the overall instances of the item. For example, in the case of the first academic item in the list – the noun *data* (see Table 1) – up to 18 percent of the total appearances of this word in the corpus include the verb *send* in some form – *sends data*, *sending data*, *data being sent*, *data to be sent*, etc.

3.2 Grammatical properties

Table 3 displays some grammatical items which appear in the absolute frequency list (the overall corpus frequency list). These are the 50 most frequent words in the corpus, which correspond to grammatical elements including articles, prepositions, conjunctions, auxiliary verbs, pronouns, non-qualifying adjectives such as demonstratives and quantifiers, and modal verbs. In addition, an adverb like *also* can be found (# 39 in Table 3), and is considered rhetorical, marking addition of information in discourse. In this respect, working as a discourse marker, the word *also* has been placed in the same list with grammatical words, and is thus distinguished from content elements. On account of its performance as a specific science and technology discourse device (cf Trimble 1985), this rhetorical adverb differs from semi-technical or academic adverbs like *effectively* or *completely* (listed in a separate relation of words, eg Table 1).

Table 3: The Top 50 grammatical and discourse items in the overall word frequency

N	Item	Frequency	Percentage
1	<i>The</i>	61.638	6,12
2	<i>Of</i>	34.991	3,47
3	<i>And</i>	24.265	2,41
4	<i>To</i>	22.068	2,19
5	<i>In</i>	20.122	2,00
6	<i>A</i>	18.276	1,81
7	<i>Is</i>	12.251	1,22
8	<i>For</i>	10.808	1,07
9	<i>That</i>	10.167	1,01
10	<i>Be</i>	6.339	0,63
11	<i>Are</i>	6.265	0,62
12	<i>On</i>	6.036	0,60
13	<i>As</i>	6.023	0,60
14	<i>This</i>	5.398	0,54
15	<i>Or</i>	5.328	0,53
16	<i>With</i>	5.286	0,52
17	<i>By</i>	5.047	0,50
18	<i>It</i>	4.296	0,43
19	<i>Not</i>	3.779	0,38
20	<i>An</i>	3.472	0,34
21	<i>From</i>	3.468	0,34
24	<i>Have</i>	2.969	0,29
25	<i>Which</i>	2.810	0,28
26	<i>At</i>	2.746	0,27

27	<i>If</i>	2.715	0,27
28	<i>Can</i>	2.691	0,27
30	<i>We</i>	2.526	0,25
31	<i>More</i>	2.387	0,24
33	<i>Will</i>	2.267	0,23
35	<i>May</i>	2.188	0,22
36	<i>Has</i>	2.167	0,22
39	<i>Also</i>	1.874	0,19
40	<i>These</i>	1.860	0,18
42	<i>I</i>	1.820	0,18
43	<i>Was</i>	1.804	0,18
44	<i>You</i>	1.804	0,18
45	<i>Such</i>	1.775	0,18
46	<i>Between</i>	1.773	0,18
47	<i>All</i>	1.728	0,17
48	<i>But</i>	1.672	0,17
49	<i>Their</i>	1.654	0,16
50	<i>Than</i>	1.649	0,16

Other articles, prepositions, conjunctions, pronouns, non-qualifying adjectives, adverbs, and modal / auxiliary verbs abound in the list that follows after item # 50. In fact, the list goes on to include adverbs that fulfill discourse marking roles, eg the function of ‘cause and effect’ (*therefore, thus*). In addition, some conjunctions and prepositions perform salient discourse features (eg signaling contrast, *whereas* and *while*). In the case of exemplifications and classifications, certain conjunctions and prepositions also combine significantly to produce discourse markers: eg *such as, for instance* and *by* + noun (in agreement with Flowerdew and Miller 1997). Finally, many examples of indexical verbs, such as *make, do, get, take, go, and give* exist in the grammatical wordlist. These lack content meaning, denoted by academic elements such as those of Table 1.

