Universidade de Aveiro Departamento de Línguas e Culturas

1999

Susan Jean Howcroft

English for Science and Technology

a Computer Corpus-based Analysis of English Science and Technology Texts for Application in Higher Education.

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Linguística, ramo Linguística Inglesa, realizada sob a orientação científica do Dr. Anthony Barker, Professor Associada do Departamento de Línguas e Culturas da Universidade de Aveiro

The Jury

President Reitor da Universidade de Aveiro

Vogais

Prof. Dr. Anthony David Barker
Professor Associado do Departamento de Línguas e Culturas da Universidade de Aveiro

Prof. Dr. Carlos Alberto Agapito Galaricha
Professor Associado do Departamento de Comunicação e Arte da Universidade de Aveiro

Profª. Drª. Maria Isabel Ponte Gonçalves de Salazar Casanova
Professora Auxiliar da Faculdade de Letras da Universidade de Lisboa

Profª. Drª. Belinda May Sousa Maia
Professora Auxiliar da Faculdade de Letras da Universidade de Porto

Prof. Dr. António Augusto de Freitas Gonçalves Moreira
Professor Auxiliar do Departamento de Didáctica e Tecnologia Educativa da Universidade de Aveiro

Acknowledgements

First of all I would like to thank my supervisor Prof. Dr. Anthony Barker for having given me such an immense amount of his valuable time and for his detailed advice on my work without which it would not be as it is.

Special thanks go to Prof. Dr. Belinda Maia of the University of Porto who found the time in a very busy schedule to read my work, I greatly appreciated her encouragement and very valuable advice on computer corpora.

I am indebted to Prof. Dr. António Moreira for his patience and vast knowledge, directing me to new sources of information, advising me on statistics and generally having an answer to the many questions I have put to him over the years I have been doing this work.

I should like to thank the students of the *Ano Comum* and my colleagues in the University of Aveiro for their support. Particular thanks must go to Dr. Silas Granjo who did so much to help me at the beginning with the technical side of the computer work.

I should also like to thank everyone in the Department of Línguas e Culturas in the University of Aveiro who helped to make this work possible.

Last but not least, my sincere thanks to my family and friends for their patience, advice and support. One friend and colleague in particular stands out, Gill Moreira, for her tolerance, humour and good sense which helped me to reach the end of this project.

Abstract

This thesis presents two analyses: first the analysis of computer corpora from undergraduate textbooks to isolate the (American) English language of science and technology they present; secondly an analysis of the English language competence of undergraduates starting their university studies in science and technology. These two analyses are contrasted in order to apply the results to the design of an English language syllabus for first year undergraduates.

A frequency and range word list was produced using a large baseline corpus to contrast with the main corpora taken from physics and chemistry textbooks on the students' bibliographies as a resource for syllabus design. Secondly, four corpora, two main and two sub-corpora produced from the physics and chemistry textbooks on the bibliographies of the undergraduates were analysed using Biber's (1988) algorithms and functions for variation across speech and writing.

The student intake was tested over five years and the results of those tests analysed. It was found that there was considerable variation in the students' levels of language competence. However, there was a close correlation between the students' competence and the number of years they had studied English in secondary school. Nevertheless there were students with extremely advanced competence and some with little or no competence in English amongst the undergraduates. Comprehension of scientific texts was generally found to correlate with more advanced competence and more years of study.

The frequency and range word list showed the contexts which are appropriate for materials to be used with these students and demonstrated variation from many of the accepted views of the language of science and technology. The computer corpora analyses varied from Biber's academic prose category. The sub-corpora demonstrated greatest variation which is believed to be as a result of specific cultural and/or literary material in the analogies used in the textbooks.

The heavy load of cultural background knowledge which the reader would need in order to work with the textbooks adequately was also found in the exercises the students were supposed to use for practice on the topic presented in the chapter. This and the interpretation of visuals in the textbooks were considered to be two principle factors that needed to be emphasised in a syllabus for first year undergraduates. However, given the time constraints on language teaching for science and technology students, a methodology which would lead to greater student autonomy is suggested using computer corpus-based studies - data-

driven learning and computer-supported distance communications and learning.

Resumo

Esta tese apresenta duas análises: primeiro uma análise de *corpora* computadorizados, criados a partir de livros dos estudantes de licenciaturas, para isolar a linguagem Inglesa (Americana) das ciências e tecnologias que apresentam; segundo uma análise dos conhecimentos da língua Inglesa que estes alunos apresentam ao iniciar os seus estudos universitários em ciências e tecnologias. Estas duas análises são postas em contraste para se aplicar os resultados obtidos ao desenho de um programa de língua Inglesa para os alunos do primeiro ano.

Foi criada uma lista com a abrangência e a frequência das palavras de um *corpus* de larga base, para ser contrastada com os principais *corpora* compilados dos livros de física e química constantes das bibliografias dos estudantes, como uma fonte para o desenho de programas. Seguidamente, quatro *corpora*, dois principais e dois subordinados, produzidos a partir dos livros de física e química referidos nas bibliografias dos estudantes, foram analisados usando os algoritmos e funções de Biber (1988) para variações entre linguagem falada e escrita.

Durante cinco anos, à entrada para a Universidade, os estudantes foram submetidos a testes e os resultados analisados. Constatou-se que havia variações consideráveis no nível de conhecimentos da língua por parte dos estudantes. Contudo, havia uma correlação apertada entre as competências dos estudantes e o número de anos que tinham estudado Inglês nas escolas secundárias. Todavia, havia estudantes com competências extremamente avançadas e outros com competências reduzidas, ou quase nulas, em Inglês. A compreensão de textos científicos estava geralmente correlacionada com os níveis mais avançados de competências e maior número de anos de estudo.

A lista com a abrangência e a frequência das palavras mostrou os contextos apropriados dos materiais a utilizar com estes estudantes e demonstrou que havia diferenças em relação a muitos dos pontos de vista aceites em relação à linguagem das ciências e tecnologias. A análise dos *corpora* computadorizados varia das categorias da linguagem da prosa académica de Biber. Os *corpora* subordinados mostram uma maior variação, que se julga ser devida a materiais específicos, culturais e/ou literário, usados nas analogias dos livros de estudo.

O grande peso dos conhecimentos de fundo de que os estudantes necessitam para trabalhar adequadamente com os livros de estudo foi, também, encontrado nos exercícios que necessitam de fazer para praticarem o que está referido nos tópicos dos capítulos. Isto, juntamente com a interpretação das imagens dos livros, foram considerados os dois principais factores a precisarem de ser relevados no programa para o primeiro ano dos estudantes. Contudo, atendendo às restrições de tempo

para o ensino de línguas a estudante de ciências e tecnologias, a metodologia que conduziria a maior autonomia dos alunos será baseada na utilização de *corpora* computadorizados (*data-driven learning*) e aprendizagem à distância assistida por computador.

CONTENTS

Figures

Tables

Abbreviations used in this thesis:

CBL Computer Based Learning

CD-ROM Computer Disk – Read Only Memory

DDL Data Driven Learning

EAP English for Academic Purposes

ELT English Language Teaching

EOP English for Occupational Purposes

ESP English for Special Purposes

EST English for Science and Technology

EGAP English for General Academic Purposes

ESAP English for Specific Academic Purposes

ICT Information and Communication Technology

LSP Languages for Special Purposes

MT Mother Tongue

OCR Optical Character Recognition

Chapter 1 Introduction

Chapter 1

Introduction

Since the early 1980's I have been fascinated by the use of computers in language teaching. From the moment the BBC computers became available, it was possible to begin to use computers in the classroom with students as computers had become small, relatively cheap and above all reliable machines. Nevertheless, first of all teachers themselves needed to find out how to work with the computers and, as there was very little software available, to design and write their own programs to use in class. However, projects were soon started to try to bring some system and principle into software design for educational purposes[1] and conferences and workshops helped to divulge information and provoke reflection on the role of computers in the classroom (Higgins 1985, Jones and Fortescue 1987, Evelyn Ng and Olivier 1987). Very soon the machines were updated and the amount of memory the new machines made available increased. This meant that the tedious, time consuming cassettes used to load programs changed to floppy disks providing relatively greater speed and also greater user-friendliness. The rate of change continued with the American IBM computers coming to dominate the education market because they were both more powerful and cheaper. This led to a state of confusion in many institutions which now had a mixture of different hardware and software, different sizes of floppy disks and programs that would not work on some computers

[1] For example, John Higgins' and then later Martin Phillips' work with the British Council Project to develop software for language teaching much of which was published in the 1980's in collaboration with Cambridge University Press, initially for the BBC computers and then for IBM compatible microcomputers.

because they had incompatible operating systems. This situation was made even more difficult as many academics had chosen the Apple Mackintosh computer as the most suitable for academic research. However, at this stage, education policy was encouraging the use of computers and the teaching of information technology as the use and application of computer technology came to be known[2]. This meant that towards the end of the 1980's teacher training courses were beginning to include CALL training (see Birnbaum 1987:19-20, Heppel 1987:20-21). Being at the cutting edge of the technological revolution was seen to be of prime importance not only to teachers (Dunn and Morgan 1987) but also to governments who believed that their future economic success in the world depended upon this change in education (see later 1.1 Science and Technology Education).

The availability of computers and their more widespread use in education also led to changes in the way that teachers prepared their work. Initially the opportunities for word processing made a big change in the preparation of materials and tests. Teachers began to write their own materials directly through the computer rather than relying on the support of a secretary or on the traditional cut and paste techniques which followed widespread use of photocopying. Predictions were made at this time that we were on the verge of the "paperless" office. It was believed that the need for paper filing systems would disappear because computer data was stored on floppy disks. It is now recognised that the contrary is true, the use of computers has led to far more paper being used as people who once would not have written anything themselves began to do so and documents can be revised much more easily, leading to more and more printouts of documents as changes are made to them in order to achieve greater accuracy or to bring documents up-to-date. As

[2] In Portugal the MINERVA Project to introduce Information Technology across the curriculum in state schools finished its pilot phase in 1989 (José Moura Carvalho, 1991).

in the example of teachers preparing their own materials through word processing, it became possible to produce much more specific, tailor-made material for individual classes and so more and more materials have been produced. Equally well, the mixture of computer operating systems has led people to be much more careful of how they store their data. Paper is much more accessible than a floppy disk which is the wrong size for the current computer system or which cannot be read by the latest machine. Contrary to popular belief some years ago (Jones and Fortescue 1987:129) computer hardware has shown that it is prone to all kinds of mechanical breakdowns and floppy disks often become corrupt at the most inconvenient moments.

However, computers were also being used to investigate language itself and huge projects were set up in universities, some on artificial intelligence (particularly in America with the ELIZA project) and others on lexicography and dictionary writing (such as the COBUILD Project in Birmingham). These projects have given way to Natural Language Processing which is basically a sub-field of computer science directly related to artificial intelligence, human computer interaction, machine translation and multimedia and the Bank of English which is now used for linguistic analysis of general language. They have been joined by such projects as the European funded ELRA project which is the European Language Resources Association which aims to include such things as recorded speech databases, lexicons, grammars, text corpora and terminological data in collaboration with European countries. There has also been a burgeoning of corpus work in many European countries themselves. The Universities of Oporto and Lisbon in Portugal are cases in point carrying out translation studies and linguistic analyses of corpora under the supervision of Profs. Belinda Maia and João Malaca Casterleiro, respectively. It has now been seen that with appropriate software (Tribble and Jones 1990) that teachers could also carry out linguistic research themselves and because of this teachers carried out work on error analysis of students' errors (for example, in Portugal, Fordham 1997) and the use of concordancing for both teaching and research developed.

Coupled with this interest in computers and computer-assisted language learning and research was an interest in special or specific language teaching. In 1986, I was required to design and teach a special course for people working for the post office (then the CTT - Correios, Telégrafos e Telefones, see Howcroft 1986). This particular course was for personnel working with computers in the head office of the post office in Coimbra and required reflection and research into the language of computers together with consideration of an appropriate methodology for a mixed ability group of foreign language learners dealing with computers. This was a special situation in which the English language was to be learnt through specific activities using a computer as part of the process of learning. The use of computers was also the product or goal of that language learning situation which made it particularly stimulating. The fact that English is the main language of the computer is important in any consideration of teaching students with or through this technology. Stubbs (1992:203) says that the Cox Report (Department of Education and Science 1989) on which he worked also argued that most interactions with computers were language experiences. (This is taken up again later in 1.4 The Dominance of English in Science and Technology) However, those that design the programs and operating systems of IBM computers are not linguists and they, therefore, cannot be expected to have taken into consideration the fact that their audience, the user, will often be a foreign language learner in a country far away. The language of the computer is often idiosyncratic from a linguistic point of view, but it has had and is increasingly having an effect on English language usage in the modern world. The number of new terms and concepts that are now employed because of the common use of computers and electronic communications is legion. The latest communications through the Internet and electronic mail have only served to emphasise this state of affairs, as has the huge increase in the amount of information available to computer users in the 1990's through the Internet and CD-ROMs containing the equivalent of whole bookshelves of knowledge. Crystal (1998) describes the Internet as a "semiotically sanitized medium" because of restricted turn taking and the fact that messages are received in order, one by one but he reminds us that we are not at the end of the technological road and that other technologies are still to come which may change all this. The language student of today faces a much greater barrage of applications and specific language which educational policy makers believe the working population should be capable of handling efficiently and rationally to maintain economic advantage in the world but this language will almost certainly be constantly changing as the technology itself changes. The types of education policy that are relevant to university undergraduates will be discussed in 1.1 Science and Technology Education and 1.2 Lifelong Education.

In order to design a syllabus for a modern day undergraduate student, whose first language is not English, all of these technological innovations and changes in language usage will have to be taken into account. Moreover, the sheer size and detail of the available knowledge on any subject which the student has to cope with also requires new learning strategies to be found. Education itself has had to change and will have to continue changing under the weight of what is now known and needs to be learnt by students today so that they can keep abreast of their subject. Political changes that have taken place such as joining the European Union have also had and are continuing to have an effect on educational policies within the member states. Harmonisation of policies, specific language training and cultural studies are deemed to be important for the European citizen (van Ek 1990, Kubanek 1998, Byram 1999). The Committee of Ministers stressed the political importance of intensifying and diversifying language learning as recently as 1998 when they reviewed the Council's earlier initiatives (Byram and Riagain 1999) The attempt to standardise courses and the subsequent qualifications obtained from them to allow recognition of academic qualifications throughout the member states is also seen as an important aspect of educational harmonisation. The whole issue of what skills and what knowledge are needed by the people who will make up the workforce in the future has led to changes in the perceptions of learning as will be discussed later in 1.2 Lifelong Learning.

Although much research has been carried out in the past, the changes that have taken place over the past two decades mean that the present situation is often quite different from the ones those studies refer to. The English of science and technology and the teaching/learning of English as a foreign language of undergraduate students studying languages as part of their courses has changed because of all the issues discussed above. Whilst some of the findings from previous research will remain valid, other findings will be called into question. This is especially so because of the possibility of conducting research on specific situations using up-to-date computers with their much greater memories, speed and capabilities. In some cases, as will be argued later in 1.10 Appropriate Text Types, there has been little scientific rigour or little information that would show the relevance of the work carried out for the undergraduates on science and technology courses in Portugal. Swales argues against overgeneralising and applying solutions found to work in situation x to situations y and z as he (1985:188) suggests "there are rarely global solutions to local problems". In other cases, there has been no published research at all on the specific kind of English that the undergraduates will come into contact with (see later 2.5 The Corpus Analysis Approach).

The undergraduate of today is also in many ways a different entity than undergraduates in the past. Many more students will arrive at the university with expectations about and knowledge of computers and their applications. This may be because of stimulating projects that were carried out in their schools in different subject areas and through computer clubs or because of the availability of computers in their homes and visits to ‘cyber’ cafés. Not only has this aspect of education changed the profile of the average undergraduate but also changes in language studies and other subjects in schools will affect the undergraduates’ profiles. Moreover, the universities themselves are just as vulnerable to change as the schools and the students are from the impact of computers and their applications. In fact, the universities lead in this change by influencing the kind of training teachers are given (see 1.5 The Situation in Portugal). New possibilities for different learning systems have been opened up and explored by universities using modern communication systems. What and how language should be learnt in universities by undergraduates studying science and technology must therefore be considered in the light of the above changes. Some form of quantification of the changes that have taken place in both the students and their subjects must be carried out to take principled decisions about the syllabus that would be suitable for these particular undergraduates.

The need for this particular piece of research therefore arose with the advent in the University of Aveiro of the combined first year for students of science and technology. These students had English as one of their core subjects in this first year and so it was felt that a syllabus had to be designed to meet these students’ needs. As Swales (1985:188) says,

> If those of us in ESP have thought long and hard about how best to serve our students' interests, it is simply because circumstances have tended to make us do so. In circumstances of restricted educational opportunity we have been forced to search out ways of providing maximum educational value.

The belief that English is of great importance for students of science and technology as discussed later (in 1.4 The Dominance of English in Science and Technology) led to the need for research that would benefit the students and provide them with "maximum educational value". This was particularly important as the discipline was to last only one year and the students' needs were expected to go far beyond this limited time scale right through their working lives. The questions that needed to be answered therefore were:

- What was the level of English of the students taking the course? In other words, what did the students already know or what had they already learnt prior to starting their undergraduate studies?
- What English did these students need to know? In other words, what kind of interaction with the English language could these students be expected to have and what was the nature of that English?

The difference between the results of these questions should give the answer as to what it is that the students need to learn. The information obtained must, in turn, form the heart of the syllabus drawn up for these students.

The second of these questions can be answered partly by looking at what the students need English for which will have at least two aspects to it. The general situation in the country, in Europe and the world will impinge upon these students just as much as the very specific context in which they find themselves in the university will affect their needs.

Considering the first of these aspects, the more general situation, we can see that the world is changing ever more rapidly and general educational priorities are also changing to prepare people for the modern world. When, where and how people study are undergoing changes brought about by changes in the technology of communication. These changes will have a number of effects on undergraduate students who are being prepared to take their places in society. The two strands of education policy and technological change will have an effect on university courses themselves and new curricula and courses are being and will be started. Course content and even the structure of courses will change. The level or length of the courses will change and different systems, for example a modular system, and different academic timetables are being and will be experimented with. Schemes to understand and standardise different countries' credit systems for studies in order to allow movement across borders between different countries' education systems and recognition of educational qualifications or part qualifications have started and will become more usual and widespread. The materials published for use on university courses will change and modern technology will have a profound effect on those materials and even on the means by which they are delivered. Contact with professors can now by means of e-mail and lectures and notes to accompany coursework be accessed through university computer networks. Assignments carried out by students can be produced and printed in sophisticated styles, transmitted electronically to the professor and comments received through the same channel. The actual information content of the work done by undergraduates can be affected by the information that they can obtain 'on-line'. The skills undergraduates are expected to have and to develop through their courses will therefore also undergo profound changes. Stubbs (1992:220) says that

The evaluation of educational change due to the new technologies involves the analysis of changed cognitive and social relations in the classroom. We therefore need simple but powerful concepts to study the *pedagogic and cognitive* logic of such situations.

1.1 Science and Technology Education

The attitudes towards science and technology education have first to be seen in the light of educational policies. Educational policies are not fundamentally different in the countries of the developed world and, as was mentioned above, the European Union countries are endeavouring to harmonise their policies on education.

There is an increasing preoccupation with science and technology education in the world in general and in the more developed countries of the world in particular because skill in science and technology in an educated workforce is seen as a means of maintaining or obtaining a position with the economic front runners in the modern world. There was felt to be a ‘crisis’ in science education in America in the 1970’s and 1980’s when fewer students were studying science at university and there was dissatisfaction with the lack of scientific knowledge in the general population (Matthews 1994). This preoccupation has been gaining pace since the 1980s when the small, robust, inexpensive personal computer was introduced into schools in Europe and America on a large scale and also because educational policy-makers felt that changes needed to be made in order to develop a workforce which could handle such technology. The developed countries’ educational policies were widely studied to find common ground which would ensure that people were trained successfully for the challenges the future was expected to bring. European countries particularly developed many programmes to study and make recommendations on all aspects of educational policy from specific course content (for example, on language Trim, Richterich, van Ek and Wilkins 1973/80) to mutually acceptable accreditation schemes (ECTS – European Credit Transfer System). What the United States of America, as the leading nation in the world, does to try to solve the crisis in education has an effect on the rest of the world and the policies adopted and technology produced and developed in the USA will affect education in Europe.

In the U.S.A. in 1985 Frank Press, then President of the National Academy of Sciences, Robert White, then President of the National Academy of Engineering and Winston Lord, then President of the Council on Foreign Relations wrote in the Preface to Keatley (ed. 1985:iii) *Technological Frontiers and Foreign Relations*, "If one compares the world of today with that of one or two decades ago, it is clear that modern science and technology have become profound influences on economies, societies, and international relations." They also suggest that "This process certainly will continue, and may accelerate." The result of which will be that it "will deeply affect American foreign policy, whether with friends or adversaries, competitors or collaborators, rich or poor." They may not have foreseen the changes that have taken place in American relations in the ensuing years but the position they outline is still having an effect on educational policies throughout the world. In the case of the developed countries this is so that economic and financial positions can be maintained in relation to the USA. In Britain Winston Churchill is reputed to have predicted that "The empires of the future are the empires of the mind." Churchill's comment implies, amongst other things, that education must be given priority in order to be successful.

Similarly, the European Round Table of Industrialists in its report *Investing in Knowledge: The Integration of Technology in European Education* (February 1997:5) says that "European society is running the risk of an increasing mismatch between the requirements of our new environment and the capabilities of our people." This report (ibid:5) identifies particularly noticeable mismatches between intellectual aptitudes in the areas of maths/sciences/technology and behavioural aptitudes which lead to "professionalism, excellence, (and) distinctive competitive edge." Furthermore, this emphasis on learning and using science and technology is seen as a problem. Allen Luke (1992), the series editor of *Critical Perspectives on Literacy and Education*, writing in the Introduction of Halliday and Martin's *Writing Science* (1993) says that the "very dependency on corporate science and technology expansion as a means for the expansion of state power and legitimacy have translated the crises of economies and cultures into the crises of sciences." Matthews (1994) regards this as the narrow 'economistic' view of education which is designed merely to develop 'human resources' so that countries can overcome their balance of payments deficit, or stay competitive with other economies. He believes that there is a need for a much more liberal science education which endeavours to develop scientific literacy in students which includes the understanding of concepts and learning about the nature of science both through its historical and social dimensions.

Nevertheless, because science education is seen by politicians particularly in this narrower economistic way, different countries are dedicating time and money to research and development in science and technology teaching. Mike Robinson (1994) from the University of Nevada describes why American government funding is being directed towards science and technology education and the training of teachers. He says this is because "science, technology, engineering and mathematics (STEM) are usually singled out as the most pressing educational areas" as shown by the Carnegie Commission on Science, Technology and Government Report (1991): *In the National Interest: the* *federal government in the reform of K-12 math and science education.* New York: Carnegie Corporation. This report (1991:7) also mentions the problem of proficiency in the English language saying that in the year 2000 'one child in twelve will lack the English language proficiency required for learning'. Robinson suggests that the national interest is best protected by training "hi-tech" workers in order to "maintain the diminishing US technological advantage in the world economy (Office of Science and Technology Policy, 1992 *Science and Technology.* Washington: Executive Office of the President.)". The fact that the economic situation of the USA has changed for the better since 1992 might make the authors of the report change their minds about whether the US is in fact losing its technological advantage in the world, but the concern is still to produce the right profile of a technologically competent workforce. Amongst other things, in order to achieve this goal Robinson places particular emphasis on the use of E-mail and the Internet in science teaching. The combination of science and technology education and the use of modern technology is a common theme amongst educators (Laurillard 1993).

The American Association for the Advancement of Science, Project 2016 was dissatisfied with the failure of curricular changes made in the 1960s as a response to the Russian launch of the *sputnik* satellite. The curricula in this period were designed by scientists rather than teachers and became over full with facts. The report, 1989:5 states that the curriculum they recommend should have the following effect:

> To ensure the scientific literacy of all students, curricula must be changed to reduce the sheer amount of material covered; to weaken or eliminate rigid subject-matter boundaries; to pay more attention to the connections among science, mathematics, and technology; to present the scientific

endeavor as a social enterprise that strongly influences – and is influenced by – human thought and action; and to foster scientific ways of thinking.

1.2 Lifelong Education

Coupled with the preoccupation with science and technology teaching is a desire to improve the education of the workforce both through school curricula and through further education programmes to produce an elite workforce which can cope with the demands of this new technological world. One means that is seen as being capable of doing this is to develop flexibility in workers so that they can be retrained to do different work whenever the need arises. Such workers must not expect to do only one kind of work throughout their lives. The average number of jobs workers are expected to do in their working lives in the future ranges from four to fifteen rather than one lifelong career. This means that people must accept the need for sustained learning in order to face the different challenges they are expected to meet in their different work situations. Differences in working arrangements are also expected because of demographic changes. The workforce is ageing in many of the developed countries and so the size of the working population in comparison with the population as a whole is expected to fall which means that new work patterns and perhaps a longer working life will have to be accepted in the next few decades if countries are to maintain their actual economic positions. The process of retraining people throughout their careers has come to be seen as a process of Lifelong Learning; an idea which was put forward in the 1970's with the British HMSO (1973), OECD (1973), UNESCO (1973) and Council of Europe CCC (1972, 73 and 74) reports on adult education.

President Clinton (1998:59) stated in his State of the Union Address that the "Information Age is, first and foremost, an education age, in which education must start at birth and continue throughout a lifetime." However, the means to achieve this objective often appear elusive. Luke (1992) suggests that there is a serious 'time-lag' between the debate about educational change and the actual 'remaking of science education'. He discusses the work done by Lingard, Porter and Knight, (1992) who say that the post-war human capital model of education[3] in the USA and Canada, the UK and Australia, has proved resilient and recyclable, despite there being little evidence that it works. Gerald W. Bracey (1997:52) suggests that "The biggest threat to the American educational system may come not from within our schools but from the depth of our divisions over what exactly they should accomplish and how best to get them to accomplish it." He goes on to point out that there has been an enormous shift in policy which has led to 62 percent of high school graduates being deemed capable of studying at a higher level and therefore enrolling in college as opposed to 20 percent after the Second World War. The massification of education is another factor that will affect teaching and learning expectations and goals. These changes will lead to a different, wider range of courses and qualifications becoming the norm with shorter, less sophisticated and more modular courses being given as and when required by the individual.

Educational policy emphasising more students enrolling in higher education is also reflected throughout the European Union and its member state Portugal. The proliferation of new universities and polytechnics in Portugal in the last two decades is a clear example of the increasing need to provide places for the greater number of students leaving school who could benefit from further years of study and higher qualifications[4]. The recent changes in the number of years of obligatory

---

[3] This model of education is one which stresses universal education that is, education for all a nation's children without financial or other constraints.

[4] The change that took place in Britain in 1995 when all the higher education institutes became known as universities is also an example of this phenomenon.

schooling requiring students to complete the ninth year also reflects these changes in educational policy. Despite the fact that Portugal is one of the countries in Europe with the lowest rate of unemployment, further education is seen as a means of achieving greater prosperity and increasing the chances of finding a ‘good’ job. Universities themselves are in competition with each other to provide more up-to-date courses which give students the preparation necessary for the types of employment in the Europe of tomorrow. Indeed recently much forward planning has gone into attracting new faculties or universities to cities in the interior of the country, which bears witness to the importance which is attached to higher education in developing the poorer less industrialised ‘interior’ regions. Educational institutes are also seeking to forge links with industry in order to work towards providing professionally oriented training for the local workforce.

Countries throughout the world are constantly producing league tables of the most advanced economies and comparing and examining the ability of different educational systems to produce the best scores on tests of mathematics and science subjects and to equate these findings with spending on education, number of hours devoted to these subjects in school, amount of homework set, and other parameters to try to discover the most successful formula to produce the elite workforce deemed necessary in the future. The world of education is fraught with insecurity as to what constitutes a technocrat’s training and how to evaluate ‘quality’ in education. ‘Standards’ of education normally translate as the results of tests. Recently testing carried out for the Third International Maths and Science Study (TIMSS, 1997) produced a league table of nations which showed the concern that many countries feel about their results in such comparisons. Both Britain (England 25th in Maths and 10th in Science, Scotland 29th and 26th respectively) and America (28th in Maths and 17th in Science) feel that they are doing “poorly” and the *Economist* (March 29th 1997) reports that

> In a television interview in December, the French president, Jacques Chirac, described as “shameful” a decision by his education ministry to pull out of an international study of adult literacy which was showing that the French were doing badly. And in Britain last year Michael Heseltine, the deputy prime minister, brushed aside objections from officials in the Department for Education and Employment, and published the unflattering results of a study he had commissioned comparing British workers with those in France, America, Singapore and Germany – chosen as key economic competitors.
>
> The Germans, in turn, were shocked by their pupils’ mediocre performance in the TIMSS tests. Their pupils did only slightly better than the English at maths, coming 23rd out of 41 countries. In science, the English surged ahead (though not the Scots) while the Germans were beaten by, among others, the Dutch, the Russians – and even the Americans. A television network ran a special report called “Education Emergency in Germany”; industrialists accused politicians of ignoring repeated warnings about declining standards in schools.

It could be argued that if major countries perceived to be economically successful like Germany and America are dissatisfied with their placement on these scales, something is fundamentally wrong with the yardstick being used for measurement. Controversy over the yardsticks used to evaluate children in different educational settings was indeed one of the reasons that the British National Foundation for Educational Research refined the test for the latest study, giving the teachers concerned precise instructions on how the test was to be carried out and monitoring schools at random. Nevertheless, the attitudes expressed highlight the emphasis that governments are giving to education as a means of achieving success in world markets. Success in education and commercial success are seen as going hand-in-hand.

The American Secretary of Education, Richard W. Riley, trying to justify the American results, (1997:60) says, “students’ proficiency in science and math is up about one level compared to what it was a decade ago. One reason we have been behind countries such as Japan is because that nation’s public schools always have put extremely heavy emphasis on science and math. We still have a long way to go.” President Clinton, not surprisingly, in a speech to the National Association of Black Journalists (July 17, 1997) put a much more positive interpretation on the International Math and Science Test results. He claimed that recent results for 4th and 8th graders showed an improvement which in turn proved that his policies were working and that therefore America could achieve “international excellence in education”. Eight Goals have been identified by the National Educational Goals Panel which was set up in 1990, these are: Goal 1: Ready to Learn; Goal 2 : School Completion; Goal 3: Student Achievement and Citizenship; Goal 4: Teacher Education and Professional Development; Goal 5: Mathematics and Science; Goal 6: Adult Literacy and Lifelong Learning; Goal 7: Safe and Disciplined, Alcohol and Drug-Free Schools; Goal 8: Parental Participation. There is also a similar list of priorities from the U: S. Department of Education (February 1997):

“*All students should be able to*:

1. Read independently by the end of the third grade.
2. Master challenging mathematics, including the foundations of algebra and geometry, by the end of the eighth grade.
3. Be prepared for and be able to afford at least two years of college by age 18, and be able to pursue lifelong learning as adults.
4. Have a talented, dedicated, and well-prepared teacher in their classroom.
5. Have their classroom connected to the Internet by the year 2000 and be technologically literate.
6. Learn in strong, safe, and drug-free schools.
7. Learn according to challenging and clear standards of achievement and accountability.

The differences between these two lists is one of specifying in more detail when and what is necessary in education for the future by the Department of Education, such as the need to be able to afford further education and the year 2000 being given as the objective for Internet connection.

One of the policies introduced into the American education system, as point 4 on the list of priorities shows, is the testing of both teachers and pupils. A similar system to that found on the American National Educational Goals Panel website (http://www.ed.gov/pubs/StratPln/priority.html), which allows people to find out about different states and their educational achievements, is one which has been introduced into European educational systems where testing and grading of results from school to school to compare those schools that are doing well with those that are doing badly. The idea is that the better schools can be used to show what should be done to achieve the desired test results and that teachers can benefit from visiting those excellent schools to learn about their methods which they can then apply in their own schools to improve standards[5]. A similar system of analysing which education systems teach science and maths best is forecast to explain what conditions are necessary to promote effective learning.

The Organisation for Economic Co-operation and Development (OECD) has collected data on how governments spend their combined $1 trillion annual education budgets and explains that their new studies (launched December 1996) will compare how schools, colleges and universities are run in each country and analyse the implications for policy makers. The fact that some countries with low education budgets achieve high scores on the TIMSS tests has caused politicians to seek alternatives to more spending on education. Similarly class sizes vary

[5] This is similar to the European LEONARDO project which encourages the movement of professionals between countries in order to find and emulate excellence in teaching.

from country to country but results do not seem to support the contention that only small classes achieve good results in science and maths. The methodologies used to teach these subjects appear to be as important as class sizes are.

As there is no consensus on what causes optimal learning, research for practical application on undergraduate disciplines is, in the light of the above concerns, an essential prerequisite to aid success in science and technology education in a foreign language. One of the implications of Lifelong Learning is that the emphasis in teaching should be on the learning process itself and not the product of that learning, so that people learn how to learn rather than learn a particular finite body of knowledge.

1.3 The Impact of New Technology

Many other changes have taken place in the last three decades which will also influence the teaching and learning situations of students of Science and Technology. One of the principal changes is the advent of the personal computer with sufficient memory and processing speed to enable specific situation research work to take place but the implications of the personal computer go much further than this. Students of English and those involved in science will find their lives are surrounded by the specific English of the computer, the associated word processor, multimedia applications and increasingly the world of electronic mail and the Internet. It is also the case that the software and computer communication systems that dominate the world, such as the Internet, originate from America and are therefore usually originally written and manipulated through English.

Bucy describes the differences that have taken place in the concept of what a computer is. He (1985:46) explains how the advent of the silicon chip has made it possible to incorporate the computing power of a main frame computer into such products as microwave oven controls and handheld calculators. He also suggests that even by 1985 computers were thought of as relatively inexpensive machines that can be used in "any number of activities, such as education, household data storage, increased job productivity, or entertainment."

Consequently this greater capacity for data storage has also opened up the possibility of conducting individual, empirical research into the language of science and technology in a manner that was unthinkable only a few decades ago. New software has also been developed to allow this kind of research to take place as a result of work on corpora (see 2.5 The Corpus Analysis Approach). The problems associated with obtaining information through the Internet seem to be much more a problem of obtaining information about where data is available in such a vast resource and of framing the right sort of question to obtain the desired result. If the question posed is too general, the enquirer will be inundated with information which will be hard to sift through in order to locate appropriate data within a reasonable amount of time. If however the questions posed are too precise, little or no information may be obtained. Often the answers to questions surprise because the area the enquirer is contemplating does not match the results of the search. For example, a search for data on "bands" would produce results on both the musical variety, i.e. brass bands, and the electronic forms, i.e. wave bands, either of which would be inappropriate if the enquirer wished information on the other. The significance of the data obtained also has to be judged by the enquirer to ascertain if it is of an appropriate level of sophistication which in turn requires sufficient background knowledge of the subject by the person requesting information.

The prominence and utility of modern technology urges therefore the teaching of a foreign language that is somewhat different from that taught in schools, although many schools are taking part in exciting projects including distance communication and e-mail. The combination of modern technology and language use has even created new styles of language. This difference between the new style of language and other styles will also have an effect on teaching methods. Teaching techniques which according to Kelly (1969:120) were "in constant use in the language classroom right through the history of language teaching" but which recently have fallen into disuse because they appeared to be 'contrived' and inappropriate will have to be reappraised in the light of the new genre being created. An example of a teaching technique that has fallen into disuse is written dialogue, which was a common tool for the presentation and practice of language in the structural model of language teaching. Modern e-mail seems to be more like this, that is, it is more like written dialogue than formal letter writing. Written dialogue was challenged as being inauthentic in the 1970's and 80's and, therefore, not suitable as a model of actual usage but, through e-mail, it now takes on renewed significance (Leech 1997). There is also to be found on the Internet written lectures which reflect a little of both worlds, being written text which is meant to be spoken and which therefore contains comments and asides that have a specific listener in mind (Stubbs 1996, McCarthy and Carter 1994). The manner in which students interact with technology is the object of much research and often the results are disappointing to those who believed that the technological revolution would revolutionise teaching[6]. Most CALL specialists reached the conclusion that computers are an aid to teaching which, rather like other modern technologies such as the video, depend upon the ingenuity of the

[6] Seymour Papert (1980) *Mindstorms: Children, Computers and Powerful Ideas,* Harvester Press is an example of this view of the huge changes (and improvements) that the technological revolution would bring to education. Robin Goodfellow of the Open University reports (1999) *Language Learners' I.T. Strategies will they be the Death of CALL?* that with university language learners in an open-access IT environment, CALL is "vulnerable to the growth of IT sophistication in learners" . So he recommends that teachers need to turn their attention to "the IT choices that learners make when they embark on self-study" otherwise carefully prepared CALL designs will be sidelined.

teacher to make them relevant and useful resources for students (Kenning and Kenning 1990, Higgins 1988, Phillips 1985, Leech and Candlin 1986). In 1987 Eastment predicted that computers would not be found in computer rooms except for computer literacy courses but would be located in normal classrooms as part of everyday teaching. He also suggested that at the time of writing (1987:10) concordancing was limited by the rather unreliable software that was available. This problem has now been largely overcome but Eastment's prediction that we would have "pedgagogical concordances" of varying levels and language types has still to be realised. Warschauer (1999) argues that in education there is no BALL (book-assisted language learning), PALL (pen-assisted language learning) and no LALL (library-assisted language learning) because these are such powerful technologies. Therefore, he argues, it is only with the integration of computer technology into teacher education and language learning that computers could be seen to have taken their place as a natural and powerful part of the language learning process. The change towards data-driven learning (DDL) is one of the more exciting new trends which will be discussed later in relation to the use of corpora for teaching purposes (Chapter 7).

The use of computers has also had an impact on linguistics and the description of language through corpus studies and these discoveries must be exploited and integrated into the curriculum (see 2.5 The Corpus Analysis Approach). Particularly in the area of collocations, new information is more easily obtained and is available for use by the teacher and the learner. Research work on language acquisition has also suggested that 'chunks' of language are used in natural language acquisition (Hakuta 1974; Huang 1971; Brown 1973; Clark 1974; Cruttenden 1981; Wong-Fillmore 1976; Newmark 1979, Peters 1983) and so the use of materials derived from concordancing the target student texts will provide one more tool to be added to the repertoire of teaching.

It is my contention that collocations are appropriate ‘chunks’ of language that can and should be used use to teach to language learners (Tribble and Jones 1990). The collocations can be obtained from the corpora compiled from the textbooks on the bibliographies which the students are meant to consult. In this way, the language that is being studied becomes entirely appropriate for the purposes of the students. The specific lexical semantics of science and technology is being presented rather than general English. Furthermore, the language being studied could be brought under the control of the student, thereby customising the learners materials for study. These aspects will be taken up in more detail later in 7.4 Data-Driven Learning.

In Portugal, and in the University of Aveiro in particular, there are now many homepages and interactive websites which the undergraduates are encouraged to consult and even study from. Most higher education institutes, like the University of Aveiro, have home pages for each of the departments. The University has its Informatics Centre through which students can gain access to the Internet, to say nothing of the facilities the students have at home or contrive for themselves. There are also Open University and distance learning courses for undergraduates making use of the Internet. The first-year students therefore soon become, if they are not already, quite sophisticated in their knowledge, use and expectations of modern technology.

1.4 The Dominance of English in Science and Technology

With the economic supremacy of America in the world now and for most of the 20th century, the English language has also come to dominate the world of science and technology. Most research work is now published in English no matter where the research was carried out.

Kaplan (1993:156) claims that "something on the order of 85% of all the scientific and technical information available in the world today is either originally written in, or abstracted in English." Furthermore, many of the books used to teach science and technology are based on American models.

The significance of this is that in most cases the English that students encounter and the English that students therefore need will be predominantly American and it will also be language that is not specifically prepared for the student of English as a foreign language. It will, however, be predominantly written language.[7] Despite the fact that all European languages are supposed to be equally important in the European Union some are seen to be more prevalent than others. Sheer numbers of speakers have an obvious impact upon this so that the Portuguese language is not one of the languages that the scientific community sees as essential for the people who will run the businesses of tomorrow in Europe. The European Commission 1997 *Eurobarometer* reported the results of a survey conducted in 34 countries in Western, Central and Eastern Europe in which Russian was the principal language of 35% of the 555M people in these countries, English 28%, German 20%, French 17% and Italian 10% and suggests that the languages at the upper end of this spectrum, that is Russian and English, appear to be spreading whilst those at the lower end are declining. Crystal (1997:10) puts forward the financial argument for using a *lingua franca* in international bodies which is that the cost of translation can swallow up to half the budget for such organisations. The European Union has yet to come to terms with this problem.

[7] Research carried out by Prof. Drª Ana Margarida Barros of the Department of Chemistry, University of Aveiro, published in her (1998) report on the European Chemistry Thematic Network (ECTN) work on *Communication and Management Skills* shows that reading and analysing texts in a foreign language (usually English but possibly in French) is considered *indispensable* by 100% of those answering her questionnaire from Universities in Portugal, and that this activity was classified as *indispensable* by 88% of the Industries consulted and as *very important* by the other 12%.

Whilst it can be argued that Brazil, Mozambique and Angola are very important markets outside Europe, the problems that these countries are experiencing means that this potential may not be realised for some time to come, if at all, and that, therefore, the Portuguese language will not be seen to be as important at the moment as it might be in the future. Crystal (1997:7) argues that a language becomes a global language because it is the language of power, both political and military, which explains why Portuguese found its way into the Americas, Africa and the Far East during the period of colonisation. However, with Mozambique joining the Commonwealth countries there is a suggestion that it feels drawn more towards countries that had a connection with Britain and the English language. The proximity of South Africa, Zimbabwe, Zambia, Malawi and Tanzania where English is either an official language or retains some influence may also help to explain this. Crystal (1997:61) suggests that whether English becomes a global language in the twenty-first century depends upon what happens in countries with the largest populations, notably China, Japan, Russia, Indonesia and Brazil. University students who will be the leaders of tomorrow will need to learn at least one of the dominant languages. From the numbers of students studying English in Portuguese secondary schools it can be seen that the language that is often being chosen is English (Ferreira, Ramos and Braga da Silva 1999).

A significant factor in the dominance of the English language is the overwhelming expansion in the use of computers in the world. Kubanek (1998:202) points out that “the *lingua franca* function of English would become obvious” to students using the Internet and contacting websites for information. This technological revolution is having an enormous effect on education and employment.

1.5 The Situation in Portugal

Given the emphasis on the study of science and technology in the world, there is also a need to study the English of Science and Technology in order to assist students to study language of science and technology. However, secondary school English courses continue to teach the English of the Humanities which ill-prepares the students for tertiary education in science and technology and the variety of English that this represents.

The changes that have taken place in educational objectives because of, and deriving from, the European Union have emphasised multicultural education. The intention is to prepare students for European citizenship. This requires both understanding and tolerance of other cultures together with more language training so that all citizens will have knowledge of a minimum of three European languages. School programmes in Portugal reflect this preoccupation with cultural identity and acceptance of other cultures. The programmes for schools and the manuals used also reflect this state of affairs. Culture and language are seen to be mixed together in such a way that it is impossible for them to be separated (Fligelstone 1998), however, the reader or listener may be unaware of differences between cultures and therefore their understanding of what is being transmitted may be faulty (Scollon and Scollon 1995). Those students who have chosen to follow sciences are given the same programmes in English as students who wish to follow humanities courses, rather than special programmes to help them to cope with the enormous amount of scientific and technical information which is published in English and which they will certainly meet if they continue their studies to university level or in their future employment. This is not to say that the English they have learnt will not stand them in good stead in many general situations but it will not prepare them adequately for their future studies. Indeed it may even lead to confusion and misunderstanding in their scientific and technological studies. Whilst it cannot be denied that this preoccupation with "multiculturalism" is important in the context of liberal education for citizenship, it is less useful for the needs of the science and technology students who require both this 'liberal' education and more or further language support with their specific English needs.

Furthermore, the teachers, who form the bulk of EFL teachers in Portuguese secondary education and are required to teach language to students in schools, have followed typical humanities education courses themselves. Several ESP theorists (for example, Widdowson 1979, Ewer 1975, Strevens 1978, Hutchinson and Waters 1987 and Kennedy 1983) have pointed out the fact that those who are required to teach the language of science and technology feel they are themselves ill-prepared, and are therefore often reluctant to do so. Even in terms of technology, teachers with a humanities background were seen to have an antipathy to "machines". However, most undergraduates these days in this and other modern universities are positively encouraged to confront the latter problem through educational technology disciplines on their courses and by being expected to submit word processed assignments for other disciplines. The use of computers in schools however for the most part continues to be considered the province of the maths department (Moura Carvalho 1991, Stubbs 1992). There is awareness of a need to change this state of affairs but as White (1988) highlights it can be difficult to achieve innovation in schools and the process takes a long time.

> "The discrimination and adoption of an innovation – in language teaching as elsewhere – follows and S-shaped curve (...). There is an early stage during which a very small percentage of innovators decide to introduce the new idea. This is followed by a second stage during which the early

> adopters, who have noted that the innovation produces no harmful effects, take on the innovation. During the middle stage, the majority adopt quickly, influenced mainly by the innovators. At a late stage, the laggards or late adopters finally give in. A minority who never adopt lie outside the curve."

In addition to this, for innovation to arise, be taken up and successfully installed White believes that effective management must exist in the organisation.

The University of Aveiro is aware of the necessity for English language proficiency for science and technology students and has introduced innovations into the structure of the science and engineering courses in order to address the students needs. The identification of core disciplines which almost all of the undergraduates in science and engineering have to study is the main one of these. Other universities, such as Porto and Coimbra[8] continue to teach ESP within the different faculties rather than as a separate or common core subject for all of the science and technology students. This system was an earlier model used for most ESP courses, known as 'content-based' syllabi focusing on the particular requirements of specific academic disciplines. This explains why published materials are usually entitled *English for Electronics*, *English for Telecommunications* or *English for Computer Science* and so on. Very few textbooks attempt to address the English of Science and Technology today. Previous coursebooks have attempted to isolate some form of sub-technical language which was thought to underpin the language of science and technology as a whole.

There are obvious advantages to the type of course which has common core subjects not least of which is the fact that it allows efficient

[8] ESP also exists of course in institutes such as ISCAA here in Aveiro where there is English for Accountancy.

use of staff, providing staff to student ratios which can cope with the huge entry to university which is taking place in most developed countries as mentioned earlier. Another advantage of this system lies in what Laurillard (1993) describes as the need for undergraduates to develop concepts rather than merely gather facts. The undergraduates need to learn how to learn autonomously and need to be guided to that end. The individual scientific concepts that can be found in any specific science subject nevertheless present some differences from what students have been taught before. The students are often unaware that this is the case in university and this in itself can lead to a lack of success and indeed to considerable frustration if students regarded themselves as good students and now at university they suddenly begin to get unexpectedly low marks. In English language studies the students will also have to undergo this transformation and recognise that what they have learnt before is only part of the story and what may appear on the surface to be the same may in fact be quite different in this new context. The fact that the discipline has to be directed more generally to science and technology can therefore be an advantage because the students can become aware that the skills they need to acquire will stand them in good stead no matter what their subject speciality is. Content knowledge will also be acquired by the undergraduates along the course so that all of the students have a similar lack of specific subject knowledge on entering the university and can benefit from adopting certain strategies when faced with new material, especially if this new material is in English.

The study which is presented here is focused on University of Aveiro students and courses as a sample of the language needs and teaching requirements of undergraduate university students matriculated in a number of different courses preparing them for the future.

1.6 Science and Technology Undergraduates and English

The University of Aveiro introduced a foundation year for all students taking the various engineering courses of the University in 1993. Five disciplines make up the core subjects taught to most of the new students (English, Chemistry, Physics, Mathematics and Computer Science), but there are some exceptions such as the Licenciatura em Novas Tecnologias da Comunicação (NTC) which only has the English and Computer Science components, the Licenciatura em Gestão e Planeamento em Turismo (GPT) which has no chemistry or physics, the Licenciatura em Planeamento Regional e Urbano (PRU) which has no physics. This innovation has meant that the English discipline has had to be revised.

Although English had been taught for many years to many of the engineering students, the foundation year or Common Year (*Ano Comum*) as it is known, has led to a considerable change. On the one hand, there has been a loss of specificity, that is the specific English genre for the course the students were studying, for example, English for Electronics and Telecommunications. On the other hand, there is a need for a wider but nonetheless specific English genre to be taught to these science and technology students, the English of Science and Technology. This is in order to help them to cope both in the first year and in subsequent years with the English they will need for the many and varied courses and bibliographies containing books in English that they are going to come into contact with. The bibliographies for core subjects contain many books in English or originally written in English and translated (up to 75% in some cases). Although bibliographies are liable to change and change quite quickly from year to year, some core texts in English remain for a number of years and are available in the library for consultation by the first year students. Therefore, the English the students have to cope with needs to be defined in order to be able to produce a syllabus which makes the optimum use of the limited time[9] and resources available to this annual first year course.

The immediate short-term needs for these students can be identified through the bibliographies they are asked to consult in their first year science and technology courses. These include a number of books in English on the core science subjects taught in the first year of the University. As scientific literature is seen as becoming more and more incomprehensible in the latter half of this century for all but a few specialists (Hayes 1992:739-740), undergraduates will need help in reading and understanding scientific texts.

The kind of language the students require to be able to read these books successfully can be identified by detailed study of the textbooks. The physics and chemistry textbooks have been studied in order to identify their needs in respect of the syllabus and they will be presented here. However, as the students come from over 25 different courses, there is a need to provide a baseline corpus for comparison for the comprehensive syllabus to be drawn up. In order to recognise what is normal use in a particular genre a very large corpus has to be consulted, the baseline corpus, in order to avoid generalising from what may be an abnormal or exceptional example of language use found in one or a small number of texts. As the study of these textbooks is based on variation studies in order to see how far they differ from other genres or text-types, some form of comparison needs to be made to highlight the differences and to add scope to the syllabus. As was mentioned above, the science and technology courses cover a much wider field than the books on the bibliography alone can represent. What would be most appropriate, given

[9] The discipline has one two-hour class per week across the two terms of the first year for most of the students. Exceptions to this are the Licenciatura em Novas Tecnologias da Comunicação (NTC) which has a term of four contact hours of English per week in the second year and the Licenciatura em Gestão e Planeamento em Turismo (GPT) which has 5 hours of English per week for the first term in the second year.

that the students' physics and chemistry textbooks on their bibliographies are overwhelmingly American publications, would be an American general science textbook aimed at undergraduates. One textbook would nevertheless not fulfil the criteria of a baseline corpus as it would itself be liable to offer an exceptional or aberrant style for the genre so a number of American general science textbooks would be needed to analyse the genre. As such a number of suitable textbooks could not be identified for this type of tertiary level student, a multimedia encyclopaedia will be used. The advantages of such material are its wide-range and huge size to meet the demands of generality in order to identify linguistic trends and tendencies in the genre. The range would be more than adequate to cover the basic science of all of the courses included in the first year foundation course and the size runs to hundreds of millions of words rather than the tens of thousands of words to be found in one general science textbook. This will be taken up in more detail later in 1.10 CD-ROM Material.

Implicit in deciding what to include in the syllabus is what English the students have already learnt or already know. In order to answer this question the students were tested and their results analysed in order to identify areas which need to be addressed by the syllabus. Chapter 4 Test Results for New Students discusses the test used in each of the academic years from 1993-1998 and the results found for new undergraduates in those years.

1.7 The Ano Comum

As there are approximately twenty five different courses being catered for in the *Ano Comum*, the first-year foundation course, each and every one of the different specialisations will be addressed nevertheless through recourse to a multimedia encyclopaedia. As mentioned above, a

CD-ROM encyclopaedia covers all of the undergraduates subject specialities and so it is argued that generalisations about English for Science and Technology can legitimately be made because of its huge size and comprehensive nature. The study will also make a detailed analysis of published texts contained in the students' bibliographies for the first year for Chemistry and Physics[10]. Despite the fact that the language needs of the students in truth go beyond the boundaries of the first year of undergraduate studies, other language needs will not be addressed *per se*. The language appropriate for conferences and post-graduate work for example will not be considered, it is too remote and lofty an objective for an English discipline syllabus designed for the first year of undergraduate studies with the considerable time constraints imposed upon it. The use of strategies for coping with texts will certainly prove useful however in the later stages of the students' courses. The methodology used with the students and skills required by same will therefore also be of importance in enabling them to cope in the remainder of their courses with the English that they will meet.

It must be acknowledged that many of the students who take the science and engineering courses in the university will often become teachers or take up jobs in business and management and only a rare few will become pure or research scientists[11]. For this reason, a concern for undergraduate work and success within university courses in science and technology would appear to be not only a more realistic aim of this research but would be addressing a common need of the students in university science and engineering courses. Moreover, the content should

[10] It is admitted that English for Information Science is as relevant as the English contained in textbooks on the bibliographies for Mathematics. However, the language in Information Science is English even if the explanations for use (and pronunciation of the terminology) is given in Portuguese. There are also appropriate glossaries that students can consult for this discipline. The language of Mathematics is subsumed by the mathematics contained in the physics textbook analysed and has been shown to be a very restricted genre (Biber 1988).

[11] See Arroteia, Jorge Carvalho; Martins, António Maria (1997) *Inserção Profissional do Diplomados pela Universidade de Aveiro: Trajectórias Academicas e Profissionais,* Aveiro: Universidade de Aveiro.

be made appropriate to the learning purpose of the students. Wilson (1997:130) suggests that databases designed for use with language students should contain texts that relate to students' tasks and interests in other disciplines in order to make the "students' goals in the language learning programme ... coincide as far as possible with the students' wider goals." Despite this, she identifies the fact that her computer-based materials were too general as a "disappointment" as they "had none of the quality control for style and linguistic coverage that good CALL demands". In other words, Wilson reminds us that the materials that are used with undergraduates need to be carefully selected so that they are sophisticated enough and they must be tried out and improved upon or abandoned if necessary should they prove to be unsuitable.

1.8 Appropriate Text-Types

Similarly, the materials used for analysis need to be carefully selected. Despite the fact that many studies have been carried out on specific English in the past, they often include a wide range of text-types, including text-types[12] that would be appropriate for post-graduate specialists. One early example of this is Barber 1962. Barber took three texts that 'straddled' disciplines, two came from university textbooks on engineering applications of electronics and astronomy and one from a journal on biochemistry. The latter text-type would be suitable only for those (few) students who enter university to study science and technology and who continue with their studies to a very high level. Similarly, Tarone *et al* (1981) studied two papers from one Astrophysics journal. In this case it is too small a study to allow generalisations to be made. Indeed Swales (1985:192) comments that this work by Tarone *et al* is an

---

[12] *text* here should be understood to include written and spoken language.

"inadequate sample". Even Tarone *et al* (1981:191) indict themselves they say

> While extensive use of the passive is shown by frequency counts of verb tense and aspect which are performed on corpora combining texts from a variety of scientific and technical fields, significantly different results may be obtained when one compares the frequency of the passive and active voices within a single scientific or technical field.

The only study that they report they have found that was on only one field was Wingard's (1981) work on medical texts. There has been an increase in recent years in the numbers of students going on with post-graduate studies, and it may well be that at a later stage those advanced students' specific language needs must be studied to see if further language training is necessary.

This pattern of combining texts-types often including journalese or popular science texts has continued in many cases even in some corpora that are regularly used by researchers. Therefore, there is no suitable study of undergraduate textbooks for science and technology students that could usefully be used to identify the target language of these undergraduates and hence the necessity to start from the beginning to analyse specific texts appropriate for these students.

1.9 The Corpora

Added to the problem of defining exactly which text-type might be appropriate is the difficulty of obtaining a sufficiently large corpus in order to produce significant results. Some of the studies done in the past were based on very small corpora or one very small corpus which now would be considered dubious as a basis from which to draw valid conclusions (for example, as mentioned earlier in 1.8 Appropriate Text-Types, Tarone *et al* 1981 only used two texts). However, more detailed studies of a smaller corpus may show features that would be lost in a very large corpus (Robinson 1991). This will be taken up in more detail later in 2.2.3 Scientific Specificity. For this reason two sub-corpora are included in the analysis. Five corpora will be used: a large physics corpus, a small physics sub-corpus, a large chemistry corpus, a small chemistry sub-corpus and a corpus from a multimedia encyclopaedia for strictly comparative purposes. Biber, Conrad and Reppen (1998:136) go even further and suggest that studies based on very small corpora are likely to be inaccurate and a 'baseline' is needed for comparison to identify significant variation. Halliday (1993) suggests that the development of the modern corpus is that "we can now for the first time undertake serious **quantitative** work in the field of grammar" but he points out that in order to be able to do this "Quantitative studies require very large populations to work with." The multimedia encyclopaedia will provide that baseline for comparison as will the large physics and chemistry corpora when used in comparison with the sub-corpora in these same subject areas.

1.10 CD-ROM Material

CD-ROM material in the form of a CD-ROM encyclopaedia provides an extensive baseline corpus from which to draw generalisations about linguistic phenomena in the study undertaken. It also provides a wide enough range in terms of subject matter to address all of the varieties of courses included in the foundation year for undergraduates. Although it will be argued that a multimedia encyclopaedia on CD-ROM is of a less sophisticated nature than a general science textbook, this is denied by a number of researchers (Huddlestone 1971, Swales 1985, Halliday and Martin 1993). These linguists argue that encyclopaedia texts are intended to be instructional and so are textbooks. Furthermore, textbooks and encyclopaedia are seen to be of a similar level or standard. They are also aimed at a similar reader, that is one who has knowledge of the subject but is not a specialist and is in the process of learning more. Although encyclopaedias may of course be used for more general purposes, more in-depth information can be obtained if the user so desires. The multimedia encyclopaedia chosen provides reading lists for further study on any topic.

The CD-ROM encyclopaedia also shares a number of features that are particularly relevant for our tertiary level students. First of all almost all of the widely published CD-ROM multimedia encyclopaedias are in American English. This is partly as a result of Microsoft's dominance in the computer market as mentioned earlier and their marketing strategy of linking other products to the sale of their personal computers.

Secondly the length of the texts reflects new technology where 'screenfuls' of information have to be coped with. Search facilities are provided. Educational policy in Europe suggests that being able to handle searches and obtain information from new technology are skills students need to have in order to work successfully with such important applications as the Internet.

