

**UNIVERSITÀ CA' FOSCARI VENEZIA
CENTRO LINGUISTICO INTERFACOLTÀ**

**ADERNI DI RICERCA
DEL CLI**

A large, bold, black graphic logo consisting of a capital 'G' and a lowercase 'i'. The 'G' is a thick, curved shape. The 'i' is a vertical bar with a solid black circle above it. The logo is centered on the page.

**FOR FOREIGN T
RTIFICATION P**

Maria Rees

The average length of the texts is 400-500 words, but the Corpus includes both shorter and much longer texts. The team took particular care to ensure that the texts were equally distributed over the five general areas chosen at the beginning of the project. Classifying the material proved to be a difficult task as some of the texts covered more than one of the topic areas chosen at the outset.

The general level of difficulty is B1/B2, though the formal register of the working papers and abstracts means that the level is not always homogeneous.

Corpus analysis

The Teaching corpus

The Teaching corpus was analysed twice using the *Web Vocabulary Profiler* (http://www.er.uqam.ca/nobel/r12707texttools/web_vp.html). This profiler, which was originally developed by Nation and Laufer, is maintained by Tom Cobb at the Université du Québec à Montréal (uqam-canada). The *Profiler* makes it possible to analyse the percentages of lexis present in a text on the basis of the *General Service List (GSL)* and the *Academic Word List (AWL)*. It is also possible to calculate the percentage of words in the *Off List*, which contains words that appear in neither the *GSL* nor the *AWL*.

Table 1. Percentages taken over the whole Teaching Corpus (11000 words).

GSL1-1000	GSL	AWL	Off List	GSL+AWL
71.50	6.32	9.21	12.97	87.03

The analysis confirmed that the teaching corpus has very similar percentages to academic texts in general.¹⁹

The Off List

Although the overall figure for the Off-List words in the Teaching corpus as a whole was 12.97 per cent, there are interesting variations to be observed in the single texts, where the range is 5-20 per cent.

¹⁹ See Coxhead (2002:76)

Maria Rees

of Economics, Business and Trade, are among the most frequent 1000 words in English: *company law, interest rate, union member, stock market, pay deal, record growth, trade fair, fair trade, supply side, store chief, capital gains tax, company performance, capital flow, expense account, duty free, share price.*

Conclusion

There is strong evidence in the current literature to suggest that small corpora can be used to identify and quantify the essential lexis of particular fields or subject disciplines. Identifying the words that are most likely to be useful to learners is of crucial importance in ESP/LSP. Focussing the attention of learners on lexical items that have been shown to occur frequently has positive repercussions on vocabulary acquisition as well as on learner motivation.

Quantifying essential lexis makes it possible to set realistic learning goals for vocabulary acquisition in situations where time is at a premium. As the amount of time given to language teaching in the majority of non-linguistic faculties in Italian universities has diminished in recent years, it has become increasingly important to target the specific lexical needs of learners as rapidly and efficiently as possible.

This article has described the various stages leading to the creation of a corpus-based word list of essential lexis for Foreign Trade. The project involved teachers, researchers, software designers and people working in the field of Foreign Trade. Their combined experience and expertise ensured that each stage of the project was followed through with great care and attention to detail.

The research carried out in the course of this project confirms that well-designed *ad hoc* corpora occupy a crucial role in enabling ESP course designers and writers of teaching and testing materials to focus more clearly on the lexical needs of specific learners.

The Lexis of Foreign Trade

Maria Rees

Part I An Approach to Corpus Creation

The project of lexical analysis described in this article was undertaken at the University of Venice as a preliminary phase in the creation of the Treviso Certificate of Foreign Trade. Unlike the examinations certifying general language competence, the Treviso Certificate tests the receptive language skills of learners whose specific lexical needs go beyond the confines of General English.

Recent research in the field of lexis definition, acquisition and use increasingly points to the fact that what is crucial for learners in specific subject areas or disciplines is not so much a large amount of vocabulary but rather a small number of carefully selected words. A small, subject-specific technical (or sub-technical) vocabulary has been shown to be more useful for the understanding of specialist texts than a higher number of general lexical items¹.

The need to identify the areas of technical vocabulary specific to the field of foreign trade led to the compilation and subsequent analysis of two *ad hoc* corpora and made it possible to create an essential word list to be used as a lexical reference for the authors of the Treviso Certificate.

The first part of this article focuses on the definition of specialist lexis, and provides a general overview of corpora and corpus-based word lists. The second part gives a detailed account of the various procedures that were followed and the decisions that were made during the various stages of the project.

Specialist lexis

The correct use of highly technical lexis or terminology is what distinguishes the subject specialist from the non-specialist.

¹ Nation (2001: 9).

Maria Rees

But the bewilderment one feels when “blinded by jargon” does not mean that terminological or technical words are intrinsically difficult to learn.

Many everyday conversations are peppered with technical terms from specific specialist fields. However, the use of words such as *defibrillation*, *anaemia*, *oncology* or *goitre* does not necessarily go hand in hand with knowledge of the medical domain in which they are used. Knowledge of such specific lexis is acquired either through personal interest or personal need. The acquisition and correct use of specialist terminology takes place in a particular context and is closely linked to the extralinguistic knowledge a person has of a particular field or discipline. It has been suggested (Fox, 1998:27) that this type of vocabulary “can be acquired when needed”.