Grammatical, discourse, and indexical words operate critically in the texts. The most frequent elements are articles and prepositions; an example is the preposition *to*, one of the most frequent items in the corpus displayed in Table 3. The preposition has been classified according to meanings like quantity, purpose, modality and movement, which appear as the most frequent ones in 22,000 instances of corpus data. In addition, the item forms key collocations and patterns in academic prose, such as the structures *related to the* and *more likely to*. Figure 3 provides the most frequent uses of *to* in the corpus:

1. A **from # to #**
2. F **to have**
3. B **in order to**
4. C **be able to**
5. A **growth from # to**
6. D **with respect to**
7. B **due to the**
8. D **according to the**
9. F **more likely to**
10. A **from # to # percent**
12. F **likely to be**
13. D **related to the**
14. F **to ensure that**
15. D **relative to the**
16. F **to use the**
17. D **in addition to**
18. D **in response to**
19. A **up to #**
20. F **the need to**
21. F **are likely to**
23. E **access to the internet**

A = quantity

B = purpose

C = modality

D = academic collocation

E = movement

**F = nouns, adjs., advs., verb
+ TO + verb patterns**

Figure 3: Top grammatical functions (ordered by frequency) related to the preposition TO

3.3 Subject

The lists in Table 1 and Table 3 work as reference material from which the lexical data is selected for analysis. The high frequency words listed in Table 1 are mainly academic, but there are also technical items. The fact is that semi-technical items abound as academic constructions, as illustrated in Table 1 and Table 2, or as grammatical combinations, as illustrated in Figure 3. In contrast, subject-specific elements, closely associated with only certain texts, demonstrate technical or specialized senses, which have been identified by observing word behavior in limited textual areas (eg in concept definitions or explanations).

Collocations and lexical phrases are two common devices in the subjects of General Business and Law in the Business Technology domain. Key elements have been pinpointed in these two areas by comparing data in the subject with main lexical feedback from the overall corpus. The words having a 'key-ness' score higher than 25 are regarded as positive (cf Scott 1997), whereas negative key words are of no concern to the subject-specific view. Positive elements have been obtained in *WordSmith Tools* by cross-tabulation of two textual sets, and by word combination likelihood tests.

The number one key word in the General Business texts is the noun *business*, collocating with technical terms in this restricted field, for example, *business finance*, *global context of business*, *business internet*, *business health*, *business technology*, whereas the noun *court* in the subject of Law, collocates with legal language, unrelated to the semantic domain of Business technology: *the Supreme Court*, *High Court*, *High Court case*, *Circuit Court*, *High Court counsel*, *Court of Appeals*. The central issue is to identify specialized language and the relationship between thematic change and key lexical features, as language variation occurs in the form of technical words according to subject area variables.

As a result, register dimensions surface in the examination of subject-based lexical shifting. My perception of register parallels Sager (1986) and Ooi (1998), among other authors. According to Sager (1986: 2), a register is 'a language used by a particular community of speakers, say those concerned with a particular subject matter or those engaged in a specialized occupation'. Table 4 illustrates lexical variation between registers, those of Business and Law in the specific corpus:

Table 4: Lexical distinction between registers

Law register	Business register
<i>The Supreme Court</i>	<i>Business finance</i>
<i>High Court</i>	<i>Global context of business</i>
<i>High Court case</i>	<i>Business internet</i>
<i>Circuit Court</i>	<i>Business health</i>
<i>High Court counsel</i>	<i>Business technology</i>
<i>Court of Appeals</i>	<i>e-business</i>
	<i>sports business</i>
	<i>understand business finance</i>
	<i>business industry</i>
	<i>business look</i>
<i>The rule of law</i>	<i>text file</i>
<i>Under the rule of law</i>	<i>text panel</i>
<i>Corporate law</i>	<i>text form</i>
<i>Employment law</i>	<i>embedded text</i>
<i>Law firms</i>	<i>text version</i>
<i>Law and business</i>	<i>full text</i>
	<i>scroll + text</i>
<i>Civil rights</i>	<i>the New York Times</i>
<i>Civil Rights Act</i>	<i>new economy</i>
<i>Civil rights cases</i>	<i>features new Windows applications</i>
<i>Civil cases filed</i>	<i>new market opportunities</i>
	<i>new currency</i>
<i>The contracting party shall</i>	<i>information flow</i>
<i>Shall agree on</i>	<i>information technology</i>
<i>Shall provide</i>	<i>information regarding</i>
<i>Each party shall</i>	<i>business information</i>
	<i>management information systems</i>