Third, an added dimension of "range", that is, how many texts the items are used in, can be analysed from the CD-ROM so that the usual context of lexical items can be examined. A wide range means that the item is more useful as it can be used in more situations. This is consistent with Michael West's (1953) idea of a high 'surrender value' for student learning or Swales' (1985) "maximum educational value". The aspect of context is also relevant for syllabus design where an appropriate context for lexical items is essential in order to make the teaching materials designed for use on the syllabus reflect authentic or natural contexts of use in the genres of science and technology as opposed to in general English genres. Robinson (1991:20), discussing the difficulty in delimiting different forms of English as every situation overlaps with another, prefers the use of the term ‘technolect’ after Lauren and Nordman (1986) because it suggests a form of language rather than an independent language as opposed to the term ‘special language’ to describe the difference between general English and this form of English.

Multimedia technology applications are seen as a possible future for study for students at this level in particular. All the developed countries are keen to foster the use of new technology in education and the University of Aveiro is committed to this objective too. The European Union is involved in many projects in this area, including going as far as the development of virtual universities like that of the Universitat Oberta de Catalunya, Spain[13]. Furthermore, advances in language teaching methodologies also suggest incorporating technology in learning strategies. Computer driven searches of appropriate corpora which may be from CD-ROMs are increasingly being used to present to learners precise, authentic language use in specific genres where learning is seen to be data and learner driven. (cf. Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora*, London: Longman.)

Much documentation supports the spread of these technologies. The European Round Table of Industrialists Report (1997:6) says that “the number of CD-ROM drives rose from 2.7 million to 9 million between 1994 and 1995, and is expected to reach 35 million units by 1998.” This report goes on to suggest that “Not integrating ICT (Information and Communication Technology) in the education process would further

[13] This is an example of an Open University for distance learning programmes which makes use of computer technology only.

widen the gap between real life and education. Youngsters are growing up in an informatics and media world: education should respond to their cultural expectation pattern, use their language." The CD-ROM encyclopaedia fits this role, but maintains an educational rather than an entertainment perspective. Young people all too often use games on CD-ROM as their "language" and although motivation is extremely important in teaching and learning, this thesis argues that education at tertiary level should be both stimulating and demanding.

The number of CD-ROM encyclopaedias has increased in recent years but in the early 1990s there were only a few widely available ones such as the Grolier, Compton's and Encarta by Microsoft. Compton's was not very user friendly while the Encarta tended to take on a more entertainment type of format including quizzes and games. For these reasons the Grolier was chosen as it combines a suitably academic style with user-friendliness. The report written by Jeremy Fox, Anne Matthews, Clive Matthews and Arthur Rope for the British Government Employment Department Group Training Agency Learning Technology Unit by the University of East Anglia and the Bell Educational Trust, March 1990 *Educational Technology in Modern Language Learning in the secondary, tertiary and vocational sectors*, describes the GROLIER ELECTRONIC ENCYCLOPEDIA which will be used here (ibid.1990:26) as "an excellent example" of a CD-ROM encyclopaedia which "holds the equivalent of 20 bookshelf volumes plus an index of all the occurrences of every word in the encyclopaedia." and the report grades this encyclopaedia for "secondary, *tertiary* and *vocational*" levels, with emphasis on the latter two which they indicate by means of the italics used. Furthermore, the report claims that the encyclopaedia is applicable to the areas of reading, writing and vocabulary and can be used "in a hypertext-like way down a track of cross-references".

The appropriateness of hypertext in teaching Portuguese students of English has been explored by Prof. Doctor António Moreira in his doctoral thesis *Desenvolvimento da flexibilidade cognitiva dos alunos-futuros-professores: uma experiência em Didáctica do Inglês* (1996 University of Aveiro). He finds cognitive learning of this type to be successful with students in Portugal. He found that (1996:x) "hypertext systems based on an approach that uses cases which are structured in such a way that they offer multiple representations of knowledge which in turn emphasise critical interconnections between different structural and surface knowledge components can be superior in their effectiveness for the preparation of students in their use of knowledge in new and in novel situations". This form of transfer is extremely important for the students under study here who are attempting to use English in a 'new and novel' situation - that of tertiary level study in science and technology.

1.11 The Syllabus

This study will attempt to examine and define just what the appropriate specific English for these undergraduate science and technology students is. This will be found through a linguistic analysis of computer corpora, from the textbooks for physics and chemistry found on the students' bibliographies, contrasted with an analysis of the students' language needs as obtained from the results of the tests described later in Chapter 4. The areas that must be addressed in undergraduate English language studies will then be identified. The results of this research are to be applied to the development of a syllabus and teaching materials for the discipline appropriate for the entry standard of English of the students taking the discipline and for their overall course needs in terms of bibliography in English for science and technology.

Many other considerations will have to be taken into account as well such as the amount of contact hours available, the size of classes and the heterogeneity of the students in those classes. All of these features of the discipline will influence the syllabus that can be used with these undergraduates. The fact that these are undergraduates just starting their courses in university will also have to be taken into consideration as mentioned earlier as they are going to have to adapt to many new aspects of life as well as new aspects of learning in an entirely different environment from the one the have been used to up to this point. Simply adapting to the size and complexity of university life is a major difficulty for many of the new students who may also be coping with being away from home and family for the first time as well. Students cannot be seen divorced from these different aspects of their lives which will colour their learning and attitude to learning and which the teacher and syllabus designer have to take into consideration in their work. The syllabus then needs to address the state the learner is in at the beginning of the course not only from the point of view of their level of knowledge, which will vary from student to student, but also from their personal situation with regard to university life. There will be a need to draw together a number of strands to blend the classes into some form of co-operative body where the differences of level in background knowledge, both of their subject specialities and language level, together with other more mundane problems of their new lifestyle will be addressed. Simply getting the students into contact with each other and making friends is important for the well-being of undergraduates and their success on their courses (Tavares, Santiago, Lencestre, Soares 1996).

1.12 The Research

The principles of orientation underlying this research in order to produce valid materials for Science and Technology students in their undergraduate years at university must therefore take into account the following criteria:

1. The English must be the American English found in those academic textbooks in the students' bibliographies, which were written to be used to teach native speaker students at undergraduate level. Biber (1988:201) finds there are systematic differences in British and American written texts in that "American written genres are consistently more colloquial and involved than British written genres, while at the same time American written genres are consistently more nominal and jargony than British genres." This is a general observation of differences between British and American English. However, the specific differences between genres with respect to the undergraduate science textbooks has to be studied.

2. The fact that there has been a technological revolution, which has affected both teaching and learning, must also be considered. The latter is reflected in this study through the use of source material taken from multimedia encyclopaedia which will serve as comparison (variation study) with the textbook corpora and provides the necessary scope of information for science and technology students studying on a wide variety of science and engineering courses.

3. The corpora must be compared with other studies to identify how far they vary from the results obtained for other science and technology texts. Biber's (1988) algorithms on academic prose will be used for this purpose.

4. Finally, the results obtained must be compared with the language the students have already acquired or learnt, which will be established

through testing, in order to identify mismatches with the English needed by those students coming into the University of Aveiro to take up places on Science and Technology courses.

1.13 Methodology

The research undertaken follows this pattern: firstly frequency and range studies of the multimedia encyclopaedia are made and then these are compared and contrasted with scientific text-types taken from the actual coursebooks used in Aveiro University for science and technology students in the first year. Frequency lists are used as one of the bases of data on variation from which descriptions of language and therefore decisions about appropriate language can be drawn for application in the materials taught in this discipline.

Comparison of textual features are also made following Biber's (1988) variation studies methodology to provide a detailed scientific comparison for the research. As some linguists have argued (cf. Roberts, 1983) the definition of an area is complicated as there is often an overlap between disciplines and between text types, which makes it difficult to describe the features which uniquely pertain to that discipline. Variation studies overcome this complication by looking at *how far* the texts under study differ from other texts rather than at an absolute contrast between them.

The level of the English the students coming in to the first year in the university have acquired is examined through tests carried out over five years of student intakes, together with the English needs identified by the different university departments for their university students. The latest research on language acquisition and corpus linguistics will be applied in order to determine what is relevant for the teaching or learning of such students. The role of modern technology in education is also addressed in the learning strategies proposed for these students.

Chapter 2 Historical and Theoretical Background to ESP

Chapter 2

Historical and Theoretical Background to ESP

The difference between general English and specific varieties of English is a very problematic area as distinctions can be drawn either between the differences in what language is used for or the differences in the language used. Despite this, many kinds of analyses have been carried out over the last few decades. The argument has come full circle and often now revolves around what general English consists of as opposed to what specific language is in either literature or science. Recently there have been new attempts to define this general English (see later 2.5.7 The British National Corpus). The reason for the attempt to define specific varieties of English was often in order to apply the findings to syllabi for teaching purposes to provide the students of those courses with tailor-made material and information about the language they needed in their studies of a particular subject The research that has been carried out has had a much more widespread effect than this however. Mainstream English as a Foreign Language (EFL) teaching has benefited greatly from the results of linguistic thought and analysis carried out ostensibly for ESP. Before it is possible to move forward in research for ESP it is necessary to review what has gone before to find what can usefully be done to try to add to the store of knowledge that has been built up. It is also important to examine how the work done in the past has affected syllabi and the teaching of English to science and technology students. In other words a synthesis of the research and ideas that have gone before can be of enormous help in defining what should be included in a syllabus for university students of science and technology.

A methodology for the teaching of language can be traced right back to Quintilian (Marcus Fabius Quintilianus. 35 - 95 A.D.) with his *Instituto Oratoria*. He outlines the teaching of rhetoric or *bene dicendi scientia* as being made up of the study of grammar which is sub-divided into correct expression or *recte loquendi scientia* and interpretation of the poets or *poetarum enarratio* which, in turn, requires the study of writing and reading or *scribendi legendique facultas*. Quintilian was aiming to produce the perfect orator through his system of linguistic studies and states that the first requirement of an orator is that “he should be a good man”. Quintilian’s methodology was that a second (or foreign) language should be taught to children through total immersion in the target language, although he also advocated adapting materials to suit different types of learners and of motivating students to learn. This idea of different types of learner requiring different types of materials is fundamental to the modern study of languages for special purposes. Similar to Quintilian’s ideas on motivation is the modern idea that motivation is a necessary prerequisite to facilitate learning which is advocated by those involved in special language training today (cf. Hutchinson and Waters 1987).

From the seventeenth to the eighteenth and nineteenth centuries, from Locke to Horne Tooke and Humbolt theories about language led to etymological studies and then descriptions of languages being made. The emergence of a method of ‘scientific’ study of language based on empirical research continued into the twentieth century with work such as that of Bloomfield on indigenous American Indian languages. All of these strands of theoretical linguistics have had and still are having effects on syllabus design, materials and the teaching and learning of special languages like that of science and technology under study here.

2.1 English for Special Purposes

The historical background to English for Special (or Specific as it was known in the 1960s) Purposes (ESP) and in particular, to English for Science and Technology (EST)[1] has neither the Quintilian requirement of ‘goodness’ nor the possibility of using his methodological approach. This field was developed for working adults, not children, and often for those in tertiary education who were studying science and engineering and for whom English was a foreign language. Locke had already, in the seventeenth century, shown concern about language as a(n) (imperfect) vehicle for the acquisition and spread of knowledge. However, he proposed remedies for these imperfections of language so that language could be used safely for the purposes of science and philosophy. Locke’s remedies consisted of the definition of complex ideas through the use of simple ones so that communication between people was possible and a bridge between minds could be built. This process Locke believed to be similar to demonstrating a mathematical conclusion and that thereby this use of language would become adequate for scientific and philosophical discourse. As a result of Locke’s ideas, there was great interest in etymological studies and many dictionaries were subsequently written in the eighteenth century. This process cannot be seen to have come to an end and research into language use for the writing of dictionaries (see 2.5 the Corpus Analysis Approach) continues today but modern technology is now used to improve the accuracy of the information included. Swift also argued for an Academy to ‘fix’ the language for scientific purposes. In other words, historically there has been a tendency to invent a language of science which is removed from

[1] The use of the term English for Science and Technology (EST) is usually attributed to Ewer (1971)., although Trimble (1985) attributes it to Selinker.

general language. This was particularly the case when Latin stopped being the *lingua franca* of scientific thought. The idea of language as fixed is contrary to fact whether for scientific purposes or any others (White 1998 and see 5.1.5 Plurals from Latin and Greek) which suggests that there will always be a need to analyse its use both diachronically and synchronically.

Theories about language have not stopped being put forward either and language studies for application in teaching have added more to the understanding of specific varieties of English. For example Swales (1985:x) sees EST as underpinning the development of ESP. He says that "With one or two exceptions ...English for Science and Technology has always set and continues to set the trend in theoretical discussion, in ways of analysing language, and in the variety of actual teaching materials."

2.1.1 Phrasebooks

Special purpose language has a very long history with the 'traveller's language course' for those who intended to visit a foreign country and who wished to obtain a smattering of the language of the phrase book type. Course books aimed at this type of learner have been in existence since the sixteenth century according to Archibald Lyall (1932) in his *Guide to the Languages of Europe: a practical phrase book* where he quotes from '*Colloques ou dialogues avec un dictionnaire en six langues* of my earliest predecessor, Henry Heyndricx, Antwerp, 1576' (Strevens 1978). Strevens (1978:190) claims that phrase-books for foreign tourists have been in existence for four hundred years and Opitz (1983) that mariners have been using specialised bilingual maritime dictionaries for more than two hundred years. This type of phrasebook language is an example of the approach which emphasises the objective or outcome of language use or 'specific purpose' as opposed to 'special language' (Turner 1981) and is based upon the idea that there is an equivalent in one language for an item found in another language. This approach takes the view that one language corresponds directly to another language although it is in code and ignores the idea that there are cultural differences between languages which need to be coped with.

The implications for language teaching for foreign travel and phrasebook language has been significant in that different emphasis was placed in teaching on speaking and listening, although traditional teaching methods in the past would have favoured pronunciation and reading aloud of phrases. The results of theories about the reasons for teaching English has changed the methodologies and materials used for that teaching. The final example given above for mariners would be entitled EOP, English for Occupational Purposes, today and might well restrict itself to very elementary goals. Similarly, just as new dictionaries continue to be produced, the study of English for Special Purposes continues to this day and there have been four, often overlapping and interconnecting major schools of thought for the teaching of ESP and science and technology this century. These can be described as: the register analysis approach, the discourse analysis and variation studies approach, the needs analysis approach and, most recently, the corpus analysis approach[2] which I shall go on to describe in order to show how they influence the study of science and technology for university students today.

Some of these approaches derived from the need to respond to a practical crisis like the need during the Second World War for a means of teaching/learning foreign languages quickly. Others have had the benefit of taking up theoretical work done by linguists which has then been

[2] The latter also often being developed and used for the gathering and analysis of data for writing new dictionaries.

applied to teaching/learning situations. The teaching of foreign languages has been benefited by studies in special language teaching/learning and to a lesser extent special language teaching/learning has benefited from general language teaching/learning (Robinson 1991, Swales 1985).

Hutchinson and Waters (1987, Ch. 2) identify four stages of development of special language analysis, with a fifth emerging. To a greater or lesser extent, all are germane to my project. They are:

1. The concept of special language: register analysis,
2. Beyond the sentence: rhetorical or discourse analysis,
3. Target situation analysis,
4. Skills and strategies
5. A learning-centred approach.

In this analysis the last two of these stages are considered to be methodological approaches and not language research as such. Indeed, Hutchinson and Waters themselves identify this division as one of “new ideas about language and new ideas about learning” (ditto, 1987:14). The methodological implications of learning English for science and technology will be addressed later (Chapter 7 The Syllabus) after presentation and discussion of the results of the research undertaken.

I will briefly discuss each approach before showing how they feed in to my thesis.

2.2 The Register Analysis Approach

The very first research material published for teaching was not meant for special or specific purposes but rather for the general learner. The idea that students could be helped to learn languages more easily and quickly by having a select list of words came into vogue between the wars. In terms of word lists or frequency counts, the earliest which was used to provide a scientific foundation for teaching was developed in America by Thorndike. Thorndike produced a list of 5,000 words for teachers, culled from a corpus of four and a half million words, which he published in 1921 as the *Teacher's Word Book.* Subsequently, Horn published 10,000 words taken from business and personal letters in 1926. The number of words published then multiplied to 20,000 in 1931 with Thorndike's *The Teacher's Word Book of 20,000 words* and in 1944 to 30,000 words with Thorndike and Lorge's *The Teacher's Word Book of 30,000 words.* These counts were used to decide on appropriate reading materials for school children and are still popular means to decide on appropriate materials for different school levels. In Britain, Michael West published his *A General Service List of English Words* in 1953 which contained 2,000 words and a supplementary list of scientific and technical vocabulary. Similarly, the graded readers in English such as those published by Penguin are meant to correspond to levels of reading competence in EFL learners learning British English. West's work like that of Palmer was specifically addressed to EFL learners whereas the American work by Thornton and Lorge was for general reading in mainstream education systems. However, all of these word lists were based on written material but in France work by Gougenheim was taking place on a spoken corpus which was to provide the basis of *Français Fondamental*, first published in 1954. The first revision of this was published in 1959 and was composed of 1475 entries, 1222 of which were lexical items and 253 grammatical words. As will be shown later (2.5 The Corpus Analysis Approach) the whole concept of distinctions between grammatical words and lexical items is called into question through corpus research. Nevertheless lists can still be used to help define a specific type of language or special language and to underpin syllabi.

2.2.1 European Languages for Special Purposes

Just as the Second World War had served as an impetus for teaching foreign languages expeditiously and for research into teaching foreign languages in Britain and America, the European Union has been and still is a driving force behind research on foreign language teaching and languages for special purposes (LSP) in the new Europe. There have been studies published from work commissioned by the Council of Europe, namely Van Ek's the *Threshold Level (1976)* and *Waystage English (1977)* designed to specify exactly what children should be taught in their secondary school courses in English as a Foreign Language. These last two continue to be published and similar books are now produced for all the member countries' languages. The *Threshold Level* has in fact been republished as recently as 1994 and there is now a *Threshold 1990* published together with Trim which purports to set out "how a learner should be able to use English in order to function independently in everyday communication." The focus given in this kind of work is on secondary education particularly and is designed to encourage all European citizens to speak other European languages. The intention that is usually explained is that all citizens should know a minimum of three European languages. However, as was mentioned in Chapter 1, 1.4 The Dominance of English in Science and Technology, certain European languages like English are becoming much more widely used and is learnt at school by an increasing number of European citizens.

2.2.2 Methodologies

These different lists of what children should be taught on language courses in schools were produced based on different methodologies. The

Threshold Level reflects the functional/notional approach popular in the 1970's because of Wilkins' work (among others), Wilkins (1976) *Notional Syllabuses,* Oxford: Oxford University Press. The lists produced based on the functional/notional approach were not produced empirically and many other possibilities of what to include at what stage in learning exist. The variety of coursebooks produced to teach functions/notions in the late 1970s and early 1980s reflect this and authors such as Abbs at the TESOL Conference in Lisbon in 1979 admitted that things were beginning to get out of hand when students were being taught how to react angrily to situations! So as a basis for syllabus design this approach left something to be desired.

If an empirical approach to register analysis is to be adopted for syllabus design, then there is more to the production of frequency counts that needs to be taken into account. For example there is the question of the definition of the words listed, reflected in the description of the *Français Fondamental* list mentioned above. What is lexical and what grammatical? The whole question of what a word *is* must also be addressed. Moreover, the corpus from which the word lists were produced must be clearly categorised so that the results can be seen to be pertinent to the learners' needs. These aspects must be made very clear if any comparison between research undertaken can be made and scientific results verified. Therefore, further consideration will be given to these aspects later in examining the results obtained from the corpora.

2.2.3 Scientific Specificity

Work by linguists, such as Halliday & Martin, Swales and Abbott, has argued for scientific specificity in order to make research comparable and, therefore, to add more to our understanding of language and language contexts. Abbott (1980:121), although writing on error analysis, argues that any published analysis should provide enough information for another researcher to reproduce the results, as with any other piece of scientific research. Halliday (1993:103-4) wants the research to make use of the same theories and methods of analysis (understandably advocating his own systemic functional grammar analysis here, above all others), so that comparison with other studies is possible. Robinson (1991:31) wants specific information on the materials used for the research to be provided. She (ibid.) says:

> "First, the fact that research exists on the same topic or subject matter that the students are interested in is not sufficient to make that research useful. We need to know the source of the material that has been researched: its date and geographical origin. In addition, we need to know the level of the material: does it represent specialist to specialist communication, or specialist to non-specialist? What was the mode of the material? Was it originally spoken or written, prepared or unprepared? All these alternatives will have an effect on the language forms selected.
>
> Second, we need to know the size of the corpus that has been researched (.....) Larger-scale studies may be able to arrive at reliable generalisations. Smaller-scale studies, however, may be able to go into more explanatory detail."

Robinson identifies one of the major difficulties of making use of much of the early research work carried out. Many studies exist but the exact situations to which they refer and/or the material on which the studies were based were often not clearly identified. This meant that it was difficult to build on those studies and many researchers were in danger of repeating work that had gone before. An example of this is the number of studies carried out (and the range of results obtained) on the active-passive in science texts. Even distinctions between what is analysed as active or passive has been discussed and various categories like the stative (Lackstrom, Selinker and Trimble) or equative *be* as main verb (Wingard) have been isolated and investigated as sub-categories of this dichotomy. This situation of reproducing work on similar topics probably arose because researchers were trying to cope with a situation they were presented with, usually abroad, and which needed solving in a short space of time. This was often the case for many of the university level courses in English in the 1970s designed for undergraduate students on many different courses, from economics to engineering.

Porter (1976) laments the fact that much of the research that had been carried out in other places was never published which meant that valuable data may well have been lost forever. Porter (1976:77) is particularly concerned that those linguistic analyses of 'the language of science' which have been lost would have been of "inestimable value" for others engaged in constructing teaching materials. Swales (1984:1) echoes Porter's sentiments when he says that although research was taking place in England in the sixties, it was usually being carried out by students on postgraduate courses and the "resulting projects and dissertations have long been buried in departmental and university libraries and are now almost completely forgotten". Even more recently linguists like Stubbs (1996:152) assert that "it has often been difficult to find studies which build on previous work".

Robinson (1980:18) argues that any useful research carried out "tends to confine itself to very limited, clearly defined, areas" such as, for example, prepositions in chemical abstracts. There is considerable difficulty in identifying only one "English of Science", because there is such a confusion of sub-genres or modes.[3] This state of affairs has made it difficult for textbook writers to base their material on register analysis.

[3] See McCarthy and Carter 1994 Pp 4-16 for a discussion of modes and their features.

Stern (1983:131-2) argues that the study of lexis or vocabulary has received little attention from English-speaking linguists[4] because it does not lend itself easily to structural and systematic treatment in the way that syntax and phonology have done but that this is an area of research which is very important for language teaching. More recent studies (Sinclair 1991) would suggest that the learning of lexical items in isolation does not reflect actual English usage where words and their meanings are associated with particular structures and contexts. Sinclair therefore suggests that words must be studied in context in order to show their specific meanings and associated structural restrictions.

2.2.4 Syllabus Implications

Halliday put forward the idea that language shows variety in terms of its use and not its user. For example, there is language specific to food or cooking: *tomato, apple, bread, butter* or sport: *referee, goalkeeper* but the user can only show different dialects which may be regional or social and so on. This sets the course for discourse analysis which will be described later but which Halliday (1993) still refers to as Register Analysis. Investigation of varieties of English can then show what the learner needs to cope with in a specific area.

Register analysis is based on the idea that nouns make distinctions because they are used for concepts or principles (which is somewhat similar to the ideas of Condillac in the eighteenth century, *Philosophical writings of Etienne Bonnot, Abbé de Condillac,* Hillsdale, N.J.:Lawrence Erlbaum, 1982). Consequently, if the register of say biology, as distinct from other registers, can be identified, then a specific syllabus can be

[4] Stern points out that research has been carried out by French and German linguists and reference is made later (2.2.7 The Impact of Modern Technology on Register Analysis) to Hoffman who describes the research done (on English) by German linguists in the GDR for teaching purposes.

drawn up which would be more limited in range and therefore, would not diffuse students' learning energies. West describes this as the 'surrender value' of the course of study. A high 'surrender value' would mean greater efficiency in terms of meeting the student's language requirements or as Swales (1985) puts it, as getting 'maximum educational value' out of the course.

White (1975) reached the conclusion that it was a "unique constellation of features rather than any single characteristic" that made one register distinctive from another. These features however have to be identified from the kinds of materials that are likely to be used by the students so that a pedagogical selection can be made for course design. Mindt (1997:42) describes this process of designing a grammar for foreign language learners as:

- the compilation of a corpus of language data
- the construction of a didactic grammar from this corpus
- the derivation of pedagogical grammars from the didactic grammar

Sinclair and Renouf (1992) suggest that in general language courses "the main focus of study should be on:

a. the commonest forms of the language
b. their central patterns of usage
c. the combinations which they typically form"

These principles should hold for the results of register analysis of a particular variety of language usage such as that of science and technology, where the context in which language occurs would be of particular importance.

2.2.5 *Publications and Coursebooks based on Register Analysis*

Following this principle Barber (1962) published "Some Measurable Characteristics of Scientific Prose" in *Contributions to English Syntax and Phonology,* Stokholm: Almquist & Wiksell, in which he identifies three subject areas which are common to all fields: sentence structure, verb-forms and vocabulary. He analyses 350 sentences and finds 2 commands, 3 statements and commands and 345 statements which he compares with unpublished data from a colleague doing postgraduate work at the University of Leeds, W. Rumszewicz, who had carried out similar work on both English textbooks on agricultural studies and passages of prose drama. He concludes that in his colleague's scientific texts all sentences are statements but in the dramatic texts only two-thirds are statements, the remainder being questions or requests. There were also a number of articles and books published following the register analysis approach: Herbert (1965) published *The Structure of Technical English*, London: Longman, Swales (1971) based on his experience in the Middle East *Writing Scientific English*, London: Nelson, Ewer and Hughes-Davies (1971-72) based on their work in the University of Chile "Further notes in developing an English language programme for students of science and technology", *English Language Teaching*, 26, 1& 3, Strevens (1973) *Technical, technological and scientific English,* Sudhikam (1975) "Lexis" in *Guidelines Sample Materials for the Teaching of English to First Year Tertiary Level Scientific and Technical Students in Universities in Thailand*, SEAMEO/RELC (Mimeo), Friel (1978) "A verb frequency count in legal English", *ESPEMA Bulletin* 10, Wingard (1981) "Some verb forms and functions in six medical texts" in *English for Academic and Technical Purposes* (edited by Selinker, Tarone and Hanzeli), Newbury House, and Palmer (1981) *Register research design.* Rowley, MA: Newbury House. Coursebooks such as Ewer and Latorre's (1969) *A Course in Basic*

Scientific English and Herbert's (1965) *The Structure of Technical English* demonstrate this approach. Swales (1985:18) comments that Herbert's book was still in print and still being used when he published *his* book twenty years later. He (ibid.) attributes this to the fact that it "shows a highly professional concern with the language of EST" however the methodology used was rather dull and the combination and connection between the diagrams used and the accompanying text was often obscure.

2.2.6 Criticism of Register Analysis

One of the first criticisms levelled at register analysis was the fundamental one that there was no such register as a "general" register which could be used for comparison with a "special" register (see above: *Syllabus Implications).* Halliday takes this position but argues that it is useful to recognise a category of "special-purpose" language or language varieties.

A second kind of criticism of the methodology of register analysis began to appear grounded in the idea that no specific information on the corpus that had been used to derive the results was presented and what information there was showed that no spoken language had been included, which would imply that the analysis was not sufficiently representative to be able to provide generalisations about language.[5]

Swales (1984:10) claims that frequency analyses do provide evidence for generalisations about a "variety, type or style of the language" and that the people who criticise frequency work most are "those most given to making claims that such-and-such a feature is *important, frequent* or *interesting* without any evidence to support these claims". Porter (1976:77) also argues that attempts at linguistic characterization have been “surprisingly careless” and that even though Bloomfield wrote forty pages in 1938 on “Linguistic aspects of science” only one explicit example is given throughout those forty pages. This problem continues to the present day for example, although Stubbs (1996:152) argues for clarity on texts analysed, he is himself open to criticism (Hoey 1993) for not giving sufficient information on the school textbooks in his own analysis.[6]

The criticism was based on the fact that researchers appeared to be claiming that features were unique to one type of text or that one feature uniquely characterises a text. Once it was understood that the distinction was much more of degree the objections were largely overcome. As White (1975) said:

> Firstly, it became clear that ... it is not possible to take the occurrence of any specific feature as being criterial of one and only one particular register. Secondly, it was obvious that what made one register distinctive in comparison with another was a unique constellation of features rather than any single characteristic.

However, the difficulty of finding sufficiently large scale studies to work from remained.

2.2.7 The Impact of Modern Technology on Register Analysis

Before the advent of the computer it was generally believed that it was impossible to distinguish unambiguously between texts and that, therefore, such analyses were not worth carrying out, owing to the difficulty in coping with large corpora and because much of the language

---

[5] Although Robinson’s comment on small scale specific studies, see above P 55 Scientific Specificity, could still hold true.

[6] First presented with Andrea Gerbig (1993) as “Human and Inhuman Geography: On the Computer-Assisted Analysis of Long Texts.” In Hoey (ed.1993) *Data, Description, Discourse*. London: HarperCollins.

that is found in corpora is not distinctive[7]. This led to the idea that there is a core of language that is common or sub-technical (cf. Robinson 1991). Trimble (1985:129) equates sub-technical vocabulary with “those words that have the same meaning in several scientific or technical disciplines” together with “those “common” words that occur with special meanings in specific scientific and technical fields”. This approach goes hand-in-hand with the methodological principle that only students of an intermediate level of English competence could or should be exposed to an English for Special Purposes course, as these students will already have sufficient basic knowledge of the language to be able to appreciate the difference between these common forms and the language that is scientifically specific or purely technical. Trimble (1985:7) suggested that students at the tertiary level are “assumed to be fairly advanced in English” but, nevertheless, recognised that not all of the students could be assumed to be equally accomplished in all of the language skills.

In the same line of argument, Hoffman (1981:114) claims that:

> “It is the significantly frequent occurrence of certain speech elements, forms or structures that characterizes scientific writing and spoken discourse. As a consequence statistical methods play an important role in selecting an inventory for teaching purposes... It is the word and the phrase levels that yield the best results, i.e. lists of typical lexical items which may serve as a highly effective teaching/learning minimum.”

This statistical approach to language teaching, ‘lexicostatistics’ was particularly prevalent in the universities and technical colleges in the Eastern bloc countries such as the German Democratic Republic where frequency counts and terminological dictionaries were produced. These

---

7 Recent large corpus studies by Sinclair (1991) demonstrate that there is a body of very frequently occurring lexis and that if those items that only occur once are removed from the frequency list it shrinks to half its size.

days strict adherence to this sort of approach would be deemed too limiting for syllabus design, although it is generally accepted that it still has a part to play in it.

Swales (1984:1) reported that although frequency analyses found little favour in British and American ESP work, a revival of this form of study is taking place because of the fact that "frequency analysis is ideally suited to computerization". He also predicted (1984:214) that ESP would only come of age when computers and video recorders were used and the processes of technical and sub-technical vocabulary acquisition were properly investigated and not merely imagined. Computers, through concordances, can already provide learners with much better investigative tools and give access to real language use instead of inventions by the course writer. Tribble and Jones (1990:15) find that the results obtained from a concordance "will only be as interesting as the raw material on which you put it to work" so that appropriate corpora must be used to generate instances of language usage which match the students' needs.

Biber, Conrad and Reppen (1998:136) say that teachers must understand the processes by which register is understood so that they can facilitate its acquisition. However, they go on to suggest that the ability to describe and understand the differences between registers has proved to be very difficult without the use of corpus-based studies. Furthermore, the features that distinguish one register from another are rarely features unique to that register. Registers usually share many linguistic features; it is the relative use of these features that usually distinguishes one from another. Therefore what is needed is a comparative quantifying approach in order to know whether one feature in a register is rare or common.

2.2.8 Variation Studies

Ewer and Hughes-Davies' (1971) syllabus suggests that particular forms such as the present simple tense, the passive voice, conditionals, compound nouns and anomalous finites are more prevalent in scientific writing, and so many of these areas were researched in the 1970's. Some of that research led to further divisions being made, for example between passive and stative (Lackstrom, Selinker and Trimble 1972). Strevens (1973) put forward the suggestion that different kinds of language could be distinguished "within 'scientific discourse'. These were: the scientific, the technological and the technical, each of which, according to him, contains its own characteristic range of style.

Halliday (1993:37) points to nominalization as a characteristic of scientific prose. Halliday takes this further and argues that "at the level of social context, ideology is realized by genre, which is in turn realized by register", in other words he sees connections of an ideological nature within text from its register to its genre. Halliday (1998:186-7) says that grammar, "in its ideational guise" transforms "human experience into meaning" but it also "impose(s) a categorisation" on "our perceptions of phenomena". This preoccupation with ideology is now dealt with in studies of social semiotics in discourse analysis.

More recently variation studies by such people as Biber (1988) and Finnegan have postulated criteria for the identification of genres and text-types and have reached the conclusion that several combinations of factors (both present and absent in specific genres) help to pin down the linguistic features that can be found in those genres. Their research also shows that there are no absolute black and white differences between genres and that this is particularly the case in the various sub-genres of academic prose. Biber (1988) postulates a continuum as a better representation of differences between genres with speech at one extreme and formal writing at the other. Nevertheless, academic prose was found by Biber to be one of the most widely differing sub-genres in the analyses he carried out. Biber's work will be discussed in more detail later in 3.3 Biber's Methodology of Variation Studies and Corpora Analyses when making a classification of the textbooks on the undergraduate students' bibliographies.

2.2.9 Recent Studies

Work on register analysis is still very dynamic at the moment particularly with Halliday and the 'Australian school'. Halliday (1993) dates scientific discourse as having arisen following Newton, in the early eighteenth century and describes some changes that have evolved in scientific writing. Halliday (1993:58) asserts that Newton showed no coyness about using *I* in his writing and that the 'suppressed person' passive favoured by teachers and scientific editors only came into fashion towards the end of the nineteenth century. The point being that scientific language has not only been invented for the purpose but that it is also historically mutating.

Dissatisfied with register analysis as a means of identifying differences between genres, some researchers put forward the idea that what made one genre different from another was at a 'higher' level than the sentence (or supra sentential, see Discourse Analysis below) which was initially termed discourse or rhetoric analysis. The preoccupation of linguists such as Strevens (1977), White (1975) and Candlin *et al.* (1978) was the purpose of the language use. Nevertheless there is a perpetuation of register analysis within discourse analysis so that these two terms are often used indiscriminately (or together) in modern methodologies. McCarthy and Carter (1994:36) represent the division of register and genre for 'reporting genres' in the following way:

Discourse world

↕

Core generic function

↕

Genres

↕

Generic blends

↕

Registers

Where the Discourse world is divided into spoken and written; the Core Generic Function is Reporting; the Genres are for example, the Information Report, the Progress Report or the Weather Report; the Generic Blends are for example, Reporting and Predicting, Reporting and Recommending or Reporting and Evaluating; and the Registers are for example, a Weather Forecast which can be even further sub-divided into TV/Radio and Newspaper. These can then be linked to different "Prototypical linguistic features", such as "Past tense, passives, relational processes". McCarthy and Carter (1994:33) survey some of the uncertainties that are found about definitions of genre and suggest that the results are that "the notion of genre becomes as slippery as the notion of register." They ask if this distinction is necessary at all but they note that it has had important implications for discourse analysis.

Biber (1988) uses genre to refer to "categorizations assigned on the basis of external criteria" and text-type to refer to "groupings of texts that are similar with respect to their linguistic form." The term register is still used most extensively by Halliday and the Australian school.

2.3 The Discourse Analysis Approach

2.3.1 Definition

Coulthard (1978) defined the concern of discourse analysis as 'the identification and description of supra-sentential linguistic structure in written and spoken texts.' Widdowson (1971) exemplified the discourse analysis approach to teaching as asking what the students need to use the language for, its uses and functions. Students must be able to describe, explain cause and effect, make comparisons and classify. His argument is that these functions of language do not change whether students are studying geography or chemistry but that they cross the subject boundaries. Furthermore, it was suggested that since students have been taught these functions in their first language (L1), teachers can appeal to this framework and with some adjustments only need teach the English realisations of these characteristics.

Widdowson (1977) advocates a distinction of terms between 'text' and 'discourse'. He suggests that text is viewing a stretch of language as an exemplification of the structure of the language, especially of devices to indicate structuring above the level of the sentence. Discourse he describes as viewing a stretch of language as a unique piece of communication.

2.3.2 The American School

A similar and parallel school of thought developed in America with Selinker, Trimble, Lackstrom and Todd-Trimble, who were concerned about the functioning of language above the sentence level and with identifying the organisational patterns in texts and the linguistic means whereby these are realised, as compared with register analysis which had concentrated only on sentence grammar. Discourse analysis does not rule out the analysis of individual sentences and so subsumes register analysis to some extent (see Widdowson 1977 above). However, in order to discover how given grammatical structures come to have given meanings in given contexts, or how larger textual or topical constraints affect the choice of individual lexical, exophoric or anaphoric items within a given clause, discourse analysts believe it is necessary to study whole texts[8].

2.3.3 The Prague School

In the decade preceding World War II a Czechoslovakian branch of 'functional' linguistics came into being, which is now referred to as the Prague School. These linguists were led by the work of Vilém Mathesius, the Russian Nikolay Trubetskoy and the Czech born American Roman Jakobson. The dominant characteristic of the Prague School approach was its combination of structuralism with functionalism and how it stressed the importance of the cognitive, the expressive and the conative (or instrumental) functions fulfilled by language. In their approach the cognitive function of language refers to the transmission of facts, the expressive to the indication of the mood or attitude of the speaker (or writer); and the conative its use for influencing the person addressed or bringing about some practical effect.

A number of scholars of the Prague School have suggested that these functions correlate in many languages with the grammatical categories of mood and person. The cognitive function is fulfilled by 3rd person non modal utterances; the expressive by 1st person utterances in the subjunctive or optative mood; the conative by 2nd person in the imperative. In their work on criticism of literature and their stylistics

[8] Once again *text* here should be understood to mean both written and spoken language.

studies, their key principle is that language is being used poetically or aesthetically when the expressive aspect is predominant.

The Prague School also conducted a lot of research into phonology and 'markedness'. The latter was then extended into morphology and syntax and it was suggested that morphologically unmarked forms have a much wider range of occurrences and a much less definite meaning than a morphologically marked form. For example, morphologically *jumped* is 'marked' for past tense by means of the *ed* ending or suffix, whereas *jump* is 'unmarked'. *Jump* is not 'marked' for present tense and therefore occurs more widely. Similarly, in vocabulary *dog* is 'unmarked' and *bitch* 'marked'. *Dog* is a more general category and can therefore be used in many more contexts than can *bitch* which is a specific (female) type of dog.

More recently their work has been on the distinction between 'theme' and 'rheme' and the notion of 'communicative dynamism' or 'functional sentence perspective'. By the theme of a sentence is meant that part that refers to what is already known (also called 'topic', 'subject', 'old', 'given' or 'shared information' by other linguists) and by rheme the part that conveys new (or unknown) information. Halliday, although a member of the British School of linguistics (see below 2.3.4), uses both 'theme' divided into 'theme' and 'rheme' and 'given' and 'new' information in his discourse analyses. It can be seen from this how the different schools in fact interact with one another and often deviate only through their primary focus.

2.3.4 The British School

The British school of 'functionalism' was led by linguists like J.R. Firth, Michael Halliday and John Sinclair. These linguists rejected the study of language ‘by itself’ and resolved to study what people actually say in context. Firth considered three features of “typical repetitive events in the social process”, which were:

1. The participants: persons, personalities and relevant features of these.
   (a) The verbal action of participants .
   (b) The non-verbal action of the participants.
2. The relevant objects and non-verbal and non-personal events.
3. The effect of the verbal action.

Sinclair’s group pioneered discourse analysis through fieldwork on classroom discourse. They rejected the traditional linguistic terminology and referred to discourse ‘moves’ such as ‘initiation’, ‘nomination’ and ‘follow-up’ by the teacher and ‘bid’ and ‘response’ by the learner (Sinclair and Coulthard, 1975; Sinclair and Brazil, 1982). The hypothesis that has developed today (Sinclair 1992) from this work is that a three-part structure (Initiation, Response, Follow-up) is always the option in spoken interaction and is virtually obligatory in many types of discourse. All of the neo-Firthian linguists have taken the view that the analyses they have made and theories they have developed could have a practical application.

Van Dijk (1997:29) identifies this preoccupation with actual occurrences of spoken language as the basis of most of the current research in discourse analysis. However, Halliday and Martin (1993) are more interested in how the translation of one type of discourse changes its status by being translated into another discourse. In their case it is the loss of status involved in mediating of the language science in schools through speech and testing through short answer writing as opposed to through writing reports like those the students will need to interact with in order to become proficient in the language of science. Halliday (1993:202) sums up the argument thus:

> To rehabilitate literacy in science teachers and students will have to work towards a much clearer grasp of the function of language as technology in building up a scientific picture of the world. Technical language has evolved in order to classify, decompose and explain. The major scientific genres – report, explanation and experiment – have evolved to structure texts which document a scientist's world view. The functionality of these genres and the technicality they contain cannot be avoided; it has to be dealt with. To deal with it teachers need an understanding of the structure of the genres and the grammar of technicality. With this knowledge they can begin to tackle the problem of science literacy ... Without it they will continue to focus on content without taking language into account, probably with an increasing emphasis on science activities rather than science texts. The linguistic technology is the key -–not just to science literacy but to understanding and practising science itself. Ways must be devised to provide access to this technology. And the answer must not involve watering the technology down.

2.3.5 Systemic Functional Grammar

Halliday built on Firth's schema and developed his systemic functional grammar to analyse genres. His systemic functional model divides the communicative function of a text into three:

- Field of discourse (social action): the processes, purposes and subject matter with which the participants are engaged.
- Tenor of discourse (role structure): the linguistic and extra-linguistic role relationships of the participants. The linguistic roles are speaker and hearer, initiator and responder and so on. The extra-linguistic roles are the social identities and relationships of participants.
- Mode of discourse (symbolic organisation): the part that language plays in situations and how it does so. This includes such notions as

medium (e.g. written, spoken, typed) and immediacy (e.g. face-to-face or distant).

Halliday further defined these aspects of register (organisation of content) through their metafunctions (organisation of language). The choices for meaning are organised into the following metafunctions: Field is associated with Ideational meaning (resources for building content); the metafunction associated with Tenor is Interpersonal meaning (resources for interacting) and the metafunction associated with Mode is Textual meaning (resources for organizing texts).

Halliday argues that scientific texts are derived historically from the need to condense information about previous scientific discoveries. They are therefore characterised by dense nominalization as this is the best means of conveying dense information. Here, once again, is the notion that solid background (or underlying) scientific knowledge of the subject being studied is necessary. Furthermore, this background knowledge is assumed by authors of textbooks to exist in their readers. This is in conformity with Labov's (1972) suggestion that only by using the concept of 'shared knowledge' can discourse be interpreted correctly. However, Trimble (1985:114) says that his research "showed and continues to show that the majority of non-native students lack the cultural background that enables them to bring more than a very limited amount of the presupposed information to their reading of EST discourse". Here he is referring to the information presupposed by the writers of scientific and technical discourse to be 'possessed' by the reader.

Horne Tooke (1778, 1786) argued back in the eighteenth century that language contained 'abbreviations'. Like Halliday's 'condensation' of knowledge he argued that abbreviations were necessary so that thoughts could be expressed in real time and that abbreviations had been developed over time so that an (empirical etymological) analysis of language would show that abbreviations such as prepositions could be traced to their historical (nominal) roots. In the *Diversions of Purley: 9-15* he says:

> "Abbreviations are employed in language in three ways:
>
> 1. In terms
> 2. In sorts of words.
> 3. In construction.
>
> Mr. Locke's Essay is the best guide to the first; and numberless are the authors who have given particular explanations of the last. The second only I take for my province at present; because I believe it has hitherto escaped the proper notice of all."

Halliday (1993:148) argues that scientists have appropriated resources that already existed in English for their own purposes "to create a discourse that moves forward by logical and coherent steps, each building on what has gone before" and that this had emerged as the most highly valued model by the end of the eighteenth century. Nevertheless, he suggests that modern school syllabuses reflect an increasing emphasis on doing as opposed to learning science which means that "children are not taught to access the genres science has evolved to store information which leads to tremendous inefficiency in the science curriculum." Children, in his view, should be encouraged to study the text organisation of scientific texts to become literate in science.

2.3.6 Rhetorical Moves

Swales, in a seminal article, (1981) examined the introductions to 48 scientific journal articles using a "top-down" approach to the analysis of science texts. This means that he began with a consideration of overall text organisation and any statements regarding choice of structure at the sentence level or below are related to that higher level organisation. He concluded that there are certain rhetorical moves or 'macropurposes' within discourse. In the case of the introductions, he claims there are generally four such moves;

- Move 1 Establish the field,
- Move 2 Summarize previous research,
- Move 3 Prepare for present research,
- Move 4 Introduce present research.

Swales argued that these moves were more important than standard English grammar. However, in his later work (ibid. 1984:213) he says that genre-analysis has a "price to pay" in that by revealing something of the "internal logic and external language of a conventionally-constrained communicative event" it may "have little to say about other, apparently quite similar, communicative events". He gives the example of his own work when he says that there is "no such thing as *an* Introduction in academic writing" and explains that introductions "would appear to be quite differently organized in different genres such as scholarly papers, theses, projects and essays". These findings suggest that analyses of text organisation must be carried out on specific genres which are relevant for students in a particular setting in order to develop specific teaching materials that the students could use to develop their understanding and scientific literacy in the Hallidayan sense given above in 2.3.5. McCarthy and Carter (1994) give some suggestions about how this kind of text organisation analysis can be carried out in the classroom using frameworks.

There is however another practical application of this type of analysis in specific fields such as business negotiations. Johns (1991) reports that interest in Uljin and Gorter's (1990) work on discourse/rhetorical moves in business negotiations has increased with the enlargement of the European Union and the consequent enlargement of different language contexts/interfaces that must be dealt with.

2.3.7 Organisational Features

Some of the early claims made about the organisational features of texts particularly that of paragraph structure (see Ewer 1975) have been called into question by other researchers. Trimble (1985:14-18) finds that there is a ‘conceptual’ paragraph as opposed to a ‘physical’ paragraph in EST. He argues that the ‘standard’ definition of a paragraph is that it is made up of a group of sentences that express a complete thought and which are set off on a page of text by indentation or spacing. However, his analysis of written EST discourse showed that the ‘complete thought’ idea, given in the first part of this definition, might in fact be realised by two or three physical paragraphs which deal with lower-level generalisations or details supporting the main generalisation, which all together would make up the ‘conceptual paragraph’. Most linguists would now accept that the organisation and cohesion of a text is not restricted to discrete paragraphs but goes over and above that level and may be found even at the level of chapters in coursebooks. (Phillips’ unpublished Ph.D. thesis Birmingham 1985 cited in Hoey 1993). Hunstan (1994:204) argues for “unit boundaries” in discourse which, in her study of a report on experimental work, sometimes coincided with paragraph boundaries and sometimes occurred within paragraphs. She (1994:206) concludes that “Changes in status coincide with transitions from one unit to the next, while value and relevance serve to bind together sections that may cover several status categories.”

Heslot (1981) argues that studying a full text section by section can show important features of those sections that would otherwise be masked by a study of the whole text. She suggests that her research is, nevertheless, in line with Selinker and Trimble's (1976) recommendation to work at a higher than sentence level.

The 'whole text' approach has gained more and more adherents in recent years. Hoey (1991) argues for it as does Stubbs (1996). Much published work however still concentrates on detailed analyses of small fragments of texts. Stubbs (1996) recognises this fact but goes on to argue for complementing the analysis of text fragments by the analysis of long texts. McCarthy and Carter (1994:112) claim that "Matters traditionally thought of as the domain of semantics and syntax can be placed squarely at the heart of discourse analysis." They also suggest that (1994:106) a top-down approach can "assist the job of relating higher order categories in the syllabus (such as text-type) to the micro-syllabus elements of grammar and lexis".

2.3.8 Discourse Rules

Cooper (1981:426) argued that "the unpractised reader of academic discourse in a language other than his own may not know the rules of discourse for that language; or if he knows the rules (because they underlie discourse in his own language), he may not know the ways in which they are realised in that language." He suggested that it may only be necessary to encourage the student to use his existing discourse knowledge to 'decode' the second language. Trimble (1985:120) takes exception with this view which he believes leads to confusion; he cites the example of 'should' in its everyday use in contrast with its specific (different) meaning in scientific texts, he says:

> "In our experience non-native students tend to transfer their reading techniques developed for 'general English' to reading EST discourse without

realizing that adjustments are often necessary. As a result they read ‘should’ with the meaning found most commonly in ESL/EFL grammars and so assume that a choice is possible.”

The distance between Portuguese and English is not very wide compared with other pairings of languages but Scollon and Scollon (1995), discussing professional communication, warn that there may be problems of interpretation because the communication is between members of different discourse systems and it must be assumed therefore that, “knowledge, assumptions, values and forms of discourse” will not be shared and the problems that are likely to arise must be anticipated in order to achieve effective communication. Some of the dominant features of Portuguese academic prose are elaborate involved subordinate clause structures, the use of the first person plural ‘we’ verb form to describe work carried out in preference to the passive and paragraphs that may be composed of only one sentence. However, the influence of different schools of thought also underpin academic writing; the English often follow empiricist lines such as that of Firth mentioned above in the British School of linguistics whereas the French often follow more rationalist lines. It would be possible to find both of these influences in Portuguese academic prose together with the more modern integrative or ‘ecological’ approaches of the 1990s. Swales (1985:72) argues that “in cross-cultural situations we have again to take up the matter of ‘teaching’ in some way the rationale of scientific communication in English” and the idea of using a methodology based on the teaching of science in the first language put forward by Widdowson (1974) must be abandoned because that is now recognised to be a local phenomenon.

2.3.9 Coursebooks based on Discourse Analysis

A number of coursebooks began to come onto the market which reflected the principles of discourse analysis notably the Allen and Widdowson (eds.) *English in Focus* series which covered nine areas: *Mechanical Engineering* (1973), *Physical Science* (1974), *Workshop Practice* (1975), *Basic Medical Science* (1975), *Education* (1977), *Agriculture* (1977), *Social Science* (1978), *Biological Science* (1978), *and Electrical Engineering and Electronics* (1980).

Professor Widdowson assumes that students already have some knowledge of science and some knowledge of English which need to be put together. He sees the student as looking for the means to express in English what he already knows in his first language. However, despite the theoretical underpinning brought to the series by the editors, these books were not a great success which Swales (1985:72) attributes to a certain rigidity in the format and exercises which fail to take sufficient account of teachers' experience.

2.3.10 Criticism of Discourse Analysis

Despite this series, criticism was levelled at the utility of discourse analysis for teaching materials. Some felt that in the 1980s discourse analysis had only shown the relevance of coherence in text based on Halliday and Hasan's (1976) work in this area and that this was really rather too little as an entire methodology for generating teaching materials.

Furthermore, whilst it is undoubtedly true that certain characteristics like describing and comparing are essential for science and technology students, *it does not follow* that students have been taught these in their first language. Nor does it follow that the realisations of these are the same in the target language as in the mother tongue. Langkilde (1981:517) found, for example, that for undergraduate students of economics in Copenhagen long adverbials (in French) interfere with the comprehension of sentences and “disturbs the well-established patterns that the students are used to finding”. This phenomenon may well operate in the opposite direction between English and Portuguese academic prose as Portuguese corresponds more closely to French in its use of long adverbials[9]. Trimble (1985:131) believes that one of the features of scientific or technical discourse that is a “special problem for the majority of non-native students” is the use of noun compounds or strings which are Germanic in origin and so not natural in many languages which is certainly the case with Portuguese native speakers.

It is also a fallacy that there is a universal ‘scientific’ way of looking at things and that everyone with adequate intellectual gifts thinks ‘scientifically’. One of the characteristics of science is that it needs to be taught to people; it does not exist naturally. Moreover, many scientific discoveries have been shown to be rather haphazard (or in a hypothetico-deductive form rather than an inductive one[10]) and order and method have only been imposed when the scientists concerned have written up their work as a paper for other scientists to read.

Beaugrande’s (1997:44) offers an ideological critique of education and scientific training. He argues that “In theory, all citizens have the same basic human rights to freedom of speech, public education, scientific training,” and yet he claims that “in practice, the great majority are systematically excluded.” This he argues is because they cannot

[9] However, Quirk (1995:127) finds a higher proportion of adverbials in speech than expected which he says “runs counter to the widespread belief that written English is more complex syntactically than impromptu speech and that the incidence of ‘adverbial clauses’ is a significant marker of relative syntactic complexity.”

[10] An example of this is the book by Watson and Crick (1968) *The Double Helix* which describes their discoveries. Karl Popper (1972) in *The Logic of Scientific Discovery* argues that scientific method is hypothetico-deductive and not, as many believe, inductive.

understand the discourse of science and therefore are not science 'literate' as Halliday and Martin (1993) term this phenomenon. When looking into the future Beaugrande (1997:59), claims that less attention has been focused upon the 'twin *knowledge crisis* and *communication crisis'*. However, he believes that there is 'an exploding body of knowledge that is locked up in discourse accessible to only a few people concentrated in centres of wealth and power' which, he argues, needs to be made available to everyone through the results of the analysis of discourse being applied to teaching. So, for Beaugrande, discourse analysis is to be seen as the key to unlock the door of scientific language. This position seems reminiscent of the plain English group whose aim is to make bureaucratic jargon much more transparent for the average person so that they are not considered "functionally illiterate" as the Americans describe the inability to cope with filling in forms and other such language manipulation activities which the average person can be expected to meet in their daily lives. The language of science and technology as it has developed over the last two centuries would seem to be a far cry from bureaucratic jargon because it demands sufficient background knowledge of the concepts concerned to understand rather than a certain legalistic hedging of the terms used as in bureaucratic jargon.

McCarthy and Carter (1994) discuss perspectives of discourse analysis for use in language teaching and give a number of examples of how this can be done. They (ibid:122) find that students can be taught what they call 'natural interactiveness' when actual, naturally occurring language is analysed and the features that are used to achieve 'successful communication' are isolated. However, much of this analysis and description has yet to be carried out to provide a comprehensive guide for the teacher. It is possible that such work will become increasingly available for use by teachers in the classroom but the task is huge and it may be that access to data on CD-ROM will be one of the ways in which students themselves can gain insights into how the language operates naturally. Training will have to be given for this to be a feasible prospect with the teacher acting as guide so that neither overgeneralizations nor simplifications are made. With more work on discourse analysis and the analysis and use of corpora being included on teacher training courses this situation should become more viable as time goes by.

2.3.11 Educational Structures

Moreover, Martin (1993) suggests that students in schools (in Australia) are being misled about the nature of scientific register and are not being prepared to understand and write scientific English. He ascribes the lack of success in science studies to this deficiency. In other words Martin sees a mismatch between the teaching of science and acquiring a useful scientific style, leading to a lack of success in science at school which leads on to fewer students taking up science in higher education.

Adams, Heaton and Howarth (1991:2) also point out that there is a "false expectation that educational structures and systems do not differ internationally" and that students believe that foreign universities operate very similarly to their own. American textbooks written for undergraduates would, from this perspective, be working on assumptions made about the American educational system which is different from the academic situation of the Portuguese students under study in this research. The differences between these two systems would reside in the teaching that goes before, whether science subjects were taught in secondary school as combined subjects, optional subjects or 'pure' science subjects, how these subjects are linked at different stages in the curriculum, the amount of and amount of time dedicated to practical experimental work conducted by the students themselves or as demonstrations by the lecturer, the contact hours per week, staff to student ratios, funding and resources available and so on.

As with Halliday and Martin (1993), Swales and Najjar (1987) had noted a discrepancy between advice given to teachers and actual scientific texts they would meet later in their careers. Martin (1993) describes teachers being instructed to teach secondary school children science through description and personal involvement which he finds to be damaging to their later scientific writing and understanding of scientific texts, which often contain neither of these features. Moreover, Swales and Najjar (ibid.) found significant variability in the styles of writing in different scientific disciplines.

2.3.12 Student Competence

Hutchinson and Waters (1987) argue that it is the "underlying competence" which a student brings to the learning process that must be examined in order to decide what and how to teach. Widdowson (1974) argues that students being taught how to read scientific English should be encouraged to translate in order to discover the functional equivalence of what they have already learnt about science in their L1, to learn how 'certain acts of communication which are central to scientific enquiry' are realised. The level of the students' English on entering the university could thus be seen as only one side of the coin, with previous scientific learning being the other. Nevertheless, the university level that needs to be achieved must be sufficient to carry the students on to levels of understanding they do not yet possess, such as that we would expect from a university-level textbook.

One aspect of teaching that Hutchinson and Waters advocate strongly is motivating ESP/EAP students to learn. In universities in Portugal language studies are often seen as a necessary evil by both students and staff in science and technology departments. It is therefore often relegated to a minor position on the curriculum which cannot fail to reinforce the idea in some students that it is of little importance to their overall studies when in fact the ability to function effectively in several languages will often become increasingly important as far as both their courses and later careers are concerned.

2.3.13 Register and Genre Theory or Variation Studies

Strevens (1978:193-4) answers the question of what is different about scientific discourse in the following way:

> It is not the basic components of his (the author's) language that differ, it is the statistical properties of the mixture in which they occur, and the intention, the purpose, behind their selection and use.

He claims that, among other features, long sentences, long nominal groups and frequent passives are characteristic of scientific discourse.

However, Mary Todd Trimble and Louis Trimble (1981:199) declare that contrary to their initial ideas, surface syntax was not a matter of personal choice on the part of the author. They found

> that for native writers of scientific and technical discourse these grammatical choices were not arbitrary: in fact, we found them sufficiently patterned that we were able to make generalizations concerning the relationship between specific rhetorical functions and the grammar chosen to express those functions.

Thus there is no absolute division to be found between register (word and sentence) and discourse (above the sentence level) analysis and that both of these continue to be studied by linguists in order to define text types or genres.