Specialist or technical vocabulary does not only consist of terminological words. Nation (2001: 198-9) has indicated that there are four categories of technical vocabulary. The first category includes words that are almost exclusively typical of a particular field or subject area. In the case of Foreign Trade, this category could include words such as *demurrage*, *dunnage* and *cabotage*. This type of terminology is not only relatively infrequent but also has a high degree of topic specificity, which means that anyone needing to use these terms would already be familiar with the professional work context (in this case, shipping or freight forwarding) in which these words are used.

For the learner, terminological words may present relatively few difficulties because they almost always have a single meaning, whereas more commonly-used English words tend to have multiple meanings (Nagy, 1997: 65). Words with multiple meanings are either polysemes or homonyms. In the case of polysemic words, there is either an obvious common core meaning (the *wall* of a house / a cell *wall*) or a common etymology that may or may not be immediately evident: *bill*: ‘statement of charges’, ‘draft law’, ‘piece of paper money’ or ‘poster’.

In the case of homonyms, there is no related meaning: *fair* (adj.): 'just', 'blonde'; *fair* (n): 'amusement park', 'event where people display and sell goods'.²

The common core meaning in polysemic words may make them easier to learn, but homonyms can be a source of considerable frustration to teacher and learner alike.

The word forms in Nation's second category can still be considered technical vocabulary, but they are less restricted in their use as they are not exclusive to a particular area of specialisation. However, the meanings of these words change according to the specific field in which they are used. For Foreign Trade, such words include *factor*: 'an agent who receives merchandise'; *average* (n): 'damage' (in marine insurance); *duty*: 'a customs tax'. There is evidence to suggest that the words in this category require particular attention.³

Nation's third category includes word forms that are even less restricted. Words such as *assessment*, *diversion* and *salvage* have very precise meanings within the specialist field, but they can still be understood by non-specialists.

The word forms in the fourth category present a minimal degree of specialisation. Although people outside the field have few problems in using these words correctly, a specialist will provide a more precise and more comprehensive definition. In the field of Foreign Trade, *goods*, *licence* and *discount* fall into this final category.

The subdivision of technical vocabulary illustrated above shows clearly that terminological words (i.e. words in the first category) are likely to be acquired easily if and when the learner has direct experience of working in the field.

² Distinguishing polysemes from homonyms may be useful in some teaching contexts (see Mihwa Chung & Nation, 2003) but finding the core meaning and exploiting it can be a time-consuming process that may ultimately be of little benefit to learners. Laufer (1997: 52) rightly points out that the 'learnability' of a word is not going to be affected by the learner's ability to distinguish between a polyseme and a homonym.

³ See Trimble (1985) and Bensoussan & Laufer (1984).

Maria Rees

The word forms in the second and third categories, which can also be defined as sub-technical lexis, are likely to be encountered more frequently and in a wider range of areas of specialisation.

General reference corpora

At its most basic level, a corpus is a collection of written texts and / or recordings (or transcripts) of authentic spoken language. The main difference between a corpus and a random collection of texts is that corpora — whatever their size — are compiled according to specific, well-defined criteria and are stored electronically.

General reference corpora are compiled and analysed in order to make general observations about the language. The first phase in the history of large-scale corpus compilation goes from the early 1960s to the late 1970s. The *Brown Corpus* (Francis and Kučera) and the *Lancaster-Oslo-Bergen (LOB) Corpus* (Johansson) both contained 1 million words of written text, but the *Brown Corpus* dealt only with examples of American English. Transcripts of spoken English were subsequently included in the *London-Lund Corpus*.

In the 1980s and 1990s rapid advances in computer technology made it possible to devise complex and increasingly sophisticated methods for annotating words. Texts can now be tagged according to phonetics, syntax, word class, orthography, etc. Technological progress was accompanied by an equally rapid practical application of corpus data - first and foremost in lexicography; and successively in grammar and reference books. Since the late 1990s, the major learners' dictionaries on the market have all been corpus-based (Carter, 2001). Thanks to these developments, learners' dictionaries now include information about collocations and frequency that would previously have been impossible to provide.

The *COBUILD/Bank of English* is by far the largest corpus, but some smaller corpora, such as the *British National Corpus (BNC)* and the *International Corpus of English*, can also claim to be representative of general language usage.

When a corpus is used to make generalised claims about a language, the larger it is the better. However, not all corpora fall, or need fall, into a

general purposes category. There is strong evidence to suggest that sub-corpora and small *ad hoc* corpora can provide useful information about language use and lexical patterns in specific fields of study. In LSP, where the lexical needs of learners are of crucial importance, small corpora that have been compiled according to well-defined criteria not only provide examples of authentic usage, but also make it possible for teachers, learners and materials writers to access invaluable information about the lexis of a specific field or discipline.⁴

Whatever the size of a corpus, certain careful decisions have to be made before starting work on its compilation. These decisions relate first of all to the purpose of the corpus and to the types of text it will include. It also has to be decided whether the corpus will focus on written or spoken language, or on both. Deciding whether the corpus should be finite or open depends on the purpose of the corpus and the ultimate use that will be made of the data it provides. While a finite corpus is ideal for an analysis of the works of a dead writer, an open (Monitor) corpus would make it possible to add the works of a living author as and when they appear.

General and academic word lists

Corpus-based word lists fall into two broad categories: lists of general purpose words, and lists containing vocabulary specific to a particular subject area or range of disciplines.