3.4 Genre

Results such as those in Table 2 and Figure 3 have been obtained by a quantitative analysis; ie a large amount of information is sifted to establish common core lexical patterns. This process is basic for the description of main linguistic units on a semi-technical level. In contrast, when focusing on one subject, as in Table

4, a qualitative approach is favored, intended to define the lexical features of certain texts.

The same applies to the study of genre-based data, as lexical items may characterize some academic text types and not others. For instance, key words identified in Textbooks vary significantly from the items of News articles: in the former, for example, the pronoun *you* is distinctive, whereas *he* is a main word in the latter. Table 5 illustrates this and other variations. The analysis is qualitative in the sense that restricted context influence determines lexical profiling. A qualitative difference is made on the choice of lexical combinations, which change according to the stylistic demands of certain genres (eg reporting in News articles).²

Table 5: Genre-based lexical variation

<p><u>Textbooks</u></p> <p><i>Allow + you to</i> <i>where you were</i> <i>to help you understand</i> <i>so you can</i> <i>if you wish</i> <i>the page you browsed</i> <i>you need to</i> <i>as you read</i> <i>if you want</i> <i>let you control</i> <i>lets you choose</i> <i>text you're reading</i> <i>when you want</i></p> <p><u>News</u></p> <p><i>He + says that</i> <i>said</i> <i>added</i> <i>also</i> <i>adds</i> <i>When he or she</i></p>
--

In the next section, I discuss my findings by fine-tuning the data in the backdrop of the academic task. I believe this last stage to be fundamental in the final assessment of the corpus material.

4 Language learning tasks and corpus material

The classification of lexical data in section 3 serves task implementation purposes. In other words, I seek to develop a framework that integrates content and language in the form of corpus and task. In this manner, key lexical-grammatical items can be confirmed in my text analysis as long as their academic meaning is productive in the ESP context. In addition, technical language is heeded if such subject-based content is actually phrased during task performance. In sum, the corpus is assessed according to communicative aims and demands.

Monitoring task learning has been done for one academic year. During this time, students conduct the different tasks that aim to test their written and oral proficiency. The activities mainly develop as action-research; students exploit academic skills such as looking up information, planning, rephrasing, summarizing, etc. In addition, they practice their reading and comprehension skills, check knowledge of subject matter, and enhance lexical-grammatical competence. According to Nunan (1989: 50), this learning setting triggers a communicative purpose in the accomplishment of the tasks: in fact, where ‘techniques of inference, linking, skimming, anticipating’ (Nunan 1989: 73) may be applied, a communicative goal calls for ‘information gap’ centered work (problem analysis, vocabulary exploitation, discussion, note-taking, etc; *ibid*: 122).

The learning task stage was thus set up and based on observation of task fulfillment. Since lexical acquisition demands optimal task instruction, all four macro-skills in language learning – reading, writing, listening and speaking – must be activated. Tasks have been divided into two main skill-driven categories: first, compilation and summarization of information (integration of reading and writing); second, oral report delivery of specific topics (involving listening and speaking).

4.1 The written task

In this first type of task, the aim is to produce a coherent picture of a particular process or concept by synthesizing documentation. I consistently propose several web pages for the task, which contain actual readings in the corpus. The intention is to have learners exploit the corpus language on their own, with the communicative objective of coping with data in an intelligible manner. They have to explore a main topic in their Business studies, make notes of the infor-

mation available, ‘seize’ the pivotal language to describe concepts, and transmit the content by rephrasing structures in charts and tables where knowledge was summarized.