Eggins and Martin (1997:230) describe Register and Genre Theory as “linguistic approaches to discourse which seek to theorise how discourses, or texts, are like and unlike each other, and why.” They go on to define the steps that need to be taken when applying such a theory. The first step, they maintain, is to describe the linguistic patterns or “words and structures” in the texts being analysed. The second step is to try to explain the linguistic differences found between the texts being studied. In short, Register and Genre Theory is a theory of functional variation or how texts coincide or differ one from another for a particular purpose.

Eggins and Martin (1997:251) define the terms *register* and *genre* as ‘context of situation’ and ‘context of culture’, respectively and they say these “identify the two main dimensions of variation between texts.” *Register* is seen as lower level (bottom-up) realisations of variation and is constituted by lexical, grammatical and semantic choices. This theory brings together work from both Register and Discourse Analysis as described above and will be called Variation Studies (after Biber 1988) in this thesis as it is applied to the textbooks under study.

Eggins and Martin (1997) explain that genre can be seen in many different ways. There is the conventional literary model of “types of literary productions” including short stories, poems and novels. Then there is the linguistic definition Bakhtin (1986) gives which broadens genre to include everyday speech and writing with the literary genres. Genre in linguistics is also defined functionally in terms of its social purpose. Eggins and Martin (1997:236) summarise this saying “Thus, different genres are different ways of using language to achieve different culturally established tasks, and texts of different genres are texts which are achieving different purposes in the cultures", or what may more simply be described as text and talk in context.

Similarly, the needs of students and course needs have to be studied alongside these analyses and continue to be important for the study of English for Special Purposes and syllabus development. Stubbs (1996:19) criticises the work carried out on scientific research articles by Swales (1990) because he failed to relate the linguistic features he found to a theory of variation in English. Stubbs (ibid.) suggests that any study of genres "must be located in a description of variation in the language overall" and that Biber's work is a good example of how wide a range of variation there is within academic prose.

2.3.14 Discourse Analysis and Computers

Sedelow and Sedelow (1994:160) acknowledge that the work that has been carried out using computers for discourse analysis to date are "purely piecemeal approaches" to "highly restricted discourse domains". They advocate the development of a conceptual thesaurus that is based on associative semantics in order to transcend narrow domains, which they believe is essential to deal with semantic space. They examine the lexical cohesive ties described in Halliday and Hasan's *Cohesion in English* (1976) and find (1994:167) that "cohesion does form a major component of our perception and analysis of discourse". They also bemoan the fact that many computer scientists are "rediscovering, laboriously, many of the relationships already worked out" by Halliday and Hasan.

The difficulty of applying modern technology to discourse analysis is quite simply the difficulty of dealing with whole texts. The computer can handle precise searches and programmes are available for grammatical tagging but the overall organisation of a text is much more difficult to capture on computer. Biber, Conrad and Reppen (1998:106-131) put forward a scheme of corpus-based research into discourse features such as given and new information and the use of discourse maps of verb tense and voice in research articles in experimental science. They conclude (1998:131) that although discourse features cannot be investigated completely automatically, interactive computer programmes and innovative output formats will be exploited in the future in order to show patterns of discourse across texts and registers.

2.4 The Needs Analysis Approach

2.4.1 Needs and ESP

Richard West (1994:1) claims that the term ‘analysis of needs’ dates from the 1920’s in India with Michael West and his consideration of the ‘surrender value’ of learning for secondary level learners. The connotation that Michael West uses here is that of language requirements of the students studying English. After this the concept of ‘need’ does not appear again for about fifty years until ESP research started in 1960. Michael West is more closely associated with register analysis and his (1953) *General Service List of English Words* but here he represents a number of researchers working abroad in universities where they were trying to find solutions to the problem of what and how to teach students who required English for their studies, either as a second or a foreign language. Halliday, McIntosh and Strevens (1964:189) refer to ‘English for Special Needs’ although their use of the term ‘need’ is also in terms of special language or register as discussed above. The term ‘need’ only began to take on its more modern connotation of why and for what purpose students learn a foreign language in the 1970s.

2.4.2 The Development of Needs Analysis

West (1993) gives four stages in the development of needs analysis:

Stage 1 in the early 1970's which focused on English for Occupational Purposes (EOP) was concerned with target situation analysis and is exemplified by Richterich (1971/1980) on English for adults, English Language Teaching Document Unit (1970) on business English, and Stuart and Lee (1972/85) on English for industry and commerce;

Stage 2 later in the 1970's which focused on English for Academic Purposes (EAP) and was also concerned with target situation analysis and could be exemplified in the work of Jordan and Mackay (1973) and Mackay (1978);

Stage 3 in the 1980's ESP and general language teaching which covered a range of analyses, target situation analysis, deficiency analysis, strategy analysis, means analysis and language audits as exemplified by Tarone and Yule (1989), Allwright (1982), Holliday and Cooke (1982), Allwright and Allwright (1977), and Pilbeam (1979);

Stage 4 in the early 1990's with integrated/computer-based analyses and materials selection exemplified by Jones (1991) and Nelson (1993).

West (1993), naturally enough, sees this latter stage with computer-based analyses as the future of needs analysis. The use of technology in both analysing and selecting materials is purported to make the syllabus more appropriate for learners needs.

2.4.3 Needs and Syllabus Design

In 1978 Munby published his book *Communicative Syllabus Design* which discusses the questions that have to be asked (and answered) before designing a course. The size and scope of his achievement in this book have meant that needs analysis has now come to be seen as crucial in any ESP course design. Munby's theoretical bases were contemporary views on the nature of communicative competence, derived principally from Hymes (1971). In his *Communicative Needs Processor* the following parameters are identified as being pertinent to syllabus design:

0.0 Participant
- 0.1 Identity (Age/Sex/Nationality/Residence)
- 0.2 Language ((L1/L2/Present level of L2/Other L2s known)

1.0 Purposive Domain
- 1.1 ESP classification (English for Occupational Purposes (EOP) or English for Academic Purposes (EAP), if EOP, pre- or post-experience, if EAP, discipline based or school subject)
- 1.2 Occupational purpose (specific job or post/central duty/other duties)
- 1.3 Educational purpose (specific discipline/central area of study/academic design classification)

2.0 Setting
- 2.1 Physical setting: spatial (location/country/town/place of work/place of study)
- 2.2 Physical setting: temporal (point of time/duration/frequency)
- 2.3 Psychosocial setting (noisy, demanding, culturally different, aesthetic - unfamiliar)

3.0 Interaction (with others)
- 3.1 Position (role relationships - dependent on purposive domain e.g. student)
- 3.2 Role-set (other interlocutors etc.)
- 3.3 Role-set identity (number/age/sex/nationality of interlocutors thus affecting role relationship)
- 3.4 Social relationships (or role relationships e.g. superior-subordinate, peer-peer, official-member of public, doctor-patient, teacher-learner)

4.0 Instrumentality
- 4.1 Medium (spoken or written)
- 4.2 Mode (monologue/dialogue)
- 4.3 Channel (e.g. face-to-face, text for silent reading, phone)

5.0 Dialect
- 5.1 Regional (and British English/American English, etc.)
- 5.2 Social class
- 5.3 Temporal

6.0 Target Level 6.1 Dimensions (size and complexity of utterance/material (text), range and delicacy of forms and functions, speed and flexibility of communication) 6.2 Conditions (degree of tolerance of: 1. error, 2. repetition, 3. hesitation, 4. stylistic error, 5. reference) **7.0 Communicative Event** (i.e. what the learner has to do, either/and productive and receptive) 7.1 Main (macro activities e.g. waiter serving customer in restaurant, student in university seminar) 7.2 Other (micro activities e.g. taking down customer's order or student introducing a new point) **8.0 Communicative Key** (i.e. how the learner does the activities above determined by 1,2,3 - attitude factor).

Table 2.1 Munby's Communicative Needs Processor

In other words, the communicative needs profile that results from applying these parameters is a very detailed description of communicative needs but with no specification of the language items which will realise these needs. That is to say, communicative needs profiling is at the pre-language stage and is designed for curriculum development. This approach was later broadened to include practicalities and constraints, teaching methods and learning strategies and materials selection, which were areas which were not considered by Munby. Nor can Munby's needs analysis be considered learner-centred, as all the information to be collected is *about* the learners and not *from* the learners themselves.

Although Hutchinson and Waters (1987:12) recognise that Munby's needs analysis was an attempt to put the learner's needs "at the centre of the course design process", they attack Munby's concept of needs as being "far too simple" and propose an improved version of their own. They argue that 'it is necessary to examine the *underlying competence* which the learner must bring to ... the study of any specialised subject' (Hutchinson and Waters, 1980:178). They propose a classification of needs which includes:

(a) *Necessities* which are 'the type of need determined by the demands of the target situation, that is, what the learner has to know in order to function effectively in the target situation' (Hutchinson and Waters, 1987:55). Identifying these necessities is often referred to as *target-situation analysis* (see Chambers, 1980).

(b) *Lacks*. Analysis of what the learner already knows leads to recognition of the gap which exists between this and the target situation in other words the 'learner's lacks' (Hutchinson and Waters, 1987:55-56).

(c) *Wants*. These wants are the learners' perceived needs or *subjective needs*. (Hutchinson and Waters, 1987:57). The learners' needs (subjective needs) may be in conflict with the needs analysis that has been carried out and therefore may be in conflict with the aims of the course, as determined by those responsible for the course. However, in such a situation it may be possible to incorporate some of the generally perceived (subjective) needs of the learners into the course. An example of this would be to incorporate speaking tasks into courses which are predominantly designed to aid reading.

(d) *Learning Strategies*. Hutchinson and Waters (1987:60-2). The identification of the learning strategies that the learner prefers to use in order to deal with the target situation may also come into conflict with the teacher's identification of suitable strategies.

(e) *Constraints*. Hutchinson and Waters consider the external factors which may condition the learning situation such as the resources available, which would include the length of time the course will run as well as the materials, aids and methods available. These constraints are also referred to as *means analysis* (see Holliday & Cooke, 1982; Holliday, 1984).

Hutchinson and Waters criteria are highly relevant for the research carried out here on undergraduate students with (a) the *necessities*, what the students need to know, being found from the corpora analyses, and (b) the *lacks*, what the students already know, being identified from the results of the language tests carried out on the undergraduates joining the first year science and technology courses. Incorporating students *wants* would be a more difficult task given the numbers involved and certain *constraints* such as the requirements to test all of the students in the same way at the same time.

Some linguists felt that there was a need to examine language much less intuitively in order to produce a more objective description. In many cases the corpora that were built up were designed primarily for other reasons but soon it became obvious how useful corpora could be in a number of areas of research into actual language use such as differences between different text-types and genres, which as I said is now called variation studies. This led to corpus analysis research as a means of describing language through examination of actual examples of use instead of invented examples. This requirement of only employing actual language is also considered crucial in modern discourse analysis.

2.5 The Corpus Analysis Approach

Modern computers have removed many of the barriers to comprehensive corpus research and have enabled significant and far-reaching specific studies to take place. These studies are not restricted to English, although many of the early corpora developed were based on both American and British English (see later the Brown and LOB corpora). There is for example the French *Trésor de la Langue Française* which is largely historical and lexicographic in conception.

Sinclair (1991:1) suggests that traditional linguistics had been limited by the amount that one person could experience and remember and he equates the situation with that of the physical sciences 250 years before. Halliday (1993:7) sees the start of corpus-based linguistics as laying the foundations for a quantitative and qualitative breakthrough in understanding linguistic systems and of this having started in the 1960's with Randolph Quirk in Britain and Freeman Twaddell in the United States.

2.5.1 English Corpora Development

The study of computer corpus data in English was started in order to find out 'the true facts of English grammar' because it is more accurate than old manual methods and also more likely to take into account all the unpredictable features which occur naturally in speech and writing. Sinclair and Francis (1994:191) suggest that:

> "Corpus data provides us with incontrovertible evidence about how people use language. ... lists are both a continuation of a tradition, and an innovation. ... the tradition was to observe what people did and record it in a reference book. In recent years, the language teaching trade has lost the ability to appreciate the prime importance of this tradition, and has relied rather too heavily on intuition."

Originally, the idea of using computer corpora in this way was derided as being a waste of time and money because it was thought that traditional methods could provide suitable examples of English grammar drawn from native speakers' insights into how the language operates. Biber, Conrad and Rippon (1994:169) point out, however, that "...corpus-based analyses frequently show that earlier conclusions based on intuitions are inadequate or incorrect.". It should also be pointed out that intuitive decisions are even more questionable in scientific and technological language, as no-one is a native speaker of 'scientific language'. Meijs (1992:146) says that while the great traditional grammarians such as Poutsma, Kruisinga and Jesperson were ardent data-gatherers, their work could be criticised on the grounds that they had a bias for the unusual and irregular which made them overlook the larger generalities in their search for "interesting" phenomena. Stubbs (1996) compares Chomskyan principles with neo-Firthian principles. He (1996:24) argues that the Chomskyan view that "linguistics is a branch of cognitive psychology", that "it can be based on intuitive data and isolated sentences, that corpus data are unrevealing", and that "the study of language in use is essentially uninteresting" is in direct contrast with neo-Firthian views as represented by Sinclair (1991:4) who argues that corpus data provide "a quality of evidence that has not been available before" and only considers actual, authentic, attested data to be of interest. Meijs (ibid.) claims that computer corpus studies can develop "a more balanced view of the 'spread' of linguistic phenomena - lexical as well as syntactic ones".[11]

The use of computer-based text corpora has become increasingly important for research into natural language processing, lexicography and descriptive linguistics, issues relating to corpus design have also assumed central importance. Biber (1994:179) says that there are two main considerations "1) the size of the corpus (including length and number of text samples), and 2) the range of categories (or *registers)* that samples are selected from." Biber explains that corpus designs can differ in whether they are *bounded* or *unbounded* and therefore static versus dynamic; whether they are *richly encoded* or *minimally encoded* (that is, whether there is grammatical tagging, phonological or prosodic encoding

[11] Meij's 'spread' could be equated with the computer corpus concept of 'range' of an item in frequency studies.

or some kind of social characteristic tagging of the participants and situational tagging); whether it contains *complete texts* or *samples from texts*; and the selection of texts may be either made by convenience versus purposeful versus random within strata versus proportional random.

2.5.2 Corpora Use

Much of the work that has been undertaken is on general language use and has attempted to find generalisations about normal usage rather than the specific language associated with English for Special Purposes. Leech (1993:10) identifies the use of computer corpora for the following, a list which he claims is "by no means complete":

- *linguistic theory:* improving models of language generally
- *computational linguistics:* natural language processing by computer; machine translation; etc.
- *grammar:* syntax, morphology, automatic parsing
- *dictionaries:* lexicography, lexicology, word-formation
- *the study of meaning:* semantics and pragmatics
- *discourse analysis and conversational analysis*
- *language variation:* spoken and written language; language and gender; general studies of style; dialect
- *speech technology:* automatic speech synthesis and speech recognition
- *speech science:* phonology, phonetics, stress, intonation, etc.
- *historical studies of language*
- *child language acquisition, psycholinguistics*
- *applied linguistics:* language learning, language testing, etc.
- *orthography:* punctuation, spelling

Leech does not, for example, consider translation studies which is an enormous area of its own with very specific views on language although this could be subsumed under computational linguistics in general and is perhaps hinted at in machine translation. Nor does he make special mention of collocations, an area of language learning and teaching that has become extremely important in recent years (Sinclair 1991, Tribble and Jones 1990). Collocations are now seen to be the building blocks of language and can be used for vocabulary management, to disambiguate similar terms and formulate or check hypotheses about language use, to help learners to understand texts, for self-access outside the classroom and to provide teachers with suitable teaching materials. However, this list does show some of the areas within which computer corpora can be applied to the study of the English of science and technology in applied linguistic research and language teaching.

Biber (1994:180) suggests that recent debate centres around whether to use large corpora as opposed to what is known as “balanced” corpora (that is, made up of a number of similar sized texts possibly from a wide range of registers) for the design of general purpose corpora. He argues (1994:180) that “it is important to address the question of whether the varieties represented match the intended uses of a corpus” and he claims that studies of a single sub-language are “legitimately based on corpora representing only that variety”. This is the view that this study takes about corpora; they must represent the variety of English that the students are expected to come into contact with and need to understand for their studies in science and technology.

Biber (1994:11) lists the following areas of study in linguistics that corpora can help with: “individual words, grammatical features, men’s and women’s language, children’s acquisition of language, author style, register patterns” and goes on to suggest that dialect and register patterns could be investigated for sociolinguistic fields when looking at the "complex co-occurrence patterns among features in different registers" which would be difficult to do without recourse to computers on a large scale. He also (1994:12) mentions the study of styles across historical periods which could provide the opportunity of investigating the development of registers over time and emphasises the role of corpora in educational linguistics. With respect to the latter he says that "large-scale studies of use are helpful in designing effective materials and activities for classroom and work-place training, allowing us to help students with the language that is actually used in different target settings." He also recommends corpora use in language testing, that is, "making tests which conform to the actual language that students will be using on a regular basis". These conclusions form the basis of the working presuppositions of this study.

The preliminary tests that were designed for the undergraduates, described in Chapter 4 went some way towards this goal of conforming to what was seen as the target language that the students would be coming into contact with in science and technology. The tested items were from the materials that were to be taught in the discipline. They were not however, derived from corpora developed for the purpose, which, in the light of modern computer corpus methods is a weakness of the testing. The reason for this was that the testing had already got underway before the corpora used in this study had been developed but now that they are available there is no reason why they should not be used for this purpose in the future.

2.5.3 The Birmingham Corpus COBUILD

The corpus work of Sinclair, started in 1960 in Birmingham, was primarily concerned with lexicography and the production of the COBUILD Dictionary. This dictionary was published in 1987 and was based on what was then considered a huge corpus of 7.3 million words of written and a smaller corpus of about 1 million words of spoken language. The 'main' corpus was started in 1960 and subsequently smaller 'side' corpora were developed (notably the Bank of English and a corpus especially prepared for Teaching English as a Foreign Language (TEFL) textbook writing (see Willis, 1989). Sinclair and Jones (1974) report that "The first corpus, in 1961, was a mere 135,000 words". This reflects the changes that have taken place with regard to the gathering of data. Initially every text had to be transcribed onto computer manually and the original computer programs for handling the texts had to be developed. Later text which had already been transcribed on computer through word processing became available and later still the use of optical scanners (usually known as Optical Character Recognition or OCRs) simplified the transcription of text and speeded up its conversion into electronic data.

Sinclair (1987:2) describes the criteria on which the 'main' 7.3 million word Cobuild corpus was developed to be relevant "for the needs of an international user" and which the team defined as the following:

- written and spoken modes
- broadly general, rather than technical, language
- current usage, from 1960, and preferably very recent
- "naturally occurring" text, not drama
- prose, including fiction and excluding poetry
- adult language, 16 years or over
- 'standard' English, no regional dialects
- predominantly British English, with some American and other varieties.

Sinclair (1987:3) also describes the balance that was given to the components but gives no clear definition of why this should be so, he merely suggests that these were chosen for "different reasons":

book authorship	- 75% male: 25% female
English language variety	- 70% British: 20% American: 5% Other
language mode	- 75% writing: 25% speech

The development of criteria for the Birmingham corpus was grounded on clear principles as those involved describe (cf. Hoey 1996) but some areas still need explaining like those given above. It is questionable whether this balance would be the only suitable choice for “the needs of an international user” as described by Sinclair.

The spoken corpus was initially a problem because of the difficulty of obtaining permission to use tape recorded conversations where the participants did not know that they were being recorded. The university prohibited such activities which led to structured dialogues and BBC programme material being used. The latter material can be criticised as it should be considered scripted or ‘prepared’ and, therefore, the spoken corpus could be seen as not reflecting ‘natural’ speech or conversation, thereby destroying the main premise of collecting data of actual (real-time) occurrences of language and making it prone to be considered (more) examples of written English read aloud.

These early problems have largely been overcome and the current corpus, known as the *Bank of English*, now runs to hundreds of millions of words but is constantly growing. The working corpus recently (July, 1998) had 329 million words of modern English text. The written texts include fiction and non-fiction books, newspapers, guides, magazines, brochures, letters and leaflets. The 20 million words of transcribed natural speech in the corpus include everyday casual conversation, radio broadcasts, meetings, interviews and discussions. Most of the texts originate from after 1990 and are designed, according to the Collins COBUILD website (1998:1), "to provide objective evidence about the English that most people read, write, speak and hear every day of their lives".

2.5.4 The Lancaster-Oslo-Bergen Corpus

The LOB (Lancaster-Oslo-Bergen) Corpus of British English has an even sample size of around 2,000 words all taken from printed sources published in 1961 and totalling a million words (see Johansson *et al.* 1978, Johansson 1982 and Hofland and Johansson 1986), compiled in the 1970's. There are 500 text samples taken from fifteen genres: press reportage, editorials, press reviews, religion, popular lore, skills and hobbies, biographies and essays, official documents, learned writings, fiction (including general, mystery, adventure, science, and romance), and humour. The total corpus size is approximately one million words of running text. The divisions of the corpus into genres is conducted largely on intuitive criteria. Sinclair (1991:19) criticises this because "a corpus which does not reflect the size and shape of the document from which it is drawn is in danger of being seen as a collection of fragments where only small-scale patterns are accessible." The breakdown into the respective categories is as follows:

Category	**Number of Texts**	**Approx. number of words**
Press reportage	44	88,000
Editorials	27	54,000
Press reviews	17	34,000
Religion	17	34,000
Skills and Hobbies	36	72,000
Popular lore	48	96,000
Biographies and essays	75	150,000
Official documents	30	60,000

Academic prose	80	160,000
General fiction	29	50,000
Mystery fiction	24	48,000
Science fiction	6	12,000
Adventure fiction	29	58,000
Romantic fiction	29	58,000
Humor	9	18,000
TOTAL	500	1,000,000

Table 2.2 Texts, Categories and Numbers of Words in the LOB Corpus

The LOB Corpus is tagged and part of the LOB known as the Lancaster Parsed Corpus contains 133,000 words that have been syntactically analysed.

There is also now the Freiburg corpus with approximately 1 million words of British English, parallel to the LOB corpus, but compiled from material published in 1991. The fact that corpora are seen to be becoming dated means that their authority to describe modern English usage is also diminished and so many more of this type of up-to-date corpora are being prepared to keep abreast of changes that are constantly taking place in language usage. These more modern corpora, when produced using similar criteria, can be used for diachronic and other comparative studies. The other reason that more up-to-date corpora are being produced is that the techniques now available and the research carried out on machine readable or electronic texts has brought some of the original criteria into question. The insights gained from such research now implies that more modern corpora can be obtained in many more different states of tagging depending on the purpose to which they are to be put. Biber’s research which forms the basis of this study drew on some of the earlier LOB texts.

2.5.5 The Brown Corpus

The Brown University corpus of written American English (see Francis and Kucera 1982) is one of the oldest of the large scale corpora. It was started in 1961. It consists of short extracts of many genera, for research purposes. The LOB corpus is a replica of the Brown (1964/1979) corpus so that parallel text samples can be compared between British and American English. This corpus is tagged. There is also now a fully tagged subset of the Brown corpus known as the SUSANNE Corpus, which contains approximately 128,000 fully tagged words in 64 texts each about 2,000 words long from four genres of the Brown Corpus; A: press reportage, G: belle lettres, biography, memoirs, J: learned (mainly scientific and technical writing), N: adventure and Western fiction.

Leitner (1993) criticises the LOB and the Brown corpora because textbooks are unrepresented in them and he contends (1993:81) that “textbooks are a major medium for communication”. Goethals, Engels and Leenders (1990:237) find that what they describe as “the journalese style that dominates the Brown and the LOB corpora” causes distortion in their work on the Leuven English Teaching vocabulary list. This bias can be confirmed by an examination of the texts included in some of the sections other than those specifically designated as press reportage, editorial or reviews. For example, the skills and hobbies section contains many texts taken from magazines such as ‘High Fidelity’, ‘Dog World’ and ‘Hot Rod’. Similarly, the Popular Lore section contains articles from many magazines including ‘Vogue’, ‘Family Circle Magazine’ and ‘National Geographic’. Even Belles Lettres contains texts from ‘The Saturday Evening Post’, ‘The New York Times Magazine’ and many different, regional ‘Quarterlies’ and ‘Reviews’. The source of texts has obvious implications for style and discourse features as Goethals, Engels and

Leenders have found. For the purposes of the research described here these failings make the use of such corpora inappropriate. Neither the absence of textbooks nor the presence of an overwhelming amount of journalese is suitable for analysis of the language that undergraduates of science and technology need to confront and is therefore not suitable for the purposes of this study.

Furthermore, Minugh (1997:68), despite recognising that the Brown and LOB corpora were “a revolution in their time”, describes the difficulty of using such corpora for searches for neologisms because of their date of development. Minugh (1997) recommends the use of British and American Newspaper CD-ROMs for this sort of search. In other words, these corpora are also becoming dated and are therefore not suitable for finding representatives of colloquial or modern language terminology or coinings. This limitation is particularly relevant for those conducting research into speech and current news because of the ability to change quickly and reflect fads and fashions. Some of those changes will become part of the language but others will disappear almost as quickly as they came. This is the heart of the problem that dictionaries such as the Oxford have every time a new edition is published. Terms which are regarded as fashionable or corruptions are often decried by readers and reviewers as having no place in such an established reference work on the English language.

2.5.6 The London-Lund Corpus

The London-Lund Corpus of Spoken English (Svartvik and Quirk (1980), Johansson (1982) is a collection of 87 spoken British English texts[12] of about 5,000 words each. The total corpus contains approximately 500,000 words of different genera. It is divided into half

[12] Text here means a communicative event

spoken and half written material. Six major speech situations are represented: private conversations, public conversations (including interviews and panel discussions), telephone conversations, radio broadcasts, spontaneous speeches, and prepared speeches divided up in the following way:

Category	**Number of Texts**	**Approx. number of words**
Face-to-face conversations or discussions	65	235,000
Telephone Conversation	110	60,000
Public conversations, discussions, interviews	20	85,000
Spontaneous commentary (radio broadcasts)	20	55,000
Spontaneous oration	12	30,000
Prepared oration	12	35,000
TOTAL	239	500,000

Table 2.3 Texts, Categories and Numbers of Words in the London-Lund Corpus

This corpus started as Randolph Quirk's (1980) *Survey of English Usage* in a traditional paper filing system but was converted into electronic form by Jan Svartvik (1980). The corpus is tagged, that is to say, the texts are annotated grammatically. The practice of tagging is criticised by Sinclair (1991) as, in his opinion, this makes a corpus less useful. However, Biber, Conrad and Reppen (1998:31) argue that tagging makes a corpus more useful for particular kinds of searches. They give the example of automatic frequency counts for each separate grammatical word of "deal" in a corpus, that is "deal" as a singular noun, proper name, verb, plural noun, present participle, past tense verb and past participle which they explain would help to show which of these is found in which registers and, therefore, the ways in which words are used (and the different meanings they show) in different registers. However, they do accept (1998:59) the need for hand-editing in order to ensure that some searches have not included items incorrectly (in this case words ending in *-ion* or *-ity* which are not in fact nominalizations, like *nation* and *city*). They also (1998:67) discuss the difficulties of comparing analyses when different criteria may have been used to describe nouns and verbs. The different studies may, for example, have included pronouns in the nouns category or not and auxiliary verbs may or may not have been included in the verb count. Similarly with automatic grammatical tagging, Biber *et al* (1998:73) recommend using interactive techniques and balancing the results found in order to correct errors made by the tagger even though this process is extremely time-consuming. The problems discussed above are particularly prevalent with automatic tagging of speech corpora because of the interjections that occur.

There is obviously a need here to have some kind of international agreement on tagging procedures or at the very least to insist that research carried out is accompanied by very clear descriptions of what was included in each of the categories that were used. The fact that English uses a number of different schemes to describe grammar, depending on the purpose or audience the work is for, means that there is ample space for conflicting criteria to be used. Sinclair's idea of keeping corpora as simple and 'pure' as possible until these areas of grammatical tagging have been clarified is probably the safest and is the position that will be adopted with the corpora used in this study.

2.5.7 The British National Corpus

The British National Corpus (BNC) was started in 1991 at Oxford University and is a 90 million word collection of samples of written and 10 million words of spoken language from a wide range of sources, designed to represent a wide cross-section of current British English both spoken and written. Leech (1993:13) gives the following information on the composition of this corpus:

Written texts (90 million words)

Selection features

- Informative/imaginative writing
- Subject field
- Date
- Genre
- Level

Spoken texts (10 million words)

- demographic sampling (50%)
- sampling by discourse type (50%)

Written Component: Informative

Primary subject fields	*Level*
Natural and pure science (5%)	Specialist (30%)
Social and community (15%)	Lay (50%)
Commerce and Finance (10%)	Popular (20%)
Belief and thought (5%)	
Applied science (5%)	*Date*
World Affairs (15%)	1975-present
Arts (10%)	
Leisure (10%)	

Genre

Books (55-65%)
Periodicals (20-30%)
Miscellaneous (published) (5-10%)
Miscellaneous (unpublished) (5-10%)
To be spoken (2-7%)

Written Component: Imaginative (20-30%)

Level	*Date*
Literary (33%)	1960-1974 (25%)
Middle (33%)	1975-present (75%)
Popular (33%)	

Table 2.4 Composition of the British National Corpus

The spoken, face-to-face conversation corpus is as follows:

Category	**Number of Texts**	**Approx. number of words**
Face-to-face conversation	160	4,000,000

Table 2.5 Conversation in the British National Corpus

There is information for research purposes on the gender, occupational group, social background and age of the speakers (informants) taking part in the collection and it has been used as the basis for the *Longman Dictionary of Contemporary English*. These corpora can be consulted in order to obtain other data such as the differences in spoken language use between gender and age groups, which has entailed a much more rigorous description of the material held in the corpora to allow this level of specificity. Leech (1993:10) points out that corpora of the size and nature of the BNC 'are often too expensive in time and effort to be built without commercial or industrial help' and describes how this corpus was funded by the British government (hence it is British English), three major British publishers (Oxford University Press, Longman and W&R Chambers) together with the British Library and the Universities of Oxford and Lancaster.

2.5.8 The Longman/Lancaster Corpus

There is also the Longman/Lancaster Corpus containing American, British and other varieties of written English. It consists of 30 million words which is described as 'representative', by which is meant that 'a full range' of variation of the language is included. What a "full range" of language is provokes some debate. This question has been answered by the British National Corpus developers as "the English which most people read, write, speak and hear every day of their lives". A new category has emerged because of this, that of “ephemera” which includes any material that people come into contact with unintentionally, such as unsolicited mail and advertising.

The number of texts and number of words contained in the categories Academic Prose and Fiction are as follows:

Category	**Number of Texts**	**Approx. number of words**
Academic Prose	98	2,700,000
Fiction	144	3,000,000
TOTAL	242	5,700,000

Table 2.6 Number of words and texts for Academic Prose and Fiction in the Longman/Lancaster Corpus

The samples are taken from many registers from the early 1900s to the 1980s. It can be seen that this corpus also has limitations for a description of either general English or for analysis of specific varieties of English usage. It suffers from a lack of balance to provide what is described as ‘general English’ by the *Bank of English* criteria mentioned above (section 2.5.3). It also suffers from having ‘text fragments’ which Sinclair (1991) regards as a failing of many corpora. It also covers too wide a period of time for much research on either modern usage or for diachronic study purposes.

2.5.9 Other Corpora

Many projects on specific issues that researchers feel are not or are underrepresented in the established large scale projects described in more detail above are taking place in universities around the world. A short summary of some of the main areas that these cover is given below to demonstrate the trends in recent corpora studies.

An Australian corpus (ACE) produced at Macquarie University, New South Wales and an International Corpus of English (ICE) produced at University College London (http://www.ucl.ac.uk/~ucleseu/design.html) have also recently been developed to address other types of Englishes in the world. ACE contains one million words of Australian English compiled along the same lines as the Brown Corpus for purposes of comparison. ICE contains one million words from the English of Australia, Canada, East Africa, Hong Kong, India, New Zealand, Jamaica, Nigeria, Singapore and the Philippines. The Melbourne-Surrey Corpus has 100,000 words from Australian newspapers.

The Kohlapur corpus contains 1 million words of written Indian English from 1987. It uses the same categories as the Brown Corpus and LOB Corpus.

A corpus of spoken American English (CSAE) is being constructed at the University of California which eventually hopes to contain one million words.

The Northern Ireland Transcribed Corpus has about 400,000 words of spoken material from 42 locations and over three age groups.

The CHILDES Project (http://poppy.psy.cmu.edu/childes/database.html) is developing a corpus of children's spoken and written language. There is also the Polytechnic of Wales (POW) corpus of 61,000 words of children's spoken language which has been parsed using Hallidayian Systemic-Functional Grammar.

The increase in the number of corpora and such corpora as those on language development will surely have an influence on teaching and learning as they show what actually takes place rather than what some small scale studies have suggested is the case in both language acquisition and language diversity.

2.5.10 EFL Student Corpora

In addition to these, there are even specific corpora for English language learners such as the International Corpus of Learner English (ICLE) and the Longman Learners' Corpus drawn from EFL learners around the world and associated with the International Corpus of English which was directed by the late Sidney Greenbaum at University College London. This ICLE corpus is expected to reactivate certain research areas such as error analysis and could very probably lead to analyses of the relative frequency of certain features in learner English as compared with those of native speakers.

The Hong Kong University of Science and Technology also has a Learner Corpus with approximately 6 million words of written undergraduate assignments and "A" level Use of English scripts from the Hong Kong Examination Authority. This corpus is still growing.

These corpora are beginning to increase our understanding of learner English so that insight can be gained into learning stages and strategies, perhaps with comparative studies on language acquisition. Biber, Conrad and Reppen (1998:172-202) describe some studies that have been carried out but recognise the fact that computer corpus-based studies have not hitherto normally been used in these areas.

2.5.11 Specialised Corpora

Although clearly of less interest to this study, more specialised corpora such as, the Helsinki Diachronic Corpus of English (Kytö 1991) and the ARCHER corpus (Biber *et al.* 1994) have been designed to cover the development of English registers over several historical periods. The Helsinki Corpus covers the periods from Old English to Early Modern English, while the ARCHER corpus includes texts from 1650 to the present. These variation studies across time as opposed to across genres are developments which seem to be harking back to some of the other traditional (now computer assisted) studies of variation in old manuscripts.

2.5.12 Concordances

Many programmes have now become available for small personal computers that allow the teacher and students to conduct research into their own corpora either to produce accurate materials or for reference purposes. Concordances, programmes that display the immediate environment of particular lexical or grammatical items, are available that will accurately reflect real English use and which can be used by both students and teachers. Moreover, these programmes can be customised to reflect the different genres to be studied by the simple means of choosing specific text types as the basis for the concordance. Teachers and students can then consult such programmes to decide what is the correct or natural choice of, for example, a preposition after a particular word, such as *different from*, by examining all the instances of *different*. It is in this specific area that recent work has been carried out by Jones (1991) and Nelson (1993) as mentioned earlier (see section 2.4.2). Published CD-ROM's on collocations, such as by Collins, are also beginning to appear on the market. These products, together with specific material on aspects of English grammar like phrasal verbs and reported speech, have been developed from research on corpus material. They contain examples taken from the corpus and not invented ones. Many of these come from the Birmingham corpus, which can be consulted by both teachers and students. As with any information, the material extracted is only as good as the corpus on which it is based. Some of the examples, whilst claiming to represent general English usage, can be seen to be from novels from earlier periods like those of Jane Austen in the early nineteenth century. Such examples can be regarded as dated and often ‘unusual’ rather than reflections of modern-day English usage. The biggest criticism of many of the first corpora produced is precisely this, that they have already become dated and cannot be seen to be representative of modern English usage any longer. They are already caught in the trap of ‘historical’ rather than ‘current’ usage.

2.5.13 Undergraduate Textbook Corpora

Corpora built up from undergraduate textbooks in chemistry and physics produced using an OCR are still a prerequisite for specific study of how these texts differ from or conform to other discourse types, despite the wealth of material described above, for the simple reason that these textbooks represent a much more specific genre than any of those available either for research purposes or commercially, especially as these textbooks for undergraduates are often in American English. As was mentioned earlier (2.5.6) the American Brown corpus was criticised by Leitner (1993) for not containing *any* textbook material. In this case, commercially available CD-ROM texts can serve a useful purpose for comparison and contrast with more general instructional texts in order to highlight the differences between the general and the specific nature of the texts of a similar level under study here. Stubbs (1996:5) argues that there is a “need to analyse not only short text fragments, but also whole long texts; and the need for the stylistic analysis of individual texts to be based on comparisons with other texts and with corpus data which represent (however imperfectly) the language.”. Thus, I have incorporated it into my approach.

This dissertation will take up three major lines of research from the register analysis, discourse analysis and corpus analysis mentioned above which are essential to syllabus development.

First a register analysis will be carried out on the physics and chemistry books from the students bibliographies, as register analysis can be applied to syllabus design following Jones' orientation (1991), using frequency counts to identify what is lacking in any syllabus or materials for specific learners. Consideration of cognates will be made in order to fine tune these lists even further and to anticipate areas of difficulty for Portuguese native speaker students.

This will be compared with a CD-ROM multimedia encyclopaedia in order to bring out similarities and differences between texts that are of the same academic level, according to Huddlestone (1971) and Swales (1985), and which will serve to reflect the moves in education towards the use of this kind of technological resource for both student and teacher-directed learning (see Guillot and Kenning 1995:365). Work with interactive and multimedia resources such as those available on CD-ROM and through the Internet are seen as being increasingly important in education as discussed in the introduction to this study. This comparison will also provide information on the range of the items, so that the relevant context (and, therefore, specific use and meaning) of the lexis can also be determined.

The corpora from the physics and chemistry textbooks will be explored using Biber's (1988) methodology for variation studies, in order to highlight the ways in which these conform to and differ from both academic prose and general language use he classifies it. Biber's methodology is explicitly defined so that it is possible to build this study on his work in an accurate and principled manner. This is deemed to be a prerequisite of any research in corpora studies so that a precise description of the criteria used to produce the data is available which can thus be evaluated in the light of the purpose to which it is to be put.

The students' language needs will also be ascertained by using the results obtained from tests on entering the university to determine the strengths and weaknesses which will need to be addressed by any syllabus designed for these students. The results of the tests are classified into grammatical categories that correspond to those used on the corpora as far as possible and test items are exemplified to provide clearer description especially for those tests that took place before the advent of this study. By comparing and contrasting these categories it is possible to reach some conclusions about the areas that need specific attention in the syllabus designed for these students.

Finally the results of what can be seen as a data-driven description will be brought together to suggest what should to be included in any syllabus adopted for these students. At this point the research findings from other corpus-based studies have to be taken into account in both grading of material through core or key patterns of usage. The commonest forms of language use and the combinations these typically form or core patterns in these textbooks must be matched with the abilities shown by the undergraduates with them on the tests they have taken. Nevertheless, syllabus design calls into question many other methodological aspects which must be addressed. What can feasibly be achieved, despite the inherent constraints, is one of the major considerations here. It is essential that innovation of the kind mentioned above in terms of modern technology is incorporated into the syllabus and more of the same kind of teaching/learning is not carried out for these very specific students with specific goals and requirements. The wider educational implications of innovation in the syllabus proposed will be addressed.

Chapter 3 Research Methodology

Chapter 3

Research Methodology

As described in Chapter Two, much work has been carried out this century and in particular in the last thirty years to try to define exactly what makes different styles of English different. Lexis, syntax, pragmatics and discourse features have all been studied in order to discover differences and many claims have been made, some on rather slim evidence. For example Tarone *et al*'s study (1981), although very professional, was based on only two Astrophysics articles which were eight pages and seven pages long respectively (see section 2.2.3 for further discussion of this point). Swales (1985) attributes this state of affairs as existing because ESP 'practioners' that is, teachers who also produce materials , design courses and conduct research, are usually working in isolation and do not often look back to the work that has gone before, nor do they learn from work that is being conducted in parallel to their own and which might usefully contribute to their work. One thing has become increasingly apparent and that is that each learning context needs to be studied in order to provide an accurate picture if the results of such research work are to have practical applications. Nevertheless, this specific work must be related to other work in the field.

Halliday (1993:124) says:

> "There are practical reasons for analyzing scientific texts. The most obvious is educational: Students of all ages may find them hard to read, and we know from various research reports that, in English at least, the difficulty is largely

a linguistic one. So if we want to do something about it we need to understand how the language of these texts is organized."

It is especially important that the materials from which the research data is taken reflect the target material for the students who are to be taught or else inappropriate conclusions will be drawn. Many studies are weakened by the fact that the materials used for the research do not reflect only and uniquely that which is the target material of the learner. This is very often because of the difficulty in obtaining sufficient material of the right type. The dangers are that material can easily slip either into a much more general English category or into much more specific specialised English, like that of the post-graduate specialist, when the learning context is for undergraduate students. In this study the first year undergraduates are the focus of attention. These students have much more of an end of secondary education profile, which is a difference of five-years from the post graduate in both science subject specificity and maturity.

At issue here are the undergraduate textbooks contained in the students' bibliographies which will be analysed in order to help the students to be more successful in their immediate studies and to prepare them for their futures. Rosenthal (1996:31) suggests that the English language that causes students (in American higher education) problems with their science studies is not the vocabulary of science but rather in comprehending typical college science textbooks whose readability levels are beyond those attained by English as a Second Language students. Many of the textbooks recommended for further reading for undergraduates in universities here in Portugal are either in English or in translation. The translated books are all too often in Brazilian Portuguese because Portugal is considered too small a market for educational materials of this level, which suggests unfortunately that this situation is likely to continue for the foreseeable future.

Researchers often recognise that different styles of English are more prevalent at different academic levels. The difference between an undergraduate and a post-graduate science student, for example, would suggest widely different text types (from textbooks teaching the subject matter of the course to journals reporting the latest research in very specific branches of science) and, therefore, styles of English. Similarly, some English for Science and Technology (EST) practitioners believe in adopting much more popular and accessible texts which would also bring with them a considerable difference in style and content than the average science textbook. One example of this difference in style is shown by research carried out by Darian (1981) into the manner in which definitions are handled in such magazines as *Popular Science* or *Time* magazine and the *Journal of Astrophysics*. Darian finds, not surprisingly, that definitions are handled differently in popular magazines from those used in specialist journals, and that these are different again from those used in textbooks used to teach the subject. Although it can be argued that the students may be more motivated by certain types of (more popular) materials, these are not considered to be a suitable basis for an analysis of syllabus design for tertiary level students. The assumption that will be made in this analysis is that if students are taught to cope with the kinds of scientific texts that appear in their undergraduate bibliographies, they will be better able to cope later on, whether it be with the literature of their specialisation (where incidentally lexical density has not been found to be a barrier to the specialist) or in other professional outcomes of their courses[1].

[1] such as teaching cf. Arroteia, Jorge Carvalho; Martins, António Maria (1997) *Inserção Profissional do Diplomados pela Universidade de Aveiro: Trajectórias Academicas e Profissionais*, Aveiro: Universidade de Aveiro.

3.1 Frequency and Range List

In order to develop a syllabus for students it is necessary to select from the vast number of words and structures that exist (and which no-one is capable of learning in their entirety) those which will be most suitable for each particular student or class of students. This is as true at the more advanced level as at the elementary stages of learning. One method of doing this is by means of frequency counts, as described before in register analysis (Pp. 50). Although it is claimed that the most frequent words are few in number and the 1,000 most frequently used words make up about 95 per cent of the total number of words in any randomly chosen corpus of language, frequency counts can show which word or structure to choose over another and can throw up evidence of omission of items that should be included in teaching materials. Frequency counts do not, however, mean that certain words or structures can automatically be excluded from teaching materials. It would be ludicrous to argue that certain language that is appropriate for teaching purposes should be excluded on the grounds that it did not appear in a frequency count. For example it may be found that certain days of the week or months of the year are not found or are found infrequently in the corpora. This does not imply that only certain days of the week and months of the year should be taught to students. This is often referred to as *opportunism*, by which is meant that some things are available in the immediate situation or are felt by the teacher to be useful to the students. On the other hand, without empirical studies it may not be immediately obvious which structures and items in textbooks designed for undergraduate students of science and technology are to be preferred.

The frequency counts described below go further than just providing evidence for the language to be included in teaching materials, they also show how widely used a word is across texts (its range), increasing its usefulness for teaching purposes and further specifying appropriate use. Contrasting the three corpora will also give information about *coverage*, that is, the number of things that can be expressed by any given item. Coverage and range together will provide clearer evidence for which items to include in the syllabus. Furthermore, examination of the frequency lists allows prediction of areas of difficulty for students whose first language is Portuguese.

3.1.1 Contrastive Analysis

Lado (1957) puts forward contrastive analysis as a form of language description across two languages which is particularly applicable to syllabus design and to the evaluation of the items which would lead to difficulties for students of particular languages. It was not meant to be a new method of teaching but to aid curriculum development and diagnose learning problems so that suitable materials could be prepared for teaching. Lado was influential because he set out procedures that could be applied for the comparison between languages. The detailed work remained to be done however.

Contrastive analysis was not found to be the universal panacea that it was hoped to be in language teaching, as linguistic theories of transformational generative grammar came into being in the late 1960s and ousted structuralism as the dominant theory underpinning teaching practices. Error analysis has also shown that the whole subject is considerably more complex than was first supposed. However, contrastive analysis has been reappraised (initially by Di Pietro 1968, 1971 and with Danesi in 1990 who brought contrastive analysis in line with transformational generative grammar but even more recently through reappraisal of interlanguage by Selinker 1991) and contrastive techniques can be applied to lexical corpora to highlight and predict areas that, because of similar word formation or shared Latin roots, could be easier (cognates) or, because of different roots, more difficult for Portuguese undergraduates, or examples of *false friends*. For example the word “abnormal” and its Portuguese equivalent *anormal* are sufficiently close to suggest that positive transfer could take place and the students’ ‘guess’ or semantic prediction would probably be accurate. However, words like “able” and its Portuguese equivalent *capaz* are considered to be difficult, although an alternative *hábil* could be used in some circumstances and would be closer to the English form. The latter would be more accessible provided that the students recognised the similar pronunciation of the two words rather than their orthographic form. This contrastive analysis aims to predict the learnability (Mackey 1965) of the language found in the corpora and is incorporated into this study through examination of the frequency lists from the corpora described in more detail in Chapter 5 in order to identify cognates and thereby identify the areas of difficulty for students.

More recently internationalisms have appeared where the same (or very similar word) is being used in many countries. An example of this might be computer jargon like *software/hardware* which are used in many languages and have caused the European Union to fund projects to produce *terminology banks* in various areas including that of information technology.

3.1.2 Context

The next step in the development of teaching materials, once the identifying and contrasting stages have been completed, is to provide a suitable and typical context for the items regarded as important (and often posing a certain level of difficulty for students). Bright and McGregor (1970:16) suggest that it is context that is the determinant in identifying the meaning of a word but they warn "There are even more dangerous traps when the overseas context that appears to correspond to the native speaker's context in fact differs." They suggest that students should be encouraged to pay particular attention to collocations. Sinclair (1994-98:18-19) suggests that the text itself contains everything that the reader needs but warns that there are restrictions which, with the help of the computer, can be explored to provide "models which help the text to reveal itself to us". Johns (1994-98:103) sees that the text that should be used by students should reflect the target material the student needs to get to grips with but should not be treated in a manner that would lead students to develop 'bad' reading strategies and that any simplified text will only be "used as a stepping stone to the real thing".

3.1.3 Collocations

The precise collocations to be included will be determined from concordancing of the corpus material to determine the most frequent collocations, where collocation is used to mean (after Grenouff 1991) the usual word or words found in the vicinity of the word being concordanced. Wilks, Slator and Guthrie (1996:67) regard concordances as "special scholarly tools" because they do not give explanations of meaning but only index words against their occurrence in a corpus, leaving out all information except the text citations. The teaching and learning load of collocations can also be reduced by a contrastive approach to the concept of lexical collocation (Bahns 1993). The fact that the undergraduates have already studied English at school does not negate this need to see differences in meaning from those that were learnt in different contexts. The idea that a common word can take on a specialised meaning in technical writing is discussed by several linguists (cf. Bright and McGregor 1970, Darian 1981, Weber 1981, Hoffman 1981). Weber (1981) gives the example of the word *digital* and says "it is impossible to decide whether the term denotes a special technical quality or is just an element of general language use. The denotative meaning of the word is determined by its textual venue i.e. whether we encounter it in a technical statement on computer operations or in a sales talk in a watch shop, where the word might be in a familiar juxtaposition to the word *watch*." Darian (1981) suggests that "ultimately the fullest meaning of a word lies at the discourse level, which allows for an extended definition and deeper exploration." Martin (1992:172) claims that "technical language both *compacts* and *changes the nature of* everyday words." Students need to connect words learnt at school with new contextual meanings in more specialised contexts to avoid a particular kind of "false friends" where words change their meanings in these different contexts (Hoffman 1981). Moon (1994-98:122-124) lists 'fixed expressions' from her analysis of a newspaper editorial. She (1994-98:126-7) finds the most common expressions in the lexicon as a whole to be 'functional' or 'grammatical' as opposed to 'lexical' and that (ibid.:134) examining the fixed expressions in text provides information on the message and the speaker/writer's presentation and how this relates to objective statement or subjective interpretation.

Whilst recognising that the analysis of specialist corpora will not always reveal what the researcher expected, Tribble and Jones (1990:35-36) make the following comment in relation to the utility of concordancing for teaching purposes:

> Two generalizations can be made about applications of concordance output, in spite of their diversity. Firstly, most of them favour discovery learning. That is, they present language in a way that enables learners to discover new knowledge for themselves, rather than being spoon-fed. Secondly, they do this by providing examples of authentic language. The fact that the source material for exercises is drawn from real life rather than concocted by

> teachers increases motivation, as it gives learners immediate contact with the target language in use.

The objective with collocations would be their application directly in the teaching materials, ideally being controlled by the students themselves.

3.1.4 The Baseline Corpus

The first stage of this research involves the analysis of approximately 30,000 science and technology texts taken from a multimedia encyclopaedia. The advantage in this is that these encyclopaedia are totally up-to-date in terms of student access to modern technology, that is, they are to be found on CD-ROMs (see *Integrating Communication Technology 1996).* Guillot and Kenning (1995:365), discussing staff induction at the University of East Anglia, suggest that

> “CD-ROM reference and textual databases are likely to become a major resource in language education at tertiary level in particular: the sheer magnitude of the information they make available, together with the information processing and interfacing options they offer, open a vast array of pedagogic possibilities for self- and teacher-directed learning.”

Although Guillot and Kenning see CD-ROM as particularly important in tertiary level education, the fact that technology is being introduced very much sooner in the educational curriculum means that the students will increasingly be conversant with this type of application. Students will expect to find it available in the university.

The most common, popular and favourably reviewed of these encyclopedia are in American English; this reflects the fact that America is the microcomputer powerhouse producing most of the popular CD-ROM encyclopedia. The fact that most CD-ROM encyclopedia are in American

English fits in neatly with the English requirements of undergraduate students in the University of Aveiro. One extremely useful addition is that, not only can a word frequency study be carried out, but a further dimension can be added to the research and that is the context in which each high frequency word is to be found. So, not only can useful information about lexis be obtained, but also a clearly defined use of those items in specific scientific texts and also the link between the word and its discourse setting. The number of texts that high frequency words are found in can also add to the information about which words students are most likely to encounter and, therefore, need to learn.

In addition, Rosenthal (1996:114) reports that introductory science textbooks for further education in the United States have been getting longer, broader and deeper in their coverage and reading complexity making many of them become encyclopaedic.

3.1.5 The Level of the Material in the Corpora

Huddlestone (1971) gives the following analysis of science texts:

(1) “High-brow”, e.g. scholarly journal articles;
(2) “Mid-brow”, e.g. undergraduate textbooks;
(3) “Low-brow”, e.g. popular science for the general reader.

Table 3.1 Huddlestone’s Level of Science Texts

Swales (1985) argues that this “‘level of brow’ is not as important as the expected relationship between the author and reader”. He describes these ‘mid-brow’ texts as “essentially instructional”. Similarly, Darian (1981:29-30) describes the relationship between *Material and Type of Audience*. His division is as follows:

1. Popular magazines, newspapers	Uneducated layman

2. *Scientific American* and popular books	A reader conversant in the general area (e.g. business, social science)
3. High school text	Layman - limited general knowledge and technical background information
4. Introductory college text	Layman - educated to college level of general knowledge
5. Scholarly journal, specialized book-length study (e.g. a volume on optics)	Specialist and advanced graduate student

Table 3.2 Darian's Level of Text and Audience

Darian claims that for each of his categories the writer assumes a different level of "*presupposition* or *background knowledge*" on the part of the reader. Glaser (1982:76-77) describes the difference in style between what she calls "the academic scientific and technological style" addressed to "'insiders' of a particular field of knowledge" and "the popular-scientific style" used for "a general audience composed of non-specialists". Glaser describes the specific features of each of these being governed by the fact that in the former "knowledge of the subject and the appropriate terminology, the code of formulas and symbols and the various functions of the syntactic patterns" is presupposed whereas the latter "show entertaining deviations from the specialist's topic for the purpose of motivating the reader". Furthermore, she distinguishes both of these styles from a "didactic" style which attempts to make "a job-specific problem (a scientific or technological subject) understandable to the learner" found in textbooks, handbooks and other teaching material used at schools and universities which are "subject to the didactic principle of intelligibility of the text." Similarly, Myers (1994-98: 189) finds that different styles of research articles and popularizations construct different views of science and that scientists "see their work as much more tentative and mediated than does the public." Myers (ibid.) found differences in syntax, vocabulary and organisation between these two types of 'scientific text' and he believes that teachers and students must take these differences into account to "follow the entry of students into a research community."

This thesis contends that the appropriate material for undergraduate students is the textbook, corresponding to Darian's fourth category above and Huddlestone's second category of 'mid-brow'. These students are at a stage where their bibliographies reflect "instructional" texts and therefore the encyclopaedia is an appropriate research tool as it is also 'essentially instructional'. This is because the students are in transition form secondary to tertiary education and have yet to develop greater knowledge of the subjects in their core disciplines on the *Ano Comum*. Both of these text-types also fall into the category of educational texts which will be reflected in their style.

3.1.6 Previous Studies and Text-Types

Swales (1985) takes Barber to task on the latter's (1962) paper "Some Measurable Characteristics of Modern Scientific Prose" for having lumped together two different sorts of texts for his analysis (informational and instructional from three texts with a combined total of only 22,400 words). Swales quite rightly criticises him for muddling his data.

It is essential if the right kinds of conclusions are to be drawn from research for application in syllabus design for the corpus chosen to be taken from the same kind and level of material, and for that corpus to be of a significant size for generalisations to be made. Too many researchers in the past have put different types of texts together for study, thereby confusing rather than elucidating the issue. Myers (1998:179-190) describes the differences between the focus of scientific texts and "popularisations" which are prepared for a more general audience. It is only appropriate in variation studies for a variety of text-types to be examined together, but what those texts are needs to be defined clearly. Biber (1988:208-210) describes precisely which texts he included in his variation studies on speech and writing. Many of the texts were taken from the LOB and London-Lund corpora mentioned earlier in Chapter Two. This level of specificity makes Biber's analysis an appropriate tool for this study.

3.1.7 What are words?

The information obtained for "words" from this encyclopaedia needs to be discussed. First the concept of *word* itself has to be addressed. The Grolier encyclopaedia corpus takes the view that a word is anything which is delimited by two spaces, rather in the way that a computer itself 'recognises' words[2]. This gives rise to some rather strange occurrences such as 'avant', which on closer inspection is seen to be part of the expression 'avant garde' (this could also be written with a hyphen but this changes nothing for hyphenated words are also treated as two separate words by this multimedia encyclopaedia).

Bright and McGregor (1970) discuss the difficulties involved in the decision-taking stage of presenting lists. They give ten 'problems' that need to be taken into consideration and offer some useful orientation for my analysis.

Their **'Problem 1'** refers to regular plurals, they suggest that 'The pupil who has mastered regular plurals will recognise *monuments* instantly if he knows *monument* and vice versa. The difference is lexical not grammatical.'

[2] Sinclair (1991) uses the term 'word-form' for this concept.

'Problem 2' is whether the word is a noun or a verb. Take, for example, the word 'play'. Is this to be regarded as two different words, once as a verb and again as a noun? The Grolier encyclopaedia corpus does not make any such distinction and so it is only through a more searching analysis using concordancing that such distinctions can be resolved. This is a very important issue however as Biber (1998:34-5) points out with his finding that "*deal/deals* functioning as a verb is almost twice as common as the noun use" in academic prose (from the Longman-Lancaster corpus) whereas fiction (from the same corpus) shows the opposite with "the noun use being considerably more common than the verb use". This kind of information is extremely important for ESP language learners and should be brought out in the materials designed for their use.

'Problem 3' is whether regular forms of verbs should be considered different. On this occasion they consider the difference to be grammatical and not lexical. Each grammatical form is given a separate entry in the Grolier encyclopaedia which will help to highlight some of the more obvious grammatical forms to be found in this type of text. Indeed, this could be regarded as one of the strengths of using this multimedia encyclopaedia. Sinclair (1991) says

> "It is now possible to compare the usage patterns of, for example, all the forms of a verb, and from this to conclude that they are often very different one from another. There is a good case for arguing that each distinct form is potentially a unique lexical unit, and that forms should only be conflated into lemmas when their environments show a certain amount and type of similarity."

Sinclair's concern is with dictionaries but his point would appear to be even more important for application in teaching materials where certain forms should be emphasised over others, where this is the form normally associated with the particular genre of concern to the students, in this case, in EST texts. Halliday (1993:71) also argues that it is impossible to separate the grammar from the vocabulary and that it is the 'the total effect of the wording -words and structures-' that the reader responds to.

'Problem 4' is when one word is in fact two lexical items as in Weber's example for "digital" given earlier in Chapter Two. These words are known as 'homographs' and are defined as words that have the same spelling but are different either in meaning, derivation or pronunciation. In this case access to the context is necessary to define which lexical item is prevalent. Once again this can be achieved through the use of a concordance, examining the collocations the word is associated with and so clearing up the ambiguity.