The most influential list of general purpose words is undoubtedly West's 1953 *General Service List (GSL)*. Despite its age and the shortcomings that inevitably derive from the fact that it was based on what is a small (by today's standards) corpus of 5 million words, the *GSL* has not been bettered.

Although frequency and range were certainly taken into consideration in compiling the *GSL*, West also considered other criteria, which included whether a word was easy or difficult to learn, whether it was emotional or neutral and, above all, whether it was necessary.⁵

⁴ See Gavioli (2002), Aston (2002), Bowker and Pearson (2002).

⁵ See Nation & Waring (1997) for an illustration of West's concept of "necessity".

Maria Rees

The *GSL* contains the most frequent word families that make up the general purpose vocabulary needed by learners before they can be expected to cope efficiently with more specialist texts.

It is generally accepted that a learner will need to know at least 2000 word families to cope with general texts⁶. The first 1000 word families account for 73.5-84.3 per cent of words in any kind of written text. The second 1000, however, increase the number of known words by a mere 4.6-6 per cent. In academic texts, the figures are 77 per cent and 5 per cent respectively.⁷ Even with this sort of coverage, one word in five will still be unknown. At a certain point, therefore, the specific needs of learners have to be taken into consideration⁸ as it is clearly not a question of how many words the learner needs, but which ones are likely to be most useful.

An early, systematic attempt to isolate the vocabulary needed by those intending to study in universities where English is the medium of instruction led to the compilation in 1984 of the *University Word List (UWL)*⁹. The *UWL* was the result of a combination of four separate studies carried out in the 1970s.¹⁰ The 836 word families it contains were used as the basis for the design and development of teaching and testing materials. Despite the wide use and influence of this list, its coverage of academic texts in general was seen to be insufficient.

This inadequacy was due in part to the very high proportion of scientific texts, which made the *UWL* unrepresentative of academic writing as a whole, but also to the size of the corpora on which it was based and on the different selection criteria adopted by the original researchers.

⁶ Research has shown that 3000 word families are needed before learning from context is possible with an unsimplified text (Liu Na & Nation, 1985).

⁷ Nation (2001: 16).

⁸ Willis (1990: 47)

⁹ Xue & Nation (1984)

¹⁰ Champion & Elley (1971); Praninskas (1972); Lynn (1973); Ghadessy (1979).

A more recent word list was produced by Coxhead in 1998. The difference between the *UWL* and Coxhead's *Academic Word List (AWL)* lies not so much in the slightly better coverage of academic texts provided by the latter, but rather in the size of the corpus (3,500,000 words) on which it was based, as well as on the range of disciplinary areas it covers (Arts, Commerce, Science and Law).

There are indications (Nation, 2001: 17-18) that the addition of the 570 headwords of the *AWL* to the first 2000 word families in the *GSL* increases the number of "known" words in an academic text by 9-12 per cent. This gain is much greater than that obtained by adding the words in the 2000-3000 group of the *GSL* to the learner's vocabulary.

Part II Towards a Corpus for Foreign Trade

The test-takers' assumed knowledge

General lexis

As most of the students attending the three-year degree course in Foreign Trade on the Treviso campus of the University of Venice will have studied English at school for at least five years, their level of linguistic competence can generally be assumed to be A2. The students take institutional English courses in the first two years of their undergraduate studies, so it can safely be assumed that by their third year they will have progressed to the Threshold Level (B1) as defined in the *Common European Framework*.

The lexical content of the B1 level only indicates broad topic areas for which language learners are expected to have vocabulary knowledge. The *Cambridge English Lexicon* (Hindmarsh, 1980) was used to find a more precise list of the general lexis that could be assumed to have been acquired. The *Lexicon* is a word list drawn up for use "by text editors, curriculum planners, examination setters and course writers" for the University of Cambridge ESOL examinations. The Preliminary English Test (B1 level), for example, explicitly uses a list of the most frequent 2387 words of English according to Hindmarsh's *Lexicon*; the First Certificate of English (B2 level) uses all the 4500 lexical items included in the *Lexicon*. Compiled in 1980, the *Lexicon* is already somewhat dated; for example, it does not contain lexis connected with

Maria Rees

recently-developed computer technology (*Internet, monitor, browser, server, etc.*). It did, however, serve as a basis for assumptions about the General English knowledge of the test takers.

Specific lexis

Most of the text books used in the Foreign Trade course were produced as part of an ongoing ESP project at the University of Venice. During their first year, students throughout the Faculty study *English for Economics* (Rees, 2000). *English for Economics* is based on a small *ad hoc* corpus and focuses on the use of specific lexis. The book contains authentic reading material dealing with some of the most pressing current issues in Economics, such as microfinance, fair trade, ageing societies, social security and pension reform, emerging economies and sustainable development. As well as an extensive series of exercises based on contrastive aspects of grammar and lexis, *English for Economics* also includes exercises in word formation and morphology. The reading passages are accompanied by detailed English/Italian glossaries.

The second-year English course in Treviso is based on *New Insights into Business* (Tullis & Trappe, 2000) and is more specifically business oriented. The topics covered in this course book include company organisation, advertising and marketing as well as issues related to the recruitment and management of human resources. As part of their second-year English course, the students also use two of the books in the *Professional English* series, also developed as part of the ESP project at the University of Venice, to concentrate on sector-specific vocabulary.¹¹

This gave the team a very clear idea of the general Economics and Business English terms the students had encountered in the first two years of their undergraduate studies, even before considering the additional language knowledge individual students may have acquired during their internships abroad.