Table 6 is an example of an information-gap activity on an introductory first year subject, *Small Businesses and Electronic Commerce*. First-year students are generally challenged with this work in the use of textbooks and reviews. This work therefore implies a genre-focused perspective, which is always appropriate for the academic milieu (cf Conrad 1996).

Table 6: Data summarization in introductory Business texts

Topic:	<i>Small Businesses and Electronic Commerce</i>
Material used:	Textbook excerpts and reviews
Main concepts:	<i>small firms, e-commerce, business practices, connectivity, new economy</i>
Features:	<i>conducting business over the internet, new benefits, business sectors, infrastructure of new economy management</i>
Statistics:	<i>business via the world wide web: 85 % by 2002 \$ 9 billion for network hardware in 1999 Fewer than 10 employees: more aggressive 65 % of B2B purchases in 2003 (six sectors)</i>

At this introductory stage, at least two or three genres have been handled for the written task; this mixture of different texts tends to underline the prominent position of semi-technical expressions, which result from common core approaches. Some examples are shown in Table 6, where, as can be observed, some key words are *firms, commerce, business, and internet*, recorded in the DCL (Table 1) as academic or semi-technical. Becoming familiar with this vocabulary proves to be very productive for reading and writing.

Learners generally become aware of the direct implication of the texts to increase knowledge at an early stage. The sources are approached from a common core perspective. This means that the academic items surveyed are regarded as basic, given their noticeable presence in most texts. In addition to the semi-technical vocabulary, students realize the need of specific terms related

to the subject area. Thus, expressing concepts such as those in Table 6, eg *new economy*, *e-commerce*, etc, also seems necessary.

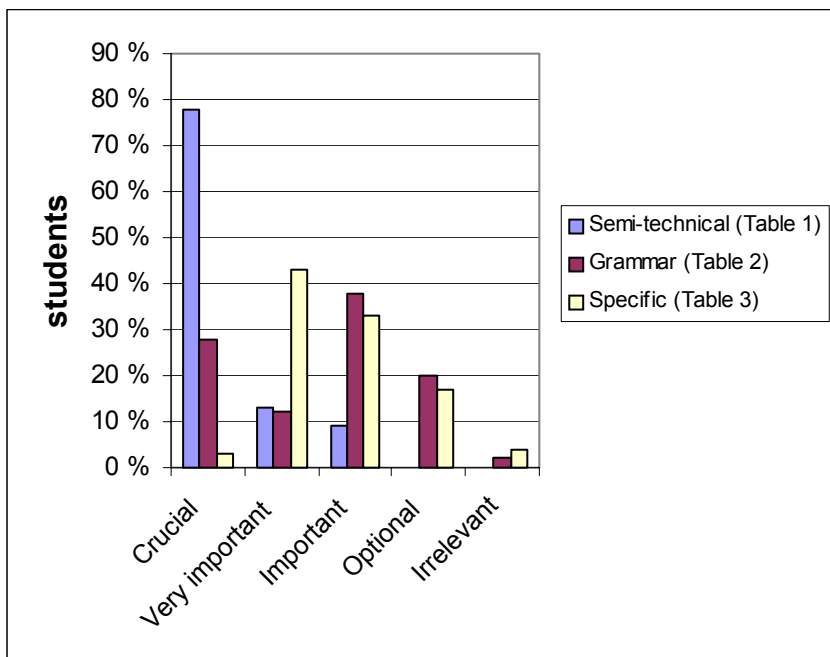


Figure 4: Students' evaluation of vocabulary in written tasks

Figure 4 illustrates the learners' judgment concerning the importance of vocabulary in the exercises. 50 students, taking Business English II as an obligatory subject (2nd year of studies), took part in the evaluation.