'Problem 5' refers to suffixes which Bright and McGregor regard as one lexical item such as 'young - younger - youngest' or 'play - player' although they recognise that the latter is prone to 'irritating but irrelevant spelling problem(s)' which they claim do not present a large 'learning load'. The high incidence of comparatives and superlatives in scientific texts suggests this is an important structure and that teaching materials would be wise to address this area of difficulty for students.

'Problem 6' refers to extending the previous argument of grammatical suffixes to cover the relationship between such items as 'permit' and 'permission'. Bright and McGregor say 'We cannot, however, usually assume a knowledge of Latin in our pupils' but in this particular instance we *can* assume certain similarities of words because of their common Latin roots whilst at the same time taking particular care to deal with the 'false friends' which occur because of the different developments or evolutions of meanings of words and their use in a specific scientific context.

'Problem 7' is that of prefixes. Bright and McGregor claim that 'any pupil will be able to jump to the meaning of such items as 'action - reaction'. Whether or not this claim (and other similar 'leaps' in understanding by students) is true would appear to depend to a certain extent on the contact that students have had with English in their schools and their understanding of discourse and shared scientific background knowledge as discussed earlier. This will be taken up again later in the evaluation of our students' test results on entering the University (see Chapter 4).

'Problem 8' discusses how compound words and hyphenated words are to be treated. Once again the Grolier encyclopaedia treats these as separate items (as mentioned earlier). This suggests that further analysis will be necessary through concordancing to analyse compound words.

'Problem 9' is concerned with what they term 'form words' such as 'a, the, and'. The Grolier encyclopaedia expressly excludes a number of such words on the grounds that they are too common. A list of these very "common" items is included at the end of each of the alphabetical lists as they occur both in this chapter and in the Appendices.

'Problem 10' is phrasal verbs which are treated as separate items by the Grolier encyclopaedia, that is to say, the verb and its particle appear separately. This is a difficulty that can only be cleared up by examination of the context of use of the main verbs found to be phrasal. Sinclair lists the phrasal verbs that account for nearly 30% of all phrasal verbs in the COBUILD corpus as "bring", "come", "get", "go", "put" and "take". Separate study of these would need to be made if this proved to be an area that the students had particular difficulty with on their test results. The results (see Chapter 4) produce mixed results in fact depending on the phrasal verb being tested.

3.1.8 Other Features of the Text and Corpus

In addition to the ten problem areas discussed above other considerations need to be taken into account such as American words and spellings, abbreviations, pronunciation conventions, Latin and Greek influence, word preferences. Each of these will be discussed below.

3.1.8.1American Words and Spellings.

Whilst many terms and spelling differences are relatively minor, some could lead to confusion especially for those students who had studied the British English model. Although many of the differences encountered might be easily deduced by students such as *colour – color* spelling differences, vocabulary differences like *autumn - fall* where the latter is often taken as a verb in English would need particular attention. Research carried out by Barber (1962:5) found that *may* was much more prevalent in American scientific texts and *can* was twice as common as *may* in British texts. Differences between some common terms for parts of a car (automobile) with such noun equivalences as *boot - trunk* and *bonnet - hood* could lead to confusion in some engineering texts, although these common terms in themselves are often more difficult than other more specialised or technical terms for students. Portuguese students are likely to find the scientific and technical terms are much closer to their own language as these often have Latin roots. One other feature that can prove particularly confusing for students is the difference in the use of prepositions, especially such items as *through - from...to.* Students already have particular difficulty with prepositions in English and to add to this confusion by having two systems might be especially daunting.

3.1.8.2 Abbreviations.

There are usually differences in the use of abbreviations between English and Portuguese (and indeed between British English and American English) which could lead to a breakdown in comprehension. English uses e.g. to indicate an example and Ex. an exercise; Portuguese students have a tendency to use ex. to indicate examples. Similarly, nº indicates number, but may be rendered nr. (near) by Portuguese students. Students are often unaware of these differences and are confused when faced with the appropriate form in English. Some of the common abbreviation differences between British English and American English are in measurements such as ml. for mile in British English which is rendered, probably more appropriately to avoid confusion with millilitre (milliliter in American English spelling), as mi. in American English. Added to this difficulty is the increase in the use of abbreviations in specific genres. Some uses of abbreviations are particularly idiosyncratic, like the pronunciation system found in the CD-ROM material described where sounds are rendered by groups of letters like 'ahl', 'ahn' and 'ahr' (see Chapter 5, Table 5.1).

3.1.8.3 Pronunciation conventions.

Encyclopedia texts written for native English speakers often contain indications of pronunciation which do not follow the dictionary (usually international) phonetic notation which some students may well have come across through reference work in their studies. These appear in corpora and would almost certainly lead to considerable confusion for non-native language learners. The sound groups would be totally inappropriate for Portuguese speakers who approach pronunciation using their own language's conventions. An example of this problem can be observed through exclamations like *Ah* and *Ha* which may be pronounced in very much the same way for a Portuguese speaker but quite differently by a native English speaker.

3.1.8.4 Latin and Greek Influence.

There has been extensive use of Latin in scientific writing and historically Latin was the *lingua franca* of educated people and most scientific studies were written in Latin into the Renaissance and thus could be understood by scientists in other countries. In modern scientific writing these Latin and Greek roots do still exist, but the use and knowledge of Latin and Greek is no longer a pre-requisite for most branches of science. Stubbs (1996:70-1) sees “Graeco-Latin loan words” as having been used to build up the vocabularies of institutions which in turn leads to “differential access to subjects on the school curriculum”. Stubbs is concerned with the “authoritative knowledge” that is expressed by such features of texts as examples of power relations and the way that writing is always aligned. Examination of just how Latin and Greek terms are employed in modern texts can be an important tool in deciding what should be included in the syllabus for today’s undergraduates. White (1998:290) claims that the “strangeness” produced in scientific texts through the use of Latin and Greek derived terms is a deliberate measure taken to ensure that the reader recognises that these terms are not to be taken as the normal view of reality construed through the use of the vernacular. Thus the need to construct different views of reality, which are not related to “common sense”, is essential for an understanding of (the language of) science to take place. Laurillard (1993) gives a number of examples of how the concepts used in science do not match those found through “common sense” and how lack of success in building adequate conceptual frameworks by students can occur in many different ways. She claims that higher education has not yet, found a means of coping with this as yet other than through the tutorial question and answer system to draw out where and when the misconceptions occur.

Strevens (1978:193) maintains that Latin and Greek roots and affixes combine to form an extremely large number of words which are 'science-specific'. He cites the roots *aqua-*, *cyto-*, *hydro-*, *plasma-*, *pyro-*, and the prefixes *ante-*, *anti-*, *poly-*, *post-*, *pre-*, *sub-*, and suffixes *-fer, -ite, -logy, -valent.* Strevens (ibid.) maintains that this scientific vocabulary makes up a 'normal part of the training of all scientists'. Portuguese students are fortunate in that they have a Latinate language which may go some way towards providing them with knowledge of and insight into the scientific applications of Latin roots.

3.1.8.5 Word Preferences.

Quirk (1995) has shown that some words are preferred in certain texts or registers even though there may well be a very similar synonym. "Ancient" and "old" for example may exist in almost equal numbers of texts (range) and frequencies in the corpora whereas "attempt" as opposed to "try", and "change" as opposed to "alter", may exist in different frequencies showing preference for one form over the other. Quirk (ibid.) argues that these kinds of choices, although apparently arbitrary, can indicate formality in texts and may therefore be representative of the particular genre they are found in. Lemke (1998:92) suggests that choices of lexis contribute to the "attitudinal stance of a text to its audience, to its content, and to other text-embodied viewpoints". McCarthy and Carter (1994:104-5) suggest that vocabulary choice is just as discourse sensitive as grammatical choices and that if language is to be considered as discourse "vocabulary must be a concern as much as any other aspect of language form".

A similar position is adopted by Biber, Conrad and Reppen (1998:43-54) who demonstrate that "big", "large" and "great", which are often presented to students as synonyms, are usually used in quite distinct patterns and with specific meanings. They find (1998:51) that fiction and academic prose have different preferences for these words with "big" being more common in fiction and "large" in academic prose. While both registers use "great" with "deal" as a collocate, fiction uses many more senses of "great" than does academic prose. They account for these findings by suggesting that "fiction texts contain frequent physical descriptions" and "more varied descriptions" whereas academic prose texts "deals with size" and "specific measurements". They go on to examine the collocates associated with these words in the two registers. Similarly (1998:98-99) they examine the preferential use of "begin" and "start" in fiction and academic prose and discover that the intransitive use of start is the most common in both registers but is more prevalent in academic prose. In contrast "begin" is usually used intransitively but in fiction it is used mainly with a *to* - clause. In other words, they argue that the patterns of language use are not synonymous across registers. Thus the vocabulary preferences found in the corpora are significant both as representations of the style of the texts and as a means of demonstrating a model of authentic usage to students.

3.1.9 Optical Character Recognition

Burnard (1992) points out that the optical scanner or optical character recognition machine (OCR), which in this study was used to compile the physics and chemistry corpora, can only recognise what is visibly present on the page and that it cannot undertake any kind of editing nor can it distinguish structurally different components of the printed page even if these are visually distinct, such as footnotes and headings. Any corpora searched will not provide these distinguishing features so that an analysis of the texts themselves as published can often reveal other interesting and important features of those texts (see later 5.2).

Laurillard (1993-97:27) describes academic knowledge at university level as a process of 'mediating learning' because the students have to learn what others have given insights into rather than what they can have direct experience of. She suggests that because academic knowledge "has this second-order character, it relies heavily on symbolic representation as the medium through which it is known" and although the medium may be language it may also be "mathematics symbols, diagrams, musical notation, phonetics, or any symbol system that can represent a description of the world." Therefore, students in university have two problems to overcome, the first that of handling the representation system and the second the ideas which they represent. Some features that must be taken into consideration therefore are the use of typographics, titles, subtitles, summaries and conclusions, drawings and diagrams, and formulae, numbers, equations and tables. These features should add to the student's understanding of the text, provided that the student is aware of the conventions used and has been trained in recognising the multimodality of texts. Lemke (1998:95) suggests that scientific text is not meant to be read in a linear manner and for him it represents a "primitive form of *hypertext*" where "footnotes represent an optional branch for readers, so do figures and their captions, and the parenthetic or main-text expressions such as '(Table 3)' or 'as seen in the first table' which point to them." In contrast speech is linear in this respect. The number of dimensions that are then available to the reader is much wider and access to them is much more open, the reader can choose what information to access then from the different textual and visual information present in scientific texts. Nevertheless, students have to have background knowledge of the canonical forms used in science in order to be able to understand and interpret the information available.

3.1.9.1 Typographics.

The punctuation and use of italics are also vehicles of information in a text. Darian (1981) refers to the use of typographics as vehicles of definition. The equals sign, the colon, pairs of commas, parentheses, the dash, quotation marks and italics can all be used to give or signal definitions in scientific and technical writing and are constantly interacting with the text itself. In this way it is possible for the definitions in a text to be either “overt” or “covert” (Darian 1981:36). Often the word which is to be defined in a covert way is flagged, that is it can be located in bold, a convention which Halliday and Martin (1993) have explored in *Writing Science* and White (1998) discusses in *Reading Science,* but which nevertheless harks right back to Locke in the seventeenth century and his concern with definitions. Lemke (1998:95) suggests that typography serves to orient as well as organise and that “the use of italic and boldface types signals emphasis or importance, as does the relative point size of type in titles, headings, abstracts, footnotes, captions, labels, etc.” while “paragraphing and sectioning of text, and geometric relations of figure space to caption space indicate to us which elements are to be preferentially read in relation to which other elements; what goes with what.” I would go further than this. This easily accessed presentational mode affects our ability to bring out latent meaning in texts. It is interesting to note how the use of computers themselves are even influencing this kind of discussion. In the past, before the widespread use of word processors, neither teachers nor students would have been so ready to discuss “point size” or “font”.

The use of a number of symbols through computers has taken this further in modern materials and, if over used, these may serve to irritate rather than encourage as is the case of the ubiquitous, perfidious ‘smiley’ to indicate a joke or other attempt to be friendly or light-hearted. Attempts to make materials for learners more attractive may require a clear statement at the beginning of how these symbols will be used in any text. If learners skip past these early explanations in the textbook, they may be in danger of missing many of the connections the author intends to make.

3.1.9.2 Titles, Subtitles, Summaries and Conclusions.

Van Dijk (1997:10-11) claims that discourse topics define the overall ‘unity’ of discourse and are “typically expressed in such discourse segments as headlines, summaries or conclusions.” He also claims that they “also happen to be the information that we usually remember best of a discourse,” which, if the case, means that these features are especially important for study purposes.

Increasingly in modern textbooks the flow of the text is divided into small sections or paragraphs with titles or subtitles used to indicate the topic discussed. These also serve to allow topics to be easily located within the text. The use of a system of colour coding of these titles and subtitles is also prevalent which together with diagrams and drawings produce a much more attractive and much less dense appearance than the old dry textbook styles of some decades ago. Visual display of the kind described above is now an essential element in making teaching materials attractive and with a proper pedagogic basis can facilitate understanding of the topic discussed.

Similarly the statement of aims and objectives at the start of a topic and summaries at the end of topics help students to focus on what they are going to learn and what they should have learnt from the texts that went before. In this way they serve to prepare the learner for the task and serve as a check on learning. These types of activities are known as ‘wrap around exercises’ to assist with text processing and to enable the learner to monitor progress on their own at home.

3.1.9.3 Formulae, Numbers, Equations and Tables.

The use of formulae, numbers, equations and tables in scientific texts features are extremely important. They are the means whereby an alternative or additional representation of the information contained in the text is provided. Certain conventions need to be understood such as the arrow or equals sign, but, fortunately, these conventions enjoy international standing and recognition. Lemke (1998:96) warns however that particularly in tables “readers are expected to supply the canonical semantic relations of thematic terms which are often underspecified or omitted”.

Similarly, the chemical symbols for elements and compounds are standardised through international convention and should not prove to be a stumbling block provided that they are understood in Portuguese, especially as the formulae follow the English order; for example NaCl, sodium chloride, in Portuguese would be reversed and read *chloride of sodium* (chloreto de sódio). Thus, it can be argued that the number and range of formulae should add to understanding rather than obscuring it, provided that there is shared background scientific knowledge. However, some mathematical working could prove confusing (see 5.4 Mathematics).

Lemke (1998:90) found that in most of the theoretical physics articles that he studied the running verbal text would make no sense without the integrated mathematical equations “which could not in most cases be effectively paraphrased in natural language” even though they were meant to be read as part of the verbal text "in terms of semantics, cohesion and frequently grammar".

3.1.9.4 Diagrams and Drawings.

As with the use of other textual features, the use of diagrams and drawings should enhance the understanding of the surrounding text, provided that the referencing to these is understood. In general, visual material in the text was seen as a form of redundancy as it reiterates what is being discussed. However, Lemke (1998:104) disagrees with this position and claims that visual figures and mathematical expressions add important or necessary information and so complement or complete the main text. Modern discourse analysis sees multi-modality in texts as an essential feature of study in discourse semiotics. Kress, Leite-García and Leeuwen (1997:257) say that "producers of texts are making greater and more deliberate use of a range of representational and communication modes which co-occur within the one text" and that the reader has to take these into account in order to "read texts reliably".

Van Dijk (1997:6) suggests that "in these times of multi-media communication ...an analysis of the visual dimension of discourse is indispensable." Van Dijk is much more interested in non-verbal signs or semiotics but, nevertheless, the visual element in student's textbooks should be an aid to understanding the discourse of the text if the students can interpret them accurately. Lemke (1998:87) argues that semiotic systems such as language, tables, graphs, images and diagrams do not just "add-on" meaning to a text but actually create new orders of meaning thereby "multiplying meaning". Furthermore, Kress, Leite García and van Leeuwen (1997) suggest that it is important to see visual images as independent vehicles for meaning in their own right. If the students can make the connections between visual images and text or 'read' visual images in scientific texts accurately, this would help the students to ascertain the meaning in those texts. The question of whether students can do this successfully is taken up again later in 5.2.4.

3.1.10 Comparison with other published data.

Bright and McGregor (1970) claim that technical texts contain a large number of words that are 'outside simplified English'. They list twelve such words: '*absurd, adequate, adjoining, aggression, alert, alternative, amateur, ample, apparatus, apprehensive, automatic, available*'. Of these words only six do in fact occur with significant frequency (defined as 100 occurrences or more) in the Grolier encyclopaedia (see word list in Table 5.1) and of those six only two could be considered sufficiently different from Portuguese to warrant attention (*apparatus* and *available*). It would seem that from this small comparison that the claims made by corpus linguists that intuition and the reality found through empirical research of computer corpora of naturally occurring texts (as opposed to specially written examples) are at odds is essentially correct.

3.2 Needs Analysis

3.2.1 The Students' Level of English

In deciding which words should be excluded from the list, several considerations were borne in mind. First and foremost was what could reasonably be expected from the students' previous contact with English in the schools. When the new students were tested each year they were asked to state how many years of study of English they had had. The results were as follows:

Table 3.3 Students' Number of Years of Study of English

Years of study	1993/4	1994/5	1995/6	1996/7	1997/8
0	4.7%				
1					
2	2.7%	2.8%		1%	0.32%
2.5	0.7%				
3	18%	13%	14%	8%	5.73%
4	3.3%	0.6%	1%	1%	1.91%
5	38%	26.5%	29%	30%	23.25%
6	1.3%	4%	3%	2%	5.09%
7	28%	47.5%	47%	53%	54.78%
8	2.7%	4%	6%	4%	7.01%
9				1%	0.96%
10	0.7%	0.6%			
11	0.7%	0.6%			
15					0.64%
18					0.32%

(3.57% of those who took the test did not answer this question at all.)

Consideration of the results for the academic year 1993/4 shows that, as some comments offered by the students remind us, there can be many instances which do not follow what would have been predicted for students leaving secondary school and starting university. The 1997/98 results also demonstrate that the question on the test paper may have been given a cursory glance and understood as *Quantos anos tem?* and the essential *years of study* have been overlooked to produce the answer 'eighteen'. [3]

It was expected that the students would have studied English for an average of three, five or seven years, with some time gap between the years in which they studied English and university. What was most worrying for the academic year 1993/4 was the percentage of students with no English at all embarking on the course in conjunction with a significant number of students who had studied English for seven or more years.

[3] When a student who gave an answer like this was questioned about it later, she admitted that she had in fact given information about her age and not her studies. However, this particular student justified her answer by explaining that she had in fact spent most of her life in America so she felt that the English language had indeed been part of her entire life.

The results for the academic year 1994/5 were a little more encouraging in that they show no students with absolutely no English, nevertheless, there are still a significant number with very little English.[4]

The figures for the academic year 1995/6 show much more clearly the expected breakdown into three, five and seven years of English. Some of the intermediate figures could be accounted for by students who have had to repeat a year at school, which, if true, would suggest that those percentages were students who could be considered weaker than others in the same broad categories.[5]

The figures for the academic year 1997/98 show how there is a general trend for students to be stronger in English than before, and the answer 'fifteen years of study' reflects students who had been brought up abroad in English speaking countries. Certain courses, such as *Novas Tecnologias e Comunicações* - New Technologies and Communication (NTC), appear to be attracting students who are generally stronger in English which is not perhaps surprising given the nature of this course which has a slightly more 'humanities' or 'arts' bias than the other *Ano Comum* courses. These students also continue with their English studies for a further year unlike most of the science and technology students in the university.

[4] Chatel (1999:246) records a similar change from 1988 to 1994 for sociology students in the University of Coimbra.

[5] Drª Maria Adelaide de Araújo Nunes of the University of Evora (1999:258) describes the "uncongenial environment" for ESP with students who "have had only a few years of English at secondary school and/or having systematically failed the subject there" and so "feel at a loss and are understandably reluctant to study a subject that they hoped they would never encounter in their lives again".

Figure 3.1 Pie Graph for the Academic Year 1993/94 showing the Students' Number of Years of English

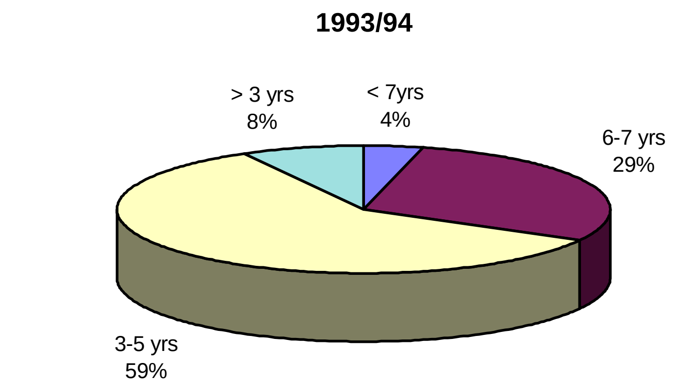


Figure 3.2 Pie Graph for the Academic Year 1994/95 showing the Students' Number of Years of English

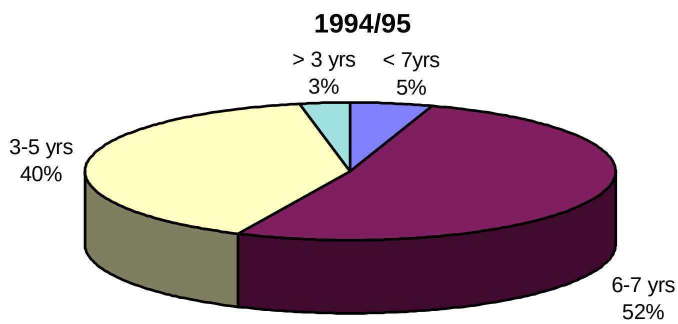


Figure 3.3 Pie Graph for the Academic Year 1995/96 showing the Students’ Number of Years of English

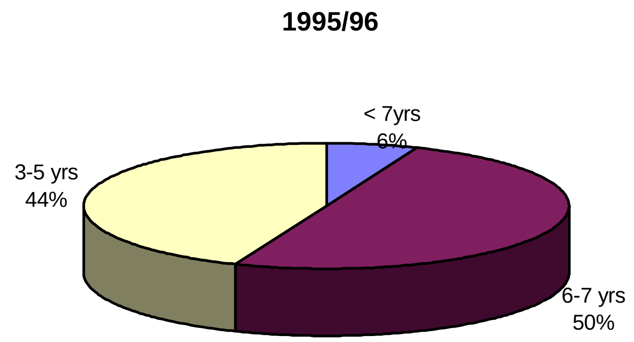


Figure 3.4 Pie Graph for the Academic Year 1996/97 showing the Students’ Number of Years of English

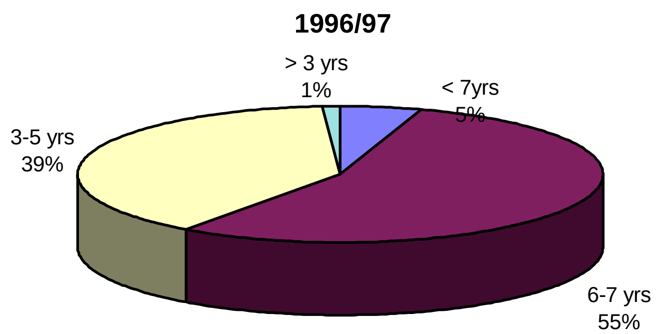
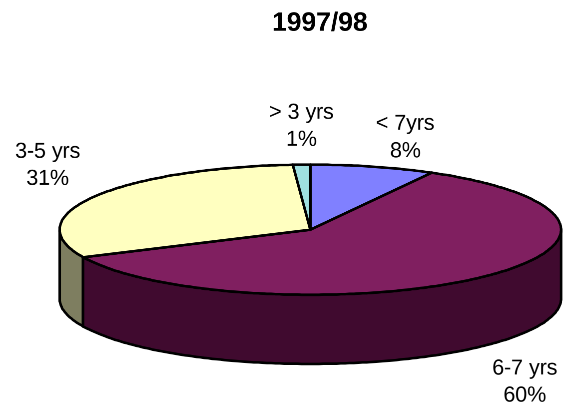


Figure 3.5 Pie Graph for the Academic Year 1997/98 showing the Students' Number of Years of English



One other factor could be affecting these figures and that is that in the first year all the students took the test but in subsequent years students were advised that they need not take the test if they felt that their English was not of a high enough standard. However, in recent years students have changed their attitude and appear to treat tests as a kind of lottery where they hope that through some stroke of luck they will pick the winning combination of answers. They perceive that at any rate they have nothing to lose by trying. It may also reflect a change taking place in schools where an increasing number of students opt for English as their main foreign language at an earlier age and so feel that they are of a higher level. There are also more and more private language schools opening up all over Portugal and some of their students must now be coming through to university in increasing numbers.

Over the years some of the students have felt compelled to add comments to the question they were asked about how many years they had studied English. Rather like the example above, some students gave an explanation for their answers. This could be because they had repeated a year, as surmised above, or that they had done all of their studies in

English in another country, for example, Australia, South Africa or America. However, a small number of students gave value judgements about the quality of the teaching they had received, one student replied "*três anos e pessimos*" whilst others inadvertently showed the difficulties they had with English by answering "I am three years" to this question. Other students explained that their studies had taken place a number of years before they had taken up their university place and so their English was 'rusty' and, yet others, that they had studied both at school and at private language schools, thus completing 'double years' or that they had taken the Cambridge University examinations in English.

These figures suggest that assumptions made about the level of the students' English could be wildly inaccurate, although there is a general trend for the students to have studied English for longer in the secondary school[6]. One other consideration is that, although the students may have studied English for seven years, most are unlikely to have studied it in the final year of their secondary school course as they will have chosen to follow science subjects and not the humanities. Given the kinds of problems explained above and the increasing pressure on grades for university entrance, it is also possible that some of the students had not studied English for more than two years because of repeating the final year to improve their grades.

The fact that students have studied English in secondary school does not negate the fact that their knowledge of language is limited to what was taught on a general English course which is likely to have concentrated on spoken English and more 'literary' kinds of comprehension and composition, and to have dealt little, if at all, with the

[6] Drªs Ana Maria Ferreira, Dulce Ramos and Fátima Braga da Silva from the University of Porto in their paper "Evaluation des Curricula de FLE au Portugal" (1999:333-337) show the numbers of students studying French, English, German and Spanish in the central region of Portugal in the academic year 1994/95. The figures for English demonstrate clearly that a significant number of students continue their language studies into the final three years before university (approximately half of those studying English in the 3rd cycle - 7th to 9th years of school).

language of science and technology. Langkilde (1982:523) describes the barriers students in the Copenhagen School of Economics were found to have because “unless they are made aware of the necessity of developing a particular method for dealing with specialized texts they will for a long time go on treating an economic text in the same way as they treated a chapter from Balzac or a scene from Molière in grammar school.” Tavares, Valente and Roldão (1996:62) say that the English Programmes for schools do mention types of texts but these are given as “dialogue, interview and advertisement” (I:42) and discourse organisation as “descriptive, narrative and argumentative”[7] (I:48)P. These authors suggest that cultural identity and understanding, within the general development of the pupil as a responsible citizen, are the main concerns in the programmes for modern languages in Portuguese schools at the moment. They also point out that teachers need to be up-to-date with their training if they are to be able to cope with the requirements of the programmes, an issue that has been mentioned many times in relation to teaching science and technology. The school teachers themselves usually come from a ‘literary’ or humanities background and are, therefore, unlikely to feel comfortable teaching English for science and technology.

Research carried out with students of the fourth year of the teaching degrees in Portuguese and English, and English and German, demonstrate that these future teachers have difficulty with numbers in English just as the students entering the university have (see Test Results for New Students, Chapter 4). This situation would therefore seem to be self-perpetuating as teachers are generally unwilling to teach something that they themselves find difficulty with. Swales (1973:9) describes how teachers found it “almost impossible to view their Science students’ interests as different from their own” and therefore assumed the students would find boring what bored them.

[7] My translation of “dialogo, entrevista, anúncio” and “descrição, narração, argumentação”

Overall these results would suggest that increasingly the students could be expected to have an intermediate level of English but with no science subject specialisation in English. The *structure* or *form* words mentioned earlier should be quite well known to the students but, as will be shown later in the test results, some discourse markers are less well understood. The syllabus design implications of these findings are to complicate the issue of the level at which to pitch the instruction. The needs of those students in the bottom 1% with less than three years of English can hardly be met, and this will lead to their virtual exclusion from most of the activities designed for the majority of 60% with six or seven years of English. Equally well the top 8% may find the level pitched beneath their capabilities and so lose motivation. These more able students must be included in the activities carried out in such a way that they feel stretched and that they are also making progress. It might be possible to engage these students in helping their classmates to reach a higher standard and incidentally help to create bonds between students in this new environment which is seen to be necessary for successful learning (Tavares *et al.*1996).

The knowledge of science and technology that the students bring with them to the first year of university is also variable. Some students will have chosen to study physics in their final year at school and some chemistry, some will have studied more mathematics than others and so on. This implies that homogeneity in terms of subject knowledge cannot be guaranteed either in the students entering the foundation year disciplines. This fact will have repercussions on all of the strategies and skills that these students require in order to be able to perform well in their studies of the subject matter in a foreign language.

3.3 Biber's Methodology of Variation Studies and Corpora Analyses

The corpora from the Physics and Chemistry textbooks on the students' bibliographies will be examined using Biber's (1988) methodology of text variation to try to see what must be taught to the students in the university that is specific to this text-type and, thereby, to make the course relevant and to 'fill in the gaps' that the students bring from their studies in school. Biber was conducting research into variation between speech and writing but he provides a very explicit methodology for the description of the linguistic characteristics of the range of genres in English that he included in his study, which will allow comparison of the physics and chemistry texts under study here with his results for Academic Prose.

Biber's goal was to include all the 'potentially important linguistic features' of the different genres included in his study in order to identify the 'linguistic parameters along which genres vary, so that any individual genre can be located within an 'oral' and 'literate' space, specifying both the nature and the extent of the differences and similarities between the genre and the range of other genres in English'. It is this identification of the differences that needs to be studied in order to identify those areas to be included in a syllabus for undergraduate science and technology students. Biber claims that it is 'bundles' of linguistic features that occur together in texts that 'work together to mark some common underlying function'.

Biber identifies 67 features from previous research, which can be grouped into sixteen major grammatical categories: (A) tense and aspect markers; (B) place and time adverbials; (C) pronouns and pro-verbs; (D) questions; (E) nominal forms; (F) passives; (G) stative forms; (H) subordination features; (I) prepositional phrases, adjectives and adverbs; (J) lexical specificity; (K) lexical classes; (L) modals; (M) specialised verb classes; (N) reduced forms and dispreferred structures; (O) coordination; and (P) negation. He gives very precise definitions of each of these features (see Appendix A) together with the functions that other researchers have ascribed to these features. For example in Tense and Aspect Markers he suggests that past tense forms are usually taken as the primary surface marker of narrative; perfect aspect forms are associated with narrative/descriptive texts and certain kinds of academic writing, and that these co-occur with past tense forms as markers of narrative; and that present tense verbs can be used in academic styles to focus on the information being presented and remove focus from temporal sequencing.

By using Biber's features it will be possible to analyse both the register and discourse of some of the texts in the undergraduates' bibliographies in order to apply the results to designing an appropriate syllabus for these students.

Chapter 4 Test Results for New Students

Chapter 4

Test Results for New Students

4.1.1 Student Numbers

In the first academic year of the *Ano Comum* and of the Preliminary Test, that is, *1993-1994*, there were approximately *1350* students studying English in the *Ano Comum*. and in *1994-1995* there were approximately *1200* new students entering the *Ano Comum* and about *180* repeating this discipline. These numbers have continued much the same for the academic years 1995/96, 1996/7 and 1997/8.

As I showed earlier most of the students entering this discipline could have studied English for either three, five or seven years in their secondary schools. What the students have learned, have learned incompletely or have not learned at all in their secondary schools is crucial for syllabus design, so the results of the Preliminary Test were analysed for it to be possible to decide what needs to be given particular attention in their proposed syllabus.

4.1.2 The Preliminary Test

Using local knowledge and experience, we realised that some of our students were likely to have studied in an English speaking country (probably 1-2%) or to have studied English at one of a number of reputable language teaching institutes (up to 15%), and, therefore, be generally proficient in the language. Some students, in replying to the survey mentioned above, volunteered the information that they had already obtained the Cambridge University Certificate of Proficiency in English which is a qualification regarded as a minimum English teaching qualification by the Ministry of Education in Portugal. An innovative decision was therefore taken to test all the potential students immediately at the beginning of the academic year and give all those who were deemed to have a sufficient knowledge already the opportunity of being excused from the discipline altogether. This decision was applauded by the student body who suggested unsuccessfully that they would like it extended to other core disciplines. The effect of the decision to innovate in such a way was to reduce class size a little in an attempt to give the less proficient students more time and attention in class and to permit those students with greater proficiency to concentrate their efforts in other areas where they might not be so proficient. Many students started learning English (mainly in private schools) whilst they were still in primary school and this early teaching has also come to be seen as beneficial in state education in Portugal. Changes have been introduced in the curriculum to permit different schemes of foreign language study often also extending this to the final years of secondary schooling for all children no matter what their core curriculum. Innovation for undergraduates on science and technology courses has also been the focus in tertiary education. Students who were found to have great competence in English were also considered to be likely to be demotivated by being in a mixed ability class with over forty other students. In actual fact some students who had expected to be excused from the discipline were surprised to learn that it was their knowledge of science that let them down in the test. The specific English being tested went beyond the mundane day-to-day usage of children and required a more mature, informed view from students

As was mentioned earlier, all the new students coming in were tested to see if some of them could be offered the chance of not taking this discipline at all because their English was considered to be of a sufficiently high level. This level would correspond to already knowing enough English to be able to pass comfortably the kind of test that they would be given at the end of the year after studying specific English for science and technology in large mixed ability groups for two hours per week for one academic year. In other words, a proficiency test was needed to evaluate the student's knowledge. It was decided that an adequate initial standard of English would equate with a mark of fifteen or above out of twenty.[1]

The test had to be one that could be administered and marked easily, given the numbers involved. A multiple choice format was chosen as an objective test and so that a template could be used for ease of marking. A short paragraph was also included to verify the results of the multiple choice test. This was changed to a reading comprehension test in the fifth year of the test as it was felt that this area of competence needed to be checked so that we could feel reasonably confident that the students who passed the test well were capable of coping with the reading that they would have to do in English on their courses. The ability to write well in English was also considered to be less important to the students' immediate curricular needs. Four versions of the same test were produced in order to avoid copying, this was later changed to two versions of the test because this was found to be both much easier for the writers of the test and yet sacrificed nothing of the security aspect of the testing. The test was made up of both grammatical and vocabulary items as both of these areas were deemed pertinent and specific to the language of science and technology which the students would need to cope with in their studies.

[1] It was considered that the students would be disinclined to accept our offer if their mean mark was lower than this, which would defeat the object in view of reducing class size and not wasting student study time.

The results of the test were analysed and those students who obtained a grade of fifteen or more were duly informed that they need not attend classes and indeed had already obtained a final mark for the course (fifteen and above). This does not mean that all the students thus informed (about 10% of those who took the test) decided to accept this result. No bar was placed on these students attending the course, if they so wished, and indeed some did choose to attend. The students could also choose to take the examinations at the very end of the academic year if they felt that they could do better than they had initially. Some students felt that this was possible after they had had access to the materials used in the discipline from which they could then study some of the relevant scientific and technical English which they perhaps felt they were unsure about initially.

This test was analysed and refined for use in future years but results for the first years show that the major discriminators were specific vocabulary and grammatical items such as the present perfect, second conditional, gerunds, phrasal verbs and specific lexical items. Increasingly items have been included in this preliminary test which reflect the syllabus of the first year, items on pronunciation and numerical knowledge for example.

4.1.3 Test Results 1993/94

The test that was used in the first academic year, 1993-1994, consisted of fifty multiple choice questions and a short, 100 word, paragraph on a given topic. The reason for this format was first and foremost that it would be very time-consuming to administer any other sort of test to such a large body of students *and* be able to publish the results early enough so as not to take up too much of limited teaching time. The written part of the test acted as confirmation of the result obtained in the multiple choice test. The topics given on the first test were:

i) The importance of computers for students at university.

ii) Why science students should study English at university.

iii) The importance of the course you have chosen to study.

As the students had to write only 100 words on one of these topics, they had to be extremely concise. Writing such a short amount is often more difficult than permitting the students to write as much as they wish. Indeed, many students attempted to go beyond the specified limit whilst others did not even attempt the written section at all. The questions also required an expository or argumentative style of writing.

Although discrete item tests are not considered very valid, they do have the advantage of being reliable. As Weir (1988) points out, the test can also be made more valid by taking into account the needs of the students on their individual courses. The different departments took the optimal view that students would need all four skills of reading, writing, speaking and listening in order to pursue an academic career (see later *4.2 Needs Analysis by University Department)* but the constraints imposed by the length of time available for the discipline meant that the goals would have to be somewhat more short-term and reflect the arguably more receptive skills of reading and listening. The latter could only represent a small percentage of the whole syllabus even so. Therefore, the syllabus that the students' would subsequently pursue could not be considered communicative in any modern sense of that term. Reading and some listening would form the bulk of the syllabus and these would be approached in a way that could give the students 'enabling skills' in the hope that given time they might build on what could be taught/learned in such a small space of time. In other words, the methodology used would be as learner centred as possible in order to meet the needs of the individual students as far as this could be achieved. The test then had no reason to reflect other methodological aims. A greater allocation of course time and resources would have behoved a more comprehensive test.

The questions on this first test started with the simple present tenses, negatives and question forms and went on to modal constructions, conditionals, phrasal verbs and passive constructions. In other words the accepted ‘easy structures’ to the more complex. Some specific vocabulary questions were also included. The results showed that approximately ten per cent of the students enrolled had achieved a mark of fifteen or more and could then be released from the discipline. However, evaluation of the test results also showed that *ten per cent* of the students could not competently handle what are considered basic structures. For example, present simple question forms, present continuous and present perfect tested in the following way:

Questions of the type:

3. ____________________ coffee?

A Do she like B Likes she C Does she like D Like she

caused *nine per cent* of those tested to make an error and *eleven point five per cent* were caught out by:

4. I ____________________ English.

A am study B studying C studies D am studying

Most of the students tested, that is *96%,* could not answer the following correctly:

34. This is the first time I ___________________ Aveiro.

A am visiting B visited C visit D have visited

Certain items like specific vocabulary, prepositions and the subjunctive caused more than *75%* of the students trouble, for example:

13. You _________________ take an aspirin for your headache.

A had better B would better C will better D have better

30. It's time we ___________________ .

A go B went C going D goes

41. I will have to phone later his number was _________________ .

A occupied B engaged C talking D speaking

38. They congratulated their cousin _______ passing his driving test.

A at B by C on D with

50. The car tried to ________ the lorry while it was waiting at the pedestrian crossing.

A overtake B overlook C overpass D overcome

44. After the mission the space shuttle lands on a _______ in the same way as an aircraft.

A path B highway C motorway D runway

37. Her cupboards are full of clothes, most of _______ she never wears.

A them B which C those D that

21. They _______ in love at first sight.

A fell B felt C feel D fall

25. He took ____ Keith at once and they became firm friends.

A after B up C to D on

36. __________ the high price of meat, the family bought lamb or beef every week.

A Because B Besides C Despite D In spite

More than *50%* of the students could not manage questions on the second conditional *(67%)*, indirect question forms *(60%)*, the future perfect *(62%)*, "*suggest*" with a direct and indirect object *(70%)*, the phrasal verbs *"get over"* and *"put in for" (54% and 61%)* and a further five vocabulary questions including such items as *"traffic jam"* and *"experiment" (50% and 62%).* With such generalised difficulty, the syllabus must obviously take such language deficiencies into account as teaching syllabi always consider the average student. Extreme positions whether higher or lower are inevitably for a smaller number of students and so those who represent the middle ground, median or more properly the standard deviation of + or – 1 on the normal curve found from testing are always those taken as the 'average' students for whom any course is designed. This is contrary to many of the older systems, particularly of higher education, which aimed to teach an elite group with all others falling by the wayside. The numbers of students involved in modern education in developed countries necessitates an attempt to raise the general level of education of *all* of those involved in the education process and necessitates new methodologies to achieve this aim. Therefore those test results that show widespread difficulty but not almost total impossibility for students are taken as items that need to be included in the syllabus in order to raise the standard of English of the majority of the undergraduates.

4.1.4 Test Results 1994/95

The test that was used in the second academic year, 1994-1995, also consisted of fifty multiple choice questions and a short, 100 word, paragraph on a given topic. The reasons for this being exactly the same format as that in the first year was that the numbers involved continued to prohibit almost any other practical possibility. However, this time the test items were altered to incorporate some of the items considered fundamental to the course as it had been taught in the first year of operation of the *Ano Comum*. Other items that were considered to be inadequate discriminators, after the test results had been studied for this purpose specifically, were eliminated. Thus further validation of the test was incorporated without sacrificing either any of the reliability of the test, its objectivity or, above all, its speed of administration and correction.

Overall results were now also available about pass rates and grades of the first year on this foundation course and these results also validated the test in that the percentage for allowing students to choose not to take the course at all equated well with that of all the students of the year reaching a high grade that is, 15 or more (approximately 12%).

The items considered fundamental that were now included in the test covered both the specific vocabulary that had been taught during 1993-1994 and an attempt to assess the student's *awareness* of pronunciation. The results this time showed that approximately *13%* of the students had reached a level which was considered adequate, and could be released. The proportionate increase in the number of students released was most likely to be due to the fact that, when the answers to the query about the numbers of years they had studied English were collated, it was found that *57%* of the students had studied six years or more (*43%* had studied five years or fewer) whereas in 1993-94 *33.4%* of the students had studied six years or more and *66.7%* had studied five years or fewer (see *3.2.1 The Students' Level of English*).

Nevertheless, when the answers to the multiple choice questions were once again analysed, it was found that the students continued to have difficulties with modals *(97%),* direct and indirect objects after "suggest" *(63%),* phrasal verbs *(71% and 83%),* reciprocal pairs *(63%)* and the subjunctive *(89%).*

It was perhaps less surprising to find more than *75%* of the students having problems with those more specific vocabulary items that had been introduced. Questions like:

When light enters another medium it ________________ .

A reflects B absorbs C bends D glows

caused *89.5%* of the students to choose an incorrect answer. This could have been because of either a problem with English or a problem with the basic scientific concept included in this question. However, the research carried out on the physics and chemistry texts suggest that this item is very frequent in just such a context.

84% could not identify the pronunciation of regular past tense verbs in a question like:

32. The *ed* ending in "*showed*" is **different** from the one in

A studied B remembered C caused D asked

More than half of the students *(52%)* could not identify the sounds of the alphabet in the following:

The letter "***A***" does ***not*** contain the same sound as __________ .

A "***J***" B "***K***" C "***H***" D "***Q***"

This item was considered fundamental because of the various formulae and descriptions of scientific objects which include reference to the shape of letters such as a "T-shaped lamina" and "J curves". Added to this, if the European *Threshold Level* (van Ek 1976) for English in schools is taken into consideration, this is found to be one of the items that has been singled out as *essential* learning for secondary school students.

The items which caused the students the least number of difficulties *(25% and less)* were pronouns *(23%),* questions using an auxiliary *(23%),* telling the time *(24%),* present perfect *(22% and 25%),* the superlative *(20%),* the past continuous *(21%),* future perfect *(17% - the lowest number of errors on this test)* and the phrasal verb *"get over" (21.5%).*

There would appear to be some strange discrepancies here. If students find phrasal verbs as easy as (or easier than) telling the time or the present perfect, and if the future perfect is easier than questions using an auxiliary, something is apparently going wrong somewhere, given that, in the first test, 62% of students had difficulty with the future perfect question and 54% and 62% with the phrasal verb questions. Although it is difficult to say what the exact cause of these phenomena is, it may be attributable to the fact that more emphasis may have been placed on what is considered *difficult* in previous teaching/learning situations and so the students have *fixed* these items better. The similarity or difference between English and Portuguese in some of these structures, such as the future perfect, may explain that what is difficult for other students is not necessarily so for Portuguese students because of similarities between the languages and vice versa with other structures such as the present perfect which is not used in the same way in the two languages.

It is even possible that the idea of what is *difficult* for students to learn is in fact incorrect. McDonough (1980:311) says

> "psychologists have objected that there is no reason to assume that linguistic complexity is itself a cause of learning difficulty because many constructions that appear complex in terms of counts of elements or underlying rules are used by native speakers with no hesitation or greater difficulty in execution than apparently simpler ones, in appropriate context …. This is not to deny that constructions do differ in complexity and learnability, rather it is to claim that the only measure of learnability is actual learning and not predictions derived from linguistic description alone."

Nevertheless, in terms of course design it does indicate that certain "basic" items cannot be ignored if 25% of the students cannot cope with them adequately, nor should we assume that time has to be spent on teaching lists of phrasal verbs when the students are, in the majority, able to cope with the more common ones adequately. Indeed, work done on corpus studies for the COBUILD project suggests that six common verbs account for nearly 30% of the 'most important' phrasal verbs as mentioned earlier. The use of corpus studies to decide many such questions for course and materials design is becoming increasingly important (see Wichmann, Figelstone, McEnery, Knowles (eds. 1997), Biber, Conrad and Reppen (1998), McCarthy and Carter (1994), Stubbs (1996))

4.1.5 Test Results 1995/96

The style of the test continued as before with adjustments being made in the light of the previous year's test and with a change in the topics for the paragraph. The topics for this test were:

i) The most important discovery this century.

ii) The changes we will see next century.

Only two topics were offered, because when three topics had been offered, one was invariably largely ignored. These topics required the students to use either the past tense as in (i) or the future as in (ii).

The fifty multiple choice questions gave the following error percentages:

Table 4.1 Analysis of 1995/96 Test Results by Item.

Test Item	Percentage Error	Test Item	Percentage Error
Pronouns	33%	Conditional	52%
Present Perfect	40%	wish + past perfect	37%

Question inversion	24%	Adjective + enough	27%
Reciprocal Verbs	25%	Future Perfect	17%
Present Perfect	22%	Subjunctive	91%
Telling the Time	21%	Passive	24%
Present Perfect	11%	Pronunciation	81%
Simple Past	16%	Modal verb	60%
Comparative Adjective	17%	Pronunciation Alphabet	47%
Time Clause	25%	Conditional (advice)	32%
First Conditional	34%	Reciprocal Pairs	84%
Superlative	15%	Relative Pronoun	25%
Advice (had better)	94%	Preposition	83%
Direct Object (lack of)	74%	Phrasal Verb	50%
Modal	21%	Conjunction	36%
Second Conditional	47%	Vocabulary	54%
Infinitive	56%	Vocabulary	84%
Indirect Question	36%	Vocabulary	67%
Modal (past)	32%	Possessive Pronoun	37%
Future Continuous	73%	Preposition	61%
Irregular Verbs	74%	Vocabulary	15%
Past Continuous	18%	Vocabulary	32%
Past Perfect	46%	Vocabulary	49%
Past Tense	39%	Reciprocal Pairs	59%
Phrasal Verb	67%	Reciprocal Pairs	44%

These results continue to demonstrate the difficulties that the majority of the students have with specific vocabulary, prepositions, phrasal verbs, the conditional, giving advice, infinitive (as opposed to gerund), modals, and pronunciation recognition. Perhaps more surprising on this test is that one of the questions on the present perfect proved to be an inadequate discriminator in that it was correctly handled by almost 90% of the students taking the test. This particular question was the following:

7. She ____________ to England

A. have never been B. has never be C. has never been D. have never be

However, the following present perfect question caused 40% of the students to err:

2 They ___________ their cousin since last year.

A. haven't seen B. aren't seeing C. didn't see D. don't see

Giving advice using "had better" was the question 94% of the students got wrong. The difficulty here is almost certainly the fact that the full form "had better" as opposed to the contracted "'d better" was given. This allowed confusion between "would better" and "had better" in the distractors. The question was the following:

13. You _________ take an aspirin for your headache.

A. had better B. would better C. will better D. have better

4.1.6 Test Results 1996/97

The test for this academic year continued much as before with one or two modifications such as the topics for the paragraphs which were changed to:

i) The worst dangers of pollution.

ii) What the world will be like after the year 2000.

Table 4.2. Analysis of 1996/97 Test Results by Item.

Test Item	Percentage Error	Test Item	Percentage Error
Pronouns	7%	Conditional	38%
Past Tense	35%	wish + past perfect	36%
Question inversion	27%	Adjective + enough	25%
Reciprocal Verbs	29%	Future Perfect	22%
Numbers in Words	78%	Subjunctive	89%
Telling the Time	38%	Numbers	29%
Present Perfect	58%	Pronunciation	87%
Simple Past	25%	Modal verb	50%
Comparative Adjective	31%	Pronunciation Alphabet	53%
Time Clause	23%	Conditional (advice)	32%
First Conditional	33%	Reciprocal Pairs	67%
Spelling	19%	Graeco-Latin Plural	69%
Advice (had better)	89%	Preposition	92%
Direct Object (lack of)	76%	Phrasal Verb	48%
Possessive Pronoun	15%	Conjunction	28%
Second Conditional	43%	Vocabulary	52%
Infinitive	50%	Vocabulary	92%
Indirect Question	58%	Vocabulary	74%
Modal (past)	26%	Possessive Pronoun	39%
Future Continuous	70%	Preposition	64%
Irregular Verbs	72%	Vocabulary	53%
Conjunction	31%	Vocabulary	24%
Past Perfect	44%	Vocabulary	46%
Comparative	35%	Reciprocal Pairs	64%
Phrasal Verb	56%	Reciprocal Pairs	43%

Many of the results on this test continue to confirm what had been found in previous tests but the inclusion of new items, such as numbers in words, tested competence in other areas which pertain ever more closely to the students' future studies. These most likely were not taught at all in school. The question on numbers in words includes the different use of the comma and full stop between Portuguese and English. The comma represents a decimal point and the full stop division into thousands in Portuguese and vice versa in English.

The following question caused 74% of the students difficulty:

5. The number 1,711 reads ______________

A. one comma seven hundred and eleven

B. one point seven double one

C. one thousand, seven hundred and eleven

D. one comma seven double one

Similarly, the Graeco-Latin plurals were tested and showed that the students had difficulty here too. Once again this is probably because these plurals had not been specifically taught.

The following question caused 69% of the students difficulty:

37. Individual teachers may use different ________ when marking the test.

A. Criterions B. criteria C. criteriae D. criterii

4.1.7.Test Results 1997/98

A change was made in the Preliminary Test in the academic year 1997/8. It was decided that, although the students who were allowed not to take the course undoubtedly had a good command of English, there was little proof that these same students could indeed manage the specific English being taught on the course and could show adequate comprehension of a science text. Therefore, the Preliminary Test was written in a similar way to the final examination for the *Ano Comum* and contained forty multiple choice questions on the same type of items tested at the end of the year together with a text for comprehension.

The results of this test showed overwhelmingly that the students who were exempted from further study either did better or at least as well on the comprehension passage as on the multiple choice questions. 75% did better and 25% the same, of those students who obtained a grade of fifteen or more on the Preliminary Test. The students at the bottom of the scale (those scoring seven or less) showed the exact opposite of the above, that is 75% did worse on the comprehension than on the multiple choice questions and 25% achieved the same score.

Table 4.3. Analysis of 1997/98 Test Results by Item.

Test Item	**Percentage Error**	**Test Item**	**Percentage Error**
Punctuation	6%	Pronunciation	93%
Fahrenheit Scale	47%	Comparative	46%
Numbers in Words	71%	Passive	28%
Conjunctions	55%	Modal	53%
Passive	35%	Past Tense	54%
Fractions	35%	False Friends	90%
Formulae	44%	Pronunciation Alphabet	22%
Dates	35%	Question Form	20%
Conjunction	55%	Pronouns	31%

(contrast)			
Conjunction (reason)	15%	Graeco-Latin Plural	79%
Conjunction (reason)	48%	Relative Pronoun	17%
Conjunction (contrast)	23%	Vocabulary	77%
Conjunction (cause and effect)	32%	Vocabulary	75%
Pronunciation	35%	Conditional	26%
Vocabulary	43%	Metric/Imperial Equivalence	69%
Conjunctions (cause and effect)	39%	Reciprocal Pairs	67%
Superlative	52%	Conjunction (contrast)	17%
Adjectives	71%	Translation (false friends)	51%
Pronunciation	25%	Vocabulary	85%
Comparative	80%	Conjunction	28%

Other changes can be noted on this test. The pronunciation question on the alphabet only caused 22% of the students to stumble compared with 53% in 1996/7.

There is some evidence to suggest that after so many years administering this test and the course itself there is a change taking place within the academic community. Teachers from outside the university have asked for copies of the Preliminary Test in order to prepare students to take it, an idea which is contrary to the whole philosophy of the test, which was to find those students whose English was already of a high level and who could already cope competently with the demands made upon them by their bibliographies in English in their subjects areas. This shows the detrimental competitive nature of educational systems which can lead to emphasis often only being given to the grade obtained, rather than to ability or performance. A change of emphasis would focus on cognitive knowledge and would preclude specific ‘teaching for the test’ to be given for a certain test score to be achieved.

One other area that has obviously continued not to be given much stress in school curricula is numbers[2]. As was the case in 1996 with 74% of students mentioned earlier, *71%* of the students in 1997 could not handle the correspondence between a number and its form in words.

Of almost equal difficulty (*69%)* was metric to imperial equivalences, in this case recognising the nearest equivalent to 100 yards in metres. This, as was mentioned earlier, may have nothing at all to do with the language involved but be much more a question of cultural knowledge, recognising the difference between measurement systems in different countries. Although students studying in the *Ano Comum* were not expected to learn the equivalent measurements and conversions of metric to imperial measurements and vice versa, because this information is readily available in a good dictionary, students were expected to have some idea of the *relative* sizes so that logical assumptions could be drawn. To take an example: If a student were faced with the sentence *Scientists can calculate the distance of the earth to the moon to within six inches*. The student should recognise that six inches is not the distance from the earth to the moon as this is equivalent to a distance of approximately 15 cm. and, therefore, this sentence must rather be a discussion of accuracy of measurement and not the actual measurement of the distance mentioned.

[2] A similar test was tried on the students in the fourth year of their teaching degrees which showed that these future teachers also had difficulty with numbers and were often unaware of the contrasts between the English and Portuguese use of the comma and point in numbers. It is not surprising, therefore, that those students who had only studied English in school should find this item difficult, a situation which is unlikely to change significantly in the near future.

Similarly, *80%* of the students could not distinguish between the relative sizes of a British billion and an American billion, (10^{12} and 10^9 respectively). Students need to have the ability to question such items and not merely to assume that the same thing is meant by all those using the same word, in this case the word *billion*. However, it is arguable that this item would in fact cause any difficulty for these undergraduates because in Portuguese *um bilião* stands for a thousand million exactly like the American measurement and the aberration here is the British measurement of a million million which may be on the point of fading out of use[3].

Other frequent items like the adjective *wide* caused considerable difficulty (71% of students got this item wrong). This item appears in the frequency list 1515 times in 1876 articles and in both the physics and chemistry corpora studied here, which would suggest that this adjective is essential for undergraduates and has not been learned by almost three quarters of those entering the university.

False friends (*eventually - eventualmente*) and specific vocabulary (*clerical work* associated with offices not the clergy) items were the worst overall items causing 90% and 85% respectively of the students to make errors. Graeco-Latin plurals also caused considerable error (*79%)* although these are often seen to be significant in scientific writing (see later results *5.1.6 Plurals from Latin and Greek* for a further discussion of this).

Conditional sentences and question forms did not appear to cause undue difficulty for most of the students who took this test with *26%* and *20%* making errors on these items. Some linking devices and deictic pronouns caused more difficulty than others; this could be attributed to

[3] Recently (BBC World Business Report 8/4/99) even British use (certainly in economics) has tended to favour the American thousand million. The answer, aired on the BBC World Business Report programme, to an e-mail inquiry from a viewer confirmed that the BBC were in fact using the American billion in their reporting.

their being relatively unusual. Only *17%* of the students had difficulty with *these* but *55%* had difficulty with *thereby* and *whereas*.

It cannot be taken for granted that the students do not know some items included in 'advanced' grammar books, nor can it be assumed that the basic structures have been learned soundly. What these results do suggest however is that it would be of more use, for example, to include a list of irregular verbs for the students to take away and learn rather than a list of phrasal verbs. In other words, the students' proficiency profile must guide the choice made about what language should be taught in the first year.

In addition, corpus analysis (on the LOB and the Brown corpora) for a syllabus for use with students in Germany by Mindt (and Tesch 1990) (1997:40-50) has shown that with careful grading students can learn a higher percentage of the most frequent and therefore important irregular verbs (apart from *be, have* and *do*) even if they stop learning after a short period of time. They contrast their list with alphabetical lists of the kind normally presented for students to learn and show that after learning five of the verbs (*say, make, go, take, come)* on their list the students will be "familiar with 27 per cent of all irregular verb forms of English". The corresponding figure for the alphabetical list is 3.6 per cent (*beat, become, begin, bet, bite*). After learning ten of the verbs on their list (*say, make, go, take, come, see, know, get, give, find)* the student "has mastered 45.6 per cent of the verb patterns of irregular verbs". The combination of what the student already knows and the results of corpora analyses like the one described here can provide a much more reasoned syllabus that aims to make maximum use of the students' study time. Mindt (1997)and Tesch (1990) like Renouff (1984) also find that their corpora analyses show the discrepancies that occur between actual language use and what is presented in coursebooks. (Tesch studied *some* and *any*, Mindt modal *will* and *would* and Renouff *see*).

These are taken up in more detail later in Chapter 7. Similar research on the corpora shows that certain irregular verbs are more suitable for the students of science and technology (see 6.3) and would cover the difference in the transition from secondary to tertiary education for undergraduates.

4.2 Needs Analysis by University Department

In order to meet the demand for English language teaching for the *Ano Comum* in accordance with what the departments whose students are involved in the foundation year feel is necessary, a simple needs analysis was requested from colleagues in other university departments. Colleagues were asked to indicate their views as to why they thought it necessary for the students to study English language for their courses. This was the first stage in the needs analysis, the results of which are given below.