¹¹ The series includes *Furniture* (Gebhardt, 2002); *Clothing* (Titmarsh, 2003); *Footwear* (Pagotto, 2002); *Tourism* (Pagotto, 2003).

As the Treviso Certificate is a test of receptive skills (reading and listening) it was decided to administer an extremely rapid vocabulary test in order to check the test-takers' receptive lexical knowledge.

This test included the general Economics and Business terms found in the first 2000 most frequent words in West's *General Service List* and in *English for Economics*; common English trade terms taken from *New Insights into Business*; English trade words commonly used in Italian, selected from the De Mauro *Dizionario della Lingua Italiana*, and sector-specific vocabulary related to products and production that appear in the first 2000 most frequent words of the *GSL* and the *Professional Series* of textbooks. The typologies of items used were matching English and Italian terms, and translation of English terms into Italian. The test was administered to 170 students in the first and second years of study and the results were analysed. This rapid, but highly focused, exercise confirmed that the group of test-takers had acquired a knowledge of what could be defined as general Economics and Business lexis.

Compiling a specialised vocabulary

One school of thought¹² holds that there are two ways to compile specialised vocabularies: one is by making a frequency count using a specialised corpus; the other involves asking "experts in the field" to identify and collect the vocabulary they consider to be most relevant. In the case of Foreign Trade, the second approach might have proved problematic in that an "expert" working in the shipping department of a company might well have produced a valid and useful list for that particular area of Foreign Trade, but the vocabulary might have been too biased towards the 'business' side of trade.

Asking the lecturers on the Treviso campus might have seemed a sensible option at first sight. However, there are no professors of "Foreign Trade", nor is there a subject of that name. The subjects studied in the first year of the degree course are core subjects common to the entire Faculty of Economics. The 'foreign' or 'international' aspects only come into play later on in the course.

¹² See Nation (2001: 17)

Maria Rees

ESP teachers are not usually subject specialists and would not therefore be classed as “experts in the field”.

Even though the members of the team working on the lexical project already had considerable experience in the compilation and analysis of *ad hoc* corpora for Economics and Business, further research into the vocabulary used by professionals operating in the field of Foreign Trade was felt to be indispensable.

An extensive Internet search was carried out in order to identify and analyse the lexis used in the world of trade and trading. Four online lists of international trade terms were examined: *The Glossary of Trade and Related Terms*¹³, *The Language of Trade Glossary*¹⁴, *The Glossary of Trade Terms*¹⁵ and *The Dictionary of International Trade Terms*¹⁶. The declared aim of these lists is to “demystify” the language of trade for experts and non-experts alike. Of the four lists, *The Language of Trade Glossary* is the only one that does not include definitions. By far the most interesting and detailed list is *The Dictionary of International Trade Terms*, which contains approximately 1000 definitions. This list also includes a large number of acronyms (*WTO, GATT, ID*) as well as all the Incoterms (*CIF, FOB, EXW, etc.*).¹⁷ The entries for *trade* in this list include *trade acceptance, trade deficit, trade name, trade promotion, trade surplus, trade terms* and *terms of trade*.

Although *The Dictionary of International Trade Terms* provides clear definitions of this type of lexis, it was too long to be considered as a possible word list for the potential test-takers. The *Dictionary* is aimed at an international audience and is intended to be used mainly by people already working in the field. It was felt that Italian university students would benefit more from a word list that was tailored to their specific needs.

¹³ Available from. www.carleton.ca/ctpl/Essential/glossary.htm.

¹⁴ Available from: usinfo.state.gov/products/pubs/trade/glossz.htm#x.

¹⁵ Available from www.alberta-canada.com/altaport/glossary.cfm

¹⁶ Available from www.itds.treas.gov/printglossaryfrm.html.

¹⁷ The Incoterms are standardised terms used in the trade and exchange of goods. They were devised in 1936 by the International Chamber of Commerce with the aim of avoiding disputes between buyers and sellers. They were last revised in 1999.

The Foreign Trade corpora

Rather than rely on what might have proved to be an overly subjective list of words made by experts in the field, or on the intuition and introspection of experienced teachers in order to gather and classify the lexis typical of the field of Foreign Trade, the team decided that a corpus-based word list would offer a greater degree of reliability. The decision was made to compile two corpora: a small finite corpus to be used both as the basis of a pre-Certification refresher course and for the preparation of a draft word list; and a larger open corpus for the purpose of comparing data from the small corpus. An open corpus was felt to be better suited to the dynamic nature of the field of Foreign Trade, which like so many parts of Business and Economics is a sector in constant evolution. One important advantage of an open corpus is that obsolete information can be removed when necessary and replaced with more up-to-date material.

During the initial round of discussions the members of the team involved in the project made a list of fifteen general topics for the texts to be selected for each of the two corpora. The number of topics was later reduced to five, each of which has between two and seven sub-topics. The decision to reduce the number was made in order to focus in greater detail on the issues most closely related to Foreign Trade. The final choice of topics is listed below:

1. WORLD TRADE

Free trade: liberalisation, protectionism, globalisation

Trade agreements

Trade organisations and institutions

Customs

Law: intellectual property, trademarks, patents.