The fact that grammatical words are also important (38%) or even crucial (28%), surpassing specialized items, does not contradict my findings about the relative importance of specific vocabulary. In fact, not only learning academic and technical collocations such as the ones described, eg *small firms*, *new economy*, *e-business*, etc in Tables 2 and 5, is seen as fundamental, but also command of lexical phrases and significant colligations, where grammatical items combine critically with content words. This is mainly perceived by learners in their point-driven note-making exercises, where they must refer to common concepts

in their subject area, extensively repeated in the readings. Some examples are *conduct business over the internet, via the world wide web* (Table 7), and others like *related to the issue, in addition to the fact that, because of the notion that*, etc.

4.2 The oral task

Language generated from grammatical lexis is regarded as even more important in the case of oral tasks. Advanced learners in the third and fourth years of study come to realize this. Challenged with the oral delivery of a technical report, students are free to choose among several fields, listed in the corpus: Management, Marketing, Law, etc. Because the choices are given within the specialist area of their studies, learners feel that the task is highly specific and subject-oriented, ie in the ESAP line (English for Specific Academic Purposes). In contrast, the EGAP (English for General Academic Purposes) focus prevails in activities such as the previously mentioned ones (Table 6), where general academic skills (eg skimming, scanning, planning, etc) have been stressed.

Naturally, specialized or technical language should be relevant for specific academic tasks. However, in the case of oral reports, this importance has not been stressed by students' comments. Instead, they feel that academic English (semi-technical items) and grammatical structures serve their core learning purposes, and thus constitute the main linguistic input.

Figure 5 exemplifies oral presentation assessment. It is based on a third year student's paper focusing on the topic of Management Information Systems. This type of action-research fits in the context of the region where I teach, Extremadura, due to its low business activity. The need to adjust to weaker infrastructures for market development and the desire to promote business incentives in their own setting actually motivate students to conduct specific target situation analyses.

My student's report described computer systems in small and medium sized firms, for which he had to contrast documentation from foreign businesses with Spanish reports. Thus, he felt the demand to use English by rephrasing and doing simultaneous interpretation. His language command was assessed together with topic preparation and presentation, and is presented in Figure 5:

Work & Skill	Mark	Comments
Preparation	Good (4.2)	documented & introduced
Presentation flow	Good "	good use of notes good visual aids more definitions needed weak conclusion (implications?)
Oral skills	OK (3.5)	
Pronunciation		mistakes: <i>launched, image...</i>
Pitch		mistakes: <i>definite, advantage...</i>
Lexis	OK (7.2)	
Colligations		mistakes: <i>discuss about, related with...</i>
Collocations		mistakes: <i>see the data, make the job...</i>
Grammar		mistakes: <i>say you that, other thing is...</i>

Figure 5: Example of an oral report evaluation

The lexis category receives a higher score (between 1 and 10 points) than the other two, where points vary from zero to five. This is due to the fact that both vocabulary and grammar are included under lexis, since grammar should not be treated separately in the approach. As Halliday (1991) claims, the lexico-grammar is thus ‘a unified phenomenon, a single level of “wording”, of which lexis is the “most delicate” resolution’ (1991: 32).

In oral reports, language command is regarded as the integration of good presentation skills with lexical knowledge. As illustrated in Figure 5, the teacher’s evaluative remarks point to weak communicative aspects. Further, data gathered from the learner’s self-evaluation and peer reviews in class confirm lexical concerns and the combination of content and grammatical units as priority matter for specific language teaching.

Students’ presentations have been video-taped for subsequent viewing by those involved. Located and analyzed, errors are discussed so that learners might find alternative ways of expressing themselves. This activity is carried out in groups, where more advanced students help the less skillful to check for mistakes. The aim is to determine ‘accepted language’ in the academic community

where this type of academic / technical English is used. Two examples in Figure 5 are: to pronounce the word *advantage* correctly, and to use the collocation *view + data* instead of *see + data*.