Colleagues in other departments asked that the students be able to speak fluently, read fluently and listen effectively. Below is a representative sample of what our colleagues in other departments perceive as the English needs of their students taken from the replies received to the initial simple needs analysis which consisted of a letter from the English Area Co-ordinator to the different departments involved in the *Ano Comum* asking for their comments. Nine replies were received relating to different courses included in the foundation year. Not all departments replied. The replies received were from the co-ordinators responsible for the following courses: Mathematics (Teaching of); Applied Mathematics and Computation; Geological Engineering; Biology; Biology and Geology (Teaching of); Ceramics and Glass Engineering; Materials Engineering; Engineering and Industrial Management.

- *Inglês coloquial*
- *Inglês técnico (Matemática, Informática)*
- *Tradução corrente de textos científicos, uma vez que grande parte da bibliografia que é indicada encontra-se em inglês.*
- *O domínio da língua inglesa é essencial na formação dos nossos engenheiros. Consideramos que a disciplina de inglês dever ter como objectivo, entre outros, a preparação dos alunos para as seguintes situações: - Dialogo fluente com profissionais da área das engenharias nos contactos internacionais; - Leitura de manuais, livros e catálogos técnicos, etc.*
- *Competência na língua de modo a permitir os alunos: -"enfrentarem-se" com textos técnicos -conseguirem dialogar.*
- *Os alunos devem dominar os termos técnicos das áreas de Turismo, Economia e Gestão. Porque temos frequentemente docentes estrangeiros, surge já durante o curso a necessidade dos alunos de comunicar em inglês nas aulas.*
- *Leitura, interpretação e conversação nas vertentes, ciências naturais (geológica), física e química*

The third point given in this needs analysis, that of being able to deal with bibliographies that were largely in English, was taken as particularly pertinent to all of the students involved in the foundation year. The aims put forward by these replies are obviously ideal and would stand the students in good stead for their futures both in terms of further study (possibly abroad) and in their careers. However, reaching the ideal is limited by a number of constraints which few of these same respondents cared to acknowledge in the curriculum they had created.

4.3 Constraints

There were three major constraints on the teaching of the students in the first year of the university, which are described below.

Several administrative decisions were taken about the size and number of groups to be taught in the first year. These constraints were common to all of the subject areas included in the first year course as far as was possible (some differences existed for practical classes in physics, chemistry and computer studies).

The first decision was about the size of the groups. There were to be at least twenty-five groups of forty-five students each. This soon became at least twenty-eight groups because of the students who were repeating the disciplines. Different systems were adopted to accommodate these students, from separate classes being created to distance-learning courses on computer being provided in 1998-99 for those students repeating the year[4].

Secondly, all of the English language students were to receive two hours of language teaching per week as theoretical-practical classes. Previously some courses, such as Tourism and Management, had had far more English over two or three years, so that this represented a considerable reduction in time for some courses. In essence, the non-pure or physical sciences received less language instruction time.

[4] The distance-learning courses are run by the University of Aveiro through the Internet and are open to all students who must enrol and obtain a login to be able to access the relevant material. Working students are obviously targeted by this scheme of work besides those repeating disciplines. Other systems, like course tutor support schemes have been instigated to help students to plan their studies and cope with the psychological strain of the move to a university environment which studies (Tavares *et al.* 1996) mentioned earlier had shown to be a reason for lack of success in the first year.

The third constraint was that the students on the course were not homogeneous, either in terms of the number of years they had studied English (see earlier Chapter 3, *3.2.1 The Students' Level of English*) or in terms of the subject which they had chosen to pursue for their degree. The latter was again a change, as in the past some courses had had English as a discipline and these groups had been homogeneous in terms of their degree subject, for example, in Management and Electronics. This meant that subject specific coursebooks such as *English for Electronics* could be used for the appropriate group but that now this was no longer possible.

The overall percentages of students with more than five years of study of English has gradually been creeping up from the initial 67% with five years or fewer of study and 33% with more than five years of study, to a complete reversal of this situation in the academic years 1997/98 with 32% with five years or fewer of study and 68% with more than five years of study as mentioned earlier in section 3.2.1. This trend could be accounted for in a number of ways; the information about the foundation course year may well have become better known among those students who were hoping to avoid language study by opting to take up other courses, in other places, although this seems unlikely; it may be that the a general trend in the secondary schools to teach more English has filtered through to higher education; or it may simply be a reflection of the fact that all the students entering the science and technology courses have now decided to try their luck at the preliminary examination and so these figures are a much more complete representation of the student intake. A trend towards more years of English in secondary schools is a positive development as it will only help students at university to come to grips with studying through English language textbooks.

Nevertheless, the aim of the foundation year to give all of the student intake into science and technology courses a sound basis on which to build their studies in later years and to maximise their learning potential appears to be being subverted. It seems to have become another hurdle for students to jump so that strategies like evasion, gamesmanship and cramming are encouraged. It is interesting to speculate whether students would voluntarily study English if their departments recommended but did not insist on it. It is possible that quite sophisticated translation services would be set up on the periphery of the university if certain texts were seen to be essential and were only available in English. Students traditionally claim that they have too little time for their studies and this would be one means to avoid having to study English at all. This line of thought leads directly on to the subject of motivation in the students once again. If the subject is only seen as a hurdle to be jumped, the students' focus will necessarily be on test results rather than on achieving their maximum potential. This means that the course will have to subversively achieve this end against the wishes of those students who take this position. Students who are at the bottom of the scale with few years of English may also be discouraged from the outset or may decide to take further courses in English outside the university. One means of persuading the students to focus on learning more English is first and foremost by encouraging them to attend most of the classes. This has been successfully achieved by linking evaluation with class attendance. Students are offered different evaluation schemes depending on whether they attend a minimum of two thirds of the classes given. This is quite a normal procedure in the university for practical classes and those that use continuous evaluation of students so there is no real difficulty in applying continuous assessment to these classes. One other effect of this scheme is that the need for hundreds of individual interviews, traditionally taking twenty minutes each with all the practical administrative difficulties that this implies in a short examination period, is avoided.

The other means to try to engage the students is to make the work as appropriate as possible for them. This can be approached through the analysis of the language of science and technology and the language requirements of the books in English on the bibliographies which they have to deal with. Chapter five will take up this aspect of the students' language needs.

Chapter 5 Scientific English for Undergraduate Learners

Analysis of Results

Chapter 5

Scientific English for Undergraduate Learners

5.1 *Analysis of Results*

This Chapter will examine the English of science and technology the undergraduate students will need to cope with from the baseline corpus and the specific corpora developed from the undergraduate textbooks in physics and chemistry. The features, outlined in Chapter 3, that are seen as being salient in these texts and for the profile of the undergraduate analysed in Chapter 4 will also be discussed in the light of the data obtained. The results will be compared and contrasted with Biber's findings for academic prose and the significant features will be highlighted. A more detailed analysis will be made of the sub-corpora from the Physics and Chemistry textbooks and consideration will be given to the role of mathematics in these texts.

5.1.1 The Baseline Corpus

The Grolier Multimedia Encyclopaedia provided the following words from approximately 31,000 texts; a total of approximately 150,000 *different* words from a corpus of several million words. The figures given after the word refer to the number of articles containing the word, the *range* of the word, and the frequency of the word, respectively. Thus, it can be seen that the actual number of words contained in the

Field Code Changed

encyclopaedia is a multiple of the 150,000 different words listed. Wilks et al (1996:189) describe the Grolier as having 10 million words.[1]

A 'law of diminishing lexical returns' has been put forward which says that as the size of a corpus increases, the average incidence of word types decreases. The one million word LOB corpus yields about 50,000 different word types and the Birmingham main corpus of 7.3 million words mentioned earlier contained only 132,000 different word types.

Only frequencies and ranges of 100 or more have been included and all proper names have been excluded where possible. However, there are instances where a proper name and a noun are not immediately distinguishable, for example, hill and Hill. Those words considered equivalent or close to a similar word in Portuguese with the same meaning, that is cognates, have been marked with an asterisk. The Grolier encyclopaedia does not include a number of words on the grounds that they are 'too common'[2]. A sample with respect to the letter 'a' is : *about, across, after, again, all, along, also, although, among, an, and, any, are, around, as,* and *at.* Two other 50,000 word corpora, taken from the Chemistry and Physics textbooks *Chemistry* (Chapters 2, 3, 4 and 5 Pages 37-219) and *Physics for Scientists and Engineers with Modern Physics* (Chapters 1-5, Pages 15-216)[3] in the students' bibliographies, have also been examined to produce frequency counts of the same kind and then compared with the multimedia encyclopaedia in order to examine the similarities and differences between them.

---

[1] Minugh (1997:79) in his use of Newspaper CD-ROMs for teaching at the University of Stockholm also mentions this lack of information about the actual size of published CD-ROM material to be one of the disadvantages, although he hopes that the companies involved in producing them can be persuaded to incorporate this information in the future. He further laments the fact that "the most frequent words are classified as "noise" and cannot be searched for."

[2] These are sometimes referred to as *structure* words of which it is estimated that there are about two hundred (see Bowen, Madsen and Hilferty (1985:194)) or *form words* (see earlier 3.1.7 Bright and McGregor 1970)) or *function words* (see Biber et al. (1998:29)). However, it is clear that the Grolier includes *any* very common word and not only words such as articles, prepositions, and pronouns.

[3] Sinclair recommends that the same areas of books are not studied in case they demonstrate only one specific variety of English, for example the English associated with introductions or first chapters. See later *bibliography* for a similar justification of Sinclair's hypothesis.

When an item is to be found in these texts the word is marked with the text which contains it. For example, *able* occurs in both the chemistry and the physics corpora and so this word is marked with the letters C and P after the range and frequency figures given in the multimedia encyclopaedia listing.

Some interesting anomalies occur with the words listed including those given in the Grolier encyclopaedia as 'too common'. Under the letter 'a' *across* and *around* do not appear in the 50,000 words of the chemistry text at all, under 'b' *be, became, become, been, begun, being,* and *bibliography* do not appear in either of the corpora and only *by* appears in both corpora. It is not surprising that the word *bibliography* should not appear in these corpora as the bibliographies were not included in the corpora as they traditionally occupy a position at the end of the textbook which was not taken to be representative of either physics or chemistry texts *per se*.

The Grolier's idiosyncratic pronunciation scheme is also to be found in the list. Each of these items is identified by the abbreviation (*pronun),* for pronunciation, immediately after the entry.

Table 5.1 Grolier Frequency and Range List

Field Code Changed

able	827/1066CP
ability	753/996CP
abilities	111/127
*abnormal	116/173
aboard	131/204
abolished	177/207
abolition	110/127
above	192/1615CP
abroad	293/357
absence	262/296P
absent	123/136P
*absolute	308/486CP
*absorb	147/182CP
*absorbed	308/408
*absorption	191/328
*abstract	328/554
*abundance	137/155C
*abundant	405/535C
*abuse	126/221
*academic	273/389
*academy	756/1120
*accelerated	150/175P
accept	282/320C
acceptance	219/247
accepted	649/768P
*access	283/378
*accessible	109/116
*accident	137/169
*acclaim	174/182
*acclaimed	195/229
*accommodate	111/118
*accompanied	382/424P
accompanying	112/123
*accomplished	339/376CP
*according	1117/1392CP
account	617/824CP
accounts	311/398
*accumulated	116/130
*accumulation	113/129
accuracy	185/256P
accurate	248/309CP
accurately	130/150CP
accused	214/238
achieve	411/468C
achieved	972/1196
achievement	320/384
achievements	246/278CP
achieving	118/120
*acid	503/1213C
*acids	227/483C
acknowledged	129/135
acquire	101/111P
acquired	425/519P
acquisition	111/135
acres	167/242P
*act	1197/2198CP
acted	166/186
acting	327/445C
*action	981/1452C
*actions	297/379
*active	1094/1384CP
*actively	130/135
*activities	826/1119
*activity	937/1356CP
*actor	300/502
*actors	124/236
*actress	168/253
*acts	561/786CP
actual	388/449CP
actually	537/625CP
acute	165/232
ad	1039/1563CP
*adaptation	163/186P
*adaptations	120/151
*adapted	413/517CP
add	140/174CP
added	798/1015CP
*addition	1334/1818CP
*additional	518/603CP
address	105/143
addressed	101/111
*adequate	196/231
adjacent	310/368P
*administered	250/320
*administration	876/1271
*administrative	452/602
*administrator	151/162
*admiral	122/174
admired	285/310
*admission	122/162
*admitted	185/206
*adopt	110/118
*adopted	802/1001
*adoption	143/168
*adult	544/802
*adults	289/383
advance	277/384
advanced	632/826P
advances	289/376P
advantage	283/343CP
advantages	142/165
advent	183/193
adventure	114/140
adventures	170/204
advertising	111/216
advice	116/133
advisor	160/170
advocate	249/265
advocated	267/293
aesthetic	224/300
affair	175/205
affairs	546/689
*affect	236/282CP
*affected	367/450C
*affecting	123/136P
*affects	158/184P
*affiliated	116/120
afterward	128/134
against	2860/4578C
age	2141/3573P
aged	106/139
*agencies	241/378
*agency	364/600
*agent	280/386
*agents	255/425C
ages	847/1172
*aggressive	174/213
aging	101/167
ago	528/919P
agree	134/146P
agreed	301/378
agreement	307/426CP
agreements	108/149
*agricultural	892/1501
*agriculture	817/1584P
ah (pronun)	405/415
ahead	104/117P
ahl (pronun)	106/107
ahn (pronun)	167/169
ahr (pronun)	101/102
aid	681/1014
aided	274/317
aids	146/277C
aim	155/175
aimed	197/214
aims	111/133
air	1581/3580CP
aircraft	304/885P

airfields	177/186
airplane	115/182CP
airport	153/206P
*album	118/191
*alcohol	221/440C
*algae	152/340
alive	132/140
alleged	123/134
*allegorical	101/132
*alliance	316/472
allied	407/685
allies	282/465
allow	362/411CP
allowed	589/708C
allowing	231/253CP
allows	280/319CP
alluvial	108/149
almost	1735/2419CP
alone	421/471C
*alphabet	101/247P
already	546/653CP
alter	111/123CP
*altered	198/223C
alternate	126/137
alternating	145/190
*alternative	230/276CP
*altitude	239/359CP
*altitudes	116/143C
aluminum	287/495CP
always	783/959CP
amateur	146/227C
*ambassador	149/181
*ambitious	157/172
amendment	251/691
amino	103/286
amount	657/1036CP
amounts	490/690CP
*analysis	572/894CP
*analytical	106/129C
analyzed	124/144C
*anatomy	199/306
ancestor	152/173
ancestors	154/192
*ancestral	105/136
ancestry	120/144
ancient	1900/3090
angle	190/357P
angles	165/264P
*angular	126/188
*animal	1002/1793CP
*animals	1122/2441CP
*annexed	186/207
announced	233/277
*annual	1033/1512C
*annually	410/474C
another	2209/3383C
*anthology	113/134
*anthropology	102/218
*anti	448/599
*antibiotics	119/197
*anticipated	133/143
*antiquity	124/136
anything	114/121
apart	320/368CP
apparatus	117/152CP
*apparent	368/431
apparently	429/488P
appeal	210/261
appeals	106/134
appear	697/872CP
*appearance	647/777C
*appearances	102/107
appeared	1049/1335C
*appearing	125/129C
appears	553/643C
*application	362/470C
*applications	359/553CP
applied	1011/1315CP
applies	103/107CP
apply	159/183CP
applying	147/158CP
appointed	920/1078
appointment	168/187
approach	551/783CP
approaches	205/265CP
appropriate	327/380CP
*approval	150/194
*approved	246/296
*approximately	842/1073P
Apr	1651/1796
April	508/678
*aquatic	170/259
arc	170/314P
arch	112/209
*archaeological	263/349
*archaeologists	117/152
*archaeology	189/368
*archbishop	147/201
*architect	508/761
*architects	236/382
*architectural	392/607
*architecture	1028/2856
*archive	978/2752
*area	2756/4964CP
*areas	1820/3630CP
argued	327/394
argument	111/148P
arid	201/298
arise	183/205P
*aristocracy	123/145
*aristocratic	157/184
arm	215/270P
armed	311/413
armies	249/454
arms	362/635
army	1255/2496P
arose	276/318
aroused	154/159
arranged	361/410C
arrangement	269/335C
arrangements	180/217CP
array	155/181
arrest	112/139
arrested	190/204
arrival	233/261
arrived	338/420CP
*art	2573/7631CP
article	183/278
articles	314/421
*artifacts	155/192
*artificial	456/672
*artisans	117/141
*artist	683/
*artists	665/1236
*arts	1051/1704
ash	136/185
aside	152/169
asked	152/162CP
*aspect	224/255P
*aspects	498/617CP
*assassinated	183/217
*assassination	177/234
assembled	106/122C
*assembly	529/862
asserted	116/134
assigned	211/230C
assist	114/121P
assistance	237/326P
assistant	259/289P
assisted	157/165
*associate	182/198

Field Code Changed

*associated	1197/1526CP
*association	763/1080
*associations	150/193
assume	142/151CP
assumed	548/646CP
assumption	123/154CP
*astronomer	185/253P
*astronomical	193/310P
*astronomy	302/633P
athletic	104/151
*atlas	171/281
*atmosphere	469/977C
*atmospheric	241/377C
*atom	214/594CP
*atomic	440/1037CP
*atoms	342/971CP
attached	461/606CP
attack	526/754C
attacked	339/403
attacking	109/124
attacks	328/425
attain	126/130CP
attained	172/187
attempt	703/847C
attempted	560/658P
attempting	150/157P
attempts	613/723P
attend	113/137
attended	327/343
*attention	689/842
*attitude	116/155
*attitudes	177/230
*attorney	209/280
*attract	205/242C
*attracted	397/436
*attraction	161/187C
*attractions	101/109C
*attractive	173/204C
*attributed	271/296
audience	266/394
audiences	161/227
Aug	1710/1875
August	562/799CP
*author	732/909
*authorities	310/366
*authority	646/1000
*authorized	147/169
*authors	186/254
*autobiographical	217/264
*autobiography	496/562
*automatic	133/219
*automatically	115/139
automobile	289/420CP
automobiles	186/237C
*autonomous	190/252
autumn	153/199
availability	100/108
available	825/1150CP
avant	176/217
average	975/1647CP
averages	226/250
averaging	116/130
avoid	228/266CP
award	468/655
awarded	411/488
awards	214/291
aware	126/138CP
awareness	138/155
away	670/825CP
axis	230/454CP
ay (pronun)	250/256

The lists obtained for the rest of the alphabet are given in Appendix B with these Portuguese cognates, as defined above, removed. The words not included, on the grounds that they are too common, have been shown in italics at the end of the respective list for each letter of the alphabet.

5.1.2 Range and Context

The search facilities mentioned can provide some relevant data on the range of an item. If a comparison is made between the adjectives *long*, *wide*, *broad*, *deep*, *tall* and *high* and the nouns *length*, *width*, *breadth*, *depth* and *height* it is possible to demonstrate the considerable differences in the topic areas they are most frequently to be found in. Here is the data for the most frequent uses of these adjectives and their respective nouns in the CD-ROM encyclopaedia:

Table 5.2 Frequency and Range Results for Abstract Nouns and Adjectives.

Length 1133 Articles - 1573 Occurrences	**Long** 4041 Articles - 6701 Occurrences
20 – measurement 17 - metric system 13 – lens 11 – fish	46 - plant 45 - mammal 40 - bird 24 - flower

Width 175 Articles - 216 Occurrences	**Wide** 1515 Articles - 1876 Occurrences
7 – dendrochronology 5 - river and stream	10 - sound recording and reproduction 9 - ice hockey 8 - television 7 - mammal

Breadth 38 Articles – 38 Occurrences	**Broad** 651 Articles - 802 Occurrences
1 – anthropology 1 – dimension (mathematics)	11 - plant 7 - antibiotics 5 - Antarctica 5 - mammals

Depth 350 Articles – 496 Occurrences	**Deep** 829 Articles - 1177 Occurrences
13 – depth charge 10 – perception 8 – water wave 7 - gulf and bay	36 - deep sea life 13 - ocean and sea 13 - syntax 9 – geosyncline

Height 563 Articles – 745 Occurrences	**High** 3145 Articles - 5679 Occurrences	**Tall** 365 Articles - 485 Occurrences
14 – plant	49 - secondary education	25 - plant
14 – tree	30 - middle schools & junior high school	11 - tree
10 – atmosphere	22 - sound recording & reproduction	10 - flower
10 – statistics	21 - nuclear reactor	

Field Code Changed

At least two of these areas will be unknown to the average humanities trained teacher: dendrochronology and geosyncline. Dendrochronolgy is the science of using tree rings to date structures and events or to reconstruct past environmental conditions. A geosyncline is a large, usually elongate depression in the crust of the earth which during subsidence has accumulated very great thicknesses (thousands of meters) of sedimentary and usually also volcanic, rocks. The latter is therefore a part of geology.

From these results it is possible to conclude that the presentation and practice of these adjectives and nouns in teaching materials would have to be through different contexts or settings if a ‘natural’ use of such items was to be given. *Length* should, from this perspective, be presented in a physics context for example, whilst *long* would be more ‘naturally’ presented in a context of biology. The noun *breadth* has such a low frequency that it could be ignored but *broad* should be in a biology context while *wide* seems more ‘at home’ in an electronics context. The results for *tall* and *high* as ‘corresponding’ adjectives for *height* once again show that different scientific settings are used with each; biology for *tall* and *height* and physics and electronics for *high*, although *height* could also be included in physics, chemistry or maths.

It would be possible for any item or sets of items to be examined in this way to contextualise the setting of any of the vocabulary or syntax that has been identified for the syllabus. This would then show the use and meaning of these items in scientific settings, which as was discussed earlier in Chapter 2 , does not necessarily correspond to the use or meaning in everyday contexts.

5.1.3 American Words and Spellings.

It is interesting to note that several American words and spellings (afterward, aging, aluminum, airplane, analyzed, attorney, authorized and automobile) exist alongside some obviously British words and spellings (advisor, aesthetic, autumn, car). A possible explanation for this is the lack of a rigorous editorial practice as mentioned below (Burnard 1992). Another explanation could be that the author of *Chemistry*, Raymond Chang, is described as having been born in Hong Kong, growing up in Shanghai and Hong Kong, studying in London and at Yale and teaching in America. Perhaps these inconsistencies reflect his background in both British and American academia. It is surprising however, that the editors did not impose consistency on the finished work.

Students appear to have limited awareness of the differences between British and American English (not to mention Australian, South African, etc., of which only those few students who had studied in these places would be aware). From the books in the students' bibliographies American English dominates and so familiarity with this form should be encouraged. This may not necessarily be by contrasting it with British English although some students may ask for such a contrast to be made explicitly because of their previous learning.

This mixing of American and British English preferences can also be seen in the earlier Thorndike and Lorge's (1944) word list for teachers (in America) which is meant to have an American bias but where words such as the British *trousers* are more frequent than the American *pants*.

5.1.4 Abbreviations.

Three letter abbreviations for the months of the year appear much more often than the full words for these in the frequency lists. Such findings would obviously not imply that the students should *only* be taught these abbreviations but they do imply that these features are the natural ones to be included in the students' study materials. An obvious application here is in Tables, Figures or Graphs where abbreviations of this type are to be expected.

Field Code Changed

5.1.5 Pronunciation Conventions.

The Grolier encyclopaedia uses such sound groups as 'ah', 'ahl', 'ahn' and 'ahr' to illustrate the pronunciation of names and never uses the phonetic alphabet. This finding seem to run counter to the teaching of phonetics for referencing skills and the use of dictionaries, in particular with the international phonetic alphabet being used to indicate pronunciation in most dictionaries.[4] It is also doubtful if these pronunciation aids would really help Portuguese speakers as they would not be pronounced in the way that this encyclopaedia obviously imagines. In other words, this feature reinforces the idea that the encyclopaedia was designed with a native speaker reader in mind.

The problem of whether to teach the phonetic alphabet so that students' can make use of it when using dictionaries revives the perennial problem of how to make the most of the study time available. Dictionaries usually present the phonetic alphabet at the beginning and simply drawing attention to this fact (and conventions of representing stress) allows those students who wish to, to pursue this potentially fruitful area further. It is recognised however that unless some attempt is made to engage the learners with this data most non-linguists will ignore it altogether and thereby lose an opportunity to enhance and enlarge their knowledge of English whenever the need arises.

5.1.5 Plurals from Latin and Greek.

Just as Sinclair (1991:68-9) discovered that one form of a lemma is much more common than others, research into Latin and Greek plural

[4] It is interesting to note that phonetic transcription is not usually included in bilingual and Portuguese dictionaries which may be a reflection of the fact that there is a general idea that Portuguese is a 'phonetic' language and that, therefore, transcription is not necessary. This is an increasingly debatable proposition especially with new spelling conventions coming into being.

forms in the Grolier encyclopaedia reveals that there is usually a substantial difference between the frequency of either the singular or the plural of Latin and Greek root words. Some simply fail to appear at all, *parentheses (10)* is a case in point where the singular *parenthesis* does not appear at all in the corpus. Occasionally there is a regular ending applied to a Latin or Greek root word such as *indexes* together with *indices* and of almost the same frequency (36 and 31 respectively). However, in the case of *formulae* and *formulas* the latter is much more frequent than the former (4 and 88 respectively). There are very few singulars and plurals that appear in almost equal numbers; *nova (10), novae (10) and novas (10)* and *stimulus (146) and stimuli (122),* all the others encountered show marked preference for one or the other form.

Sinclair (1991:67ff) demonstrates that the of singular and plural forms of nouns are not equivalent, by documenting the different patterning of *eye* and *eyes*. He finds that “There is hardly any common environment” between the two word forms and they “do not normally have the capacity to replace each other”. The plural co-occurs with adjectives such as *blue, brown, covetous, manic.* The singular hardly ever refers to the anatomical object, except when talking about injury or handicap. The singular and plural also occur in different sets of fixed phrases (*all eyes will be on, rolling their eyes, turn a blind eye, keep an eye on).* It is this sort of analysis which highlights the fact that lexis and syntax are totally interdependent.

Similarly, in this encyclopaedia, the plural *algae (339)* is found to be more frequent than the singular *alga (28)* and yet *axis (452)* is found to be more frequent than *axes (142).* These findings reflect the fact that the former is being used as a generic term whilst the latter is being used to specify a particular part of a graph. It is essential to be able to make this form of distinction so that these terms can be used in their most likely contexts.

Field Code Changed

Many authors have recognised the importance of Latin and Greek terms in scientific texts but these have tended to be taught in lists of singular and plural forms rather than in specific contexts where either the singular or the plural would be most appropriate. The kind of language manipulation exercise where students produce the singular or plural of a Graeco-Latin word is generally considered unsuccessful as a teaching/learning strategy and would ignore the semantic differences inherent in scientific contexts. For example, in the Physics and Chemistry corpora only two Latin and Greek singular and plural forms exist together. These are: *bacteria* - *bacterium* and *axis* - *axes*. *Formula* - *formulas* are to be found in the Chemistry corpus but no *formulae*. All other Graeco-Latin words are only found either in the singular or in the plural only. For example, *analysis, apparatus, appendix, criterion, data, parentheses.*

The Latin and Greek roots and affixes described by Strevens (1978:193) given earlier in *3.1.8.4* are not found in their entirety in the Physics and Chemistry corpora either. His *cyto, plasma, pyro, ante,* and *post,* are found in neither the physics nor the chemistry corpora and many of the other prefixes like *hydro* and *poly* are in combinations such as *hydrogen* and *polygon* which are cognates with Portuguese. Similarly with his suffixes, *-ite* is most often found in words such as *white* and *write* and *-valent* in *equivalent.* The Grolier does present examples of all of the prefixes and suffixes mentioned by Strevens but once again closer examination shows that the majority of entries for the prefixes and suffixes are not of the type anticipated by him. For example under the prefixes *hydro-* and *anti-* the largest number of entries refer to hydrogen and antiques. It would seem therefore from these results that these items need not be heavily focused upon in the syllabus.

White (1998:275) argues for a distinction to be made between science and technology texts. He (ibid.) says that science texts show preference for non-vernacular Latin/Greek borrowings, whereas technology texts prefer elaborated nominals where all the elements are of vernacular derivation together with acronyms or provisional or 'proto-nouns'. A proto-noun is a word that is now commonly used as a noun but was originally an acronym such as scuba (self-contained underwater breathing apparatus) or laser (light amplification by simulated emission of radiation). However, White (1998:285) claims that the "fact that classical scholarship is no longer so widespread may offer a part explanation as to why Greek/Latin coinings have declined" in science, even though "they remain the norm". The textbooks studied show a combination of both the science and technology features that White found with the Latin/Greek borrowings mentioned above, together with vernacular and proto-nouns such as *scuba*.

5.1.7 Word preferences.

The word *attempt* is found to be approximately five times more prevalent in the list than *try* and the word *change* is ten times more frequent than *alter*. These choices, although apparently arbitrary, may, according to Quirk (1995), indicate formality in the texts and be representative of this particular genre

Bright and McGregor (1970:29) claim that technical texts contain both technical vocabulary and a large number of 'general' words that are "outside simplified English". They further claim that there is "good reason to suppose that anybody who is going to continue his education in English will find them (general words outside simplified English) useful". They list twelve such words: '*absurd, adequate, adjoining, aggression, alert, alternative, amateur, ample, apparatus, apprehensive, automatic, available*'. Of these words only six do in fact occur with significant frequency in the Grolier encyclopaedia (see word list in Table 5.1 above: *adequate, alternative, amateur, apparatus, automatic* and *available*) and of those six

Field Code Changed

only three could be considered sufficiently different from Portuguese to warrant attention (*amateur, apparatus* and *available*). It would seem from this small comparison that it is true that the results found through intuition and those found in reality through empirical research are different. However, it could be that the genre studied here differs considerably from that used by Bright and McGregor about which, incidentally, they give no specific information.

As mentioned earlier, Hoffman (1981), Darian (1981), Weber (1981) and White (1998) all argue that even the same lexical item can (and usually does) take on new meaning when used in a scientific text. White (1998:285) further suggests that the use of acronyms in technological texts is because these are “eminently well equipped to meet the constant need for new lexis to map the ever-unfolding reality of technological development”. The fact that only part of the lexis of science and technology can be seen to be in any way stable suggests that any study such as this will only be representative as long as the textbooks are considered sufficiently up-to-date for use with undergraduates. The textbooks that have been selected for use in this study continue to be in the students’ bibliographies. Despite the fact that newer editions have recently been published, these have not undergone any significant changes.

5.2 Other Features of the Text

The features that must be taken into consideration (already discussed in Chapter 3), as they contribute to understanding and interpreting texts, cannot be seen in the frequency lists of the corpora. These are *typographics, formulae, numbers, equations and tables* and *diagrams and drawings* these will be discussed below with reference to the undergraduate textbooks studied here.

5.2.1 Typographics.

The use of typographics in the texts must be explored because as Kress, Leite Gárcia and Leeuwen (1997:270) say “If humans make and communicate meanings in many modes, then language no longer suffices as the focus of attention for anyone interested in the social making and remaking of meaning.” Therefore they point out that it is “highly problematic to read only linguistically carried meaning”. Modern textbooks, like modern newspapers, are colourful and have many different sizes and styles of type and punctuation together with pictures and drawings and are altogether different from the dense text typology of the past. These changes bring with them the complication of interpretation and discovering meaning referred to by Kress, Leite Gárcia and Leeuwen above and which they (ibid.) believe to be “a normal condition of reading, whether of a textbook page, of any page of a newspaper, or of a television screen” in the modern world.

In Chang’s *Chemistry* technical terms are explicitly marked orthographically the first time they are used. Orthographic marking involves printing the word in boldface within the body of the text. A marked term is accompanied by a definition or rather technically an *elaboration* in the nearby text. Typically, once the term has been elaborated, it is no longer highlighted. In addition, abstracts or ‘asides’ are printed in the left-hand margin in blue type. For example Chang (1991:170-1):

A ***barometer*** is *an instrument that measures atmospheric pressure.* A simple barometer can be constructed ……

See Section 1.7 to review the definition of a newton

In SI units, pressure is measured in *pascals (Pa),* defined as one newton per square meter:

$$pressure = \frac{force}{area}$$

Field Code Changed

This equation gives the SI definition of 1 atm.

$$1 \text{ atm} = 1.01325 \times 10^2 \text{ kPa}$$

In Serway**,** *Physics for Scientists and Engineers with Modern Physics*, the orthographic marking involves printing the word in boldface within the body of the text and the elaboration is indicated by means of italics in the nearby text.

For example in Serway (1992:202) there is the following:

"Effusion

Whereas diffusion is a process by which one gas gradually mixes with another, ***effusion*** is *the process by which a gas under pressure escapes from one compartment of a container to another by passing through a small opening.* Figure 5.20 shows the ….."

White (1998:267) claims that science and technology differ in the way that they define; science locates definitional structures "in a systematised set of taxonomic relationships", while technology "typically acts to identify the functionality of the items". The chemistry text definitions described above do both of these things, the barometer definition is like technology, as defined by White, in that it describes its function, and the newton and atmosphere definitions follow his description of pure science definitions locating these terms in systems and taxonomies.

It is important that these presentational elements are included in the syllabus and that the different representations in pure science and those of technology are expressed in the materials designed for use with the students.

5.2.2 Titles, Subtitles, Summaries and Conclusions

As was mentioned earlier (section *3.1.9 Optical Character Recognition*), the optical scanner or optical character recognition machine (OCR), which in this study was used to assemble the physics and chemistry corpora, can only recognise what is visibly present on the page and that it cannot undertake any kind of editing nor can it distinguish structurally different components of the printed page, such as footnotes and headings, even if these are visually distinct (Burnard 1992). Any corpora searched will not provide these distinguishing features so that a human/visual, graphical/layout analysis of the texts themselves as published can often reveal some other important features of those texts (see later *5.2.4 Diagrams and Drawings).*

Both of these textbooks use a system of keywords to emphasise those items considered important in the text and to allow cross-referencing to these. Similarly, the text itself can be sub-divided into body text, text describing figures, footnotes and asides. As mentioned earlier, the Chemistry text employs the use of marginal notes in blue type and it is only in the chemistry text that extensive use of footnotes is to be found, reaching up to 20% of the total page text on occasions.

The division of text into shorter sections is considered important by Stubbs (1996:84) who suggests that the author is showing by it a particular attitude towards the expected reader. He notes that school textbooks are traditionally divided up into small sections which may suggest that the author expects a limited attention span from readers. Division into smaller sections is also a technique of the Physics and Chemistry textbook authors under study here to improve the readers' interpretation and understanding of their essentially instructional texts and for ease of finding relevant sections of the textbook. Chang (1991:xxii) describes this as 'readability' and says that he tried to provide flexibility

Field Code Changed

for in-class assignments whilst making smooth transitions from topic to topic.

Each Chapter in these textbooks is summarised and there are lists of keywords on the topic of the chapter at the end of every chapter. One of the most distinguishing features of these textbooks is the long section of questions based on the theory presented and exemplified throughout the chapter. It is the sections which exemplify, in different coloured boxes in the body text and the individual sections with 'real world' applications of the theories discussed, and the questions at the end of the chapter which distinguish this genre and make it essentially instructional in nature. These are the sections nevertheless that often display more culturally loaded information with their references to particular American sports and activities which are an attempt by the authors to present the topics through activities that the (American) students can relate to their everyday lives (see *5.3.3 The Physics and Chemistry Sub-Corpora*).

5.2.3 Formulae, Numbers, Equations and Tables.

In science texts the use of formulae, numbers, equations and tables are essential features, as they often represent the only clear representation of the information contained in the text or an alternative interpretation of the text. Certain conventions need to be understood, such as the arrow or equals sign in order to interpret formulae and equations correctly but, fortunately, these conventions enjoy international standing and it can be assumed that the students had been taught their use in school in Portugal so that shared background knowledge exists in respect of these conventions. They do, however, involve the integration of semantics, cohesion and frequently grammar in order to be read as if they are part of the text (Lemke 1998). In this way the students' understanding must go beyond the simplest interpretation of these features. Tarone *et al.* (1981:201) find in the astrophysics journals that they studied that:

> One of the most striking characteristics of the sentences used in these papers is the fact that lengthy equations are embedded within them, and must be arranged in such a way as not to interfere with the reader's processing of the basic grammar of the sentence. Because of end-weight such equations are often placed at the end of clauses, and the use of active or passive verb forms is often conditioned by this requirement.

In Chang there are a number of equations that follow Tarone *et al.*'*s* findings, being placed at the end of clauses such as:

The molar mass of H_3PO_4 is given by

$$3(1.008\text{g}) + 30.97\text{ g} + 4(16.00\text{ g}) = 97.99\text{ g}$$

However, there are other grammatical constructs used such as conditionals and either... or clauses:

If we had used the empirical formula HO for the calculation, we would have written

$$\%H = \frac{1.008g}{17.01g} \times 100\% = 5.926\%$$

$$\%O = \frac{16.00g}{17.01g} \times 100\% = 94.06\%$$

In terms of factor-label method, we can write the unit factor as

Field Code Changed

$$\frac{1\ \text{mol}\ O_2}{2\ \text{mol}\ CO} = 1 \quad \text{or} \quad \frac{2\ \text{mol}\ CO}{1\ \text{mol}\ O_2} = 1$$

In Serway there is a similar situation with an entire paragraph containing equations which must be read as part of the sentence grammar:

The defining equation for acceleration,

$$a = \frac{dv}{dt}$$

may also be written in terms of an integral (or antiderivative) as

$$v = \int a \cdots dt + C_1$$

where C_1 is a constant of integration. For the special case where the acceleration *a* is a constant, this reduces to

$$v = at + C_1$$

Similarly, the chemical symbols for elements and compounds are standardised through international convention and should not prove to be a stumbling block, provided that they are understood in Portuguese especially as the formulae follow the English order. Thus, it can be argued that the number and range of formulae should add to understanding rather than obscuring it, provided that there is shared background scientific knowledge. Formula would also help to explain any common term used in English for a particular chemical compound such as common salt rather than its chemical name. However, some mathematical working could prove confusing (see later *5.4 Mathematics)*.

Formulae can represent almost 20% of the text on a page in both the Physics and Chemistry textbooks on average. Lemke (1998:89) finds that it is "actually unusual to find high concentrations of both equations and graphics in the same article or on the same page". This is untrue of the undergraduate physics and chemistry textbooks being studied here which often integrate graphics and equations. The example from Serway given above ends with a graph showing the velocity versus time curve for a particle moving with a velocity that is proportional to the time which was represented by the formulae contained in the paragraph quoted and is followed by a further paragraph extrapolating further and containing formulae as part of the sentence grammar. Lemke (ibid.) suggests that experimental-empirical reports tend to have more graphics, whilst theoretical analyses have more equations. This might explain why the textbooks have both together in order to survey the experimental work which has been carried out and also to represent the theoretical contribution made to the field by that work.

Tarone *et al.* only considered active and passive voice in the combination of clauses and equations but in these textbooks many other structures were found such as: conditionals (see example above), contrasts and comparisons (*Similarly,..., are compared as follows, however, not* etc.) alternative either...or constructions (see example above), exemplifications (*that is, in other words, is given by, as follows* etc.), logical conclusions (*we can now write ..., we can write this as ..., we obtain ..., we would thus write*) and the expression of laws as formulae (*This is called the associative law of addition:* $A + (B + C) = (A + B) + C$, *and is known as the commutative law of addition:* $A + B = B + A$, *Charles' law:* $V \propto T$...etc.). The syllabus for the undergraduates coming into contact with these types of texts must take account of the integration of grammar and formulae and equations, exploring the different types of sentence structure that they are usually found and the means by which these are expressed in words. Recognition of the oral expression of formulae and equations is important for understanding lectures given in English and also for note taking.

Field Code Changed

5.2.4 Diagrams and Drawings

The use of diagrams and drawings should enhance the understanding of the surrounding text or even create new orders of meaning but, in either case, it is essential that the referencing to these be understood for this to take place. In general, visual material in the text is a form of redundancy, as it reiterates or expands upon what is being discussed. However, some problems may arise when the text and the diagrams do not correspond exactly. An example of this occurred in some materials produced for use with undergraduates in the *Ano Comum* in the university. An exercise was produced using diagrams and text taken from an encyclopaedia explaining the four-stroke petrol engine. The students were asked to match the descriptions of the four strokes with the diagrams which had been placed in a random order. The relevant part of the text gave the following list:

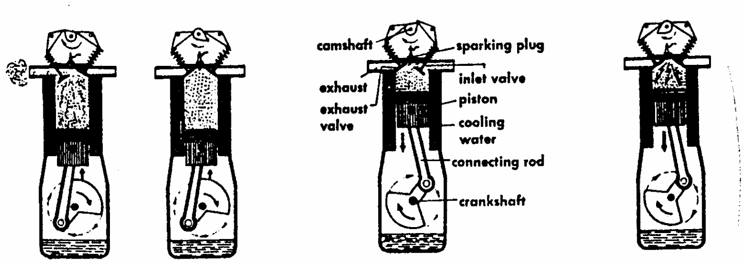
1st stroke: induction stroke: while the inlet valve is open, the descending piston draws fresh petrol-and-air mixture into the cylinder.

2nd stroke: compression stroke: while the valves are closed, the rising piston compresses the mixture to a pressure of about 7-8 atm.; the mixture is then ignited by the sparking plug.

3rd stroke: power stroke: while the valves are closed, the pressure of the gases of combustion forces the piston downwards.

4th stroke: exhaust stroke: the exhaust valve is open and the rising piston discharges the spent gases from the cylinder.

The diagrams were presented in the order shown below:



The students could have adopted a number of strategies to find the correct sequence. One of the obvious features was the fact that the labelling had been retained in image three, which would therefore, conventionally, make this the first diagram. The answer to this problem as given by the encyclopaedia in the original was, stroke 4, stroke 3, stroke 1, stroke 2. The discussion of this order invariably revolved around the fact that there was inconsistency between the second image and the fourth image. The second image does not conform to “forcing the piston downwards” whereas the fourth image does. The students often failed to take note of the ‘spark’ that image four displays but which is difficult to detect in a black and white image like those shown above.

In other words the relationship between diagrams and the supporting text is not as self-explanatory as may at first be thought. This suggests that the syllabus must reflect this difficulty by presenting a number of different types of relationships between visuals and the text. As was mentioned earlier in 5.2.3 this could include equations and visuals being read as part of a paragraph in the physics textbook with which the students would need practice and would need to develop strategies to overcome some of the difficulties encountered like the mismatch described above. This is an area that is increasingly important with the change towards much more use of visual representation in modern life both in textbooks and in computers and specific discourse strategies which encourage students to explore the relationships set up by visual representation is increasingly important in education (Carter, Goddard, Reah, Sanger and Bowring 1997).

Analysis of the Physics and Chemistry textbooks shows that photographs, diagrams, graphs and drawings of apparatus are all found together with formulae and examples in highlighted boxes. The relative composition between pictures and text on each page is, on average, one

Field Code Changed

third visual to two thirds text, although these values can range quite widely from between 70% for pictures and 30% text, and 90% text and 10% diagrams.[5]

Chang (1991:xxiii) says that their use of a “5-color design” will help students “to visualise the appearance of compounds and various chemical processes”. They also mention that in this edition they have added “many full-color photographs and line drawings” and have introduced “A number of marginal arts” to “enhance discussion and to accompany worked examples”. Moreover, they have attempted to be consistent in their use of colour to illustrate similar concepts “wherever appropriate”. In other words, a principled approach has been taken by the textbook editors, which the students need to be aware of, in order to benefit from the insights these features should bring.

5.3 The Undergraduate Textbooks

The physics textbook recommended for students of the combined first year (*Ano Comum*) of the science and technology courses of the university of Aveiro is written by an American, Raymond Serway, published in 1992 and entitled *Physics for Scientists and Engineers with Modern Physics*, 3rd Edition. There is now a fourth Edition (1996) available of this textbook which claims that it has been successfully used by over 700 colleges and universities. The chemistry book for this year is by Raymond Chang, published in 1991 and entitled *Chemistry*. Both of these books appear, to a certain extent, to be trying to overcome the pedagogic limitations outlined by Halliday and Martin in their (1993) book *Writing Science* who claim that science textbooks pay too little attention to

[5] It is interesting to note that, even in some translations of textbooks into Portuguese, the labelling of diagrams remains the same.

solid pedagogic principles. Chang (1991:xxii) claims to use a "Problem-Solving Pedagogy" where students are "asked to examine the reasonableness of the answer" they give to problems in the end-of-chapter exercises. Students are expected to explain the "why" of chemistry through the review questions, the "how" of chemistry through the problems and to identify the "concept, topic or technique to be applied" through the miscellaneous problems.

The way that these particular textbooks have attempted to overcome Halliday and Martin's ascription of pedagogic limitations is by dealing with topical issues or everyday situations. Each of the chapters in this physics textbook contains an essay on a topic of more general interest, but with a very specific physics focus, and each chapter is followed by a number of questions or problems based on laws dealt with in the chapter, but which attempt to put the scientific theories into more popular situations such as those of travel, sports, nature and so on. Chang (1991:xxii) explains that "to define complex terms in a clear manner, and to explain difficult concepts carefully" use is made of analogy, for example downhill skiing and dynamic chemical equilibrium are used to introduce a chemical concept. Similarly, the chemistry textbook claims in the preface (1991:xxii):

> "**Real-World Applications.** One of the joys of learning chemistry is seeing how chemical principles can be applied to everyday experience. The Chemistry in Action sections show the relevance of chemistry to biological, medical, technological, and engineering fields, as well as current news topics."

Glaser (1982:78) points out that "analogies from the learner's everyday experience are often used by way of comparison" in what she describes as the didactic style of ESP textbooks and that this has implications for the linguistic structure of the text. However, Laurillard (1993:59-60) warns how unsuitable analogies are often used by students

Field Code Changed

and that this process of imagining concrete analogies is not “a reliable way of gaining access to the experience of academic knowledge”. She (1993:59) also claims that “Physics is notorious for alluring concrete analogies that lead you falsely” and even suggests that teachers themselves use inappropriate analogies in their teaching. Whilst it is not being suggested that these textbooks lead students astray with their extensive use of analogies they may be falling into the trap that Laurillard warns about because the students are not capable of thinking scientifically and may therefore draw the wrong conclusions from the analogies given. It is in this way that even teachers can fall into traps as Martins and Veiga (1999) explore in their study on training primary school teachers to teach science to pupils by exploring contexts from their daily lives. They found that the teachers themselves often needed to learn how to think scientifically in order to overcome misconceptions before they could help their classes to do the same.

Some of the titles of the real world or “Chemistry in Action” analogies in the chemistry textbook show the diversity of the subjects the students will encounter, for example these range from “The Scientific Method and the Extinction of the Dinosaurs”, “Salvaging the Recorder Tape from the *Challenger*”, “Black and White Photography”, “Breath Analyzer”, “Scuba Diving and the Gas Laws”, “Fuel Values of Foods and Other Substances”, “How a Bombardier Beetle Defends Itself”, “Determining the Age of the Shroud of Turin”, to “The Thermodynamics of a Rubber Band” and many more. The scope that these provide for misconceptions is therefore quite broad depending on the ideas the students already have on these diverse topics.

In addition, this textbook details nine supplements for use with it including video and computer programs, these are: Student Solutions Manual, Microscale and Macroscale Experiments for General Chemistry, Study Guide, Instructor’s Manual, Test Bank, R-H Test, Overhead

Transparencies, Chemistry at Work Videodisc, Micro Guide. These supplements are not available to the undergraduates through the library which contains multiple copies of the textbooks themselves in English. The latest edition of Serway references a world wide web site at the University of Texas which will provide answers to students questions. It is likely that this site was intended for those hundreds of colleges and universities in the United States which he claims successfully use the textbook. The tendency to provide more and more support material through computers (Serway also has accompanying computer simulations and spreadsheets) cannot be ignored. The syllabus for these undergraduates will have to encompass computer literacy in order to keep in step with the developments that are taking place in educational materials for science and technology students.

5.3.2 The Physics and Chemistry Algorithms and Functions compared with Biber's Academic Prose

Halliday (1991) makes a case for the validity of examining not only lexical frequency in text but also grammatical frequency. He claims that grammatical frequency is even more powerful than lexical frequency because the system is closed and the number of choices is small, so that significant probabilities can be calculated.

Biber standardised or "*normalized*" grammatical features found in texts, that is, he standardised the raw data to reflect the frequency in a thousand word extract by dividing the number of occurrences of a certain grammatical feature by the total number of words in the text and then multiplying by one thousand[6]. By observing this same level of scientific rigour any text can be compared with Biber's results in order to draw conclusions about its position on the continuum of variation and thereby,

[6] In Biber, Conrad and Reppen (1998:33) they refer to this process as a *normed count.*

Field Code Changed

for the purposes of this research, to draw conclusions about the nature of the text concerned and its consequent difficulty for students. The specific definitions of each of the sixty-five algorithms used by Biber and in this research can be found in Appendix A.

The values presented in the table below include normalised frequency values and the chi-square test (χ^2) to show which values are significant. The degrees of freedom value (df) is one and, at the five percent level, the critical value is 3.84. Yates correction was used on all values of less than five.

Table 5.3 Normalised Frequencies from the Main Corpora compared to Biber's Academic Prose with Statistical Significance Values (chi-square χ^2)

Linguistic Feature	**Physics Text**	**Chemistry text**	**Biber's Academic Prose**
Past Tense	1.5 χ^2 = -19.95	1.3 χ^2 = -20.33	21.9
Perfect Aspect	0.9 χ^2 = -4.13	1.0 χ^2 = -3.95	4.9
Present Tense	44.2 χ^2 = -5.97	69.2 χ^2 = 0.47	63.7
Place Adverbials	3.7 χ^2 = 0.26	1.3 χ^2 = -1.06	2.4
Time Adverbials	1.8 χ^2 = -0.8	1.4 χ^2 = -1.29	2.8
First Person Pronouns	7.6 χ^2 = 0.63	8.7 χ^2 = 1.58	5.7
Second Person Pronouns	2.5 χ^2 = 16.2	0.9 χ^2 = 0.2	0.2
Third Person Personal Pronouns	4.5 χ^2 = -4.89	2.6 χ^2 = -7.68	11.5
Pronoun *it*	4.5 χ^2 = -0.61	3.8 χ^2 = -1.14	5.9
Demonstrative Pronouns	8.4 χ^2 = 13.92	5.3 χ^2 =3.14	2.5
Indefinite Pronouns	0.1 χ^2 = -1.8	0 χ^2 = -2.45	0.2
Pro-verb *DO*	0.1 χ^2 =-1.73	0.1 χ^2 = -1.73	0.7
Direct WH-questions	4.3 χ^2 = 14.44	3.3 χ^2 = 7.84	0
Nominalizations	51.8 χ^2 = 7.15	49.8 χ^2 =5.47	35.8
Gerunds	6.8 χ^2 = -0.34	7.0 χ^2 = -0.26	8.5
Total Other Nouns	155.7	159.1	188.1

	χ^2 = -5.58	χ^2 = -4.47	
Agentless Passives	19.5 χ^2 = 0.36	11.9 χ^2 = 1.53	17.0
By - Passives	3.6 χ^2 = 0.61	2.8 χ^2 = 0.05	2.0
BE as Main Verb	20.4 χ^2 = -0.48	19.5 χ^2 = -0.78	23.8
Existential *there*	0.5 χ^2 = -5.04	1.0 χ^2 = -0.94	1.8
that Verb Complements	4.7 χ^2 = 0.31	3.6 χ^2 = -0.003	3.2
that Adjective Complements	0.1 χ^2 = -1.6	0.1 χ^2 = -1.6	0.4
WH - Clauses	0.3 χ^2 = -0.83	0.1 χ^2 = -1.63	0.3
Infinitives	0.1 χ^2 = -13.61	9.2 χ^2 = -1.01	12.8
Present Participial Clauses	2.3 χ^2 = 0.19	1.2 χ^2 = -0.28	1.3
Past Participial Clauses	0.1 χ^2 = -1.6	0 χ^2 = -0.85	0.4
Past Participial WHIZ Deletion Relatives	3.2 χ^2 = -1.5	0.6 χ^2 = -5.4	5.6
Present Participial WHIZ Deletion Relatives	2.8 χ^2 = -0.02	1.3 χ^2 = -1.16	2.5
that Relative Clauses on Subject Position	1.5 χ^2 = 3.2	2.0 χ^2 = 8.45	0.2
that Relative Clauses on Object Position	0.5 χ^2 = -0.8	0.4 χ^2 = -1.01	0.8
WH Relative Clauses on Subject Position	1.5 χ^2 = -0.98	1.5 χ^2 =-0.98	2.6
WH Relative Clauses on Object Position	2.1 χ^2 = -0.08	1.1 χ^2 = -0.98	2.0
Pied-piping Relative Clauses	0.8 χ^2 = -0.77	0.5 χ^2 = -1.3	1.3
Sentence Relatives	0	0	0
Causative Adverbial Subordinators	0.7 χ^2 = 0.03	1.2 χ^2 = 0.53	0.3
Concessive Adverbial Subordinators	0.3 χ^2 = -0.98	0.4 χ^2 = -0.72	0.5
Conditional Adverbial Subordinators	4.2 χ^2 = 1.22	2.2 χ^2 = -0.08	2.1
Other Adverbial Subordinators	2.2 χ^2 = -0.01	1.5 χ^2 = -0.35	1.8
Total Prepositional Phrases	127.5 χ^2 = -1.03	125.4 χ^2 = -1.43	139.5
Attributive Adjectives	29.8 χ^2 = -28.84	36.6 χ^2 = -21.12	76.9
Predicative Adjectives	6.1 χ^2 = 0.24	3.1 χ^2 = -1.15	5.0
Total Adverbs	6.4 χ^2 = -39.79	8.5 χ^2 = -36.19	51.8

Type/Token Ratio	51.9 $\chi^2 = 0.03$	38.3 $\chi^2 = -2.99$	50.6
Word Length	5.8 $\chi^2 = 0.05$	6.0 $\chi^2 = 0.1$	4.8
Conjuncts	2.4 $\chi^2 = -0.4$	4.5 $\chi^2 = 0.33$	3.0
Downtoners	0.9 $\chi^2 = -1.76$	1.7 $\chi^2 = -0.68$	2.5
Hedges	0.1 $\chi^2 = -1.8$	0.1 $\chi^2 = -1.8$	0.2
Amplifiers	1.2 $\chi^2 = 0.35$	1.6 $\chi^2 = 0.06$	1.4
Emphatics	2.2 $\chi^2 = -1.0$	2.2 $\chi^2 = -1.0$	3.6
Discourse Particles	0.3 $\chi^2 = -0.04$	0.1 $\chi^2 = -0.16$	0
Demonstratives	7.2 $\chi^2 = -1.55$	5.8 $\chi^2 = -2.75$	11.4
Possibility Modals	5.5 $\chi^2 = -0.001$	5.1 $\chi^2 = -0.04$	5.6
Necessity Modals	1.6 $\chi^2 = -0.22$	1.5 $\chi^2 = -0.65$	2.2
Predictive Modals	3.8 $\chi^2 = -0.04$	2.8 $\chi^2 = -0.53$	3.7
Public Verbs	1.8 $\chi^2 = -3.4$	2.9 $\chi^2 = -1.91$	5.7
Private Verbs	11.5 $\chi^2 = -0.08$	7.2 $\chi^2 = -2.25$	12.5
Suasive Verbs	0.3 $\chi^2 = -4.41$	0.3 $\chi^2 = -4.41$	4.0
Seem/appear	0.1 $\chi^2 = -2.96$	0.5 $\chi^2 = -1$	1.0
Contractions	0.2 $\chi^2 = -1.6$	0 $\chi^2 = -3.6$	0.1
Subordinator *-that* Deletion	0.2 $\chi^2 = -1.23$	0.1 $\chi^2 = -1.2$	0.4
Stranded Prepositions	0 $\chi^2 = -2.33$	0 $\chi^2 = -2.33$	1.1
Split Infinitives	0.1 $\chi^2 = -0.16$	0.1 $\chi^2 = -0.16$	0
Split Auxiliaries	0.4 $\chi^2 = -5.59$	1.0 $\chi^2 = -4.31$	5.8
Phrasal Coordination	22.5 $\chi^2 = 75.43$	20.1 $\chi^2 = 56.46$	4.2
Independent Clause Coordination	0.7 $\chi^2 = -1.52$	0.4 $\chi^2 = -2.11$	1.9
Synthetic Negation	0.6 $\chi^2 = -1.11$	0.6 $\chi^2 = -1.11$	1.3
Analytic Negation	1.6 $\chi^2 = -2.38$	2.0 $\chi^2 = -1.82$	4.3

The algorithms that are found to differ significantly from Biber's Academic prose for each of the corpora are the following:

Table 5.4 The Physics Main Corpus: Significantly Higher and Lower Results

Significantly Lower Results	**Significantly Higher Results**
Past Tense	Second Person Pronouns
Perfect Aspect	Demonstrative Pronouns
Present Tense	Direct WH-questions
Third Person Pronouns	Nominalizations
Total Other Nouns	Phrasal Coordination
Existential *there*	
Infinitives	
Attributive Adjectives	
Total Adverbs	
Suasive Verbs	
Split Auxiliaries	

Table 5.5 The Chemistry Main Corpus: Significantly Higher and Lower Results

Significantly Lower Results	**Significantly Higher Results**
Past Tense	Direct WH-questions
Perfect Aspect	Nominalizations
Third Person Personal Pronouns	*that* Relative Clauses on Subject Position
Total Other Nouns	Phrasal Coordination
Past Participial WHIZ Deletion Relatives	
Attributive Adjectives	
Total Adverbs	
Suasive Verbs	
Split Auxiliaries	

Taking the two main corpora together there are eleven algorithms found to be significantly lower or higher in both the physics and the chemistry corpora from the sixty-five examined by Biber, these are: *Significantly Lower in <u>both</u> Main Corpora*: Past Tense, Perfect Aspect, Third Person Personal Pronouns, Total Other Nouns, Attributive Adjectives, Total Adverbs, Suasive Verbs, Split Auxiliaries.

Field Code Changed

Significantly Higher in <u>both</u> Main Corpora: Direct WH-questions, Nominalizations, Phrasal Coordination.

Biber comments that past tense forms are usually taken as a surface marker of narrative, and that there is a co-occurrence of the past tense, third person pronouns and perfect aspect verbs in narrative, reported styles, typical of fiction. As these features are found to be significantly lower in the corpora this suggests that overall these texts emphasise an extreme 'academic' text profile. This in turn suggests that they would be even more difficult for students because dry, academic texts are seen to be much more abstract and more difficult to understand.