2. COMPANIES AND MARKETS

Kinds of companies and alliances

Kinds of market

Maria Rees

3. PRODUCTION, PRODUCTS & SERVICES

Agriculture

Industry

Tertiary production: handicrafts, fashion, furniture, footwear

Services: Tourism, telecommunications, R & D

Quality control

4. MARKETING, SALES AND DISTRIBUTION

e-commerce

Customers and customer service

Brands

Trade fairs

Advertising

Cross-culture business

Transport, shipping and freight

5. CURRENT ISSUES

Fair trade (with least developed countries)

Consumers and consumer protection

Workplace issues (equal opportunities, unions, discrimination,
work/life balance, social responsibility)

The Teaching Corpus

The purpose of the teaching corpus was to design material to be used during a short pre-Certification course for the test-takers. Online self-access material was also designed so that students could prepare for the test format and for test navigation.

Size

The teaching corpus contains 20 texts, all taken from Internet, for a total of just over 11000 running words. The texts vary in length from

Maria Rees

The teaching corpus, which has since been published in book form (Rees *et al.*, 2004), includes an extensive series of comprehension exercises and vocabulary development. Each passage is followed by notes explaining the difference between the general and specific uses of some of the words in the reading passages (e.g., *to peg*: 'to fix prices or an exchange rate', as contrasted with the meaning 'to put clothes on a washing line'). Particular attention is given to the English/Italian glossaries that accompany each text. The glossaries are divided into general vocabulary and vocabulary that is specific to the field. The words included in the specific glossaries formed the basis of the first draft word list for Foreign Trade.

The Monitor corpus

Parallel to the creation of the teaching corpus, other team members, who had direct experience of working in Foreign Trade, were involved in the creation of the Monitor corpus. The initial phase of painstaking selection and classification of texts took place over a six-month period.

In line with the overall aims and purposes of the Treviso Certificate, it was decided to use texts from Internet, given that the Web is now the prime source of information for those wishing to keep abreast of current issues and trends in the field.

The team first made a list of the main specialist newspapers and magazines available online. The main sources were:

<i>Financial Times</i>	http://ft.chadwyck.co.uk
<i>The Economist</i>	http://www.economist.co.uk
<i>BBC Business News</i>	http://news.bbc.co.uk
<i>Business Times Online</i>	http://business.timesonline.co.uk

A preliminary selection of texts was made by using key words to search the archives of the news providers. Some of the material found during the search was rejected because it was not considered to be sufficiently related to Foreign Trade. Particular attention was paid at all stages of the selection process to the level of difficulty of the language used in the texts.

Given that students of Foreign Trade also need to find reference information and academic material on the various subjects they study, the team also added three of the major specialist Internet sites frequently consulted by people in the field of general Economics:

The World Trade Organisation <http://www.wto.org>

The World Bank Group <http://www.worldbank.org>

The International Monetary Fund <http://www.imf.org>

The other major sites consulted were international commercial and governmental organisations.

UK Government website <http://www.tradepartners.gov.uk>

Italian institute for Foreign Trade <http://www.italtrade.com>

Italian Trade Commission Webpage <http://www.subconitaly.com>

The official EU website <http://www.eu.europa.int>

The International Chamber of Commerce <http://www.iccwbo.org>

The links provided in each of the websites listed above enabled the team to find other sources of material.

Genres

In line with the criteria and the format of the test, three main genres were selected:

1. Short news items, taken from online news providers. In most cases these consist of a headline followed by a short summary of the news item.
2. Newspaper articles. These are longer news items and contain figurative language and comments.
3. Reports, articles and working papers. These are characterised by a more formal register and deal more with the theoretical aspects of Economics and international events. Some of the material also includes tables and graphs.

No alterations were made to the texts included in the Monitor corpus, and the original length of each text was maintained as far as possible.

Maria Rees

The average length of the texts is 400-500 words, but the Corpus includes both shorter and much longer texts. The team took particular care to ensure that the texts were equally distributed over the five general areas chosen at the beginning of the project. Classifying the material proved to be a difficult task as some of the texts covered more than one of the topic areas chosen at the outset.

The general level of difficulty is B1/B2, though the formal register of the working papers and abstracts means that the level is not always homogeneous.

Corpus analysis

The Teaching corpus

The Teaching corpus was analysed twice using the *Web Vocabulary Profiler* (http://www.er.uqam.ca/nobel/r12707texttools/web_vp.html). This profiler, which was originally developed by Nation and Laufer, is maintained by Tom Cobb at the Université du Québec à Montréal (uqam-canada). The *Profiler* makes it possible to analyse the percentages of lexis present in a text on the basis of the *General Service List (GSL)* and the *Academic Word List (AWL)*. It is also possible to calculate the percentage of words in the *Off List*, which contains words that appear in neither the *GSL* nor the *AWL*.

Table 1. Percentages taken over the whole Teaching Corpus (11000 words).

GSL1-1000	GSL	AWL	Off List	GSL+AWL
71.50	6.32	9.21	12.97	87.03

The analysis confirmed that the teaching corpus has very similar percentages to academic texts in general.¹⁹

The Off List

Although the overall figure for the Off-List words in the Teaching corpus as a whole was 12.97 per cent, there are interesting variations to be observed in the single texts, where the range is 5-20 per cent.