Great importance is given by learners to being able to build ‘good sentences’, and to maintain the ‘discourse flow’ in oral tasks. In other words, learners point to the ability to create cohesion in discourse. Most errors are, in fact, related to rhetorical marking. For instance, a syntactic construction such as *another issue involves* marks transition in the speech. A common flaw, in turn, is grammatical deviation related to the learners’ use of incorrect verb colligations, eg *say you that...* (Figure 5). Other common grammatical mistakes are *after* (pause) *I show that...*, *my idea is based in...*, etc.

Figure 6 illustrates the percentages of learners’ opinions regarding oral task production. In terms of language command, vocabulary building is seen as a strong demand. The black curve signals this need in the graph at 56 percent. Only the use of visual aids receives a higher percentage: 98.

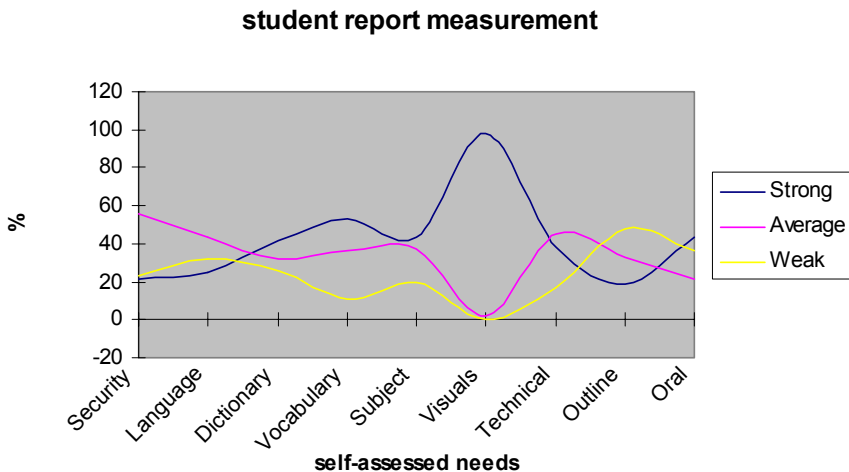


Figure 6: Learner’s self evaluation of needs and lacks in various aspects of the oral presentations

Knowledge of technical concepts and terminology is considered an average need (43 %), less strongly in demand than vocabulary. Experience of subject matter closely parallels this assessment (with 41 % of students judging it as a

strong need, and 40 % as average). Dictionary use is assessed similarly: 42 percent of the learners deemed it of paramount importance for oral tasks, whereas 38 percent considered it an average requirement.

The feedback obtained from task performance points to the learners' recognition and awareness of the chief lexical categories examined in section 3: semi-technical, technical and grammatical. In addition, the contextual features of registers and genres influence lexical behavior, and particular subjects in Business Science and Technology underline this influence.

The high position given to semi-technical lexis by students leads to considering this lexical set as fundamental academic vocabulary. The items are identified in the corpus as the DCL words, namely common core nouns, verbs, adjectives and adverbs. In contrast, the technical language described in the subject-based sets of texts is not seen as crucial for task communication. Lexical profiles in the corpus registers (Table 4) are specialized; nonetheless, the learners tend to already have a suitable command of this terminology. The inclusion of the lexical profiles in the corpus need not be as monitored and regulated as that of semi-technical language.

As a result, I believe that register dimensions should not be explicitly explained in the ESP setting. In contrast, the evaluation of specific language learning should shift to assessing performance in specific academic tasks, where, at a semi-technical level, acquisition of lexico-grammatical units best satisfies communicative needs.

Finally, I consider grammatical word combinations as highly important language, since learners tend to appeal to the need of grammatical command in the tasks, eg a good order in sentences. This group of items comes from the overall frequency word list in the corpus (Table 3). Like the study of semi-technical constructions, the analysis of grammatical collocations leads to examination of key elements in context. Thus, colligations convey grammatical features, eg the preposition *by* + gerund indicates instrumentalization. For ESP tasks, these grammatical combinations become very productive, not only as function words in grammar, but also as discourse markers (see Figure 3).