Biber reports that third person personal pronouns mark relatively inexact reference to persons outside the immediate interaction and in previous studies (1986) has found that they co-occur frequently with past tense and perfect aspect forms as a marker of narrative, reported, (as opposed to immediate,) styles. These texts are then more 'immediate' in style than would be expected. Biber (1988:137-8) describes the non-narrative purposes as including "(1) the presentation of expository information, ...(2) the presentation of procedural information, ...(3) description of actions actually in progress." These features of texts must then be respected in the materials used with students in order to prepare them for their studies. Equally well the combination of these findings suggest that fiction would have no place in the discipline if only appropriate texts are to be used with the students. However, this assertion will be challenged with respect to the sub-corpora findings which follow.

5.3.3.The Physics and Chemistry Sub-Corpora

As mentioned above both the Physics and the Chemistry textbooks contain essays in every chapter which are meant to illustrate or explain the subject of the chapter in either a more stimulating or in a real-world application. These essays were analysed using the same criteria given above because it was felt that they could represent difficulties for the foreign language student. The full texts are given in Appendix E.

Chapter One of Serway (1992) *Physics for Scientists and Engineers with Modern Physics* contains an essay entitled *Scaling* - *the Physics of Lilliput* written by Philip Morrison of the Massachusetts Institute of Technology. The principal objective of the essay is to examine the hypothesis that it is possible to have either very much larger or very much smaller creatures who could look just like us. People are said to come in 'all shapes and sizes' but if this suggestion is examined more closely it can be seen that this variety really only operates within strict upper and lower limits and the extremes recorded in the *Guinness Book of Records* have varied little over the last century (between four feet or one metre and seven feet or two metres approximately). In other words it is not possible to have the minute six inch creatures or the giants twelve times Gulliver's size as described by Swift in his *Gulliver's Travels*. Morrison reaches the conclusion that "Lilliputians must be a hungry lot, restless, active, graceful, but easily waterlogged." because of the necessity to produce enough energy they would be constantly searching for and eating food but, if it rained, their relative surface area would cause them to be weighed down by the water covering their bodies, rather like a fly.

The essay itself is about two and a half thousand words long and gives a ratio of the total number of words to the number of different words or token/type ratio of 55.4. Biber (1988:104-5) in *Variation across speech and writing* argues that the higher the *type/token* ratio the higher the lexical variety and the more abstract the text concerned is. This figure would therefore represent a difficult text for the student, given that approximately two in every three words is different.

The Chemistry textbook presents similar but generally shorter essays and there are more of them in each chapter than the usual one

Field Code Changed

long essay per chapter in the Physics textbook. The essay which has been used for comparative study from the Chemistry textbook is entitled *Salvaging the Recorder Tape from the Challenger* and it appears in the Chemistry in Action section of the third chapter of the textbook and is thus part of the overall corpus taken from this textbook. There are two other Chemistry in Action essays in this chapter illustrating the chemical reactions which have been described in the preceding chapter but these are extremely short with lots of equations and lots of photographs respectively. These essays were therefore rejected as being too limited in scope for the purpose of comparison. This particular essay is concerned with the crash of the space shuttle Challenger in 1986 and the subsequent recovery of the tape of the flight and the chemistry used in order to be able to listen to the seawater damaged recording that had been made of the fateful flight. The essay is only about 450 words long and presents a token/type ratio of 51.1, which is considerably higher than for the chemistry corpus as a whole.

The results for each of the linguistic features as described in Appendix A for these two essays are given below:

Table 5.6 Normalised Frequencies from the Sub-Corpora compared to Biber's Academic Prose with Statistical Significance Values (χ^2)

Linguistic Feature	Physics Essay	Chemistry Essay	Biber's Academic Prose
Past Tense	7.6 χ^2 = -9.34	20.4 χ^2 = -0.1	21.9
Perfect Aspect	4.9 χ^2 = -0.05	0 χ^2 = -5.95	4.9
Present Tense	35.0 χ^2 = -12.93	24.9 χ^2 = -23.63	63.7
Place Adverbials	1.7 χ^2 = -0.6	0 χ^2 = -3.5	2.4
Time Adverbials	2.8 χ^2 = -0.09	4.5 χ^2 = 0.51	2.8
First Person Pronouns	17.3 χ^2 = 23.61	0 χ^2 = -5.7	5.7
Second Person Pronouns	5.2 χ^2 = 101.25	2.3 χ^2 = 12.8	0.2

Third Person Personal Pronouns	16.3 χ^2 = 2.0	6.8 χ^2 = -1.92	11.5
Pronoun *it*	8.3 χ^2 = -5.32	4.5 χ^2 = -0.61	5.9
Demonstrative Pronouns	2.1 χ^2 = -0.32	4.5 χ^2 = 0.9	2.5
Indefinite Pronouns	0 χ^2 = -2.45	0 χ^2 = -2.45	0.2
Pro-verb *DO*	0.7 χ^2 = -0.36	0 χ^2 = 0.06	0.7
Direct WH-questions	0.7 χ^2 = 0.04	0	0
Nominalizations	13.9 χ^2 = -13.4	38.5 χ^2 = 0.2	35.8
Gerunds	5.9 χ^2 = -0.8	6.8 χ^2 = -0.34	8.5
Total Other Nouns	164.8 χ^2 = -2.89	226.2 χ^2 = 7.72	188.1
Agentless Passives	10.4 χ^2 = -2.56	22.6 χ^2 =1.85	17.0
By - Passives	2.8 χ^2 = 0.05	4.5 χ^2 = 2.0	2.0
BE as Main Verb	23.2 χ^2 = -0.02	18.1 χ^2 = -1.37	23.8
Existential *there*	1.0 χ^2 = -0.94	2.3 χ^2 = 0	1.8
that Verb Complements	3.8 χ^2 = 0.003	6.8 χ^2 = 3.0	3.2
that Adjective Complements	1.0 χ^2 = 0.03	0 χ^2 = -2.03	0.4
WH - Clauses	3.8 χ^2 = 30.0	0 χ^2 = -2.67	0.3
Infinitives	7.6 χ^2 = -2.11	20.4 χ^2 = 4.51	12.8
Present Participial Clauses	1.0 χ^2 = -0.49	2.3 χ^2 = 0.192	1.3
Past Participial Clauses	0 χ^2 = -2.03	0 χ^2 = -2.03	0.4
Past Participial WHIZ Deletion Relatives	1.7 χ^2 = -3.46	11.3 χ^2 = 5.8	5.6
Present Participial WHIZ Deletion Relatives	1.4 χ^2 = -1.02	2.3 χ^2 = -0.2	2.5
that Relative Clauses on Subject Position	0.7 χ^2 = 0	2.3 χ^2 = 12.8	0.2
that Relative Clauses on Object Position	1.4 χ^2 = 0.01	0 χ^2 = -2.11	0.8
WH Relative Clauses on Subject Position	4.5 χ^2 = 0.75	0 χ^2 = -3.7	2.6
WH Relative Clauses on Object Position	2.1 χ^2 = -0.08	9.1 χ^2 = 21.78	2.0
Pied-piping Relative Clauses	0.3 χ^2 = -1.73	0 χ^2 = -2.49	1.3
Sentence Relatives	0.7	0	0

Field Code Changed

	$\chi^2 = 0.04$	$\chi^2 = 0$	
Causative Adverbial Subordinators	1.7 $\chi^2 = 2.7$	0 $\chi^2 = -2.13$	0.3
Concessive Adverbial Subordinators	0 $\chi^2 = -2.0$	0 $\chi^2 = -2.0$	0.5
Conditional Adverbial Subordinators	3.5 $\chi^2 = 0.39$	0 $\chi^2 = -3.22$	2.1
Other Adverbial Subordinators	1.4 $\chi^2 = -0.45$	0 $\chi^2 = -2.94$	1.8
Total Prepositional Phrases	112.4 $\chi^2 = -5.26$	131.2 $\chi^2 = -0.49$	139.5
Attributive Adjectives	30.2 $\chi^2 = -28.36$	76.9 $\chi^2 = 0$	76.9
Predicative Adjectives	14.9 $\chi^2 = 19.6$	6.8 $\chi^2 = 0.65$	5.0
Total Adverbs	14.9 $\chi^2 = -26.29$	15.8 $\chi^2 = -25.02$	51.8
Type/Token Ratio	55.4 $\chi^2 = 0.46$	51.1 $\chi^2 = 0.01$	50.6
Word Length	5.6 $\chi^2 = 0.02$	6.0 $\chi^2 = 0.1$	4.8
Conjuncts	5.9 $\chi^2 = 1.92$	9.1 $\chi^2 = 10.45$	3.0
Downtoners	1.7 $\chi^2 = -0.68$	2.3 $\chi^2 = -0.2$	2.5
Hedges	0.3 $\chi^2 = -0.8$	0 $\chi^2 = -2.45$	0.2
Amplifiers	6.6 $\chi^2 = 19.31$	2.3 $\chi^2 = 0.29$	1.4
Emphatics	6.6 $\chi^2 = 1.74$	2.3 $\chi^2 = -0.9$	3.6
Discourse Particles	0.3 $\chi^2 = 0.04$	0	0
Demonstratives	5.9 $\chi^2 = -2.65$	4.5 $\chi^2 = -4.18$	11.4
Possibility Modals	10.4 $\chi^2 = 4.11$	4.5 $\chi^2 = -0.46$	5.6
Necessity Modals	4.2 $\chi^2 = 1.02$	0 $\chi^2 = -3.31$	2.2
Predictive Modals	8.3 $\chi^2 = 4.54$	2.3 $\chi^2 = -0.98$	3.7
Public Verbs	3.1 $\chi^2 = -1.69$	0 $\chi^2 = -6.74$	5.7
Private Verbs	7.3 $\chi^2 = -2.16$	2.3 $\chi^2 = -9.16$	12.5
Suasive Verbs	0.7 $\chi^2 = -3.61$	0 $\chi^2 = -5.06$	4.0
Seem/appear	0.3 $\chi^2 = -1.44$	0 $\chi^2 = -2.25$	1.0
Contractions	1.0 $\chi^2 = 1.6$	0 $\chi^2 = -3.6$	0.1
Subordinator *-that* Deletion	1.0 $\chi^2 = 0.03$	0 $\chi^2 = -2.03$	0.4

Stranded Prepositions	0 χ^2 = -2.33	0 χ^2 = -2.33	1.1
Split Infinitives	0	2.3 χ^2 = 3.24	0
Split Auxiliaries	1.4 χ^2 = -4.14	6.8 χ^2 = 0.17	5.8
Phrasal Coordination	17.0 χ^2 = 36.02	0 χ^2 = -5.26	4.2
Independent Clause Coordination	1.4 χ^2 = -0.53	22.6 χ^2 = 214.76	1.9
Synthetic Negation	2.8 χ^2 = 0.77	4.5 χ^2 = 5.61	1.3
Analytic Negation	8.0 χ^2 = 2.38	0 χ^2 = -5.36	4.3

The algorithms that are found to differ significantly from Biber's Academic prose for each of the sub-corpora are the following:

Table 5.7 The Physics Sub-Corpus: Significantly Higher and Lower Results

Significantly Lower Results	**Significantly Higher Results**
Past Tense	First Person Pronouns
Present Tense	Second Person Pronouns
Pronoun *it*	WH-clauses
Nominalizations	Predicative Adjectives
Total prepositional Phrases	Possibility Modals
Attributive Adjectives	Predictive Modals
Total Adverbs	Phrasal Coordination
Split Auxiliaries	Amplifiers

As the object of the study of these sub-corpora is to see how far they differ from the main corpora (and Biber's findings), the results which show a significant difference from *both* the main corpus *and* Biber's findings are those of interest. The algorithms which differ from Biber's findings in both the main physics corpus and the physics sub-corpus are the following:

Significantly Lower: Past Tense, Present Tense, Attributive Adjectives, Total Adverbs, Split Auxiliaries.

Significantly Higher: Second Person Pronouns, Phrasal Coordination.

Field Code Changed

Nevertheless, the fact that the sub-corpus shows a greater occurrence of certain features makes the essay significant in terms of syllabus design, where certain features should be included in the syllabus because of their presence in typical materials that the students will come across in their studies. McCarthy and Carter (1994:112) say that “whatever aspects of lexico-grammar we choose to look at, we cannot really separate them from the concerns of creating discourse”. In other words, these features make up the whole and cannot be taken out of context without misrepresenting natural language use, in this case the style of the science textbook in exemplifying real-world situations. If we want students to cope with these kinds of texts, we must bring the students into contact with the specifics of those texts.

The features that differ in the physics sub-corpus need not be compared to those found for the chemistry sub-corpus as they are, in themselves, a deviation from the norm of the main (physics) corpus and so deserve study in their own right.

The results for the chemistry sub-corpus are given below in Table 5.8.

Table 5.8 The Chemistry Sub-Corpus: Significantly Higher and Lower Results

Significantly Lower Results	**Significantly Higher Results**
Perfect Aspect	Second Person Pronouns
Present Tense	Total Other Nouns
First Person Pronouns	Infinitives
Total Adverbs	Past Participial WHIZ Deletion Relatives
Public Verbs	*that* Relative Clauses on Subject Position
Private Verbs	WH Relative Clauses on Object Position
Suasive Verbs	Conjuncts
Phrasal Coordination	Independent Clause Coordination
Analytic Negation	Synthetic Negation
Demonstratives	

The chemistry sub-corpus differs from the chemistry main corpus in one crucial way; it contains many more significantly high features than the main corpus. That is to say, it contains many more examples of features that are not so prevalent in either the main corpus or Biber's findings for Academic Prose.

The algorithms that the main and sub corpora share are the following:

Significantly Lower: Perfect Aspect, Total Adverbs, Suasive Verbs
Significantly Higher: that Relative Clauses on Subject Position

Although the points in common are few, this finding is even more significant as it shows the wide degree of difference between the text as a whole and the essay studied here. This can only reinforce the conviction that this kind of differentiation in the text will cause some students to have greater difficulty than ever with the attempt by the author to exemplify what is being studied through 'real world' situations. That is to say, what is intended by the author to provide pedagogical enlightenment can prove to be linguistic obfuscation for the non-native speaker learner.

Like Sinclair, Biber (1988:238-9) comments on the fact that the longer the text the fewer new word types there are to be found, so that if the entire length of text is considered, as in the figures calculated for the main corpora above, such an accurate description of the difficulty of a particular text especially for comparative purposes, is not demonstrated. Furthermore, Biber himself (1988:48) suggests that "academic prose is contextualized in that it crucially depends on shared (academic) background knowledge for understanding". However, as Bloor and Bloor (1991:2) point out there is a "false expectation that educational structures and systems do not differ internationally", which means that we would do well to anticipate differences in the students' academic background from

Field Code Changed

the background assumed in the textbook being examined. Halliday and Martin (1993:2) suggest that native speaker students of science are “alienated” by the language of science. If all of these conclusions are true, how much more alienated will the foreign language learner be by both a combination of the (foreign) language of science and the lack of a shared academic background to the subject. This is especially the case with the essays under discussion, with their dense text and exophoric appeal to native speaker background understanding in scientific, general and literary knowledge.

5.3.4 The Physics Sub-Corpus: Gulliver’s Travels

A simple frequency listing of the text immediately highlights a number of lexical items in the text which we would not have predicted in a physics textbook despite the fact that technical discourse is supposed to use repeatedly “a small set of technical vocabulary to refer to the exact concepts and entities intended” (Grabe 1984).

In the physics text there are words relating to animals (not counting *man*, *family*, *human* and so on) such as: *animal (1), animals (1), bee (1), bison (3), cattle (1), deer (2), dinosaurs (1), dog (1) and dogs (1), elephant (2), fish (1), fly (2), frogs (1), gazelle (5), horse (1), insects (1), lamb (2), mammals (1), mouse (2), whale (3), whale’s (1) and whalelike (1)* and others relating to plants: *agriculture (1),grass (1), trees (1).* Other ‘biological’ references include; *body (11), bone (8), bones (7), breathing (1), digest (1), fingers (1), flesh (1), forearm (1), leg (5), limb (1), limbs (1), muscles (1), ribs (1), senses (1), skeleton (1), skin (3), tendons (1), and warm-blooded (2)* not to mention the much more difficult to classify *giant (1)* and *giants (5).*[7] Readers could be forgiven if at this point they became confused as to whether this was indeed a physics text or whether they had not come across the discourse

[7] The figures given in brackets refer to the number of times that each word appears in the text, that is, their frequency.

of biology by mistake. The questions which immediately follow this essay also continue in the same vein and *hummingbird, elephant* and *guinea pig* are used in the follow-up work.

The fact that certain items are used many times over in a text leads to a lower type/token ratio, despite the fact that they may be difficult in themselves, especially for non-native science students[8]. This is further complicated when these human biological structures are compared to the structures of buildings and *braces, columns* and *cables* are encountered being compared with *muscles* and *tendons*, with an earlier analogy being drawn between the strength of a *wire* or a *rope*. The collocations for *braces* are as follows:

the skeleton - supported by various *braces* and cables which are muscles and tendons

the strength of his columns and *braces* is proportional to their cross-sections

The specific vocabulary and specific grammar of texts are now seen to be inseparable. Halliday and Martin (1991:4) point out that

> "technical terms are an essential part of scientific language, it would be impossible to create a discourse of organized knowledge without them. But they are not the whole story. The distinctive quality of scientific language lies in the lexicogrammar (the 'wording') as a whole, and any response it engenders in the reader is a response to the total patterns of the discourse."

Halliday (in Ajmer & Altenberg eds. 1991:32-3) goes on to attack a school of thought that suggests that grammatical frequency has no validity. He argues that:

[8] It is debatable if many native speakers would be able to draw an adequate distinction between gazelle and deer. The *Oxford Advanced Learner's Dictionary* gives the following definitions "**gazelle** small, graceful antelope, **deer** any of several types of graceful, quick-running, ruminant animal, the male of which has antlers"

Field Code Changed

> “it does not make sense to condone relative frequency in lexis but deny its validity in grammar (...) the concept of the relative frequency of positive: negative, or of active: passive is no more suspect than the concept of the relative frequency of a set of lexical items. It is, on the other hand, considerably more powerful, because the relative frequencies of the terms in a grammatical system, where the system is closed and the number of choices is very small (typically just two or three), can be interpreted directly as probabilities having a significance for the language as a whole.”

He goes on to say that these “grammatical choices may mean different things in different registers”. This is obviously the case here where not only is there a very specific (and unusual for a physics text) set of lexis but this is coupled with significant grammatical variation from other academic prose. The syllabus must therefore consider these aspects and include suitable practice in both.

5.3.5 Comparison with other Genres in Biber’s Variation Studies

Biber (1988) identified a number of grammatical features in texts and tried to demonstrate that the variation between different text types is on a continuum rather than consists in any absolute distinction. Nevertheless, he also demonstrates that it is both the absence as well as the presence of certain grammatical features that distinguish different genres. This leads into areas that Biber has discussed in relation to what he called ‘Academic prose’ (a very diverse grouping including texts from the humanities to texts from medicine), specifically, the use of pronouns. Biber (1988:193) claims that academic prose, particularly those texts dealing with technology/engineering and the natural sciences, does not generally contain third person pronouns as this genre shows “non-narrative concerns”. (He finds humanities prose to be something of an aberration in this respect "showing a topical concern for concrete events and participants"). Lemke (1990:440) disagrees and claims that

> "Technical discourse is also dominated by third person forms. No "I" speaks to a "you", no space for dialogue, disagreement, or differing points of view is opened in this way either. Even the solidary (inclusive) "we" is absent, and only the authoritative authorial (exclusive) "we" of multiple authorship is allowed. The world of technical discourse is a closed world which admits no criteria of validity outside its own."

There are eight examples containing *he* in the *Gulliver's Travels* text which would suggest that it is untypical of its genre according to Biber's results as mentioned above. Lemke is probably referring to the use of "it" rather than "he" or "she" when he says 'third person forms', as later he goes on to say (1990:440) that technical discourse is 'independent of the particular human agent who has happened upon "the facts".' Biber, however, ranks the 'pronoun IT' as a separate feature from 'third person pronouns' in his text analysis.

On examining other pronouns in this short essay, one finds that there are two examples of *I* (both of which appear in a quotation) 30 examples of *we,* four examples of *us,* eleven examples of *our* and two examples of *ours*. Biber explains that first person pronouns are usually treated as demonstrating personal involvement in a text and are often associated with cognitive verbs and ego-involvement. Moreover, they have been used as a comparison between spoken and written texts. McCarthy and Carter (1994:14) have developed a scale based on Smith's (1986) work which allows examination of mode variation between reader-listener and writer-speaker where 2nd person pronouns indicate the presence of the reader and listener and first person pronouns indicate the presence of the writer and speaker. Passive constructions and third person references are

Field Code Changed

seen to indicate the 'absence of reciprocity' of senders and receivers. As the examples of *I* are both from quotations, it would be wrong to suggest that this shows Philip Morrison's ego involvement in the text, but the unusual presence of 30 occurrences of *we*, eleven examples of *our* and four examples of *us* in a scientific text should alert to what this implies in terms of involvement. This text shows features that would more usually be associated with other genres than the one being studied and suggests a greater degree of informality in the text.

There are seven examples of *you*, which as Biber suggests requires a specific addressee and indicates a high degree of involvement with that addressee. This is perhaps less surprising in a text that is meant to be instructional. Biber has used second person pronouns as a marker of register differences and so once again there is evidence of involvement in this text. There is one example of *one* as a pronoun, *If one wishes*. McCarthy and Carter (1994:15) find the pronoun 'one' to be a marker of 'absence of intimacy' in either the spoken or the written mode.

There are sixteen examples of *it*. In a previous study Biber (1986) suggested that a high frequency of this pronoun marked a relatively inexplicit lexical content due to strict time constraints and showed a non-informational focus, in other words *it* in high frequencies is used in text-types like telephone conversations, face-to-face conversations, personal letters, spontaneous speeches and interviews. Others (Kroch and Hindle 1982) have also associated greater use of this pronoun with spoken situations which is clearly not the case here. This text breaks away from the general situation in the textbook being studied which suggests that there must be provision made in the syllabus for a sufficiently wide range of text-types so that these features can be studied in an appropriate context, which would nevertheless not be face-to-face conversation. However, McCarthy and Carter (1994) recommend finding texts that combine the discourse features that students need to study even if these are in another text-type which can be used as an appropriate vehicle for studying those particular discourse features. Halliday and Martin (1993) say that there is a lot of discussion in the science classroom to clarify the language of the scientific textbook being studied even though the students usually only write short sentences and definitions. The authors of the textbooks being analysed here are anticipating that their books will be used mostly in classes with teachers. Non-native speaker undergraduates however are expected to have to read these textbooks alone. It may therefore be entirely appropriate for texts from these textbooks to be used in language classrooms where discussion of the texts can take place with a language teacher rather than a science teacher. This would go some way towards reproducing the expected mode of use of the textbooks and perhaps thereby help the students to examine both the language and the scientific discourse they need to cope with.

There are 21 examples of *his* and 9 examples of *their*. Biber reports that third person personal pronouns mark relatively inexact reference to persons outside the immediate interaction and, in previous studies (1986), has found that they co-occur frequently with past tense and perfect aspect forms as a marker of narrative, reported (as opposed to immediate), styles. There are no examples at all of either *she* or *her*. The gender deficiency found in this work confirms the findings of linguists like Halliday, Martin and Beaugrande who claim that the language of science is the domain of white, middle-class, adult males. However, the use of *one* and *you* in the same essay points to a certain confusion of usage of pronouns and the use of *one* as a pronoun is not included in Biber's analysis at all. Once again this essay would appear to be atypical of its genre according to Biber's findings. Serway suggests that he uses informal language in order to make his work clear and penetrable for students but this confusion does not support his proposition.

Field Code Changed

The relevant normalizations for pronouns in this essay are as follows:

First Person Pronouns 17.3

Second Person Pronouns 5.2

Third Person Pronouns 16.3

Pronoun *it* 8.3

Biber's averages for these features were 5.7, 0.2, 11.5 and 5.9 respectively.

Examination of Biber's findings reveals that for first person pronouns the Gulliver's Travels text is closer to the averages Biber found for *Religion* (*16.6), Biographies* (*22.1),* and *Science Fiction* (*22.2).* The use of biographical data on the scientists whose theories are discussed is common in scientific textbooks and given the subject of *Gulliver's Travels* fiction is also included to some extent in this essay.

For second person pronouns *Prepared Speeches* (5.*2)* and *Hobbies* (*4.2).* For third person pronouns *Hobbies* (*14.1).*

For pronoun IT *Humor* (*8.2)* and *Prepared Speeches (8.9)* and *Press Reviews (7.9).*

If these texts provide an accurate picture of the use of these features, it would be possible to widen the scope of the materials used with students by including some of these as alternatives and making the work more varied and interesting whilst sacrificing none of the relevancy.

Biber, Conrad and Reppen (1998:61) find that "academic prose uses nominalizations to treat actions and processes as abstract objects separated from human participants." and that academic prose "more often refers to a process with a stative nominalization, where fiction and the spoken corpus describe a specific person's action with a verb or adjective." In other words (1998:75) academic prose shows "a preference for static rather than dynamic packaging of information." They find that six different nominalised words are very common in academic prose, that is with frequencies of over 500 per million words. These are *movement, activity,* *information, development, relation* and *equation.* In contrast, no nominalisations were found to occur in fiction or speech this frequently. All of the nominalisations mentioned above are found in both the physics and the chemistry corpora studied here.

Abstraction is seen by Halliday and Martin (1993) to be one of the reasons that science writing is so different from other writing that students come into contact with and it is this abstraction factor that leads to the difficulty experienced with science texts[9]. However, Biber's definition of nominalisations is somewhat different from Halliday and Martin's. Biber includes all words ending in *-tion#, -ment#, -ness#, or-ity# (plus plural forms)* only, whereas Halliday and Martin allow anything which can function as an element in another clause. Halliday and Martin (1993:15) say

> "Isolated instances of this (nominalization) would by themselves have little significance; but when it happens on a massive scale the effect is to reconstrue the nature of experience as a whole. Where the everyday 'mother tongue' of commonsense knowledge construes reality as a balanced tension between things and processes, the elaborated register of scientific knowledge reconstrues it as an edifice of things. It holds reality still, to be kept under observation and experimented with; and in so doing, interprets it not as changing with time (as the grammar of clauses interprets it) but as persisting - or rather, persistence - through time, which is the mode of being of a noun."

This is confirmed by Sinclair (1997:36) in his work on corpora. He has found that if a word exists as both a noun and a verb the more concrete meaning will be associated with the noun and the more abstract with the verb. He gives the example of *combat* which

[9] Halliday and Martin argue that students actually enjoy the technical terminology of science texts and do not have difficulty with it as long as it is presented systematically.

as a noun means “actual physical fighting” and as a verb “means something like ‘struggle against’”.

Other features identified by Biber as representative of academic prose are passives and the use of the past tense and, as mentioned earlier, it is the co-occurrence of some of these features, (for example third person pronouns plus past tense plus perfect aspect forms), that is important in positioning the text on the continuum of the genre. In order to compare these factors with those Biber obtained, it is necessary to *normalise* the text to a standard 1,000 words as Biber did in his research. The figures obtained from this *normalisation* process are as follows:

Past Tense = 7.6   Biber found a mean of 21.9 for this feature.

This feature is only a third as frequent as in the Biber findings putting it more on a par with *Professional Letters* (10.1) in Biber’s study. Biber comments on the fact that his category of academic prose contains wide variations and he suggests that Humanities prose shows a high score because of its ‘topical concern for concrete events and participants’ while engineering/technology prose reflects ‘concern with abstract concepts and findings rather than events in the past’ and therefore has a low score. This is borne out by the results in the main physics corpus (1.5) which is even lower and not matched by any of Biber’s categories. It is interesting to note that the Chemistry sub-corpus is very close to Biber’s finding (20.4 to 21.9 respectively) but that overall the main corpus is even lower (1.3) than the physics main corpus.

Agentless Passives = 10.4   Biber found a mean of 17 for these

Passive Voice = 2.8   Biber found a mean of 2 for this feature

Passives are taken as characteristic of writing and when the agent is dropped there is a static, more abstract presentation of information. In the case of agentless passives this text is more on a par with *Press Reportage* and *Popular Lore* in Biber's study (11 and 10.6 respectively). Svartvik (1966) calculated the number of passive clauses per 1,000 words of running test for his 320,000 word corpus of eight text types. His results showed an average of 11.3 and a range of 3.0 in advertising to 23.0 in science. A comparable average in the physics text under inspection would be 13.2, once again considerably lower than that referred to by other investigators. It could be argued that it is the attempt by the authors to reach (involve) the readers (students) that causes this finding.

Perfect Aspect Verbs = 4.9 Biber found a mean of 4.9 for this

Biber notes that these verbs have been associated with narrative/descriptive texts and with certain types of academic writing. It is interesting that this text is exactly the same as Biber's finding for academic prose, whereas the main physics corpus is only 0.9. Nevertheless, this is the lowest mean score for this feature found in any of the text types examined by Biber which makes academic prose in a category of its own as regards the use of the perfect. For syllabus purposes this is particularly significant and must be explored.

Nominalizations = 13.9 Biber found a mean of 35.8 for these.

This score is almost matched by that for *Hobbies* in Biber's study (13.1), closely followed by *Science Fiction (14.0)* and then *Humor (12.1)* and *General Fiction* (10). Perhaps this finding is not so unexpected, given that the discussion contained in the text examined here deals with the science in *Gulliver's Travels*.

Nouns = 164.8 Biber found a mean of 188.1 for this feature.

In Biber's study *Adventure Fiction* most closely matches this mean score (165.6) followed by *Mystery Fiction* (165.7) and *General Fiction* (160.7). This may reflect the nature of the subject matter once again or the attempt to make this physics text more amenable to its audience.

Prepositions = 114.4 Biber found a mean of 139.5 for these

Field Code Changed

The prepositions examined by Biber are taken from Quirk *et al.* (1985:665-7). Biber finds that prepositions tend to co-occur frequently with nominalizations and passives in academic prose and other informational types of written discourse. The closest mean score for this feature in Biber's study is for *Hobbies* (114.6) and *Popular Lore* (114.8).

These results would place this text in rather different company than that given in Biber's results for academic prose, however, these are mean scores and considerable variation has to some extent been integrated into Biber's study by including all kinds of academic prose and not only the academic prose of science and technology which is of prime interest for the students who have to study such texts in the University of Aveiro. Nevertheless, the precise description of the study undertaken by Biber has allowed a number of features in this physics text to be compared with his findings and so allows a scientific comparison and interpretation to be made, which in turn can be the basis for a reasoned approach to the relative difficulty of such study material for our students and, consequently, a clearer definition of the approach that needs to be adopted in teaching such students to cope with their textbooks. In this case, some of the features of abstraction are not present to any significant degree, as defined above, but that the attempt to be more accessible will, in fact, lead to even greater difficulty for the foreign language student of physics at university level.

5.3.6 The Chemistry Sub-Corpus: Salvaging the Tapes from the Challenger

The relevant normalizations for the same features discussed above for the physics sub-corpus, together with the nearest text-type mean found by Biber are as follows:

First Person Pronouns 0 *Academic Prose (5.7)*

Second Person Pronouns 2.3 *Broadcasts (2.7) and Religion (2.9)*

Third Person Pronouns 6.8 *Professional Letters (8.7)*

Pronoun *it* 4.5 *Press Reportage (5.8) and Academic Prose (5.9)*

Past Tense 20.4 *Academic Prose (21.9) and Broadcasts (18.5)*

Agentless Passives 22.6 *Official Documents (18.6) and Academic Prose (17)*

Passive Voice 4.5 *Official Documents (2.1) and Academic Prose (2.0)*

Perfect Aspect Verbs 0 *Academic Prose (4.9)*

Nominalizations 38.5 *Official Documents (39.8) and Academic Prose (35.8)*

Nouns 226.2 *Broadcasts (229.8) and Press Reportage (220.5)*

Prepositions 131.2 *Biographies (122.6) and Broadcasts (118.0)*

As with the Physics textbook, the Chemistry textbook includes essays which purport to bring the science which is being taught in the chapter into the realm of the everyday world. As with the physics textbook, these essays would appear to be more complicated for the foreign language student of science and technology, although the findings in this case conform much more often with Biber's findings for Academic Prose. In some cases, such as Pronoun *it* and Perfect Aspect Verbs the results found for this essay were much lower than in Biber's study but Prepositions were much higher, suggesting that this essay is still outside the usual range found in Biber's work.

McCarthy and Carter (1994:91) argue that "pronouns ... stand in a direct relationship with noun phrases and the demonstratives at the discourse level" and that noun phrases and demonstratives 'topicalize' entities whereas pronouns "simply continue topics already raised to the status of current focus". The combination of higher and lower features found above therefore help to define the identity of this sub-corpus of chemistry, distinguishing it from the main chemistry corpus and

Field Code Changed

highlighting the even greater difference found in the physics corpora studied here.

One of the features that has to be taken into consideration in this sub-corpus is the cultural and historical aspects of the case of the crash of the Challenger space shuttle. Americans could be expected to 'remember' this event as it was a tragedy for a nation who have traditionally found failure difficult to accept. Foreign students could not be expected to share such a collective consciousness on this topic and indeed Chang's essay does not appear in a Portuguese translation.

5.4 Mathematics

Whilst there is no separate mathematical corpus, as this was deemed to represent little in terms of language that could be analysed to provide significant data for materials writing, mathematics does represent extra-linguistic features which must be taken into account to understand the learning difficulties posed by mathematical formulae and numbers in texts. Lemke (1998:104) says mathematics is "more powerful than visualisation, even though it is less intuitive, because it can represent patterns that cannot be visualised, and allow them to be compared, manipulated, combined, etc."

Both the Physics and the Chemistry corpora produced approximately four pages of four columns of figures before the word lists presented in Appendices C and D. The frequency of numbers in the corpora suggest that they are of some significance in the textbooks under analysis. Particularly in physics Serway recommends that students coming to his book should have already have studied calculus and if they have not then at the very least they should be studying it concurrently.

Therefore some observations on numbers and their significance for the students' understanding of text must be made. As was mentioned before, the use of diagrams, figures and pictures usually reinforce the content of the text they are associated with and the same is usually true of numbers and formulae included in the texts; they reiterate the commentary of the text, exemplify or complement the meaning in some way. However, English and Portuguese do not follow the same resolution of mathematical problems. A simple example will illustrate this difference. If one number is divided by another the 'working' of the calculation will be different even though the result should turn out the same. Take, for example, 1526 divided by 32. In English this would appear as follows:

```
      47.6875
32) 1526
     128
      246                    4/128
      224                    7/224
        220                  6/192
        192                  8/256
          280                5/160
          256
            240
            224
              160
              160
               000
```

As can be seen from the above, the answer to this problem, the quotient, is given above the line at the very top of the calculation, 47.687, the divisor is on the left, 32, and the number to be divided inside the frame to the right, 1526. Each subtraction is shown below the number to be divided in a series of steps. The indication that the result is a decimal is given by the punctuation 'full stop' between the whole numbers and the decimals and the necessary 'working' is given to one side of the calculation itself (in this case on the right although there is no hard and fast rule about this positioning).

Field Code Changed

In Portuguese this calculation would look something like the following:

```
1526.000) 32______
 246       47,6875
  220
   280
    240
     160
      00
```

The number to be divided is on the left with a number of zeros added to allow for the decimal places to be represented, 1526.000, and the divisor, 32, on the right, the quotient is given under the line on the right,47,6875, and only the results of the various steps of subtraction are given on the left in the ‘working’ of the problem. In addition, the punctuation used to distinguish whole numbers from decimals is a comma, but a point is used for thousands.

Most Portuguese students when faced with the English style of presenting the ‘working’ of such mathematical calculations find it difficult to work out what is going on with each of the steps, although they are of course essentially the same, the wealth of information provided seems to confuse rather than enlighten.

As will be discussed later in relation to the students and can be observed from both television and film subtitles, numbers are often not readily translated from English into Portuguese. This observation holds true even though the numbers are not being used in expressions such as *six of one and half a dozen of the other,* which would be understandably more difficult to translate as this would obviously require the translator to interpret the phrase to suit the action in the film or television programme. Added to these difficulties are those of the metric and imperial systems of measurement which will be especially marked in these American texts because America still adheres to the Imperial system. Despite the fact that Britain is now almost completely metric, other differences still remain between British and American measurements an American "ton" is lighter than a British "ton" and is, of course, different again to its metric equivalent "tonne" and a British gallon is more than an American gallon. International scientific convention, and European Union regulations, would require metric measurements to be given for everything but, as was described above for the corpora, the fact that the authors have attempted to bring their observations to bear on the everyday world and common American pursuits invites the use of imperial measurements which make up part of that world. The words "foot" and "feet" and "inches", and "miles" and "acres" and "pounds" and "tons" are indeed found in the corpora and will probably cause difficulty as with the confusing billion and ton which are also there. Serway says that he uses metric measurements in all but the engineering sections which he nevertheless keeps to a minimum.

5.4.1 Mathematics in the Gulliver's Travels Text

As mentioned above, one area of potential confusion for the foreign student of science is the use of the imperial system of measurements. In this essay these are evidenced by the use of "inch", "inches" and "foot" as in:

Lilliputians were a little under 6 *inches* high, on the average, and
all built on the scale of one *inch* to the *foot*

The idea of the 'magnitude of things' in the imperial system will continue to be of major importance while America is a major trading nation and world power.

Other mathematical expressions found in the text which could conceivably cause confusion is the use of "square" as in:

Field Code Changed

In other words, the breaking strength of a wire or rope is proportional to its area of cross-section, or to the *square* of its diameter
Because the strength of his columns and braces is proportional to their cross-sectional area and thus to the *square* of their linear dimension

These items are often much more accessible to foreign language students if they are expressed as mathematical formulae when it becomes obvious that these uses of ‘square’ are referring to a number multiplied by itself and not to the shape of an object.

Biber suggests that although “all academic sub-genres are characterized by the features of highly informational production” (frequent nouns, prepositions, attributive adjectives, long words, and high lexical variety), mathematics texts have an even higher score because their subject matter is technical and often non-linguistic, using mathematical expressions instead. With respect to visualising meaning, the *Gulliver's Travels* text does contain some illustrations of relative bone sizes, which should help the student to envisage the relative proportions being described in the text, but the level of abstraction is nevertheless still extremely high as illustrated by the examples given for *square* above.

Chapter 6 Discussion of the Results

Chapter 6

6.1 Discussion of the Results

> *"There is a widespread consensus that language is never neutral and texts are never innocent. Things can always be formulated differently, any linguistic expression of the facts chooses some aspects of reality and downplays others, and all choices are political (Martin, 1985). Representations are always from a point of view, and express group interests. Such points of view are not usually explicit, are often denied and may not be directly observable, because they are often a matter not of individual words, but of patterns of distribution and frequency." Stubbs (1996:235)*

Just as the results from the frequency studies and corpora analyses will only be as relevant to students as the material they are based on is relevant to them, the results of the tests on the students' level of English may well vary with time. The readability levels of scientific texts have been found by Hayes (1992) to have been increasing this century. A study carried out at the University of New York by Dr Linda Hirsch (reported in Rosenthal 1996:116-7) also found that the reading skills necessary to read recent introductory science texts "required reading skills at the college level or beyond, levels difficult to attain for many ESL or post-ESL students" in American universities. One of the introductory science texts that Dr Hirsch included in her comparison was the physics textbook written by Serway studied here. She gave this textbook a reading level of 14 which corresponds to grade 14 or the second year of an American college-level course. Despite the fact that Dr Hirsch considered the physics textbook to be somewhat easier than that

Field Code Changed

of chemistry, this level of readability will be extremely difficult for the students considered by this thesis.

The fact that the reforms being carried out in schools will have an effect on the scientific knowledge the students bring with them to their undergraduate studies must also be taken into consideration. Added to this is the manner in which those subjects were taught. Textbooks produced for students from other cultures build on the understanding of the educational background that the students are perceived to have. If science subjects in schools were taught in isolation or as integrated subjects, such as earth science rather than geography and chemistry and so on, this will have an effect on the type of textbook written. Older books and textbooks from different educational cultures may therefore be at odds with the student profiles which will cause a subsequent increase in difficulty for the undergraduate in a different cultural setting.

Many of the changes that are taking place in education have underlying principles that are based on Europe-wide perceptions of what skills and knowledge the educated person of the future will possess. As students come through to the university from these reforms in the schools both the students' backgrounds and expectations will have changed and the university has to be prepared to meet these different expectations and aims. Moreover, the learning styles and previous educational experiences of the students coming in to the university must be taken into account. If the students' had not studied science in a collaborative way with discussion of the issues raised, this will provide a mismatch with the style of the American textbooks studied which build on this educational background in students rather than one in which rote learning and memorisation are the norm with students not expected to ask questions in class. Flexibility and the ability to reflect on and study these various parameters must therefore remain one of the most important aspects of university education.

The findings from the frequency analyses and corpora studies must therefore be examined in order to suggest what implications they bring to the teaching of undergraduates at university.

6.2 Coursebooks and Multimedia Encyclopaedia Frequency and Range Results.

The results of the comparison between the multimedia encyclopaedia and the Physics and Chemistry coursebooks show some surprising variations, particularly in the area of common words. Some of the words excluded by the Grolier as being too common are not to be found at all in the Physics and Chemistry corpora. Similarly, certain word forms are to be found only in one or other of the Physics and Chemistry corpora. These findings are consistent with the view that distinctions between texts are caused by the relative frequency of use of certain features rather than absolute and unique differences[1]. It must also be remembered that only part of each of the textbooks was included in the corpora so that there is always the possibility that these findings would have a slightly different emphasis if the whole of the textbooks had been included. For example, the frequency of the noun 'bibliography' would have been greater if the bibliographies from the textbooks had been included in the corpus. The results therefore show tendencies and not absolute values.

These frequency lists can be used to show which items are appropriate to include in materials for students and which require particular emphasis. For example, the differences in the usage of the Latin and Greek singular and plural forms found would suggest that

[1] Biber et al. (1998:136) gives the example of 'balls' and 'strikes' being used as countable nouns only in broadcasts of baseball games, as the exceptional, rare situation where these features are found only in that one register rather than being shared with other registers to a greater or lesser degree.

Field Code Changed

teaching materials should reflect this difference rather than being prescriptive and suggesting that only one plural is 'correct' usage when the corpora suggest that actual usage is other than this[2]. The advantage of having access through the corpora to the context of these forms and their range across texts also provides information on the most useful items to be used in each particular situation. The differences found between Physics and Chemistry on words that would have been predicted to be essential for science provide clear guidelines for the context that these should be presented and studied in as was described in 5.1.2.

Stubbs (1996:40) reaches a number of conclusions about work on lexico-grammar that are relevant here. He makes the following points:

1. Any grammatical structure restricts the lexis that occurs in it, and conversely, any lexical item can be specified in terms of the structures in which it occurs.
2. Such restrictions are typically not absolute, but clear tendencies: grammar is inherently probabilistic.
3. Meaning is not constant across the inflected forms of a lemma.
4. Every sense or meaning of a word has its own grammar and each meaning is associated with a distinct formal patterning. Form and meaning are inseparable.
5. Words are systematically co-selected: the normal use of language is to select more than one word at a time.
6. Since paradigmatic choices are not made independently of position in syntagmatic chain, the relation between paradigmatic and syntagmatic has to be rethought.
7. Traditional word-classes and syntactic units also have to be rethought.

Native speakers have only limited intuitions about such statistical tendencies. Grammars based on intuitive data will imply more freedom of

2 Peters (1998:6-12) reports on the Langscape Project of Macquarie University, Sydney, Australia on the Langscape 1 questionnaire on spelling by age group and nationality. The Langscape 4 questionnaire is

> combination than is in fact possible. Grammar is corpus-driven in the sense that the corpus tells us what the facts are. Some of these facts may seem intuitively obvious in retrospect. But they cannot be predicted in advance and they certainly cannot be exhaustively documented from intuition.

The implications of these points is that the use of the corpus material for syllabus design becomes essential rather than merely desirable if the students are to be taught the actual meanings and usage of scientific English and thus helped to cope with their bibliographies in these areas. It is difficult to find a substitute for scientific texts that show the particular syntax and semantics of scientific English which could be used for undergraduate study materials without confounding these forms and thus misrepresenting what scientific English is and means. The frequency lists would be a starting point but collocations and referring back from the lists to the actual texts concerned would be essential. A teaching methodology based on the analysis of comparable or contrastive uses of lexis in physics and chemistry for example would help to highlight the different uses of lexical items in scientific texts. White (1998:268) argues that there are differences in the specialist lexicons of science and technology, science is characterised by lexicon "revaleurisation" that is, after Martin (1993), it establishes categories which reconstrue common sense experiences of reality; technology on the other hand is characterised by lexicon extension which neither challenges nor displaces the vernacular system. In the textbooks studied both of these features are found together, suggesting that these textbooks really represent the two discourses of both science and technology at the same time. This situation is probably a result of the different influences that there have been in science education policy in the United States this century. Matthews (1994) describes this as moving from the practical

---

investigating the issue of the preferences for the plurals from Latin.

Field Code Changed

application of technology and 'general' science giving way to discovery learning then moving on to a much more elitist 'pure' form of science study and finally recently to a more liberal study of science which includes the history of science and discussion of the moral, social, cultural and ethical aspects of the application of science. Matthews also describes how the pedagogic aspects of the pure science curriculum were not taken into account and teachers were not involved in the design of the school curricula which were dictated by scientists alone. This was especially the case after America felt that it lagged behind the USSR when Sputnik was launched in 1959. However, this has since been superseded, as mentioned above, and there is no longer such a centralised curriculum as prevailed at that time and the twin technology (applied science) and pure science elements are now integrated in the modern curriculum.

Furthermore, White (1998:276) argues that technology extends the everyday sense of terms which are possible because "the polysemous nature of much vernacular lexis means that different phenomena may be referenced by the same lexical item". The use of polysemy to extend the sense of lexis makes it an important area to concentrate on in teaching to demonstrate and sensitise the students to this phenomenon in their reading of such textbooks as these.

McCarthy and Carter (1994:102) advocate teaching by means of a contrastive approach to specialised and general texts in order to bring to light the important distinctions and conventionalised patterns in the use of tense, voice and aspect in specific genre. They suggest developing student awareness by analyses of the discourse of the texts presented as a teaching strategy which leads into useful activities that students can employ themselves when faced with a number of similar types of textual pattern like that of the report. In other words they argue that it is possible to teach students how to deal with particular kinds of texts and when they have mastered the techniques taught they will have developed suitable strategies which they can apply to other texts that they need to understand.

The use of the corpora to develop materials for use with students also helps teachers to overcome any preconceptions they might have about what is, or is not, scientific English. The actual instances of use are what is important with no need to invent examples which may be entirely inappropriate. The teaching strategies recommended by McCarthy and Carter described above would also increase the teacher's own understanding of the nature of scientific English, which is often an area that is overlooked. Furthermore, these insights into usage and semantics must be used in the testing that is carried out on the students to ensure that they truly have a command of the English of science and technology rather than a command of the perceived ideas of what the English of science and technology is, from a humanities trained teacher's point of view.

6.3 Textual Features Compared with Biber's (1988) Variation Results.

Biber was dissatisfied by the search for absolute distinctions between speech and writing. He (1988:93) decided that there were no absolutes between these but it was possible to grade texts on a number of factors that were characteristically present or absent in particular contexts. He gave these groups of factors the headings;

Dimension 1 'Involved versus Informational Production',
Dimension 2 'Narrative versus Non-Narrative Concerns',
Dimension 3 'Explicit versus Situation-Dependent Reference'
Dimension 4 'Overt Expression of Persuasion'

Field Code Changed

Dimension 5 ‘Abstract versus Non-Abstract Information’
Dimension 6 ‘On-line Informational Elaboration’

Dimension 1 is concerned with to what extent a text contains those elements that can be associated with spontaneous speech. A high score on this dimension would show discourse that was interactive, fragmented, affective and generalised, associated with not having time to prepare what is going to be said. This is contrasted with text that has a lot of information which has been carefully integrated into it with very precise lexical choice which would require careful planning and time to accomplish and be much more clearly pre-prepared which would be shown as a negative score.

Dimension 2 is very straightforward and deals with the presence or absence of the characteristics of narrative, defined as describing past events and referring to participants in those events. The opposing, negative factors on this dimension are those of a more static, descriptive or expository discourse.

Dimension 3 measures the extent to which a text shows devices for the explicit, elaborated identification of referents. The negative factors on this dimension show reference to places and times outside of the text itself. In other words this dimension distinguishes between highly explicit, context-independent reference and non-specific, situation-dependent reference.

Dimension 4 shows features that function together to mark persuasion: either explicit marking of the speaker's own persuasion (point of view) or argumentative discourse designed to persuade the addressee.

Dimension 5 examines the factors that show a text to be abstract, formal and highly technical.

Dimension 6 distinguishes discourse that is informational but is produced under real-time conditions so that it displays fragmented presentation of information with tacking on of clauses rather than carefully integrated presentation of information.

Biber (1988:94) explains how he standardised the frequencies of the features in each factor so that those features that occurred with great frequency would not have an inordinate influence on the factor score. Applying the same calculations to the results obtained from the physics and chemistry main and sub-corpora, these texts can be compared with the corpus as a whole which Biber examined. All of the features were standardised to a mean of 0.0 and a standard deviation of 1.0. The results are as follows:

Table 6.1. Mean scores of each of the Dimensions compared with Biber's Academic Prose corpus results

	Physics Main Corpus	Chemistry Main Corpus	Physics Sub-corpus	Chemistry Sub-corpus	Biber's Academic Prose
Dimension 1	- 7.65	- 8.03	- 1.8	- 5.96	- 14.9
Dimension 2	- 5.06	- 6.02	- 3.23	- 2.41	- 2.6
Dimension 3	5.34	2.75	1.89	- 1.07	4.2
Dimension 4	- 5.42	- 4.79	1.56	- 2.37	- 0.5
Dimension 5	5.72	3.53	4.45	11.44	5.5
Dimension 6	- 0.73	- 1.57	0.94	- 1.31	0.5

The following figures show the results of the corpora examined in this study compared with Biber's main texts mean scores for each of the six dimensions.

Field Code Changed

Figure 6.1 *Dimension 1 ‘Involved versus Informational Production’*

40	
	telephone conversations
35	face-to-face conversations
30	
25	
20	personal letters
	spontaneous speeches
	interviews
15	
10	
5	
	romantic fiction
	prepared speeches
0	adventure fiction
	mystery fiction
	general fiction; **physics sub-corpus**
	professional letters
	broadcasts
-5	
	science fiction; **chemistry sub-corpus**
	religion; **physics main corpus**
	humor; **chemistry main corpus**
	popular lore
-10	editorials; hobbies
	biographies
	press reviews
	academic prose; press reportage
-15	
	official documents
-20	

6.3.1 Discussion of Dimension 1 'Involved versus Informational Production'

Factor 1 contains features that are present and absent in certain types of texts. The features that make up Factor 1 on the positive scale are:

private verbs
THAT deletion,
contractions
present tense verbs
2nd person pronouns,
DO as pro-verb,
analytic negation,
demonstrative pronouns,
general emphatics,
1st person pronouns,
pronoun IT,
BE as main verb,
causative subordination,
discourse particles,
indefinite pronouns,
general hedges,
amplifiers,
sentence relatives,
WH questions,
possibility modals,
non-phrasal co-ordination,
WH clauses and
final prepositions.

The negative features on this scale are:

nouns,
word length,
type/token ratio, and
attributive adjectives.

Biber explains that the negative features on this factor are all associated with careful, precise presentation of informational content, which is not usually a characteristic of speech, whereas the positive features are characteristic of "on-line" information that is to say,

Field Code Changed

information that is produced immediately or what teachers usually refer to as 'thinking on one's feet' and show involvement and interactive or affective purpose. Biber (1988:132) does not however see this dimension as a distinction between speech and writing *per se* but rather as "the interpretation of involved real-time production versus informational, edited production". White (1998:289) argues that 'hedges' mark one of the differences in the lexico-grammar of the scientific and vernacular technological systems of valeur which would suggest a much more subtle refinement would have to be made between the corpora included in Biber's study in order to separate the scientific from the technological.

Figure 6.1 shows that the results for both the main and the sub-corpora are generally in the same direction as Biber's findings. This is not surprising as the texts are highly informational. There are some features included in this Dimension however, that might help to explain why all of the results are higher than those Biber found for academic prose. For example, one of the factors was WH questions which were found to be significantly higher in both of the main corpora (see Chapter 5, Tables 5.4 and 5.5). This is one of the features found in large numbers in both of the textbooks for undergraduates studied here, which contain several pages of problems for the students to solve at the end of each of the chapters.

The Physics sub-corpus shows a significantly higher frequency of pronoun *it* and analysis done by McCarthy (1994-98:275) provides a tentative conclusion on the uses of *it, this* and *that* in texts, seen in this dimension as demonstrative pronouns and pronoun *it*:

(1) *It* is used for unmarked reference within a current entity or focus of attention.
(2) *This* signals a shift of entity or focus of attention to a new focus
(3) *That* refers across from the current focus to entities or foci that are non-current, non-central, marginalizable or other attributed.

McCarthy (ibid.) sees this kind of finding as raising fundamental questions about “how writers (and speakers) structure their arguments, create foci of attention in texts and signal desired interpretations.” The tentative interpretation I would make here is that the physics sub-corpus displays different argument structures than the other corpora.

In three of the corpora; the Physics main corpus and the physics and chemistry sub-corpora, second person pronouns are also found to be significantly higher than in Biber’s findings (see Chapter 5, Tables 5.3, 5.6 and 5.7). This is because of the essays that are used to demonstrate real-world applications of the theories discussed in the chapters (see 5.3.4 and 5.3.6 for discussion of the essays used in the sub-corpora). The intention of the authors is more ‘involved’ and ‘affective’ in order to teach the reader. Biber, Conrad and Reppen (1998:149-150) suggest that “first- and second person pronouns, *wh*-questions, emphatics, amplifiers, and sentence relatives can all be interpreted as reflecting interpersonal interaction and the involved expression of personal feelings and concerns.” Glaser (1982:78) found that “emotive features and figures of speech alongside with the visual code are predictable characteristics” of the ESP style of using analogies from the learner’s everyday experience.

On the other hand, factors such as nouns and attributive adjectives, which were seen as negative factors for this dimension, were significantly lower in both of the physics and chemistry main corpora with the exception of both the sub-corpora for nouns and the chemistry sub-corpus for attributive adjectives. The effect of this would be to raise the result more towards the centre of the scale, as can be verified in Figure 6.1. The physics sub-corpus then shows affinity with fiction rather than academic prose which is understandable given the topic of *Gulliver’s Travels* as mentioned earlier.

The implications of these findings for teaching and syllabus design is to reconsider whether some other text types and cultural topics should

Field Code Changed

not be included in science and technology courses both for the subjects covered and for the textual attributes that pertain to them. Sports are used consistently as a means to involve the (student) reader, as can be seen from the following extracts from the corpora. In the physics corpora there are references to American Football, Golf and Baseball as in:

Physics Text

> PROBLEMS
>
> 34. A quarterback takes the ball from the line of scrimmage, runs backward for 10 yards, then sideways parallel to the line of scrimmage for 15 yards. At this point, he throws a 50-yard forward pass straight downfield perpendicular to the line of scrimmage. What is the magnitude of the football's resultant displacement?
> 36. A novice golfer on the green takes three strokes to sink the ball. The successive displacements are 4 m due north, 2 m northeast, and 1 m 30º west of south. Starting at the same initial point, an expert golfer could make the hole in what single vector displacement?
> 4. A golf ball is hit off a tee at the edge of a cliff. Its x and y coordinates versus time are given by the following expressions:
> In addition, the spin of a projectile, such as a baseball, can give rise to some very interesting effects associated with aerodynamic forces (for example, a curve thrown by a pitcher).

However, the use of the terms and expressions “quarterback”, “scrimmage”, “golfer”, “green”, “strokes”, “sink the ball”, “tee”, and “cliff” would be particularly dense for foreign language students of science and technology.

The Chemistry corpus contains the following reference to Table Tennis which also requires considerable processing to understand:

> 5.28 A dented (but not punctured) Ping-Pong ball can often be restored to its original shape by immersing it in very hot water. Why?

This corpus also contains references to basketballs and tennis balls but without the same level of difficulty displayed above. The Chemistry corpus also contains such items of a ‘real-world’ type as the following:

5.29 Discuss the following phenomena in terms of the gas laws: (a) the pressure in an automobile tire increasing on a hot day, (b) the "popping" of a paper bag, (c) the expansion of a weather balloon as it rises in the air, (d) the loud noise heard when a light bulb shatters.

5.30 Nitric oxide (NO) reacts with molecular oxygen as follows The heat generated in this reaction helps melt away obstructions such as grease, and the hydrogen gas released stirs up the solids clogging the drain.

The predicted areas of difficulty in the questions illustrated above are American “tire” (British “tyre”), “light bulb shatters”, “grease” and “solids clogging the drain”.

As mentioned above, these items occur most often in the ‘problems’ section at the end of the chapter and are characteristic of this kind of textbook. As such, they need specific attention to help students to cope with this written but interactive style. The fact that these are nothing to do with the language of science implies that the undergraduate textbooks do not conform to what many expect in terms of scientific language. The idea that the language of science excludes other types of discourse is a failing of many ESP coursebooks which must be remedied in the syllabus.

Field Code Changed

Figure 6.2. *Dimension 2 'Narrative versus Non-Narrative Concerns'*

| romantic fiction
7 |
|
|
|
|
6 | mystery, science and general fiction
|
|adventure fiction
|
|
5 |
|
|
|
|
4 |
|
|
|
|
3 |
|
|
|
|
2 | biographies
|
|
| spontaneous speeches
|
1 | humor
| prepared speeches
|
| press reportage
| personal letters
0 | popular lore
|
|
| face-to-face conversations
| religion; editorials
-1 | interviews
|
|
| press reviews
|
-2 | telephone conversations
| professional letters
| **chemistry sub-corpus**
| academic prose
| official documents
-3 | hobbies
| broadcasts; **physics sub-corpus**
|
|
|
-4 |
|
|
|
|
-5 | **physics main corpus**
|
|
|
|
-6 | **chemistry main corpus**

6.3.2 Discussion of Dimension 2 'Narrative versus Non-Narrative Concerns'[3]

The results for Dimension 2 are once again in keeping with the general tendency for academic prose, although the results for the main physics and chemistry corpora are an exaggeration of the tendency towards non-narrative concerns as Figure 6.2 shows. The features that Biber grouped under the heading 'Narrative versus Non-narrative Concerns' were:

past tense verbs,
third person pronouns,
perfect aspect verbs,
public verbs,
synthetic negation and
present participial clauses.