¹⁹ See Coxhead (2002:76)

The *Off List* was examined carefully in order to find out exactly what type of words it contained. As might be expected, the *Off List* was mainly made up of the names of countries, people, companies, etc., as well as a high proportion of adjectives of nationality. It also contained brand names and names of specific products (*Marlboro, aspirin, cocoa*); acronyms and abbreviations (*ICC, FX, ad., etc.*); weights and measures (*tonnes, lbs, bushel*); names of chemical substances (*asbestos, formaldehyde, polyethylene*). As well as a fairly high number of compound words (*shareholders, waterways, workforce*) there are several examples of prefixed words such as *transnational, multilateral, intergovernmental*.

However, the *Off List* also contained some words that can be considered specific to the field of trade and commerce. The majority of the following words were subsequently included in the word list.

<i>arbitration</i>	<i>barrier</i>	<i>brand</i>	<i>cargo</i>
<i>competitors</i>	<i>consignment</i>	<i>defer</i>	<i>conveyance</i>
<i>discharge</i>	<i>exportation</i>	<i>fleet</i>	<i>forwarder</i>
<i>formalities</i>	<i>freight</i>	<i>import</i>	<i>pallet</i>
<i>protectionism</i>	<i>storage</i>	<i>treaty</i>	<i>warehouse</i>

The Monitor Corpus

Analysis of the Monitor Corpus was carried out using in-house profiling software that was designed and developed by one of the experts at the University of Venice Language Centre (CLI) in the later stages of the project. The TM Corpus Profiler²⁰ will ultimately make it possible to consider frequency, range and distributional consistency much more precisely²¹.

²⁰ The TM Corpus Profiler was devised and developed by Gino Schiavinato of the Centro Linguistico Interfacoltà (CLI).

²¹ Some preliminary results are in an unpublished article by Loretta Piai, who was responsible for the compilation and analysis of the Monitor Corpus.

Maria Rees

Frequency and range

The frequency with which a content word occurs within a single text generally gives quite a precise idea as to the subject of that particular text. However, the frequency with which a content word occurs within a single text does not necessarily mean that the word is suitable for inclusion in a word list of essential lexis. For example, in one of the texts in the Teaching corpus, the word *bead/beads* occurs 20 times in a passage of 615 running words. Although this high occurrence indicates that the main topic of the text has something to do with beads, the word does not appear anywhere else in the two corpora. In other words, its range is too restricted for it to be included in a word list of essential lexis.

Range is clearly important, especially when an entire corpus is taken into consideration. Particularly narrow range, when set in relation to a word's frequency of occurrence, can determine whether a word is terminological (Nation, 2001: 200). However, if the corpus is related to a single discipline and / or to a limited variety of text types, range may assume a different degree of importance.

What emerged from the analysis of the Monitor corpus was that there were very strong similarities, though slightly different degrees of frequency and distributional consistency, among the first 200 words in each corpus. These findings confirmed the validity of the work carried out on the teaching corpus and made it possible to complete the editing process of the word list for Foreign Trade.

Editing the word list

The first draft word list was based on the specific terms in the glossaries in the teaching corpus. The decision to exclude certain words had already been made. The words excluded *a priori* included acronyms, Incoterms, names of countries and adjectives of nationality, brand names and proper nouns. Other high-frequency words, such as *million*, *billion*, *dollar* and *euro* were also excluded.

The preliminary draft list had included other high-frequency words (such as *agriculture*, *inform*, *innovate*, *quality*, *international*, and *feedback*), which were removed from the list on the grounds that they

would present no problems of comprehension for native speakers of Italian with a basic knowledge of General English.

During this initial editing stage, some other words were excluded because they were felt to be too specific for the needs of the potential test-takers. These words included *barge*, *gondola* (as used in rail transport), *stock-picked* (adj.) and *dry-box sea container*.

Word families

The team opted for word families as lemmatisation did not seem necessary or worthwhile for the immediate purposes of the word list of essential lexis.

A word family, as defined by Nation (2001: 8) includes the headword, inflected forms and closely related derived forms. These closely related forms include the headword plus affixes such as *-ness*, *-ition*, *un-* and *-ment*. Although prefixes and suffixes are acquired at different stages of the learning process, it can be assumed that learners at B level would know the majority of the most frequent affixes. The word families in the Foreign Trade word list also include what Nation (2001: 266) calls "less transparent derivatives". Thus, the word family for *compete* includes *competes*, *competing*, *competed*, *competition*, *competitor/s*, *(un)competitive* and *competitiveness*.

Final considerations

The 200 word families in the Word List for Foreign Trade, which can be found at the end of this article, make up the essential lexis for students of Foreign Trade and for the potential test-takers of the Treviso Certificate. The List does not contain terminological words as these are acquired and used within a specific professional context. Many of the words in the List are also found in the first 1000 most frequent words of the *General Service List*. One reason why so many of these words appear in this corpus-based Foreign Trade List is due to the fact that so many of them change their meaning according to the different contexts in which they are used. Another reason for not excluding these words from the final Word List is that a substantial proportion of the *GSL* 1000 group are used as noun modifiers: for example, all the words in the following expressions, which can clearly be defined as lexis specific to the fields

Maria Rees

of Economics, Business and Trade, are among the most frequent 1000 words in English: *company law, interest rate, union member, stock market, pay deal, record growth, trade fair, fair trade, supply side, store chief, capital gains tax, company performance, capital flow, expense account, duty free, share price.*

Conclusion

There is strong evidence in the current literature to suggest that small corpora can be used to identify and quantify the essential lexis of particular fields or subject disciplines. Identifying the words that are most likely to be useful to learners is of crucial importance in ESP/LSP. Focussing the attention of learners on lexical items that have been shown to occur frequently has positive repercussions on vocabulary acquisition as well as on learner motivation.