5 Conclusions

This paper has described the support of learner task performance data for specific corpus analysis. Corpus-based information in ESP, in my view, should depend on learning situation and academic factors, such as students' wants, or even institutional constraints in the types of tasks preferred. The texts recom-

mended or required at the university should thus serve as reference, subject to change according to learning criteria.

Two conditions in the specific setting have been met in my approach: subject-based knowledge, deriving from individual technical proficiency, and language factors, resulting from optimal communication at the academic level. Lexical command is highly significant in both cases, as learners realize the need of specific words to suit specific contexts. Written and oral tasks are main devices that test these learning traits. Good linguistic skills demonstrate lexical knowledge in the tasks, while subject competence contributes to acceptable performance. In this gradual acquisition of effective communicative abilities, realizing lexical behavior is the first step (word reception), and development of items in discourse constitutes active resolution (word production).

Two types of tasks have been described: reading charts and topic presentation. Semi-technical words are highly demanded for both; students feel the need of academic language that is common core in Business texts. Technical items are likewise pinpointed as important; however, since they are better known by students in the area of specialization, subject-specific terms are valued less highly. Instead, semi-technical language is given great consideration for activities involving summarizing and paraphrasing skills. In addition, rhetorical-grammatical devices play a significant role in these exercises. Communicative fluency is achieved through the effective use of grammatical items and discourse markers in oral reports.

In this study, corpus language is exploited and assessed according to how word behavior changes in the specific context. For instance, the differentiation between genres such as academic Textbooks and Conference papers determines the degree of lexical variation for specific purposes. In the process, a representative corpus is essential, which implies that both the subject area and learning situation will work as reference criteria for the selection of content.

This paper has considered the critical ESP factors of language, content, learning interests, and learning situation for the design of a representative corpus of English for Business Science and Technology. The three stages of adopting a method, analyzing results, and assessing language use are clearly linked in this approach. The ESP corpus cannot evolve or be useful without the application of all three phases. In this respect, as Hutchinson and Waters (1987: 107) claim, the ESP learning situation is approached as a process. Likewise, in this context, the design of the corpus should be consistent and coherent with the development of language competence for specific purposes.

Notes

1. Other key thematic areas in Business Science and Technology, such as Accounting or Economics, have been discarded in my corpus, since they offer a smaller focus on the subject area of Business and Information Technology. In other words, they fail to fulfill topic-driven criteria.
2. In my courses, I have dealt less with genre-based language, such as that pointed out in Table 6. The contrast of this data with task application findings has been insufficient, albeit likely to be used for future research. I may expect, however, that a significant percentage of my learners recognize the important use of reporting phrases in written composition to mark reference in discourse (eg *he said that, he added*, etc).

References

- Breland, Hunter M., Robert J. Jones, and Laura Jenkins. 1994. *The College Board Vocabulary Study. College Board Report No 94-4*. New York: College Board Publications.
- Burgmeier, Arline, Gerry Elfred and Cheryl B. Zimmerman. 1991. *Lexis. Academic Vocabulary Study*. Englewood Cliffs: Prentice Hall.
- Conrad, Susan. 1996. Investigating Academic Texts with Corpus-based Techniques: An Example from Biology. *Linguistics and Education* 8: 299–326.
- Curado Fuentes, Alejandro. 2000. *A Lexical Common Core in English for Information Science and Technology*. Cáceres: Universidad de Extremadura, Servicio de Publicaciones.
- Farrell, Paul. 1990. *A Lexical Analysis of the English of Electronics and a Study of Semi-technical Vocabulary*. Dublin: Trinity College.
- Flowerdew, John and Lynn Miller. 1997. The Teaching of Academic Listening Comprehension and the Question of Authenticity. *English for Specific Purposes* 16 (1): 27–46.
- Gains, Jonathan. 1999. Electronic Mail – A New Style of Communication or Just a New Medium?: An Investigation into the Text Features of E-mail. *English for Specific Purposes* 18 (1): 81–101.
- Guest, Michael. 1998. Spoken Grammar: Easing the Transitions. *The Language Teacher Online. The Japan Association for Language Teaching* 22 (6): 35–46. <http://langue.hyper.chubu.ac.jp/jalt/pub/tlt/98/jun/guest.html>.