There were no significant negative features used in the calculation of this dimension. Although present tense and attributive adjectives were found to be negative weights, these were not included in the calculation for this dimension.

The features that contributed to the exaggerated negative effect of this result were the significantly lower results found for past tense, perfect aspect verbs and third person pronouns on both of the main corpora, past tense on the physics sub-corpora and perfect aspect verbs on the chemistry sub-corpus. In addition, public verbs were significantly lower on the chemistry sub-corpus. It is interesting to note that if Biber had retained the present tense and attributive adjectives as negative factors on this dimension[4], both the chemistry and the physics main

[3] In Biber, Conrad and Reppen (1998:148) this dimension is relabelled "Narrative versus non-narrative discourse".

[4] The factors for these two were eliminated by Biber because he included each feature on only one factor score in order to maintain their independence although he (1988:89) found them to have factorial scores of - . 47 for present tense verbs and - .41 for attributive adjectives which he regarded as salient in his calculations.

Field Code Changed

corpora for attributive adjectives and the physics and chemistry sub-corpora for present tense would have the effect of emphasising the tendency towards lower results than those Biber found. This implies that the effect of including these features in the calculations on this dimension would have been to produce negative weightings on this dimension with the result that the corpora would have shown an even more extreme negative trend and would have increased the distance from any of Biber's findings even further.

Biber (1988:137-8) sees non-narrative purposes as

> "(1) the presentation of expository information, which has few verbs and few animate referents; (2) the presentation of procedural information, which uses many imperative and infinitival verb forms to give a step-by-step description of what to do, rather than what somebody else has done, and (3) description of actions actually in progress."

These findings would conform to the more conventional view of scientific method and therefore of 'scientific' discourse. However, Trimble (1985:126) presents a much more semantic analysis of the use of these tenses in scientific discourse:

> "if writers use the past tense in reporting research done previously by themselves or by others then that research is of secondary importance to the current work being reported on. If, on the other hand, the writer uses the present perfect or the present tense, then the research is of more direct and primary importance to the writer's current work. Also, the present tense is often chosen when a discussion follows the initial citing of a reference to their own or the others' research and/or when important generalisations are being expressed."

Trimble (1985:123-4) suggests that there are three areas where the non-temporal use of tense regularly occurs in written EST discourse and these are: 1. when apparatus is described, 2. when reference is made to a visual aid, and 3. when previously published research is referred to.

Points 1 and 2, describing apparatus and making reference to a visual aid are significant in the corpora studied with such exhortations in the physics corpus as:

> "See Problem 54 for definition"
> "see Appendix B.2"
> "we see from Equation 3.14 …"
> "Figure 1 shows the …"
> "Fig. 2.14 summarizes the signs of …"

But we also have the exhortation to:

> "Try it and see!"

Which would seem to be a somewhat less formal style than would be expected from scientific prose. The implications of these findings are that the relationship between the stylistic representation of information in academic prose like this needs to be given particular attention. Laurillard (1993-97:27), discussing the fact that academic education is concerned with 'mediating learning', points out that it 'relies heavily on symbolic representation as the medium through which it is known' and although this medium is often language it may also be 'mathematics symbols, diagrams, musical notation, phonetics, or any symbol system that can represent a description of the world, and requires interpretation'. Aiding students in this interpretation process in the scientific and technical fields, as well as with the language itself as it is a foreign one, is particularly important. Laurillard (ibid.) laments the fact that although

Field Code Changed

interpretation has been subject to research at secondary school level, very little has been done at university level in this respect. Further research is necessary to try to clarify the issue of how students interpret graphical and symbolic information, as it is apparent that misconceptions do occur and with understanding of the problem suitable solutions could be found.

Figure 6.3 *Dimension 3 'Explicit versus Situation-Dependent Reference'*

| official documents
7 |
| professional letters
|
6 |
|
| **physics main corpus**
5 |
|
| press reviews; academic prose
4 |
| religion
|
3 |
| **chemistry main corpus**
| popular lore
2 |
| editorials; biographies; **physics sub-corpus**
| spontaneous speeches
1 |
|
| prepared speeches; hobbies
0 |
| press reportage; interviews
| humor
-1 **| chemistry sub-corpus**
| science fiction
|
-2 |
|
|
-3 |
| general fiction
| personal letters; mystery & adventure fiction
-4 |
|
|
-5 |
| telephone conversations
|
-6 |
|
|
-7 |
|
|
-8 |
|
|
-9 | broadcasts

Field Code Changed

6.3.3 Discussion of Dimension 3 'Explicit versus Situation-Dependent Reference'[5]

Dimension three is derived from both positive and negative features on factor three. The positive features are:

WH relative clauses on object position,
pied-piping constructions,
WH relative clauses on subject position,
phrasal coordination and
nominalizations.

The negative weights are:

time adverbials,
place adverbials and
adverbs.

The features that are found to be significantly at variance with Biber's academic prose on this dimension are Total Adverbs, Nominalizations and Phrasal Co-ordination. Total adverbs are found to be significantly lower in all of the corpora while Nominalizations and Phrasal Co-ordination are both found in significantly higher numbers in the main corpora. Phrasal co-ordination is more complex with a discrepancy being found between the physics and the chemistry sub-corpora. The physics sub-corpus has a significantly greater number of these, whilst the chemistry sub-corpus has a significantly lower number. The chemistry sub-corpus also has less analytic negation. The combination of analytic negation, *be* as main verb and non-phrasal co-ordination according to Biber (1988:106) can all be associated with "a fragmented presentation of information, resulting in low informational density." This result then argues for high information density in the chemistry sub-corpus.

[5] In Biber, Conrad and Reppen (1998:148) this dimension is relabelled "Elaborated versus situation-dependent reference" because it is characterised by "highly explicit, context-independent reference versus situation-dependent reference".

Biber (1988:110) says that WH relative clauses together with phrasal co-ordination and nominalization show referentially explicit discourse which is usually integrated and informational. He (ibid.) suggests that this dimension distinguishes between endophoric and exophoric reference (Halliday and Hasan 1976). This would place the chemistry sub-corpus in a different category from all the other academic prose categories and from the main corpora.

Biber, Conrad and Reppen (1998:153) describe the use of *wh*-relative clauses (including pied-piping constructions) as specifying “the identity of referents within a text in an explicit and elaborated manner” whereas time and place adverbials “are used for text-external references to the physical context of the discourse. The following extracts are taken from the beginning and the end of the sub-corpus essay to demonstrate these features:

> When the space shuttle Challenger exploded in flight on January 28, 1986, the crew cabin separated from the rest of the orbiter and broke up when it hit the water. The cabin was equipped with tape recorders to collect shuttle data and record conversations among the crew. However, there was no "black box" to protect the tapes as is used in airplanes. Thus, when the tapes were found six weeks later in 90 feet of water they were considerably damaged by exposure to seawater and resultant chemical reactions. The tapes were described as "a foaming, concretelike mess, all glued together."
>
> The recording showed that at least some of the crew members were aware in the final seconds that the shuttle was in trouble. The impressive fact about this tape-salvaging project is that the principle involved is no more complex than what you would encounter in an introductory chemistry experiment!

Field Code Changed

The extracts show the frequent use of exophoric referencing for example: *space shuttle Challenger; the orbiter; "black box"; The recording; an introductory chemistry experiment.*

And clauses such as: *When the space shuttle Challenger exploded in flight on January 28, 1986; when it hit the water; when the tapes were found six weeks later in 90 feet of water; what you would encounter in an introductory chemistry experiment.*

In his discussion of adjectives and adverbs Biber (1988:237) refers to his earlier work as finding that the distribution of these together with prepositions and subordinate features were varied. He finds that prepositional phrases are most frequent in formal abstract styles and subordination in highly interactive, unplanned discourse. Here again only the physics sub-corpus is significantly different in this respect with a higher figure for prepositional phrases than Biber found for academic prose. All of the corpora show significantly fewer adverbs than Biber's corpora did and therefore, a more negative weighting in this dimension.

The implications of these findings are that the sub-corpora are seen to be difficult to understand if the context is not properly understood. The need for sufficient background knowledge is made particularly apparent in this dimension if the texts are to be understood. Whether the students have sufficient background knowledge to enable them to interpret these texts correctly is then the problem to be investigated. The focus for teaching in this case would be in identification of the referencing in these essays and for learning would be practice with identifying different types of referencing within texts such as these.

Figure 6.4 *Dimension 4 ‘Overt Expression of Persuasion’*

4	
	professional letters
3	editorials
2	romantic fiction
	hobbies
	personal letters
	physics sub-corpus
1	interviews; general fiction
	telephone conversations; prepared speeches
	spontaneous speeches; religion
0	official documents
	face-to-face conversations; humor; popular lore
	academic prose
	biographies; mystery and science fiction; press reportage
-1	
	adventure fiction
-2	
	chemistry sub-corpus
	press reviews
-3	
-4	
	broadcasts
	chemistry main corpus
-5	
	physics main corpus

6.3.4 *Discussion of Dimension 4 'Overt Expression of Persuasion'*[6]

The features used for this dimension are:

prediction modals,
necessity modals,
possibility modals,
conditional clauses,
suasive verbs,
infinitives and
split auxiliaries.

Biber (1988:148) sees all of these functions as "overt markers of persuasion in one way or another". Biber (ibid.) regards professional letters and editorials as "opinionated genres", which therefore have a high score on this dimension.

Both of the main corpora in this study show extremely negative scores compared to Biber's findings. However, the physics sub-genre shows a much more positive score than most of Biber's genres, making this a much more opinionated text. The *Gulliver's Travels* text provides the following extract which shows the use of modals and conditionals in this sub-corpus.

> Within our present technology our scaling arguments are important. If we design a new large object on the basis of a small one, we are warned that new effects too small to detect on our scale may enter and even become the most important things to consider. We cannot just scale up and down blindly, geometrically, but by scaling in the light of physical reasoning, we can sometimes foresee what changes will occur. In this way we can employ scaling in intelligent airplane design, for example, and not arrive at a jet transport that looks like a bee - and won't fly.

---

[6] In Biber, Conrad and Reppen (1998:148) this dimension is relabelled "Overt Expression of Argumentation" because (1998:155) this dimension "marks the degree to which persuasion is marked overtly, whether marking the speaker's point of view, or the speaker's attempt to persuade the addressee."

Predictive modals are used to refer to the future and consider events that will or will not occur (e.g., *what changes will occur, won't fly),* possibility modals and conditional clauses are used to consider different perspectives on the issue (e.g., *If we design a new large object, may enter, We cannot just scale up and down blindly, we can sometimes foresee, In this way we can employ scaling)*

In contrast, the other corpora are all to be seen as not involving opinion or argumentation at all. The fact that science texts neither show doubts nor allow alternative points of view or argumentation of the facts presented to the reader may be one of the reasons that they are said to exclude (see Halliday and Martin 1993). On the other hand, once again (see Glaser above) the physics sub-corpus is an example of the author's attempt to be open-ended and include the student reader in the discourse.

The problems that need to be dealt with as seen from the main corpora are therefore similar to the problems that native speakers would have with scientific texts, which is understanding the concepts developed by the authors or as Laurillard (1993-7:27) says "the problems stem from the fact that the two worlds, of everyday knowledge and academic knowledge, are not as synergistic and inseparable as Vygotsky suggested, but are contrasting and separate." Students need to learn what experts are telling them rather than what they can observe from everyday experience, thus they need to develop academic knowledge of the world.

Figure 6.5 *Dimension 5 ‘Abstract versus Non-Abstract Information’*

11 | **chemistry sub-corpus**
|
|
|
10 |
|
|
|
9 |
|
|
|
8 |
|
|
|
7 |
|
|
|
6 |
| **physics main corpus**
| academic prose
|
5 |
| official documents
| **physics sub-corpus**
|
4 |
|
|
| **chemistry main corpus**
3 |
|
|
|
2 |
|
| religion
| hobbies
1 |
| press reviews
| press reportage
| professional letters; editorials
0 | popular lore
|
| humor; biographies
|
-1 |
|
|
| broadcasts
-2 | prepared speeches; interviews
|
| general, science and adventure fiction; spontaneous speeches
| mystery fiction; personal letters
-3 |
| romantic fiction; face-to-face conversations
|
| telephone conversations

6.3.5 Discussion of Dimension 5 'Abstract versus Non-Abstract Information'[7]

The features included in this dimension are:

conjuncts,
agentless passives,
adverbial past participial clauses,
by-passives,
past participial WHIZ deletions and
other adverbial subordinators.

Biber interprets this dimension as distinguishing genres with an abstract and technical focus from the other genres. In this case all of the corpora subscribe to this difference, although the chemistry main corpus and the physics sub-corpus are not as pronounced as the others, they are nevertheless considerably higher than any of the other genres except official documents. The chemistry sub-corpus is much higher than any of the others and the extract given earlier (Dimension 3) demonstrates why, with the large number of passives: *was equipped, were found, were considerably damaged by exposure to seawater, tapes were described, were aware* and the WHIZ deletion *the principle involved is no more complex*. The fact that the agent is dropped or demoted is interpreted as resulting in a "static, more abstract presentation of information". Agentless passives are four times more common than by-Passives in the corpora which suggests that the writers state facts not "hypotheses for which some named source is responsible, and which are open to different interpretations" (Stubbs 1996:147). Biber, Conrad and Reppen (1998:76) find that there is a preference in academic prose for "generalised states

[7] In Biber, Conrad and Reppen (1998:148) this dimension is relabelled "Impersonal versus non-impersonal style" because this dimension marks "informational discourse that is impersonal, technical, and formal in style versus other types of discourse."

Field Code Changed

and processes” rather than “the descriptions of specific people performing actions” which will be found in fiction and conversation.

As was mentioned above for overt expression of persuasion, the problem to be approached here is the same as that found in native speaker students in higher education having to learn the academic representations of the world which are presented as facts rather than negotiable concepts where argumentation is possible. Many teaching materials produced for students of science and technology have treated the use of the passive as a transformation of the active voice in sentences, which is to misrepresent the *meaning* of the passive voice. Discourse analysis has shown how power and authority are confirmed through the use of the passive and foregrounding of information in texts (see van Dijk 1997, Stubbs 1996). The use of the passive is deliberate in science and technology texts to achieve this authoritarian position *vis-à-vis* the reader, who is therefore not permitted to question the “laws” put forward in the text.

Figure 6.6 *Dimension 6 'On-Line Informational Elaboration'*

| prepared speeches
|
|
| interviews
3.0 |
|
|
|
| spontaneous speeches
2.5 |
|
|
|
|
2.0 |
|
|
|
|
1.5 | press editorials; professional letters
|
|
|
|
1.0 | religion
| **physics sub-corpus**
|
|
|
0.5 | academic prose
|
| face-to-face conversations
|
|
0 |
|
|
| bibliographies
|
-0.5 |
|
| hobbies; **physics main corpus**
| popular lore
| press reportage; official documents; telephone conversations
-1.0 | press reviews
|
| romantic fiction
| broadcasts; **chemistry sub-corpus**
| personal letters
-1.5 | humor
| general fiction and science fiction; **chemistry main corpus**
|
|
| mystery and adventure fiction
-2.0 |
|
|
|
|

Field Code Changed

6.3.6 Discussion of Dimension 6 'On-Line Informational Elaboration'

The features that have high positive weights on this dimension are:

THAT clauses as verb complements,
demonstratives,
THAT relative clauses on object positions, and
THAT clauses as adjective complements.

Biber interprets this dimension as indicating informational elaboration under strict real-time conditions, which would explain why speeches have a high score on this dimension. Biber (1988:159) also suggests that the features on this dimension "enable a direct encoding of attitude or stance" which would explain the relatively high results for professional letters, editorials and religion. The physics sub-corpus is just below the latter which suggests that it is more attitudinal or takes a particular stance more than is usual in academic prose. One of the other characteristic features that Biber found for this dimension but did not include (because its factorial weight was only .32) was existential *there*. This feature is found to be significantly lower in the physics main corpus than that found by Biber. Stubbs (1996) finds existential *there* to be used with an unattributed source which is therefore impersonal and no responsibility is given for the information presented in this way. The code sources are of objective, public knowledge. The implication is that the texts generate their own truth in this way without the reader being involved in interpretation.

Examples of THAT complements and demonstratives from the physics sub-corpus are given below:

> Whereas, if the size of a body be diminished, the strength of that body is not diminished in the same proportion, indeed, the smaller the body the greater its relative strength. Thus a small dog could probably carry on his

> back two or three dogs of his own size; but I believe that a horse could not carry even one of his own size.
>
> If we go far enough toward the very small, surfaces no longer appear smooth, but are so rough that we have difficulty in defining a surface. Other descriptions must be used. In any case, it will not come as a complete surprise that in the domain of the atom, the very small, scale factors demonstrate that the dominant pull is one which is not easily observed in everyday experience. Such arguments as these run through all of physics.

It is interesting to note the use of "I believe" here which reinforces the idea of expression of attitude in this sub-corpus. This example appears in an imagined dialogue with Galileo included in the essay. Quotation is often a source of unusual examples in corpora as they may refer to some period of time from which modern language usage has altered. Recognition of 'dated' language, although obvious to the native speaker, would require advanced knowledge of the foreign language student. This stylistic device in the text would only serve to confuse the foreign language student with the use of such features as the subjunctive, which was seen through the tests to be one of the features that most of the undergraduate students coming into the university in the first year science and technology courses were not much acquainted with.

All of the other corpora examined produced negative results on this dimension which suggests that they neither express attitudes nor opinions nor had any kind of time constraints on their production. Biber, Conrad and Reppen (1998:75) find that there is a preference for *to*-clauses rather than *that*-clauses in academic prose because "academic prose has a preference for static rather than dynamic packaging of information". Academic prose presents the idea that 'this is so and has always been so', rather than the idea that change is taking place constantly.

Field Code Changed

6.4 Academic Prose Sub-Genres

Biber found considerable variation amongst the sub-genres of academic prose. He examines the following sub-genres: Natural Science, Medical, Mathematics, Social Science, Politics/Education, Humanities, and Technology/Engineering for the six dimensions. Consideration of the relationship between the corpora studied here and Biber's sub-genres will also provide further insights into where these text-types fit in the pattern of the academic prose genre as studied by Biber.

Table 6.2 shows the main physics and chemistry corpora compared with the sub-genre studied by Biber on the six dimensions. Table 6.3 shows the sub-corpora relative to Biber's sub-genre which were taken from the major LOB or London-Lund sub-categories. The LOB corpus has been the subject of criticism for causing difficulties with the studies carried out to produce the Leuven English teaching vocabulary list because of its journalistic leanings (and also the Brown Corpus - Goethals, Engels and Leenders 1990 mentioned earlier). This may also have an effect here. White (1998:282) finds there are differences between what would be expected in popular technological magazines in such things as their use of explications of acronyms, although he found popular science magazines met expectations in this regard. A word of caution must therefore be given in the interpretation of these findings, which may be affected by inappropriate text-type style being included in Biber's study taken from the LOB corpus..

Table 6.2 The main physics and chemistry corpora compared with Biber's Academic Prose sub-genres.

	Physics Main Corpus	Biber's sub-genre	Chemistry Main Corpus	Biber's sub-genre
Dimension 1	-7.65	-4.4/-14.0 Maths/Soc. Sc.	-8.03	-4.4/-14.0 Maths/Soc. Sc.
Dimension 2	-5.06	-4.1 Tech/Eng.	-6.02	-4.1 Tech/Eng.

Dimension 3	5.34	5.1 Soc. Sc.	2.75	2.7 Nat. Sc.
Dimension 4	-5.42	-2.1 Nat. Sc.	-4.79	-2.1 Nat. Sc.
Dimension 5	5.72	7.3 Med.	3.53	3.4 Soc. Sc.
Dimension 6	-0.73	-0.8 Nat. Sc.	-1.57	-0.8 Nat. Sc.

Table 6.3 The physics and chemistry sub-corpora compared with Biber’s academic Prose sub-genres

	Physics Sub-Corpus	Biber’s sub-genre	Chemistry Sub-Corpus	Biber’s sub-genre
Dimension 1	-1.8	-4.4 Maths.	-5.96	-4.4 Maths.
Dimension 2	-3.23	-3.1 Maths.	-2.41	-2.6 Nat. Sc.
Dimension 3	1.89	2.7 Nat. Sc.	-1.07	---
Dimension 4	1.56	2.6 Pol./Ed.	-2.37	-2.1 Nat. Sc.
Dimension 5	4.45	3.7 Pol./Ed.	11.44	9.7 Tec./Eng.
Dimension 6	0.94	0.9 Pol./Ed.	-1.31	-0.8 Nat. Sc.

The following figures show the results of the corpora examined in this study compared with Biber’s mean scores for the academic sub-genres for each of the six dimensions.

Figure 6.7 *Dimension 1 ‘Involved versus Informational Production’ for the Academic Prose Sub-Genres*

```
0      |
       |
       | physics sub-corpus
       |
       | Mathematics
-5     |
       | chemistry sub-corpus
       | physics main corpus
       | chemistry main corpus
       |
-10    |
       |
       |
       |
       |
       | Social Science; Technology/Engineering
-15    | Humanities; Politics/Education
       |
       | Medical
       | Natural science
       |
-20    |
```

Field Code Changed

6.4.1 Discussion of Dimension 1 'Involved versus Informational Production' for the Sub-Genres

Biber finds the differences on this dimension to be relatively small for the academic prose sub-genres that he considered and indeed the corpora studied here also fall into a generally negative grouping on this dimension. Biber (1988:193) argues that Mathematics has a higher score on this dimension because its subject matter is "technical and sometimes non-linguistic, using mathematical expressions instead". All of the corpora studied here fall closer to the Mathematics sub-genre rather than any other and the extensive use of non-linguistic information in these textbooks has already been discussed earlier (*3.1.9.3 Formulae, Numbers, Equations and Tables*).

The distinction which White (1998:267) made between science and technology lexicogrammar does not appear to have had any significant effect on the overall results for the corpora studied here. However, other distinctions that White made between science and technology such as the use of Latin/Greek borrowings in science and the use of acronyms (or provisional or proto-nouns) in technology are not specifically included in the parameters used by Biber and therefore were not applied in this dimension on the corpora studied. This is an area that could be researched in the future to see to what extent the corpora represent a balance between the lexicogrammar of 'pure' science and the lexicogrammar of technology as defined by White (1998).

Figure 6.8. *Dimension 2 'Narrative versus Non-Narrative Concerns' for the Academic Prose Sub-Genres*

0	
-1	
	Medical
	Humanities
-2	
	chemistry sub-corpus
	Natural Science
	Social Science; Politics/Education
-3	Mathematics
	physics sub-corpus
-4	Technology/Engineering
-5	**physics main corpus**
-6	**chemistry main corpus**

6.4.2 Discussion of Dimension 2 'Narrative versus Non-Narrative Concerns' for the Sub-Genres

Biber finds this dimension to exhibit considerable variation among the sub-genres. He (1988:193) argues that the higher score found for Humanities shows "a topical concern for concrete events and participants" characterised by historical and biographical studies which describe and analyse events in the past. In contrast, the Technology/Engineering sub-genre shows "concern with abstract concepts and findings rather than events in the past" characterised by philosophical and analytic studies which deal with abstract, conceptual information. The differences found in the corpora used in this study have

Field Code Changed

already been mentioned above, showing the significant difference in the use of the past tense in these corpora. It is not surprising that the sub-genre that are most closely associated with the both main corpora findings and the physics sub-corpus finding are closest to Technology and Engineering as these areas are specifically mentioned in the applications that the textbooks discuss to give analogies with the theory mentioned. The Chemistry textbook says that these real-world applications are from technology, engineering, biology, and medicine.

Figure 6.9 *Dimension 3 'Explicit versus Situation-Dependent Reference' for the Academic Prose Sub-Genres*

6	
	physics main corpus
5	Social Science
	Technology/Engineering; Politics/Education
4	Medical
	Mathematics; Humanities
3	
	Natural Science; **chemistry main corpus**
2	
	physics sub-corpus
1	
0	
-1	**chemistry sub-corpus**
-2	

6.4.3 Discussion of Dimension 3 'Explicit versus Situation-Dependent Reference' for the Sub-Genres

Biber interprets the difference found on this dimension as relating to the explicit and elaborated identification of referents as exhibited by the Technology/Engineering prose, contrasting with the lack of this in

Natural Science which shows situation-dependent reference rather than inexplicit reference. Biber (1988:193) argues that texts taken from disciplines such as geology, meteorology, and biology deal with "specific aspects of the physical environment and thus make extensive reference to that environment." The main corpora studied here divide equally between these two positions, the physics main corpus being on a par with the Technology/Engineering prose and the chemistry main corpus on a par with Natural Science. This may not be at all surprising but it will have to be taken into consideration in developing the syllabus. The subject matter of the Physics sub-corpus *Gulliver's Travels* was shown earlier (5.3.4) to have extensive biological referencing in it which may explain why it is even further away from the Technology/Engineering sub-genre and closer to Natural Science than the main physics corpus.

6.4.4 Discussion of Dimension 4 'Overt Expression of Persuasion' for the Sub-Genres

Figure 6.10 shows the results for dimension 4 'Overt Expression of Persuasion'. Biber (1988:194) finds dimension 4 to show how Political/Education texts are quite persuasive or argumentative, while Social Science prose is "more typical of academic exposition in being non-persuasive". This difference he attributes to the purpose and style of the sub-genre. The higher scores show the "extent to which an author considers alternative points of view and argues persuasively for a particular perspective." The more the study relies on experiment or empirical data the less it depends on "the logical comparison of alternatives and the use of persuasive form". Once again the physics sub-corpus demonstrates the difference in stance adopted by the author when he relates the experimental or empirical data given in his exposition with the real world applications of the facts.

Field Code Changed

Figure 6.10 *Dimension 4 ‘Overt Expression of Persuasion’ for the Academic Prose Sub-Genres*

3 |
|
| Politics/Education
|
|
|
2 |
|
|
| **physics sub-corpus**
|
|
1 |
|
|
|
|
|
0 |
| Mathematics
| Technology/Engineering
|
| Humanities
|
-1 |
|
|
|
| Social Science
| Medical
-2 |
| Natural Science
| **chemistry sub-corpus**
|
|
|
-3 |
|
|
|
|
|
-4 |
|
|
|
|
| **chemistry main corpus**
-5 |
|
| **physics main corpus**
|

Field Code Changed

Figure 6.11 *Dimension 5 ‘Abstract versus Non-Abstract Information’ for the Academic Prose Sub-Genres*

11 | **chemistry sub-corpus**
|
|
|
10 |
| Technology/Engineering
|
|
9 |
| Natural Science
|
|
8 |
| Mathematics
|
| Medical
7 |
|
|
|
6 |
| **physics main corpus**
|
|
5 |
|
| **physics sub-corpus**
|
4 |
| Politics/Education
|
| Social Science; **chemistry main corpus**
|
3 |
| Humanities
|
|
|
2 |
|
|
|
1 |
|
|
|
0 |

6.4.5 Discussion of Dimension 5 'Abstract versus Non-Abstract Information' for the Sub-Genres

Biber (1988:194) describes the difference between the sub-genres on this dimension as demonstrating the distinction between the strictly technical and abstract which therefore "do not deal with specific participants or events" and those which are less technical in nature. He (ibid.) argues that this strictly technical style where "empirical studies are factual, and therefore faceless and agentless" from a linguistic point of view might be seen to result from the style that all scientists and engineers are explicitly taught. Humanists Biber argues "are taught (and teach) that passives are dispreferred constructions and that good writing is active", a position that is echoed by word processing grammar checkers! From this point of view the Chemistry main corpus is less abstract or technical (or more humanistic) whilst its sub-corpus is extremely abstract and technical. The wide range to be found within this one textbook argues for considerable variety in the text-types that should be used with undergraduates to present and practice language through.

Investigation into the styles adopted by teachers in schools in Portugal to teach scientific writing would also help to shed some light on the cultural expectations of undergraduates with regard to the need to use passives in science and technology. Some undergraduates have considerable difficulty in reorganising information into a series of logical steps like those found in the report of an experiment. This could indicate a difference in report writing styles between different cultures.

Field Code Changed

Figure 6.12 *Dimension 6 'On-Line Informational Elaboration' for the Academic Prose Sub-Genres*

```
      |
      |
      |
      | Mathematics
3.5   |
      |
      |
      |
      |
3.0   |
      |
      |
      |
      |
2.5   |
      |
      |
      |
      |
2.0   |
      |
      |
      |
      |
1.5   |
      |
      |
      |
      | Medical
1.0   |
      | Politics/Education; physics sub-corpus
      |
      | Social Science
      |
0.5   |
      |
      |
      | Technology/Engineering
      | Humanities
0     |
      |
      |
      |
      |
-0.5  |
      |
      | physics main corpus
      | Natural Science
      |
-1.0  |
      |
      |
      | chemistry sub-corpus
      |
-1.5  |
      | chemistry main corpus
      |
      |
      |
```

6.4.6 Discussion of Dimension 6 'On-Line Informational Elaboration' for the Sub-Genres

Biber (1988:195) explains how his earlier description of this dimension taking the *that* compliments to be a characteristic of discourse that could not be carefully planned and integrated does not preclude their use in logical development in a text or for emphasis which indeed characterise "all academic prose sub-genres ... to some extent" and demonstrates their primary use in such genres. For this reason the Mathematics sub-genre is seen to be divergent because of the use of formulae and argumentation in this sub-genre which requires *that* compliments to mark logical relations within the text. The corpora studied here do not show such an extreme position as that found in the Mathematics sub-genre but, nevertheless, conform to Biber's idea that these features are to be found in all academic prose to some extent.

6.5 The English of the Students in the First Year of University.

Two clear trends are apparent from the test results obtained for the new students in the first year of the university. First, it seems that the students are studying more English at school. This result is confirmed by other factors such as the increase in the number of students studying the English language in schools particularly in the fifth and sixth years, despite the changes in demographics that the schools are undergoing. The second feature that can be observed is that the more years of English the students have had at school the better their results are on the preliminary test. This may seem obvious, however the view that the students are not learning anything, and that the same material is repeated over and over again in each school year as a result of this, is often bemoaned at conferences on teaching and meets with

Field Code Changed

widespread sympathy from teachers. These results run counter to that general idea. The students are indeed learning more English when they have more time dedicated to the study of the language. This is not to say that the students have learnt the specific language of science and technology but it does suggest that strategies for dealing with the comprehension of texts have been developed and that these strategies can be applied by the more advanced students in other situations. Rosenthal (1996:19), reviewing the research on second language students and academic success in further education, finds that the level of language proficiency necessary to ensure academic success takes five to seven years to develop. She is discussing the American system which has both immigrants whose first language is not English and foreign students in higher education and notes that there are many different systems in operation within the universities and colleges in the United States to teach English to such students. However, she also points out that this time factor cannot be overlooked but is a considerably longer period of time than that allotted to any of the language programmes in the colleges and universities. Added to this she recognises that the fact that faculty members in the other academic disciplines often have no idea how students learn English which leads to a separation of English and the study of the academic subject matter. She recognises that this situation is no longer appropriate because the acquisition of English occurs best when students are using the new language purposefully but that many mainstream faculty members have unrealistic expectations of what students can achieve within the confines of the language classroom. Despite the very different circumstances between the American system she is describing and the Portuguese system examined here, the latter unrealistic expectation holds true. In other words, time is of the essence but so is contact with the language in purposeful contexts.

The results that show that the students have a greater ability with some of the structures traditionally regarded as difficult or more advanced and struggle with those traditionally considered simple call into question these distinctions. Examination of the irregular verbs in the frequency lists for the chemistry and physics corpora reveal that the results of Mindt's corpora analyses (1997:47-49) are not duplicated in the results of this study. Mindt's top ten irregular verbs (apart from *do, have* and *be*) are *say, make, go, take, come, see, know, get, give* and *find*. Of these *say, go*, and *come* are not found with any frequency in either of the main corpora and *take* and *get* are only found in one of them. On the other hand, *show* and *write* are extremely common in these corpora. Halliday (1993:19) working on the original COBUILD corpus provides a list of the first 25 most common verbs, once again without the most frequent ones of all *be, have* and *do*. This list does not coincide with Mindt's list entirely either. The most common irregular verb forms as given by Halliday (ibid.) include *think* and *tell* before *find* and *give*. Added to this is the problem of how many of these verbs are actually found in the past tense forms. Halliday (1993:21) finds that the use of the present and the past tenses is approximately the same, however, the corpora studied show a preference for the present tense. These findings suggest that the syllabus must be built on the evidence contained in the frequency listings of the corpora used here, rather than on any other arbitrary corpus, or indeed other studies conducted for other purposes on other material, in order to be both useful and relevant for the students. (see McCarthy and Carter 1994:20).

Similarly, the grammatical structures taught should appear in actual contexts of use so that they reflect the meanings and common usage of these kinds of textbooks. Biber, Conrad and Reppen (1998) mentioned above, Trimble (1985), Stubbs (1996) and McCarthy and Carter (1994) (among others) have all demonstrated how both the

Field Code Changed

grammatical choices and their meanings differ between genres and text-types. The students were not specifically tested to see if they could distinguish between general English usage and the specific science and technology usage of certain grammatical items. Nevertheless, their test results do show that many of the students have not grasped the general usages they should have learnt in school.[8] Of particular importance is the specific use of modals, conditionals and irregular verbs and their meanings in scientific texts.

[8] Mindt (1997:43-45) argues that the grammars used in schools (in Sweden) do not show the most common forms of usage of *some* and *any* (amongst others) and so misrepresent the language as used by native speakers in any case. Therefore he argues for a new approach to didactic grammar based on corpus analysis which would show the most frequent usages for students to learn in a graded manner.

Chapter 7 The Syllabus

Chapter 7

The Syllabus

The language of science and technology is an area where new information is being introduced more and more quickly; inevitably new discourses will also develop quite quickly. Changes are being introduced in schools because of new information and because of new attitudes to the world engendered by a new understanding. A more environmentalist and interdisciplinary approach to learning new discourses (Stubbs 1994, 1996, Kerridge and Sammells eds. 1998, Pope 1998 gives an overview) is being put forward and these changes will have to be studied by educators and students alike. Areas such as e-mail can be seen to be an altogether new, recently invented genre, whereas many other new discoveries can be seen to have arisen from other genres and share some of the characteristics of those other genres. For example, in technology there are many new inventions the descriptions of which nevertheless follow some of the discourse characteristics of the genre of technological style. What then can be done to prepare students to study these and other new genres? Stubbs (1992:206) says

> IT extends literacy and therefore requires new study skills for school and work: the newer technologies can provide access to learning resources, such as documentary films and databases, and therefore lead to an increase in collections of authentic language products for teaching about the uses of language. The key question is therefore accessibility. The diversity of self-access teaching materials and study packages, and the increased availability

Field Code Changed

of CD-ROM, video disk, satellite television, etc. will require an increase in study skills to access such materials.

7.1 Study Skills

ESP can be broken down into other sub-divisions like EAP, which is broken down further by Jordan (1997:3) into ESAP (that is, English for Specific Academic Purposes and EGAP (English for General Academic Purposes), although he reports that the more usual model in the USA is to break EAP down into EAP and EST. EAP is seen by Jordan to cover both the more specific focus of a subject such as engineering and also the skills and proficiency in formal academic style and register that the students need for study purposes. In either division, EAP can be taken as the short term objective of the undergraduates studying on the science and technology courses analysed here. This is particularly relevant as a teaching objective given the constraints imposed by both the size of the classes, their heterogeneous nature and the paucity of time available. The objective of the discipline would be to cope with the students' immediate course concerns. However, the medium and longer term needs of the students cannot be ignored and a process whereby the students can be given the skills to continue learning (or the "learning to learn" dimension) must not be ignored.

Jordan (1997:7) suggests the following breakdown of study skills:

STUDY SITUATION/ACTIVITY	STUDY SKILLS NEEDED
1. *lectures/talks*	1. listening & understanding 2. note-taking 3. asking questions for: repetition, clarification and information
2. seminars/tutorials/discussions/supervisions	1. listening and note-taking 2. asking questions - as above 3. answering questions; explaining 4. agreeing and disagreeing; stating points of view; giving reasons; interrupting 5. speaking with(out) notes: giving a paper/oral presentations, initiating comments, responding; verbalising data

3. *practicals/laboratory work/field work*	1. understanding instructions: written and spoken, formal and informal 2. asking questions; requesting help 3. recording results
4. private study (journals and books)	1. reading efficiently: comprehension and speed 2. scanning and skimming; evaluating 3. understanding and analysing data (graphs, diagrams, etc.) 4. note-making; arranging notes in hierarchy of importance 5. summarising and paraphrasing
5. *reference material/library use*	1. using the contents/index pages 2. using a dictionary efficiently 3. understanding classification systems 4. using a library catalogue (subject and author) on cards, microfiche and computer 5. finding information quickly (general reference works and bibliographies) 6. collating information
6. *essays/reports/projects/case studies/dissertations/theses/research papers/articles*	1. planning, writing drafts, revising 2. summarising, paraphrasing and synthesising 3. continuous writing in an academic style, organised appropriately 4. using quotations, footnotes, bibliography 5. finding and analysing evidence; using data appropriately
7. *research* (linked with 3-6 above)	*in addition to 3-6 above:* 1. conducting interviews 2. designing questionnaires 3. undertaking surveys
8. *examinations:* a) *written*	1. preparing for exams (techniques) 2. revision 3. understanding questions/instructions 4. writing quickly: pressure of time
b) *oral*	1. answering questions: explicitly, precisely 2. explaining, describing, justifying

He describes the skills necessary for these activities to be the following:

Skills generally applicable:

1. organising study time efficiently, i.e. time management
2. logical thinking: constructing arguments - use of cohesive markers and connectives; recognising weaknesses and bias in arguments; balance; critical analysis
3. accuracy
4. memory: recall; mnemonics
5. using computers/word processors

Field Code Changed

Jordan (1997:8) notes that the term *reference skills* is sometimes confused with the generic term *study skills*. In Jordan's description of study skills there are a number of repetitions e.g. listening, note-taking and asking questions, in the various situations that students can be expected to find themselves. Jordan's framework was produced with students studying in an English-speaking environment in mind, so some of these items will not be the most crucial for undergraduate students in Portugal, although in the initial needs analysis from other departments one mention was made to listening to lectures given in English which would suggest that in some circumstances there may be greater similarities with an English-medium educational situation. Moreover, the use of questions was found to be of particular importance in the textbooks studied so written questions will need specific attention on the syllabus for these specific undergraduates. Stubbs (1992:208-9) points out that Jordan's point 5 above – reference skills may be different using CD-ROM material than for conventional library selection of material. He questions what novel uses and representations of language and knowledge are brought about by computer systems and concludes that "Computers require precise and accurate instructions, and the production and interpretation of clear and precise information is an important goal of English teaching."

Jordan assumes that study skills can be transferred from one area into another, that is, if the students are proficient in study skills in their mother tongue, they can transfer these skills to the foreign language situation. He (1997:5) does accept that students might need help with this and also with adjustment from school to a different academic environment. He also recognises that there are other elements of EAP that may need addressing, such as style, and that some students will not be proficient in study skills even in their mother tongue. In this particular instance, it is both the textbook which represents and relates to students from a different culture and academic environment and the fact that most of the students on the first year discipline are also in the university environment for the first time.

Laurillard (1993:2) argues that at undergraduate level it is unrealistic to expect students to take control of their own learning, but goes on to show how she sees the student developing academic knowledge through mediated learning. Waters and Waters (1992:264) say that in their experience "what students frequently lack is not only a knowledge of study skills, but, more fundamentally, the underlying competence necessary for successful study - self-confidence, self-awareness, the ability to think critically and creatively, independence of mind and so on." The general development that is taking place in undergraduates who have just made the transition from school to university must also be addressed together with the more straightforward aspect of the lack of study skills. One means of achieving this development in students is through different methodologies in the classroom which encourage increasing confidence through working on different levels of material with success and interacting with classmates in pairs or small groups to avoid shyness and ridicule and to encourage the sharing of knowledge between students. More general discussion encouraging students to explore their own ideas and opinions on topics which are given proper consideration and treated with respect by the teacher and other students can aid self-confidence and creativity. Different schemes which allow students to interact with the lecturer on a more personal basis such as outside the classroom in attendance hours and through different means of directing questions which may be personal (through e-mail) or taken up with the whole group by the lecturer if they are found to be more generalised can also lead to success and have the added advantage that the lecturer reflects upon what has been successfully achieved and what has not be grasped by the students.

Field Code Changed

Rosenthal (1996:150-174) describes some approaches adopted in the United States which encourage lecturers to evaluate their own performance and laments the fact that often the principles that guide scientific research are not applied to science teaching. She (1996:178) sums this up thus:

> Few researchers would maintain the same protocol if experiment after experiment failed. They would reevaluate their hypotheses, reassess the literature, check for errors, develop new procedures, and redesign their experiments. Yet, when it comes to their teaching, some of the same individuals resist any form of change, any innovation, which might improve student interest and achievement in science.

Rosenthal admits that it might seem rather odd to be comparing and contrasting such different fields as science and second language acquisition but she feels that these two merge in the students who are trying to study science (in America) and who have limited English. Despite the need being greater in students studying science where the language of instruction is English, many of the problems she outlines apply equally to foreign language students who need to understand English for their science studies and so reflection by the lecturer on their teaching to aid success in science learning is essential.

7.2 Student Needs and the Syllabus

If decisions are taken based on the needs analysis provided by other departments and the student profiles gained from testing, the following areas will have to be addressed in the discipline:

i) The classes will have to be taught in English, thus exercising the students' listening skills and preparing them for lectures given in English in some of the Departments. The use of video recordings of actual lectures given in this or other universities in English would be important to provide the students with an appropriate context for language study. This practice would also pave the way for attendance at conferences where English would almost certainly be one of the conference languages. [1]

ii) Both American and British English will have to be included especially for comparative purposes as students may have come from different educational backgrounds where one or the other of these Englishes will probably have been taught.

iii) Basic scientific reading texts must be used as a core for the discipline and text attack strategies will have to be taught because the students are also coming to grips with the more advanced scientific subject matter (even in Portuguese). Discourse and text analysis - study of such aspects of scientific texts as cohesion, pronoun referencing, deixis, linking words, cause and effect, definition and classification. It would be possible to use corresponding texts in different subject areas with respect to these (McCarthy and Carter 1994) or even to conduct analyses of Portuguese and English texts on the same subject (Leech 1997).

iv) Mathematical symbols, formulae and numbers in British and American English will have to be taught, including contrasting Portuguese and English use, as these are likely to have been omitted on school courses but are extremely important to the understanding of the scientific texts on the students' bibliographies. Formulae would also involve the revision of the alphabet and learning how equations are put into words or form part of the text being studied. The patterns created by the use of mathematical functions, graphs and tables are

[1] Many of the Departments in the University of Aveiro promote conferences aimed at undergraduates, usually in the final years of their courses, which may have international speakers. The Departments of Management and Tourism (amongst others) also often have visiting scholars some of whom were American and lecture undergraduates in English.

Field Code Changed

considered by Lemke (1998:102) to be “important in the value-scheme of natural science” which the students must be able to perceive in order to understand the relationship between the patterns and assumptions made in scientific theory. Weights and measures in both the metric and imperial systems must be examined including British and American differences. Consciousness raising and alerting the students to areas that are different will include looking at some cultural differences between the two languages as mentioned earlier.

v) Note-taking and summarising should be included in the course as these are skills that will be necessary throughout the students courses. The summaries and notes produced by students might well be in Portuguese if they are for personal use for understanding and storing information for use in other disciplines. The study of comparative texts mentioned earlier might help here to highlight signalling devices in discourse that the students must be conscious of to organise their notes.

vi) Reference skills such as dictionary work will have to be included to provide knowledge of sources of information for the students to further their studies independently but increasingly these reference skills will have to be extended to the use of CD-ROM material and the Internet. Dictionary work focusing on abbreviations, countable and uncountable nouns, spelling and pronunciation is one simple means of providing the students with a means of discovering more about the language when they need to. An attempt to equip the students with the means to proceed further in their studies on their own initiative can be approached through more detailed study of types of definitions; both those used in reference materials and those to be found in the textbooks on the bibliography of both an overt and covert type (Darian 1981) with the corresponding text-type signalling.

vii) The students will have to be taught to interpret graphs, tables and diagrams in English and to recognise the referencing to these in the text and the means by which they complement or add information to the main text. Listening activities where the students have to complete graphs and tables and interpret them into another form must be included. These activities could usefully be carried out in a language laboratory. Laurillard (1993-97:112) rates the combination of audio and visual material as one of the most productive because it gives greater control to the student and could be used to set tasks that “enhance and interpret students’ experience of the world”. She (ibid.) suggests that the visual part need not necessarily be printed material (which is however the most flexible medium), but may be an object which the student has to observe or a situation could be created whereby the audio material guides the student to perform some other operation, such as work on a computer. Lemke (1998:93) argues that the juxtaposition and combination of visuals in texts will multiply the meanings so that “we can mean more, mean new kinds of meanings never before meant and not otherwise mean-able.” The way in which this is achieved has to be explored as “the user must integrate visual and verbal realisations of objects, concepts, relations and processes in the joint interpretation of text and figure.”(Lemke 1998:110).

viii) Appropriate specific vocabulary will have to be taught in appropriate contexts with their usual collocations taken from the corpora, together with the pronunciation of these and the semantic variation that science and technology texts cause in lexis and their associated grammatical structures. Specific grammar will have to be taught with its usual realisations in scientific texts as necessary for the effective realisation of the tasks undertaken. The development of more student autonomy in using the corpora for their own difficulties with science and technology language is important given the differing student backgrounds.

7.3 The Students' Background Knowledge and the Syllabus

It would appear from the list given above that some of the items would already have been taught in MT. However, although teachers can most definitely appeal to frameworks taught in school, it is not a sound notion that everyone thinks in the same "scientific way". An example of this is the "choke" on a car. In Portuguese, in older models of cars, the driver *abre o ar* literally translated as "opens the air " on a cold morning in order to start the car. In English exactly the opposite is done, the "choke " is pulled out, thereby cutting off (choking off) the air to the petrol mixture. The basic scientific principle, of enriching the petrol mixture, is the same but it is not expressed in the same way. In other words, on the surface of this expression there is a different scientific explanation of what takes place. Halliday and Martin (1993:16) argue that there are some minor variations among different languages of how grammar construes phenomena into a scientific theory. They (ibid.) suggest that English and French are different not so much because the grammar of scientific theory is different between them, but because the English language constructs reality more along empiricist lines whilst the French language constructs reality along rationalist lines in scientific theory. Dr Catherine Middlecamp, Director of Chemistry at the University of Wisconsin-Madison (whose report is included in Rosenthal's survey of science teaching for language minority students in the USA, 1996) argues that Western scientists are more inclined to use categories than others. She explains that although the categories into which chemistry is usually broken down such as organic, inorganic, analytic biophysical etc. appear to be culture-free, in reality they are not. Furthermore, she argues that even if two cultures are similar in their tendencies to categorise the world "there is no guarantee that the lines will be drawn in the same places". The question needs to be raised as to whether there is a Portuguese 'scientific way' and if so, what this is. Kaplan (1966:15) describes different scientific discourse patterns in paragraph development employed by different linguistic systems. He finds that there is a difference between English and the Romance languages (which would include Portuguese) because Romance languages include digressions and include extraneous information. The differences that exist between different linguistic systems has significance for Portuguese students who may well have recourse to many of these (scientific) systems. The students may be unaware that there are distinctions between scientific approaches when they are consulting books in different languages. The naming of processes and theories are also often different between the languages of science the French, for example, have not always adhered to the International Scientific (SI) system, preferring to coin their own terms.

Equally well, the fact that reading skills are transferable should be used to help to get the students to read effectively in English. However, a number of the students on this course undoubtedly opted for science and technology because they did not like, or demonstrated less aptitude for, foreign languages. This being the case, it will be an uphill struggle to create the conditions necessary for successful transfer of skills. Halliday and Martin (1997:49) discuss the "ongoing apprenticeship of students into science discourse" which implies that what has gone before has to be taken into account to decide what will follow. An attempt to motivate students, who may well have an active dislike of English, must be made to encourage them to engage with and enjoy the study of science and technology through English. As Hutchinson and Waters (1987:141) say,

Field Code Changed

> Enjoyment isn't just an added extra, an unnecessary frill. It is the simplest of all ways of engaging the learner's mind. The most relevant materials, the most academically respectable theories are as nothing compared to the rich learning environment of an enjoyable experience. This is an aspect of pedagogy that is taken for granted with children, but is too often forgotten with adults. It doesn't matter how relevant a lesson may appear to be; if it is bores the learners, it is a bad lesson.

One of the factors that could add to enjoyment, and therefore to learning, is variety and materials must also be chosen to provide this. However, one of the problems to be overcome in this situation is the fact that the students have different backgrounds in English and different interests in scientific or technological subjects, therefore a range of materials would be necessary to provide practice and to meet the different levels of English which the students bring to the task. The results presented in Chapter 6 point to the different subject areas in academic prose that it would be appropriate to use to provide variety without sacrificing any of the appropriate discourse features that the students have to learn to cope with. Furthermore, it is likely that some of the students stopped studying English more than a year before coming to university and are therefore somewhat 'rusty'.

Halliday and Martin (1993:71) identify the following as the difficulties that students have with scientific English:

1. interlocking definitions
2. technical taxonomies
3. special expressions
4. lexical density
5. syntactic ambiguity
6. grammatical metaphor
7. semantic discontinuity

The approach that they suggest to deal with these difficulties is to analyse any text and to relate it to the context in the discourse, however they point out that this must be done in a principled way. Scientific texts should be analysed and interpreted in contrast with other texts so that the similarities and, more importantly, their differences can be studied. If the tasks are varied, it would be possible to use 'authentic' texts taken from the corpora and at the same time avoid difficulty for those students who have a lower level of English whilst still engaging the students whose English is more advanced. The gradual build-up of both linguistic knowledge and scientific knowledge could then develop with the use of schemata theory and other text attack strategies. Steffansen and Joag-Dev (1984:54) say that no text is explicit and that students need to use “guessing” strategies to understand the message, however, if the student does not share the same background information as that of the author they will “re-interpret vague aspects to conform with their own schemata and will be unaware of other possible interpretations which in fact conform to the author’s schemata.” This danger can be avoided, they suggest, by “establishing a correspondence between what is known …and the givens in a message” whereby the students themselves can “monitor their comprehension and know whether they have understood a text.” This is along the lines of Krashen’s (1982:16) theories of ‘monitoring’ where students can apply what they have learnt about a language to their own production of language. However, he believes this is only possible provided that the students have plenty of time available, correctness of language use is considered important and they can remember the ‘rules’.

The knowledge the students bring to a task could be pooled to benefit the whole group and active student-to-student collaboration encouraged in order to overcome some of the difficulties of teaching such heterogeneous groups. The benefits of students collaborating with each other is that they will be more aware of where the difficulties lie, both in terms of scientific knowledge and language problems and students may

Field Code Changed

find it easier to expose their difficulties to a colleague rather than to the teacher in a large class. Laurillard (1993-97:187) suggests that it is "impossible for teaching to succeed if it does not address the current forms of students' understanding of a subject." She argues that as university education becomes less elitist teachers will recognise that students cannot succeed unaided and there is a need to try to improve the situation by investigating students needs and using the data obtained to design (better) future teaching. The point, as Laurillard (1993:93) says, "is not just to change what is taught, but also how it is learned."

Bazerman (1998:21) suggests the "Difficulties that students and others have with scientific language are in the recognition and appropriate manipulation of the verbal objects which correspond to conceptual objects" and furthermore, "once the object is given a stable name, its details, problems and material peculiarities and relations to other objects in its network vanish in a higher level abstraction which becomes difficult to unpack once made." If this is so, the development of new conceptual frameworks which the students are undergoing in both the science and technology they are studying and the interpretation of these same concepts and their unpacking in English must be borne in mind and allowed for in the syllabus through discussion and examination of what students understand about the different concepts contained in the topics covered.

It is clear from the study carried out here that the language of science and technology is different from that of general English and that these differences need to be addressed by the course (see Chapter 6). The means this thesis proposes for doing this is through data-driven learning.

7.4 Data-Driven Learning

As syllabus design should always be responsive to learning theory, it is important to consider what has been discovered about language acquisition in recent years. Over the last two decades there has been considerable research done on language acquisition which has shown that children acquire language in unanalysed “chunks” or as some researchers term it they use “prefabricated language” in certain, predictable, social contexts. That is to say, children use a kind of formula of undifferentiated morphemes for many situations they encounter and only later gradually refine their perception and use of these “chunks” or “prefabricated pieces” of language. An example of this might be “what is that?” which is rendered as /hw①sdæt/ or /hw①sdæ/ (See Hakuta 1974; Huang 1971; Brown 1973; Clark 1974; Cruttenden 1981; Wong-Fillmore 1976; Newmark 1979, Peters 1983)[2]. The next stage in language acquisition is the analysis of these phrases and the recognition that they are made up of a number of words. Many people can remember an example in second language learning when something was misunderstood because the words used were undifferentiated or differentiated incorrectly. An example of this from Portuguese would be confusion between nouns and verbs with such close items as *a baixa* - *abaixa* and *a travessa* - *atravessa*. Other examples may occur even in the mother tongue when what is heard and how something is written is confused. An example of this would be the word ‘misled’[3], although this is a slightly different aspect of the same phenomenon of undifferentiation.

Some of the phrases used by children are found to be more productive because they allow substitution to a greater or lesser extent. Different items can be inserted in ever greater complexity. An example of this kind of substitution would be ‘Modal + *you* + VP’ as in *could you pass the salt?, could you hand me that pencil?, would you lend me a*

[2] Even in language acquisition there has been a move towards the use of corpora for recent studies. Biber, Conrad and Reppen (1998:172-202) report on analyses of 8-12 year olds using the CHILDES corpora as they found that previous research was often limited in scope, used only one or two subjects and focused on a small number of linguistic features and often a single register.

Field Code Changed

dollar? (Nattinger and DeCarrico 1992:18). Aston (1997:56) argues that "rather than generating and interpreting utterances by combining or analysing morphemes on the basis of generalised grammatical and pragmatic rules, users appear to make use of larger memorised chunks associated with particular types of problem in particular types of contexts, instantiating these as necessary with relatively simple modifications." He claims that work in corpus linguistics by Sinclair (1991) has provided evidence for this view, showing the extent of recurrent co-selection of lexicogrammatical forms with relatively simple variation.

Furthermore, researchers such as Wilson and Sperber (1986 and 1988) suggest that high frequency items require less processing effort on the part of the hearer and that there is a tendency for those phrases which are used very often to become fixed. In English this finding can certainly be confirmed quite simply by looking at the shifts that take place in individual words. For example, some years ago 'to-day' was hyphenated in this way yet now 'today' is written as one word with no hyphenation. Originally it was two separate words, 'to' and 'day'. This trend can be seen over and over again in such words as 'albeit', 'nevertheless' and more up to date still in such words as 'alright' which was written as two separate words until quite recently. Some of these latest changes can often give rise to arguments about 'correct' language use. There are always those who object to any change occurring in language use or spelling and, of course, in other cultures there are even institutions designed to try to stop the rot. The French Academy, for example, attempts to discourage or even prevent the adoption of English words and expressions like *le weekend*. White (1998:285-6) reports that a similar process of shift takes place in the language of technology where

---

[3] The word is read in combination *mi+sle+ d /mizld/ rather than two parts miss + led /mis'led/. The confusion arises because of the overgeneralisation of the -ed ending being seen as a past tense suffix added to the verb which would then be *misle.

the “existence of a lexically minimal term - a single word form - to reference a given category is generally seen as evidence that the category is stable and salient within its ideational domain.”

Nattinger and DeCarrico (1992) suggest that, although much of the research done was concerned with language acquisition in children, there is no reason to believe that adults would go about the language-learning task any differently, and indeed misunderstandings like those mentioned above confirm this. Nattinger and DeCarrico (1992) go even further and suggest that “It is our ability to use lexical phrases, in other words, that helps us to speak with fluency.”

On the other hand, there is increasing evidence from computer corpus-based research that language itself occurs in a largely predictable way. That is, the commonest forms of language occur in overwhelmingly high frequencies and collocations. Collocation here means the co-occurrence of certain words within a short space of each other in a text. A certain word or ‘node’ is the focus of attention and the words to either the left or the right of it are studied, these are called the collocates. The use of a concordance which focuses on the node can reveal important language patterns in texts. Often the position in the sentence can also be revealed by this type of concordance study of language patterns which is an important piece of evidence for students to observe for their own use in writing in English. One of the most important aspects of this computer-based research is that it is reflecting natural language use, that is, it is descriptive and not prescriptive and examples of use are not invented ones which can be, as Sinclair (1991) points out, “extremely unlikely to occur in speech or writing”. Researchers in these areas report that the commonest forms are in the majority in the frequency and collocation studies they have done, no matter how large the corpus they are using. Sinclair says that if those words that occur only once in a corpus were removed the corpus would be reduced by half. He also

Field Code Changed

suggests that "grammatical and lexical distinctions may be closer together than is normally allowed".

What are these common forms and what does this research imply in terms of language learning and teaching? This research does indeed suggest that language is a much more finite system than has hitherto been believed. The commonest language is used most of the time in predictable or "prefabricated" chunks and it should be this language that students should be provided with, in order to give them a rapid, fairly comprehensive grasp of naturally-occurring language in the shortest possible time scale. The idea of teaching what is most frequent has been around for a long time (at least since West in the 1920's); the only danger is that what is most frequent today will not be the same as what is most frequent tomorrow and decisions about what needs to be taught should be based on the most up-to-date data from corpora that are made up from that specific language that is the students' target.

Some of the findings from computer corpus-based research run counter to what intuition about language would suggest and, more importantly, run counter to what coursebook writers believe to be the case. This is true of both meaning, form and usage, as the following examples demonstrate. A list of the commonest meanings of the verb 'see' would include 'using the eyes', 'looking at', 'meeting', 'grasping with the mind or imagination', 'discovering or checking', 'experiencing or witnessing', 'other meanings e.g. accompany or escort' and phrasal verbs (taken from the Oxford Advanced Learners Dictionary in that order).