Quantifying essential lexis makes it possible to set realistic learning goals for vocabulary acquisition in situations where time is at a premium. As the amount of time given to language teaching in the majority of non-linguistic faculties in Italian universities has diminished in recent years, it has become increasingly important to target the specific lexical needs of learners as rapidly and efficiently as possible.

This article has described the various stages leading to the creation of a corpus-based word list of essential lexis for Foreign Trade. The project involved teachers, researchers, software designers and people working in the field of Foreign Trade. Their combined experience and expertise ensured that each stage of the project was followed through with great care and attention to detail.

The research carried out in the course of this project confirms that well-designed *ad hoc* corpora occupy a crucial role in enabling ESP course designers and writers of teaching and testing materials to focus more clearly on the lexical needs of specific learners.

FOREIGN TRADE

WORD LIST

accede	collect	dispute
acknowledge	commerce	distribute
acquire	commodity	domestic
advance	company	durable
advise	compete	duty
affect	comply	economy
agree	concern	employ
amend	consign	endorse
apply	consume	entrepreneur
arbitrate	container	environment
arise	contract	estimate
arrange	convert	exchange
asset	core	exhibit
assure	cover	export
available	crop	face
average	currency	facility
bank	customer	fair
bill	customs	fee
brand	deal	finance
broker	debt	firm
bulk	declare	flow
business	decline	foreign
campaign	defer	former
cargo	delay	forward
carry	deliver	free
cash	demand	freight
challenge	develop	gain
charge	discharge	global
clear	display	goods

Maria Rees

grant	manage	property
grow	manufacture	protect
handle	mark	provide
handle	market	purchase
haul	means	rail
hazard	merchandise	range
health	mislead	rate
hold	money	receive
household	negotiate	record
hub	output	regulate
import	overseas	require
improve	own	restrain
income	pack	restrict
industry	pallet	route
insure	partner	rule
invest	party	safe
issue	patent	safeguard
joint	pay	schedule
label	penalty	scheme
labour	perform	secure
lack	perish	sell
launch	policy	service
legal	pollute	set
lend	practise	settle
liberal	premium	share
license	price	shift
load	process	ship
mail	produce	skill
major	profit	specify

spread	takeover	transport
stock	tariff	truck
store	term	trust
strategy	threat	undertake
supply	trade	value
support	trademark	warehouse
sustain	transaction	work
switch	transfer	worth

Maria Rees

BIBLIOGRAPHY

ASTON, G., 1997, "Small and large corpora in language learning", in B. LEWANDOWSKA-TOMASZCZYK and P.J. MELIA (eds.), *PALC 97 Practical applications in Language Corpora*, Lodz University Press, pp. 51-62.

ASTON, G., 2002, "The Learner as Corpus Designer", in B. KETTEMANN and G. MARKO (eds.), pp. 9-25.

BENSOUSSAN, M. and LAUFER. B., 1984, 'Lexical guessing in context in EFL reading comprehension', *Journal of Research in Reading*, 7, 15-32.

BORN-LECHLEITNER, I., 1992, "Words, words, words...dealing with vocabulary in ESP", in *Vienna Working Papers*, Vol.1, No. 2, May, pp. 65-80.

BOWKER, L. and PEARSON, J., 2002, *Working with Specialised Language: a practical guide to using corpora*, London, Routledge.

BURNARD, L., 2002, "The BNC: Where did we Go Wrong?", in B. KETTEMANN and G. MARKO (eds.), pp. 51-70.

CARTER, R. and NUNAN, D., 2001, *Teaching English to Speakers of Other Languages*, Cambridge University Press.

CARTER, R., 2001, "Vocabulary", in R. CARTER and D. NUNAN (eds.), pp. 42-47.

COOK, G., 2003, *Applied Linguistics*, Oxford Introductions to Language Study, Oxford University Press.

COXHEAD, A., 2000, "A New Academic Word List", *TESOL Quarterly*, 34 (2), pp. 213-238.

COXHEAD, A., 2002, "The Academic Word List: A Corpus-based Word List for Academic Purposes", in B. KETTEMANN and G. MARKO (eds.), pp. 73-89.

Dictionary of International Trade Terms, <http://www.itds.treas.gov/printglossaryfrm.html>

DUDLEY-EVANS, T., 2001, "English for Specific Purposes", in R. CARTER and D. NUNAN (eds.), pp. 131-136.

DYCK, G.N., 1999, "Concordancing for English Language Teachers", Paper presented at the annual session of TESL Manitoba.

FORTIS, D., 2001, *Il Testing Computerizzato: una prova oggettiva di lingua inglese per studenti di Economia*. Unpublished thesis, Università "Ca' Foscari" di Venezia.

FOX, G., "Using corpus data in the classroom", in B. TOMLINSON (ed.), pp. 25-43.

GAVIOLI, L., 2002, "Some thoughts on the problem of representing ESP through small corpora", in B. KETTEMANN and G. MARKO (eds.), pp. 293-303.