- Halliday, M.A.K. 1991. Corpus Studies and Probabilistic Grammar. In K. Aijmer and B. Altenberg (eds). *English Corpus Linguistics*. London: Longman, 30–43.
- Halliday, M.A.K. and Ruqaiya Hasan. 1985. *Language, Context, and Text: Aspects of Language in a Social-semiotic perspective*. Geelong, Victoria: Deakin University.
- Huddleston, Robert D. 1971. *The Sentence in Written English*. Cambridge: Cambridge University Press.
- Hutchinson, Tom and Alan Waters. 1987. *English for Specific Purposes. A Learning-Centred Approach*. Cambridge: Cambridge University Press.
- James, Gregory and James Purchase. 1996. *Corpus-based Lexical Analysis of English in Business Studies and Economics*. Hong Kong: Hong Kong University of Science and Technology.
- Kennedy, Christopher and Robert Bolitho. 1984. *English for Specific Purposes*. ELTS. London: Macmillan.
- Lewandowska-Tomaszczyk, Barbara, and Paul Melia (eds). 2000. *PALC '99: Practical Applications in Language Corpora*. New York: Peter Lang.
- Lozano Palacios, Antonio. 2000. *Vocabulario para los estudios de Bibliotecología y Documentación*. Granada: Universidad de Granada.
- Luzón Marco, María José. 2000. Collocational Frameworks in Medical Research Papers: A Genre-based Study. *English for Specific Purposes* 19: 63–86.
- Nelson, Mike. 2000. A Corpus-based Study of Business English and Business English Teaching Materials. Unpublished PhD Dissertation. Turku: University of Turku. <http://www.kielikanava.com/thesis.html>.
- Nunan, David. 1989. *Designing Tasks for the Communicative Classroom*. Cambridge: Cambridge University Press.
- Ooi, Vincent B.Y. 1998 *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.
- Pedersen, James. 1995. The Identification and Selection of Collocations in Technical Dictionaries. *Lexicographia* 11: 60–73.
- Sager, Naomi. 1986. Sublanguage: Linguistic Phenomenon, Computational Tool. In R. Grishman (ed). *Computational Linguistics: An Introduction*. Cambridge: Cambridge University Press, 1–17.
- Scott, Mike. 1997. PC Analysis of Key Words and Key Key Words. *System* 25 (1): 1–13.

- Scott, Mike. 1999a. Reverberations of an Echo. In B. Lewandowska-Tomaszczyk and P.J. Melia (eds), 49–68.
- Scott, Mike. 1999b. *WordSmith Tools 3.0*. Oxford: Oxford University Press.
- Starfield, Sue. 2001. ‘I’ll Go with the Group’: Rethinking ‘Discourse Community’ in EAP. In J. Flowerdew and M. Peacock (eds). *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press, 132–148.
- Tribble, Chris. 1997. Improvising Corpora for ELT: Quick and Dirty Ways of Developing Corpora for Language Teaching. In B. Lewandowska-Tomaszczyk and P.J. Melia (eds), 106–117.
- Trimble, Louis. 1985. *English for Science and Technology: A Discourse Approach*. Cambridge: Cambridge University Press.
- WWW Resources in Economics*. 2001. WebEC. Helsinki: Faculty of Social Sciences, University of Helsinki. <http://www.helsinki.fi/WebEc/>.