Glossary of Trade and Related Terms, <http://www.carlton.ca/ctpl/Essential/glossary/htm>.

Glossary of Trade Terms, <http://www.alberta-canada.com/altaport/glossary/cfm>.

GOULDEN, R., NATION, I.S.P. and READ, J., 1990, "How Large Can a Receptive Vocabulary Be?", *Applied Linguistics*, Vol. 11, No. 4.

HAMMOND, J. and DEREWIANKA, B., 2001, "Genre", in R. CARTER and D. NUNAN (eds.), pp. 186-193.

HAMP-LYONS, E., 2001, "English for Academic Purposes", in R. CARTER and D. NUNAN (eds.), pp. 126-130.

HARWOOD, N., 2002, "Taking a lexical approach to teaching: principles and problems", *International Journal of Applied Linguistics*, Vol. 12, No. 2.

HEWINGS, M., 2002, "A History of ESP through *English for Specific Purposes*", *ESP World* - Web-based Journal.

HUTCHINSON, T. and WATERS, A., 1987, *English for Specific Purposes*, Cambridge University Press.

KETTEMANN, B. and MARKO, G. (eds.), 2002, *Teaching and Learning by Doing Corpus Analysis*, Amsterdam & New York, Rodopi.

Language of Trade Glossary, <http://www.usinfo.state.gov/products/pubs/trade/glossz/htm#x>.

Maria Rees

LAUFER, B., 1997, "What's in a word that makes it hard or easy? Some intralexical factors that affect the learning of words", in N. SCHMITT and M. MCCARTHY (eds.), pp. 140-155.

LUDBROOK, G., 2004, "Il lessico specialistico: definizione, quantificazione e uso didattico", in *Didattica delle Lingue Straniere, Testing e Multimedialità*, Quaderni di Ricerca del CLI, Cafoscarina Editrice, Venezia.

MCCARTHY, M.J., 1984, "A new look at vocabulary in EFL", *International Journal of Applied Linguistics*, Vol. 5, No. 1.

MCCARTHY, M.J., AND CARTER, R.A., 1997, "Written and Spoken Vocabulary" in N. SCHMITT and M.J. MCCARTHY (eds.), pp. 20-39.

MEARA, P., 1995, "The Importance of an Early Emphasis on L2 vocabulary", *The Language Teacher*.

MIHWA CHUNG, T. and NATION, I.S.P., 2003, "Technical vocabulary in specialised texts", in *Reading in a Foreign Language*, Vol. 15, No. 2, October 2003.

NAGY, W., 1997, "On the role of context in first- and second language vocabulary learning", in N. SCHMITT and M.J. MCCARTHY (eds.), pp. 64-83.

NATION, I.S.P., 2001, *Learning Vocabulary in Another Language*, Cambridge University Press.

NELSON, M., 2000, A Corpus-based Study of Business English and Business English Teaching Materials. Unpublished PhD Thesis, University of Manchester.

NELSON, M., "A Worldly Experience", *The Guardian*, March 20, 2003.

NATION, I.S.P., and WARING, R., 1997, "Vocabulary size, text coverage and word lists", in N. SCHMITT and M.J. MCCARTHY (eds.), pp. 6-19.

O'DELL, F., 1997, "Incorporating vocabulary into the syllabus", in N. SCHMITT and M.J. MCCARTHY (eds.), pp. 258-278.

QUINION, M., 2002, *Ologies and Isms* (A Dictionary of Word

Beginnings and Endings), Oxford University Press.

REES, M., 2000, *English for Economics*, Cafoscarina Editrice, Venezia.

REES, M., 2004, *English for Business*, Cafoscarina Editrice, Venezia.

REES, M. (ed.), 2004, *English for Foreign Trade*, Cafoscarina Editrice, Venezia.

RENOUF, A., 2002, "The Time Dimension in Modern English Corpus Linguistics", in B. KETTEMANN and G. MARKO (eds.), pp.27-41.

ROBINSON, B., 2000, "Words, words and knowledge of words", (An Interview with Norbert Schmitt), Selected Interview, *The GRETA Magazine*, Vol. 9, No. 1.

SCHMITT, N. AND McCARTHY, M.J. (eds.), 1997, *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge University Press.

SINCLAIR, J., 2000, "Introduction", in *Collins COBUILD English Dictionary for Advanced Learners*, III edition, Glasgow, Harper Collins Publishers.

SWALES, J., 1990, *Genre Analysis: English for academic and research settings*, Cambridge University Press.

TOMLINSON, B.(ed.), 1998, *Materials Development in Language Teaching*, (Cambridge Language Teaching Library), Cambridge University Press.

TRIMBLE, L., 1985, *English for Science and Technology: a Discourse Approach*, Cambridge University Press.

TULLIS, G. and TRAPPE, T., 2000, *New Insights into Business*, Harlow, Longman.

WARING, R., 2002, "Basic Principles and Practice in Vocabulary Instruction", *The Language Teacher*.

WEST, M., 1953, *A General Service List of English Words*, London, Longman, Green & Co.

WILLIS, D., 1990, *The Lexical Syllabus: A new approach to language teaching*, Collins COBUILD.

Maria Rees

XUE, G. and NATION, I.S.P., 1984, "A University Word List", in *Language Learning and Communication*, 32, pp. 215-219.