The actual findings in percentages, however, show that the most common (53% of the Birmingham corpus), examples are in the sense of 'I see' and 'you see'[4]. When coursebooks were examined these were found

[4] Similarly, Brown (1994:61-79) examines the inter-relationships between the sense of a verb and the various syntactic patterns in which it can be found and which are often absent in the *Oxford Advanced Learner's Dictionary*. Nevertheless, he (1994:77) regards understanding "the kinds of mechanisms that can be employed in texts to convey more than is explicitly asserted" as essential for advanced students.

to account for only 10% of occurrences. Biber, Conrad and Reppen (1998:80-82) describe a similar misrepresentation in ESL textbooks in their representation of subject position *that*-clauses. Biber et al (ibid.77) find that *that*-clauses in subject position are rare in all genres (only 5-10 occurrences per million words) but that these are virtually non-existent in the spoken corpus they examined. One of the ESL textbooks examined, however, had two exercises for the students to use subject position that-clauses orally. Biber et al. (ibid. 81) conclude that the results from corpus analyses could improve textbooks in two ways;

> "First, books could emphasize those constructions most commonly found in the target register. Students typically study English for particular purposes - for example, conversational English for fluency in everyday interactions, or academic English for proficiency in educational reading and writing tasks. Textbooks can build on these goals to teach the grammatical constructions that students are most likely to encounter, given their communicative goals."

Added to this phenomenon is the fact that many coursebooks encourage students to view different forms of a word in groups suggesting that there is some kind of affinity between them but research suggests that there is much more reason to see these as different. Sinclair and Renouf (1987) say:

> "From a lexical point of view, it is not always desirable to imply that there is an identity between the forms of a word. ...But often, particularly with the commoner words of the language, the individual word forms are so different from each other in their primary meanings and central patterns of behaviour (including the pragmatic and stylistic dimensions), that they are essentially different 'words', and really warrant separate treatment on a language course."

Field Code Changed

An example of this that Sinclair gives is the lemma *move*. He says (1994:20), "The forms 'moving' and 'moved' share some meanings with 'move', but each form has a very distinctive pattern of meaning. Some of the meanings found elsewhere in the lemma will be realised, and some will not. In the word 'moving', for example, there is the meaning of emotional affection which is quite prominent." Similarly, Sinclair (1997:32-38), drawing on the more recent Bank of English, gives examples of not only the most usual meanings and uses of language but also the most usual collocations. He shows that 'nice' although a very neutral adjective has very strong patterns of language associated with it. Sinclair finds that *nice*

> "selects the indefinite article *a* and most emphatically rejects the definite article *the.* When in predicative position, it attracts strongly a modifier such as *very, pretty, extremely.* When attributive, it is commonly found with another adjective with which it combines in meaning, so that *a nice relaxing time* is nice because it is relaxing. Where *nice* immediately precedes a noun, and has no modifier itself, the nouns it goes with seem to be frequently selected from a few short lists – *day, evening,* etc., *boys, girls,* etc., and *surprise.* Often there are set phrases.

Sinclair (ibid.) also argues that teachers should concentrate on teaching meaning and suggest that if this is done, then it is obvious that often one meaning is associated with one structure whilst another associated with a different one. He claims that the more concrete, narrow meaning goes with the noun and the more figurative, vaguer meaning goes with the verb. He gives the example of *combat* which as a noun means actual physical fighting but as a verb means 'struggle against' usually with abstractions like inflation, recession etc. Sinclair suggests that these distinctions are easy to appreciate when they are pointed out, but they are not always distinguishable grammatically. Hopper (1997:93) argues that there is little terminology that the modern linguist uses that would have been unfamiliar to Quintilian and that, because of this, some integral parts of the language which, as Stubbs (1993:17) says, “lie somewhere between word and group ... are missed both by current grammatical descriptions and also by conventional definitions of collocation”. Hopper (ibid.) suggests that this situation is also the case with the English verbal expression. He (1997:94) uses Firth’s sentence “She kept on popping in and out of the office all the afternoon” as an example of the difficulty of identifying the verb in such sentences. He (1997:99) concludes that corpus linguistics is showing that the “category of Verb itself might be more in the nature of a cluster or family-resemblance category rather than a simple word class” or “folk category”. He (1997:101) recommends the use of discourse as a data source so that this can be made evident.

According to Sinclair (1997:37) rather than making the language limited, the fact that regular linking of grammar or form and meaning will not only cut down on the load the learner has to cope with but it will make the curriculum more interesting and will allow the learner to ‘develop unique and personal utterances which are almost guaranteed to be acceptable’. The example that he gives here is the structure

‘a(n) X of Y’

where X can be measures such as *pint, yard, ounce,* etc.; informal portions *blob, dash, lump, shred,* etc.; shapes *shaft, stick, tuft,* etc.; flows of liquid *dribble, jet, spurt,* etc.; containers *bag, bucket, tank, tub,* etc.; formal collectives *herd, flock, team,* etc.; and informal collectives *bunch, clump, group,* etc. If *-ful* is added to some things which are not normally seen as containers such as *bag* to become *bagful* then almost anything can become a container - a skirtful, a houseful, a shipful, etc. Sinclair argues that this is what language is like and therefore, this is what

Field Code Changed

should be taught. He provides the following checklist for the language teacher:

Present real examples only
Know your intuition
Inspect contexts
Teach by meaning
Highlight productivity

Intuition, as Sinclair uses it here, means for teachers (or learners) to be able to give the meaning of words in isolation and to pronounce upon the well-formedness of sentences in isolation.

He claims (1997:38) that corpora will “clarify, give priorities, reduce exceptions and liberate the creative spirit.” Biber, Conrad and Reppen (1998) have also reached the same conclusion in their studies of synonymous words and structures. They find that words have preferred collocates and senses and that structures are also used in certain ways in certain registers. They conclude that students should be made aware of these differences and taught accordingly rather than that there are synonyms and synonymous structures that can be used indiscriminately.

McCarthy and Carter (1994:38) argue for a discourse view of language and suggest that this involves “examining how bits of language contribute to the making of complete texts....” exploring “the relationship between the linguistic patterns of complete texts and the social contexts in which they function ....”, considering “the higher-order operations of language at the interface of cultural and ideological meanings and returning to the lower-order forms of language which are often crucial to the patterning of such meanings.”

Kjellmer (1991) suggests that

> "lexical items should not be taught and learnt in isolation but only in their proper contexts. This means shifting the emphasis from individual words to the collocations in which they normally occur.
> It is only when the student has acquired a good command of a very considerable number of collocations that the creative element can be relied on to produce phrases that are acceptable and natural to the native speakers."

Mindt (1997:40-50) describes the experiences teachers in Germany have had with using corpora for teaching in Germany. He (1997:41) questions the content of the grammatical syllabus and suggests that tradition is not sufficient reason for including grammatical knowledge. He believes that research is the only reliable source of information for what should constitute the grammatical syllabus as there is no comprehensive grammar for the teaching of English as a foreign language. He argues that first a corpus must be compiled and then a didactic grammar should be constructed from the corpus; finally a pedagogical grammar should be produced from the didactic grammar. He and Tesch have been conducting research to try to produce a grammar for their students in Germany. One of the first areas studied was the teaching of *any*. Far from being concerned with the rules as given in the books they use in schools where 'any' is contrasted with 'some' and rules like "*Some* is generally used in affirmative sentences, 'any' in questions and negations" are given. They find three types of *any* as in:

> Any 1: I thought any fool would know
> Any 2: I shan't get any scripts from the assistants before then
> Any 3: But is there any truth in it?

and suggest that these occur in the following situations and frequencies:

> Any 1 generally occurs in affirmative and declarative sentences and applies to a referent whose existence is presupposed. Type 1 makes up more than 50% of all cases of any.

Field Code Changed

> Any 2 occurs in negative and declarative sentences and applies to the referent whose existence is not presupposed. This type covers between 30 and 40 per cent of all instances in authentic texts.
> Any 3 occurs in affirmative and interrogative sentences and applies to a referent whose existence is not presupposed. This type makes up about 10 per cent of all cases of any.

Mindt (1997:44) points out that although 'any 1' as defined above is the most frequent form of 'any' this is rarely mentioned in teaching materials and is rarely mentioned in grammars of contemporary English. However, he (ibid.) notes that in the English textbooks he examined this usage was present in the same frequency but it was never explicitly taught in any of the exercises on 'any' which restricted the teaching to types 2 and 3.

Based on these findings Tesch (1990:345f) proposes a new approach in the teaching of *some* and *any*. The grading she suggests is not assumed to take place within one lesson but would normally spread over several teaching units but would include the use of the 'missing' meaning of any which is the most frequent and where:

> The traditional opposition of *some* and *any*, which is normally introduced as the first distinction and at the very beginning, only occurs in step 5 (for fast learners it would be possible to combine steps 4 and 5). The new grading emphasises the main uses of *some* and *any* contrasting them step by step with their appropriate counterparts in a new and unconventional way which is consistent with the use of *some* and *any* by native speakers.

Similarly, in a large number of grammars *will* and *would* are treated within the same framework of reference. *Would* is generally considered as the past tense form of *will*. Mindt (1997:47) makes a distinction between temporal meaning and modal meaning. The modal meanings of *will* and *would* can be divided into five principal meanings, three of which make up 97% of all cases of *will* and four of them 95% of all cases of *would*: certainty/prediction, volition/intention, possibility/ high probability, hypothetical event or result, and habit.

The results were as follows:

	will	would
certainty/prediction:	71%	31%
volition/intention:	16%	
possibility/ high probability:	10%	33%
hypothetical event or result:		18%
habit:		13%

Mindt argues from these results that because of their different semantic profiles *will* and *would* should be treated separately in teaching materials.

Similar work has been carried out in Portugal by Prof. Casanova from the University of Lisbon who argues (1995:100) that most English grammars (and therefore language teachers) give inaccurate explanations of English grammar which makes them inadequate or unusable. In the case of the present perfect, the emphasis that is normally found in grammars is on an incomplete action which was started in the past represented by the verb tense but as Prof. Casanova shows this is simply not correct and causes many exceptions to need to be cited. One of the examples Prof. Casanova uses to demonstrate the inadequacy of this explanation is the difference between *John has lived in Paris* and *John has lived in Paris for ten years*. In these cases it is the adverb of time that indicates that the action is incomplete rather than the verb tense. In the former case he no longer lives in Paris yet in the latter he does which is expressed by *for ten years* rather than the verb tense.

Mindt (1997:46) suggests that his research work emphasises the importance of distributional data in grammars for teaching purposes. Without distributional data there can be no informed grading of the functions of a grammatical form in a language course. The absence of distributional data in almost all preceding grammars results in a grading

Field Code Changed

that is based on intuition rather than on empirical evidence and very often does not reflect the actual use of English. Halliday (1993:1) argues that it is only with the development of the modern corpus that "serious **quantitative** work in the field of grammar" can take place, the results of which can show the probabilities of one grammatical pattern occurring rather than another. The results that are obtained from such quantitative research, Halliday (1993:6) suggests, are important for "learning and teaching languages". Through his work on the COBUILD corpus, Halliday (1993:20-21) argues that positive and negative occur in English on a ratio of 9:1 and that the 25 most frequent forms mostly occur only as verbs whereas in the next 25 a large number of the forms function as both noun and as verb. Francis (1991:145), working on the same University of Birmingham COBUILD corpora, finds that "different senses of a noun display different grammatical behaviour". Todaka (1996:13) working from the UCLA Oral Corpus and the Brown University Corpus finds that the difference in usage of *between* and *among* can better be explained by regarding their difference as a "distinction between 'individual' and 'collective'", that is, if the items in the NP objects are seen individually, *between* is used, if not *among* is used. Added to this, the sentence construction most often used with *between* is *between* A+B+(C...) whereas that with *among* is most often *among* plural noun. He notes however that when either of these could be used the preference for one or the other depended upon the discourse register (formality) and the prescriptive rules. He (1996:13) suggests that learners of English can apply his findings to "everyday uses of these prepositions". Despite all these studies, there is, as yet, no work that is available for either teachers or learners that describes English language usage comprehensively.

Minugh (1997) argues that, whenever school grammars use the words 'usually' or 'often', students should be encouraged to go to a corpus and examine a series of instances. In this way, he says they could gain insight into the fact that the rules in school grammars are 'necessarily overly simplistic and categorical'.

Johns (1997:102) says that working with data leads to not only "a radical revision of preconceived ideas about what one should be teaching" but also "how one might teach it."

a) The simple principle 'It is probably not worth teaching anything that does not occur at least x times in a corpus of y million words' (x and y being redefinable taking into consideration the level of the learners) makes it possible to exclude immediately much that is traditionally enshrined in classroom tradition.

b) *Pari passu* the work suggests ways of dealing with areas of language which have traditionally been poorly taught or regarded as unteachable (e.g. article usage) and reveals areas of language structure (e.g. the contextual patterning of nouns) that have been neglected both descriptively and pedagogically.

c) The data controls not only which features of the language are taught, but which exponents are presented and which meanings are taken as primary (e.g. in Academic English, *may*, showing an estimate of probability based on 'experience').

d) More fundamentally, the traditional division between independent 'levels' of language (e.g. lexis-syntax-discourse) appears increasingly untenable once one starts to place at the centre of one's concern the ways in which words behave in context. As a result, although the materials have for the most part a syntactic/functional starting point they could (as the students themselves have observed) as well be labelled 'Remedial Vocabulary' as 'Remedial Grammar'.

Work done by Phillips (1985) and later by Hoey (1991) suggest that discourse can also be explained better by means of lexical phrases. Nattinger and DeCarrico (1992) say "Lexical phrases are parts of language that have clearly defined roles in guiding the overall discourse. In particular, they are the primary markers which signal the direction of

Field Code Changed

discourse, whether spoken or written." Although corpus-based research can aid teachers to see what is natural language use, they must be careful to bear in mind not only the date of the corpus but also to make a clear distinction between spoken and written language. Much of the work done has shown a contrast between the two and has even gone as far as noting differences between different age groups. There are dictionaries based on computer corpora which clearly demonstrate the most frequent meanings and collocations of words together with an explanation of differences between spoken and (general) written language use (Longmans, Collins, Cambridge etc.). The Longman's Dictionary of Contemporary English (LDOCE) claims to have 25,000 fixed phrases and collocations. The editors say:

> "The English Language is made up of building blocks or chunks of words. When we produce a phrase, for example "*if you don't mind me saying so*", we don't think about each individual word and then link them together: we automatically think of the phrase as one block of words working together, one chunk of language. Through extensive corpus analysis, we have identified these chunks of English, these fixed phrases. So now students too can have access to these phrases..."

There are a number of books published giving collocations and examples of use of such items as prepositions, phrasal verbs and other structures through concordance samples which can be used to generate exercises. Kennedy (1989) shows that two generalised frames for prepositional phrases with *at* 'at + (the) + Proper N denoting place' and 'at + Personal Pronoun' account for 63% of the occurrences of this preposition. Similarly, the corpora used here could form the basis of exercises for inclusion in the materials for undergraduates. Goethals, Engels and Leenders (1990:231-268) have even developed an automatic exercise generator based on electronic texts, as has Wilson (1997). Wilson produced exercises on pronouns, participles and appropriate words in context which are somewhat simplistic as yet because of the difficulty of getting the computer to perform more complex tasks but which, nevertheless, demonstrate a trend that would considerably simplify the work of the teacher wishing to develop data-driven learning exercises in the future.

Most personal computers can be used to examine text and a simple concordance can be bought or indeed written for use on an average computer. An example of a KWIC concordance teachers can write themselves is given in Tribble and Jones (1990:84-89) *Concordances in the Classroom*. Others are available on-line (from ICAME and even OUP) so that they can be downloaded by teachers for their own use. They also give some very useful guidance on the use of concordances including using them to analyse the students' own work in order to bring out the contrasts between natural language use and student use. Work of this kind could not only personalise student correction but would foster more learner autonomy. McCarthy and Carter (1994) give examples of classroom activities contrasting discourse features in texts including students' own work in order to raise the students' awareness and help student improvement. In terms of correction, just using a spelling checker with a word processor helps to highlight the student's individual difficulties.

A number of CD-ROM encyclopaedias have search devices which will allow research into particular uses of words or alternatively texts can be saved and then a simple concordance can be used. This has the advantage of identifying very specific language use in specific subject areas for more advanced language study. That is to say, these encyclopaedia help not only to identify the frequency of a word but also the range of the word, that is, how widespread the use of the word is in a

Field Code Changed

more general domain. The most commonly used language can be found and also that language which is used in different registers.

7.5 Methodological Implications

What are the implications of the research in terms of methodology, once the forms and meanings that are to be presented to the students have been identified? First of all the course materials to be used must be designed to ensure that they do not mislead the students about words and structures and their meanings and uses but do reflect natural language use, as in the case of the most common meanings and usage as discussed above.

The work done on language acquisition suggests that the pattern practice exercise is necessary and useful at an early stage in learning and that later substitution exercises should be used to reflect the developing awareness and adaptability of certain language patterns and the use of more natural chunks of language to replace certain words or expressions (Sinclair 1997). Frameworks for paragraphs have shown themselves to work and students can indeed improve if they are given exercises of the type which include cohesive devices which the students then use as a basis for their own work. An example of this is where students are given information which they have to reorder into a framework such as; There are several problems such as (1).........and (2)........ However, (3)...... Therefore, (4)........ In addition, (5)....... Finally, (6)... . An example is given initially and then a similar paragraph has to be produced based on new data. The research mentioned earlier rather than suggesting that these techniques are ‘old hat’, encourages their use as reflecting normal language acquisition. Model opening and closing paragraphs for different types of text could also usefully be presented and analysed at a more advanced level. Indeed these features of text must be analysed to ensure that students are aware of the discourse features that are associated with them, which may include a transition of the tense used as Biber *et al.* (1998:128) have shown in their study on research articles. McCarthy and Carter (1994:58) suggest that students might be encouraged to produce text frames which map the article or text being studied.

Hoey's work (1991) demonstrates that students should be taught to recognise cohesive devices in order to understand texts and that in order to write more natural texts in English they should be aware of and use different forms of repetition. Nattinger and DeCarrico (1992:60) say that lexical phrases "signal the direction of discourse" whether the information to follow is in contrast to, is in addition to, or is an example of information that has preceded and, therefore, students should recognise and practise this. An obvious way of doing this is through the use of Cloze exercises which highlight the fact that only certain words are possible and reflect the limited nature of most of the language that native speakers use (with the exception of poetry and other forms of imaginative creative writing which deliberately extends or breaks the rules). Pronoun referencing and deictic features of text are very specific to academic prose as Biber *et al.* (1998) and McCarthy and Carter (1994) show and their specific use should be studied once again by taking actual examples from the corpora and getting students to work on them in a number of ways.

Sinclair and Renouf (1987) suggest that "the main focus of study should be on:

a) the commonest word forms in the language
b) their central patterns of usage
c) the combinations which they typically form"

All these are available through computer corpora but even Sinclair and Renouf allow that the use of a grammatical table "may improve the

Field Code Changed

learning process" by shedding "light from a different angle" and support an 'eclectic' position.

Three final principles that can be deduced from the research described are:

1. Building on what students already know
2. Extending that knowledge further and
3. Remembering to ensure that there is adequate reinforcement of language through recurrence of words, phrases and frameworks.

This first idea has also been backed up by psycholinguistic research into the manner in which learners remember vocabulary. It is suggested that schema are used and that these schema or word groupings are referred to in order to enlarge upon and refine understanding. The second principle is that recycling of items can lead the student to extend the range of the word and gain insight into its use and facets, thereby refining the meaning of the word in specific contexts. The third principle, that is, repetition of items, is something which coursebook writers often fail to do or fail to do consistently and which teachers must make an effort to remedy. The better a teacher knows the materials that are being used on the course, the easier this is. Students often take the stance that work done in an earlier part of a course is no longer relevant later in the course. This may be a response which has been produced from school activities which divide up the material to be taught into convenient sections which are then tested (and forgotten?) and not referred to directly again later in the course. The detrimental effect that testing can have on teaching leading to 'teaching for the test' has to be avoided especially at this tertiary level where the students are learning ever more detailed information in fewer subject areas and so cannot afford to 'forget' the earlier concepts on which the more specialised work is based. This position cannot be applied to language learning either because the process is clearly cumulative.

Finally, Portuguese corpora are being produced and when these are available there should be an even more valuable tool to help teaching. Bahns (1993:56) describes work carried out on lexical collocations between German and English. Through contrastive analysis of “tens of thousands” of lexical collocations the students are helped to identify equivalent phrases and observe where differences occur so that they can avoid errors in English. There is a need to reduce the learning load for students through analysing and isolating the differences and similarities between the two languages so that the students can be helped to produce natural language and to avoid specific types of errors. It might also be possible to find parallel texts (from European sources) in both English and Portuguese which would be useful for examining differences and similarities in scientific and technological discourse. Leech (1997:21) describes such texts being developed through the C.R.A.T.E.R. (McEnery *et al* 1994) and Multext (Ide and Vérons 1994) research projects. He suggests that the fact that these texts are often highly specialised and technical is a drawback but this may be a positive aspect for our undergraduate students.

Using parallel texts would also accept the fact that often students would be producing some kind of summary or translation from English into Portuguese for their own use. Halliday (1993:125) suggests that when texts are translated the translator does not normally alter the discourse structure of the text that is being translated so this can help the students to analyse scientific and technological discourses. Carter (1993:146) argues that this kind of contrastive analysis can help to produce awareness of socio-cultural meaning which is an extremely important need if the textbooks continue to be based on the American models as studied here. He (1993:147) goes further than this however,

Field Code Changed

and suggests that greater language awareness of this kind increases learner autonomy and gives learners greater control over their learning, which, for university undergraduates, is an essential part of the educational process. Aitcheson (1994:95) suggests that understanding words is “not just a case of sorting out the meanings of individual lexical items” but that, to understand something fully “involves understanding the mental models of a culture.” Adams, Heaton and Howarth (1991:11) suggest that understanding “how cross cultural problems arise can help the course designer, the teacher and the student to make reasoned choices at the rhetorical and stylistic levels.” It has been argued that recognising the different meanings of technical words in scientific discourse is one of the basic skills that the reader needs in order to understand that discourse but it is obvious from the research mentioned above this is a somewhat simplistic view and the respective culture that underlies the text must also be taken into account. Brumfit (1994:32) suggests that the emphasis must be on knowledge as a process rather than as static information and that it is essential for teachers to be sensitive to the different understandings developed by particular cultural and linguistic groups in order to be able to help students with their individual needs.

Computer corpora can be used in at least three ways in teaching. Fligelstone (1993:98) identifies these as:

- Teaching about (i.e. the principles and theories of corpus linguistics)
- Teaching to exploit (i.e. the practical aspects of corpus study)
- Exploiting to teach (i.e. deriving language-teaching materials from corpora)

There is a fourth activity however which results from teaching itself Renouf (1997:256):

- Teaching to establish resources (i.e. designing and creating the corpus)

The first of these principles will take some time to develop and might more easily be used on mainstream language course with students who have more time to focus on language itself. As with the debate about teaching students the phonetic alphabet time is the principal constraint in teaching about language. The second and third of these principles would require the expertise of the lecturer working with the students on the corpora and is feasible now that the corpora have been produced. The fourth principle is something that could be applied if students had specific areas of their studies which they felt needed addressing so that the corpora could be built up or driven by what the students perceived they needed to work on.

In conclusion, the data gathered through this study can provide the examples of natural science and technology language of explanation and exposition of the science textbook together with the means to present actual language use of this medium to students on the English discipline in the first year of university. Initially the corpora would be exploited for teaching (and testing) resources but at a later stage, with adequate resources available, could lead on to being exploited by the students themselves on an individual basis to solve their individual language problems or difficulties.

7.6 Modern Technology and the Syllabus

CALL - Computer Assisted Language Learning lends itself particularly well to developing reading skills in English, and the use of computer corpora, CD-ROM material, the Internet and e-mail should be included in the English discipline for undergraduates. Johansson (1991:305-6) believes that in the future teachers will be able to draw on

Field Code Changed

vast data sources to select material from, but that there is still a need for smaller corpora because these can be “analysed exhaustively in a variety of ways”. Computers have had and are having more and more influence over life today and one of the features that more and more students are coming across and enjoying is electronic mail. In Britain e-mail is included in the National Curriculum for I.T. but it is an area that would lend itself well to the EFL classroom. The students could use e-mail as part of a tutoring system to help with individual needs and more advanced students could be involved in projects on the corpora themselves, analysing particular areas that they wish to find out more about. Correspondence of this kind contains certain features which will have to be taken into account in teaching . McCarthy and Carter (1994:5-10) discuss how far spoken and written modes can be distinguished by informants and go on to consider how ‘writerly’ a text is or the degree of ‘spokenness’ of a text together with the degree to which it is monologic or dialogic with the reader. In other words these more modern genres are often a mixture of modes, including features that would normally be associated more formally with one or other of the spoken or written modes. E-mail has a number of distinguishing features such as the fact that it is a written dialogue including many of the features of speech and the fact that the same messages are often repeated over and over again if the dialogue continues for any length of time. One earlier taboo in teaching was to present written dialogue as representing true conversation but now it can be seen that a situation that goes back to the very beginnings of foreign language teaching must once again be returned to in order to help students to cope with the future.

These types of activities would also be a means of developing and encouraging strategies for the students to carry over into the future. Carter (1993:139) defends the idea that students should learn about the language as it “is valuable in its own right and has ‘educative’ potential in the broadest sense ... it can enhance learning 'through the language' about the cultures and ideologies which inform the target language and its uses." There are some difficulties attached to activities like the use of e-mail however, as Johansson (1991:307) points out; it is a medium somewhere between speaking and writing and for this reason it is more prone to error than more studied, revised writing. He (ibid.) suggests that it is also more "playful and creative, less bound by conventions" so that it would be a means for students to feel less inhibited in their use of written language but it would not be a suitable means for encouraging accuracy in language use.

The repetitive aspect of e-mail "conversations" through computers also requires some reading skill techniques like scanning and skimming. However these would be conducted in an even more interesting, on-screen situation where the text scrolls up and down. The replies given and further discussion of points raised have to be picked out from the repeated material and signings on and off in the electronic conversations that take place on chat pages.

Computers change the roles that normally pertain, in that the reader may become the writer or editor of the text and can control the amount and type of information that is displayed on the monitor (Gill and Whedbee 1997:160). The Internet is another source of material which students can access and which they could then edit for their own purposes. Substantial editing is necessary especially if information gained through e-mail or the Internet is to be incorporated in the students' own documents, and practice in doing this would be required. With more of these activities taking place there would be a need to revise the strategies used and to feed the insights gained through such activities back into teaching materials, thus keeping flexibility and openness to change a basic requirement of the syllabus.

Field Code Changed

Electronic dictionaries are already available, including in Portuguese, which would be useful tools for both teachers and learners to test their hypotheses about general language use and perhaps through special glossaries of specific terms for their subject requirements. However, for specific language use the corpora produced for this analysis should be made available so that the actual, immediate, authentic situation can be examined and studied by the students themselves.

There are administrative considerations that have to be taken into account for corpus-based studies to be carried out with students. Appropriate computational infrastructure has to be provided first but another consideration is the level of confidence that teachers feel with computers so that they can be exploited fully and appropriately (Hughes 1997:292-307). Renouf (1997:255-266) describes her experiences with teaching corpus linguistics to teachers of English (at post-graduate level). She suggests that things have improved considerably over the last decade in terms of the technology available for teaching classes and discusses the need for and possibility of providing continued support through distance contact. Her students were post graduates who after the course returned to their own teaching environments abroad but through computer communications they could keep in touch and obtain support for their work. The Council of Europe attaches particular importance to distance learning and to the use of new technologies for modern language learning in Europe (Trim 1998:208). In the Appendix to Recommendation Nº R (98) 6, adopted by the Committee of Ministers on 17 March 1998, article 21 in Section E Adult Education agrees to:

> 21. Support the provision of national and international structures so as to ensure the widest availability of facilities for distance education (including the use of communication and information technologies), in order to promote the development of diversified advanced

communication skills, where possible linking autonomous learning to institutionalised learning.

This agreement could be met through the methodology described above for teaching undergraduates in Portuguese universities. Distance education is certainly a possibility in the University of Aveiro which already has considerable expertise in this area through the Department of Didactics and Information Technology. The first year undergraduates can also since 1998 available themselves of first year courses on-line. It is entirely feasible that lecturers in all departments begin to 'converse' with their students on an individual basis through e-mail as the university's computer network already covers all departments. Increasing expertise in this area would suggest that certain issues would present themselves for remedial teaching. The FAQs (Frequently Asked Questions) that those working in distance learning report (Laurillard 1993, Rowntree 1992) serve to alert the lecturer to common problems which can then be dealt with generally rather than on an individual basis. The only foreseeable disadvantage to this system is the number of messages that could be received, at any time and that require time to answer. Speed typing skills may become a necessity for tutors swamped by student's queries. Computer communication systems using e-mail connecting students with individual mentors (scientists) have been tried in some education systems (New York 1993) to aid success in education. The building up of such relationships appears to be fruitful for both the scientists involved and for the students. The scientists gain insights into the student's problems and the students feel comfortable enough to ask for guidance on any number of problems, academic and personal. As with the results from computer corpora which will change our view of how language operates in fact and will cause new theoretical positions to be developed, the whole nature of relationships between students and their

Field Code Changed

lecturers will undergo profound changes because of communications by computer. We must be prepared to meet these new situations in the near future.

Chapter 8 Conclusion

Chapter 8

Conclusion

My results show that although there is general agreement with some of the dimensions studied by Biber (1988), the accepted attributes of academic prose and the undergraduate textbooks studied here are by no means congruent. I believe that these mismatches occur because of the pedagogic nature of the texts and intention of the authors to interact with and 'teach' or instruct the anticipated, native speaker, undergraduate reader. The analogies with real-world applications used in both the physics and chemistry textbooks is where this mismatch is most often apparent. These real-world analogies appear both in essays and in the problems provided at the end of chapters for the students to use to practice the scientific topic presented in each chapter. American English speech is also found to be replete with sports analogies from basketball, baseball and American football like those found in the problems for students to solve at the end of the chapters, so these textbooks are in this sense merely reflecting the culture from which they come (Rosenthal 1996:105).

I would argue that many of the variations identified could be attributable to the mixture of the scientific and the technological found in the textbooks with their specific and different language usage. I call into question some of the previous studies and corpora for their lack of rigour which meant that they leant more towards a popular or journalistic representation of science and technology rather than an academic science and technology setting. Furthermore, the problems and real-world analogies in the textbooks involve much more culturally-specific

Field Code Changed

information which is more difficult for undergraduates, who are simultaneously foreign language learners, than general scientific writing.

The level of English shown by the undergraduates studied here varied, from very competent to quite weak, although there was more evidence that recently undergraduates had generally studied English for more years than in the first years of the foundation year course. The stronger students showed that their reading comprehension skills had also developed sufficiently well to allow them to apply strategies to what for them were new English language situations like that of a (popular) scientific or technological text which they had not studied in language classes in school. The weaker students showed that their reading comprehension skills were very poor or non-existent. Further research is necessary however to determine if the students who did well on the tests could cope equally well with test material drawn specifically from the corpora developed for this study, that is, from the materials on the bibliographies. It is now practical for this to take place using the corpora built up for this study. This test material should also include a component focusing on the different combinations of typographics, footnotes, formulae, numbers, equations, tables, diagrams and drawings found in these kinds of textbooks to check comprehension and interpretation of these features by students.

Further research into the lack of success of students would involve two distinct parameters, the difficulties with English *per se* and the difficulties with the language of science and scientific concepts, so that it is possible to evaluate which of these is proving to cause most difficulty for students so that a strategy could be developed for coping with it in the discipline. Knowledge about the students' competence in science and technology on entering the university might be determined through collaboration with other departments in the university. Looking at the entry marks in specific science subjects would not necessarily be sufficient to assess the students' future success in their first year as undergraduates. The number of different combinations of circumstances that the students present on entering the university is vast and knowledge about those competencies, seen as core competencies in science, but which many of the students lack would also aid in targeting the syllabus for these students.

The wide variety in levels between the strongest and the weakest students suggests that new strategies must be found to cope with large groups of such a heterogeneous nature. The suggestion that is put forward here is that these new strategies should be based on materials and evidence obtained from the frequency counts and variation studies carried out on the undergraduate textbooks. The teachers would then have appropriate and relevant materials, and good information, so that the focus would be on the items that would be most useful to students. The use of computers and corpus analysis in the discipline would allow the students themselves to approach their individual problems with the language of science and technology and would eventually allow self-access and distance and continuous learning to take place by means of the university computer network.

In addition, the opportunity to work with colleagues from the departments which teach the first year undergraduates in an interdisciplinary manner would help to reinforce the teaching at this level and provide a coherent framework for students to appreciate the relevance of the work being done in language classes. From this co-operation it would be possible to develop projects where the English language needed by students for their project work could be analysed and formulated from the language classes. This interdisciplinarity would also serve to motivate those students who find it difficult to perceive the relevance of their language studies to their courses. In other words, content-based EFL will provide a focus and goal for the students.

Field Code Changed

The testing of the students would then have to change. The present system does not take into account the target language of science and technology and so the corpora produced here should be exploited for testing of students, both at the preliminary stage and for the normal university evaluation tests throughout the year. In this way, it would be possible to see if the students who were released from the discipline did indeed cope with actual language from the corpora of undergraduate textbooks rather than that perceived by their teachers to be relevant. Added to this, through the use of tests available through computers, it would be perfectly possible to design a novel testing procedure which could in itself be more flexible, allowing students to attempt certain tests when they felt that they were ready. Incidentally, it should not prove too difficult to improve the speed of marking and feedback to students by having a computer-marking system, releasing the teacher for other valuable activities.

The possibility of developing further teaching materials through analyses developed by students themselves from their own interests and needs is feasible provided that the necessary resources are available. These would comprise not only up-to-date computers with network connections but also teaching staff who are confident with both the technology and corpus-based techniques. This latter knowledge would to a great extent avoid the complaint that language lecturers do not feel confident with the subject matter of the materials they are trying to use with students of science and technology as their focus would be entirely on the evidence presented from the corpora in a linguistic analysis. In other words teachers would be focusing on the language and not on science and technology *per se*.

The role of the lecturer would also undergo a change in the type of contact and interaction with students. The use of e-mail would allow a much closer one-to-one interaction between student and tutor and might develop a different relationship from that enjoyed in a large group of students meeting for a limited amount of time. The use of e-mail itself would be a means of moving forwards into the modern world of communications and language use itself, although the emphasis would be on individual support from the teacher for students. Experience from other universities (Motteram, University of Manchester 1998) who have a highly developed system of tutoring through e-mail would suggest that tutors would eventually develop a series of frequently asked questions (FAQs) which could be made available for students to consult and thereby save some of the tutors' time answering the same questions over and over again. Similarly, support material could be provided on-line for students to work on their own.

More and more corpora are becoming available on-line and on CD-ROM and with a small investment of time and money many other resources could be exploited in the language class. As was mentioned at the beginning of this thesis (see *1.5 The Situation in Portugal*), the undergraduates in the first year are on numerous, different engineering or degree courses and the use of different corpora in this way would allow for a diversity of interests which might only become apparent at a later stage in the students' courses. Use of the European Union terminological database EURODICAUTOM on-line would be one means of addressing the diverse engineering needs in the students who should be encouraged to focus on precision in language for the communication of scientific and technical data and perhaps even to make or perfect comparable terminology in Portuguese where this is lacking. Allowing for this kind of subject flexibility would also lay the groundwork for the students to take up a means of continuing their language studies beyond the end of the first year and adapting the materials they use to their actual needs.

What is missing from this work is a comprehensive analysis of the use of lectures and other spoken communication for students in the first

Field Code Changed

year (and subsequent years) of their courses. The spoken corpus would also vary widely after the first year and would require other multimedia resources. Video in particular should be exploited more to present and practice listening comprehension and note-taking. The actual kinds of lectures (or spoken communication such as papers at conferences) that the students could be expected to come into contact with should also be gathered into a database of materials for both self-access and class use. Interdisciplinary work with the other departments could allow videoing of actual lectures or parts of lectures delivered in the university in English. These lectures, or excerpts from lectures, could then form the basis for language study materials. There are examples of university lectures in science and technology available through the Internet from some American universities. The reason that these lectures are available on-line is for the students on those courses to study from and then to contact the tutor via e-mail with any queries and to deliver their assignments. A similar system could be experimented with in the way described above.

In conclusion, the corpora produced here could be exploited for use with students in the first year in collaboration with other disciplines to focus more closely on those areas identified by colleagues in other disciplines to be central to the first year students' needs. Further corpora are needed to include spoken language from science and technology. The testing of the students also needs to take into account the competencies the first year students require to cope with English science and technology texts. The use of information technology needs to be reinforced to provide the students with more resources, support and individual contact with their tutors, as well as to prepare the students for their future professional lives and as life-long learners.

Many of the recommendations made here can be realised in the short or medium term in this university with its sophisticated resources and in other Portuguese universities which want to adopt common-core courses and modern technology. What would need to be introduced to continue the relevance and utility of the language taught/learned by these undergraduates is to extend the English discipline into other years of the courses. Language classes might be provided in parallel with courses to be taken on an *ad hoc* basis as students saw fit. This would necessitate a reappraisal of the language needs of the students at later stages in their courses and the development of an appropriate syllabus, methodology and materials. The suggestion (see *7.5 Methodological Implications*) that students themselves could be encouraged to provide the materials that they need to work on which could then be turned into an electronic corpus of materials which would form the basis for the language studies carried out by the students would be relevant in this case.

There is evidence (see Chapter 4) that in the first year, approximately 10% of the new students (those with fewer than five years of English studies at school) would benefit from more hours of study to bring them up to the level of the other students and to make their science studies through English a feasible proposition. Increasing the number of hours devoted to English only for these students and including provision for them to work extensively through self-access material would equip them better for their future studies. Nevertheless, the results of the research carried out in this thesis can form the basis for specific materials for different language competencies in students by drawing on parallel texts which, nevertheless demonstrate the relevant discourse features displayed in the main corpora. In this way these students could be brought closer to understanding the texts that they are encouraged to consult through understanding of the characteristics of those text-types.

Steps are already being taken to exploit computer resources with students and to provide on-line English texts for students to work on in a

Field Code Changed

variety of ways (including pronunciation of new vocabulary). Using suitable materials in the language laboratory which accurately reflect the students' needs is also being attempted rather than continuing the tradition of decontextualised drills and pronunciation work. The results of this analysis has alerted lecturers to making their materials reflect the target material for these undergraduates and to place emphasis on interpretation of visual materials together with texts. All of these different facets are being brought together into a syllabus which recognises that much of the work has to be carried out outside the classroom by the students on their own and gives weight to oral classroom interaction in order to make the most of the contact time available. Constant reappraisal of the syllabus has always been a feature of the discipline. New insights into both the students' competence and needs and new research findings and the materials used, taking into account materials that have worked successfully with students are fed back into the syllabus for the first year students. The corpora will go further than this however, as they will serve as a guide and object of study for the lecturers themselves to use to inform their ideas and judgements of what scientific English is and more importantly is not.

Bibliography

Bibliography

Abbot, G. (1980) *Towards a more rigorous analysis of Foreign Language Learners Errors* IRAL Offprint Vol. XVIII/2 1980 Julius Groos Verlag, Heidleberg

Adams, P., Heaton, B., Howarth, P. (eds. 1991) *Socio Cultural Issues in English for Academic Purposes,* Macmillan, London.

Ahmad, K. *et al.* (1985) *Computers, language learning and language teaching.* Cambridge: Cambridge University Press.

Aijmer, K. & Altenberg, B. (eds. 1991) *English Corpus Linguistics***,** London, Longman.

Aitcheson, J. (1994) "Understanding words" in Brown, G., Malmkjaer, K., Pollit, A., Williams, J. (eds.1994) *Language and Understanding,* Oxford: Oxford University Press. Pp.83-95.

Alderson, J. C. (1988) "Testing and its Administration in ESP" in Chamberlain, R. and Baumgardner, R. J. (eds. 1988) *ESP in the Classroom: Practice and Evaluation, ELT Documents 128* London: Modern English Publications. Pp.87-97.

Alderson, C. and Urquhart, A. (eds. 1984) *Reading in a Foreign Language,* London: Longman.

Allen, J.P.B. & Widdowson, H.G. (eds. 1974) *English in Focus***,** Oxford, Oxford University Press

Allen, J.P.B. & Widdowson, H.G. (1978) *Teaching the Communicative Use of English***,** in Mackay & Mountford (eds. 1978) *English for Specific Purposes***,** London, Longman.

Allwright, J. & Allwright, R. (1977) "An approach to the teaching of Medical English" in Holden, S. (ed. 1977) *English for Specific Purposes***.** Oxford: Modern English Publications. Pp 58-62.

Armstrong, S. (ed. 1994) *Using Large Corpora***,** London: MIT Press.

Arroteia, Jorge Carvalho; Martins, António Maria (1997) *Inserção Profissional do Diplomados pela Universidade de Aveiro: Trajectórias Academicas e Profissionais,* Aveiro: Universidade de Aveiro.

Astor, C., (ed. 1997) “Voices in Education” in *Education in the United States: Continuity and Change, U.S. Society & Values Electronic Journals of the U.S. Information Agency*, Vol. 2 **Nº 5** December 1997 pp.37-39

Bahns, J. (1993) “Lexical collocations: a constrastive view”, *ELT Journal,* Volume 47/1 January 1993: Pp 56-63.

Bakhtin, M. M. (1986) *Speech Genres and Other Late Essays.* Trans. Vern McGee. Eds. C. Emerson and M. Holquist. Austin: University of Texas Press.

Barber, C. L. (1962) *Some measurable characteristics of modern scientific prose* in Swales, J. (1985), *Episodes in ESP*, Oxford: Pergamon Press Ltd.

Barros, A. M. (1998) *Communication and Management Skills,* European Chemistry Thematic Network Report.

Bates, M. and Dudley-Evans, A. (1974) “Notes on the Introductory English courses for Students of Science and Technology at the University of Tabriz, Iran, *ELT Documents (74/4),* The British Council. Pp. 12-18.

Bazerman, C. (1998) “Emerging perspectives on the many dimensions of scientific discourse” in Martin, J. R. and Veel, R. (1998) *Reading Science,* London: Routledge. Pp15-28.

Beaugrande, R. de (1997) *New Foundations for a Science of Text and Discourse: Cognition, Communication, and the Freedom of Access to Knowledge and Society,* New Jersey: Ablex Publishing Corporation.

Beaugrande, R. de (1997) *The Story of Discourse Analysis* in Van Dijk, T. A. (ed. 1997) *Discourse as Structure and Process* London: SAGE Publications.

Biber, D. (1986) *Spoken and written textual dimensions in English: resolving the contradictory findings*. Language 62:384-414

Biber, D. (1988) *Variation across speech and writing* Cambridge, Cambridge University Press.

Biber D., Conrad S. and Reppen R. (1994) *Corpus-based Approaches to Issues in Applied Linguistics***,** Applied Linguistics, Vol. 15, No.2 Oxford University Press

Biber D., Conrad S. and Reppen R. (1998) *Corpus Linguistics. Investigating Language Structure and Use,* Cambridge: Cambridge University Press.

Birnbaum, I. (1987) “IT for Better Teachers” *Educational Computing,* Sept. 1987 Vol. 8, Issue 6. Pp. 19-21.

Bloor, M. & Bloor, T. (1991) *Cultural Expectations and Socio-pragmatic Failure in Academic Writing* in Adams, P., Heaton, B., Howarth, P. (eds. 1991) *Socio Cultural Issues in English for Academic Purposes,* Macmillan, London.

Bowen, J. D., Madsen, H., Hilferty, A. (1985) *TESOL Techniques and Procedures,* Massachusetts: Newbury House Publishers.

Bracy, G. W. (1997) in Astor, C., (ed. 1997) “Voices in Education” in *Education in the United States: Continuity and Change, U.S. Society & Values Electronic Journals of the U.S. Information Agency,* Vol. 2 **Nº 5** December 1997 pp.37-39

Bright, J. A. & McGregor, G. P. (1970), *Teaching English as a Second Language***,** London, Longman.

Brown, G., Malmkjaer, K., Pollit, A., Williams, J. (eds.1994) *Language and Understanding,* Oxford: Oxford University Press.

Brown, K. (1994) “Syntactic clues to understanding” in Brown, G., Malmkjaer, K., Pollit, A., Williams, J. (eds.1994) *Language and Understanding,* Oxford: Oxford University Press. Pp. 61-79.

Brown, R. 1973. *A First Language: the Early Years*. Cambridge, Mass.: Harvard University Press.

Brumfit, C. (1994) “Understanding, language, and educational processes” in Brown, G., Malmkjaer, K., Pollit, A., Williams, J. (eds.1994) *Language and Understanding,* Oxford: Oxford University Press. Pp. 22-33.

Brumfit, C. J. and Johnson, K. (1979-81) *The Communicative Approach to Language Teaching,* Oxford: Oxford University Press.

Bucy, J. F. (1985) “Computer Sector Profile” in Keatley, A. G. (ed. 1985) *Technological Frontiers and Foreign Relations,* Pp. 46-78, Washington: National Academy Press.

Burnard, L. (1992) “Tools and Techniques for Computer-assisted Text Processing” in

Butler, C. S. (ed. 1992) *Computers and Written Texts***,** Oxford: Basil Blackwell.

Butler, C. (1985) *Statistics in Linguistics*, Oxford: Blackwell.

Butler, C. S. (ed.1992) *Computers and Written Texts*, Oxford: Basil Blackwell.

Byram, M and Riagáin, P. O. (1999) "Towards a Framework for Language Education Policies in Europe" Strasbourg: Council of Europe DECS/EDU/LANG(99) 6rev.

Candlin, C. N., Kirkwood, J. M. and Moore, H.M. (1978) "Study Skills in English: Theoretical Issues and Practical Problems" in Mackay, R. & Mountford, A.J. (eds.1978), *English for Specific Purposes*, London, Longman.

Carter, R. (1990) "Towards Discourse-Sensitive Cloze Procedures: the Role of Lexis" in Halliday, M.A.K., Gibbons, J., Nicholas, H. (eds. 1990) *Learning Keeping and Using Language* Volume II Selected Papers from the 8th World Congress of Applied Linguistics, Sydney, 16-21 August 1987. Amsterdam/Philadelphia: John Benjamin's Publishing. Pp. 445-453.

Carter, R. (1993) "Language Awareness and Language Learning" in Hoey, M. (ed.1993), *Data, Description, Discourse*, London, Harper Collins. Pp. 139-150.

Carter, R., Goddard, A, Reah, D., Sanger, K. and Bowring, M. (1997) *Working with Texts : A core book for language analysis*, London: Routledge.

Casanova, I. (1995) "Gramática Inglesa: Procura-se" in the *Actas do XI Encontro da Associação Portuguesa de Linguística Vol III Gramática e Varia*, Depósito Legal 102905/96, Lisboa, Setembro de 1996 Pp.97-103.

Chamberlain, R. and Baumgardner, R. J. (eds. 1988) *ESP in the Classroom: Practice and Evaluation, ELT Documents 128* London: Modern English Publications.

Chang, R. (1991) *Chemistry* International Edition, Fourth Edition, McGraw Hill Inc., USA.

Chattel, R. K. (1999) "ESP - the case of English for Students of Sociology in the University of Coimbra" Actas do 4º Encontro Nacional do Ensino das Línguas Vivas no Ensino Superior, Faculdade de Letras da Universidade de Porto. Pp. 245- 254.

Clark, R. (1974) 'Performing without competence'. *Journal of Child Language* 1:1-10.

Clinton, W.J (1997) in Astor, C., (ed. 1997) "Voices in Education" in *Education in the United States: Continuity and Change, U.S. Society & Values Electronic Journals of the U.S. Information Agency,* Vol. 2 **Nº 5** December 1997 pp.37-39

Clinton, W.J (1998) "Excerpt from President William J. Clinton's State of the Union Address", in *American Studies Journal,* Nº 41, Summer 1998 Pp 59-60.

Close, R. A. (1965) *The English We Use for Science***,** London: Longman

Connor, U. and Kaplan, R.B. (eds. 1987) *Writing across Languages: Analysis of L2 Text,* Reading, MA.: Addison-Wesley.

Cooper, M. *Aspects of the Structure of Written Academic Discourse and Implications for the Design of Reading Programmes***,** in Hoedt, J. & Lundquist, L., Picht, H. & Qvistgaard, J. (eds. 1982)) Proceedings of the 3rd European Symposium on LSP, Copenhagen August 1981, The Copenhagen School of Economics. Pp.403-433.

Corder, S. P. (1973), *Introducing Applied Linguistics***,** London: Penguin

Corder, S. P. & Roulet, E. (eds. 1974) *Linguistic Insights in Applied Linguistics,* Second Neuchatel Colloquium in Applied Linguistics, Brussels: AIMAV; Paris: Didier.

Coulthard, M. (1977) *An Introduction to Discourse Analysis,* London: Longman.

Coulthard, M. (1978) *Discourse Analysis in English - a short review* in Kinsella, V. (ed. 1974) *Language Teaching and Linguistics: Surveys,* Cambridge: Cambridge University Press.

Coulthard, M. (ed. 1994 - 1998) *Advances in Written Text Analysis,* London: Routledge.

Cruttenden, A. (1981) 'Item-learning and system-learning'. *Journal of Psycholinguistic Research* 10:79-88.

Crystal, D. (1997) *English as a global language,* Cambridge: Cambridge University Press.

Crystal, D (1998) "To surf or not to surf: that *is* the question" in *Network A Journal for English Language Teacher Education* Vol.1 Number 1 December 1998 The British Council.

Danesi, M. and Di Pietro, R. (1990) *Contrastive Analysis for the Contemporary Second Language Classroom,* Ontario: Ontario Institute for Studies in Education.

Darian, S. *The Role of Definitions in Scientific and Technical Writing: Forms, Functions, and Properties*, in Hoedt, J. & Lundquist, L., Picht, H. & Qvistgaard, J. (eds. 1982)) Proceedings of the 3rd European Symposium on LSP, Copenhagen August 1981, The Copenhagen School of Economics.

Davies, A. (1988) *Procedures in Language Test Validation* in Testing English for University Study, ELT Documents: 127, MEP 1988

Di Pietro, R.J. (1971) *Language Structures in Contrast.* Rowley, Mass.: Newbury House.

Dunn, S. and Morgan, V. (1987) *The Impact of the Computer on Education,* London: Prentice Hall International.

Dudley-Evans, T. (1994) “Genre analysis: an approach to text analysis for ESP” in Coulthard, M. (ed. 1994- 1998) *Advances in Written Text Analysis,* London: Routledge. Pp.219-228.

Eastment, D. (1987) *Computer-Assisted Language Learning in the Bell Educational Trust.* Cambridge: Bell Educational Trust.

“Education and the wealth of nations” in *The Economist* March 29th 1997 Pp.15-16.

Eggins and Martin (1997) “Genres and Registers of Discourse” in van Dijk, T. A., (ed. 1997) *Discourse as Structure and Process,* London: SAGE Publications.

Ellis, G. & Sinclair, B. (1989), *Learning to Learn English***,** Cambridge, CUP.

ELTDU (1970) *English for Business: research and preliminary planning report***.** Colchester: English Language Teaching Development Unit.

Evelyn Ng, K.L. and Olivier, W.P (1987) “Computer Assisted Language Learning: An Investigation on some Design and Implementation Issues” in *System,* Vol. 15, No. 1. Pp. 1-17.

Ewer, J. R. (1975) “Teaching English for Science and Technology: the specialised training of teachers and programme organisers.” In *English for Academic Purposes: Information Guide No. 2.* London: British Council.

Ewer, J. and Hughes-Davies, E. (1971-72) *Further notes on developing an English language programme for students of science and technology***,** English Language Teaching Journal XXVI/1 and 3

Ewer, J. R. and Hughes Davies, E. (1974) “Instructional English”, *ELT Documents (74/4),* The British Council. Pp. 19-24.

Ewer, J.R. & Latorre, G. (1969) *A Course in basic scientific English***,** London, Longman.

Fairclough, N. (ed. 1992) *Critical Language Awareness,* London: Longman.

Ferreira, A. M., Ramos, D., Braga da Silva, F. (1999) “Evaluation des curricula de FLE au Portugal” Actas do 4° Encontro Nacional do Ensino das Línguas Vivas no Ensino Superior, Faculadade de Letras da Universidade de Porto. Pp.333- 337.

Fligelstone, S. (1993) “Some reflections on the question of teaching, from a corpus linguistics perspective.” *ICAME Journal* **17**:97-109.

Flowerdew, J. (1993) Concordancing as a tool in course design. *System* **21 (2):** P. 213-29

Fordham, S. (1997) *Portuglish: Mistakes made by Portuguese-speaking learners of English,* Lisbon: Plátano Editora.

Fox, J. Matthews, A., Matthews, C., Rope, A., (1990) *Educational Technology in Modern Language Learning in the secondary, tertiary and vocational sectors,* for the British Government Employment Department Group Training Agency, Learning Technology Unit by the University of East Anglia and the Bell Educational Trust, March 1990.
Sheffield: Crown Copyright.

Francis G. (1991) “Nominal group heads and clause structure” WORD *Journal of the International Linguistic Association,* Volume 42, Number 2 August 1991 Pp.145-156.

Francis G. and Sinclair J. (1994)‘*I Bet He Drinks Carling Black Label’: A Riposte to Owen on Corpus Grammar***,** Applied Linguistics, Vol. 15, No.2 Oxford University Press Pp190-200.

Francis, W.N. and Kucera, H. (1982) *Frequency analysis of English usage: lexicon and grammar***.** Boston: Houghton Mifflin.

Friel, M. (1978) *A verb frequency count in legal English***,** ESPEMA, 10, Spring 1978

Gavioli, L (1997) "Exploring Texts through the Concordancer: Guiding the Learner" in Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora,* London: Longman. Pp. 83-99.

Gill, A. M. and Whedbee, K. (1997) "Rhetoric" in van Dijk, T. A., (ed. 1997) *Discourse as Structure and Process,* London: SAGE Publications.

Glaser, R. (1982) *The Problem of Style Classification in LSP (ESP)* in Hoedt, J. & Lundquist, L., Picht, H. & Qvistgaard, J. (eds. 1982) Proceedings of the 3rd European Symposium on LSP, Copenhagen August 1981, The Copenhagen School of Economics. Pp. 69-82.

Goethals, M., Engels, L. K., Leenders, T. (1990) "Automated Analysis of the Vocabulary of English Texts and Generation of Practice Materials" in Halliday, M. A. K., Gibbons, J. & Nicholas, H. (eds. 1990) *Learning Keeping and Using Language,* Vol. II P.231-268 Amsterdam: John Benjamins Publishing Company. Pp. 231-268.

Goodale, M., *Collins Cobuild Phrasal Verbs***,** Harper Collins 1993

Grabe, W. P. (1984) "Written Discourse Analysis." *Annual Review of Applied Linguistics* 5: 101-23.

Grabe, W. P. (1987) in Connor, U. and Kaplan, R.B. (eds. 1987) *Writing across Languages: Analysis of L2 Text,* Reading, MA.: Addison-Wesley. Pp.115-137

Granger, S. The Learner Corpus: a revolution in applied linguistics, in *English Today, 39* Vol.10 No.3 July 1994, Cambridge University Press. Pp. 25-32.

Green, G. M. (1996) *Pragmatics and Natural Language Understanding,* 2nd Edition, New Jersey: Lawrence Erlbaum

Guillot, M-N. & Kenning M-M. (1995) *Exploiting the Potential of CD-ROM Databases: Staff Induction at the University of East Anglia* in Computer Assisted Language Learning 1995 Vol.8 No.4 pp. 365-381.

Hakuta, K. (1974). 'Prefabricated patterns and the emergence of structure in second language acquisition'. *Language Learning* 24:287-97.

Hakuta, K. (1976) 'Becoming bilingual: a case study of a Japanese child learning English**'.** *Language Learning* 26: 321-51.

Halliday, M.A.K. (1985) *An introduction to functional grammar,* London: Edward Arnold.

Halliday, M.A.K. (1993) "Quantitative Studies and Probabilities in Grammar" in Hoey, M. (ed.1993), *Data, Description, Discourse***,** London, Harper Collins. Pp.1-25.

Halliday, M.A.K. (1994) "The construction of knowledge and value in the grammar of scientific discourse, with reference to Charles Darwin's *The Origin of Species"* in Coulthard, M. (ed. 1994 - 1998) *Advances in Written Text Analysis,* London: Routledge. Pp. 136-156.

Halliday, M.A.K. (1998) "Things and relations: Regrammaticising experience as technical knowledge" in Martin, J. R. and Veel, R. (1998) *Reading Science*, London: Routledge. Pp185-235.

Halliday, M., McIntosh, A. & Strevens, P. (1964) *The linguistic sciences and language teaching***.** London: Longman.

Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*, London: Longman.

Halliday, M.A.K. and Martin, J.R. (1993) *Writing Science. Literacy and Discursive Power* London: The Falmer Press.

Halliday, M.A.K., Gibbons, J., Nicholas, H. (eds. 1990) *Learning Keeping and Using Language* Volume II Selected Papers from the 8th World Congress of Applied Linguistics, Sydney, 16-21 August 1987. Amsterdam/Philadelphia: John Benjamin's Publishing.

Harris, R. and Talbot, J. T. (1989) *Landmarks in Linguistic Thought: The Western Tradition from Socrates to Saussure,* London: Routledge.

Hatch, E. & Brown, C. (1995) *Vocabulary, Semantics, and Language Education,* Cambridge: Cambridge University Press.

Heppel, S. (1987) "Trainers Reach Crisis Point" in *Educational Computing,* Sept. 1987, Vol. 8, Issue 6.

Herbert, A.J. (1965) *The Structure of Technical English***,** London, Longman

Heslot, J. (1982) *Tense and other Indexical Markers in the Typology of Scientific Texts in English***,** in Hoedt, J., Lundquist, L., Picht, H. & Qvistgaard, J. (eds. 1982)) Proceedings of the 3rd European Symposium on LSP, Copenhagen August 1981, The Copenhagen School of Economics.

Higgins, J. (1985) “Computers in English Language Teaching” ELT Document 122, London: Pergamon Press.

Higgins, J. (1988) *Language Learners and Computers,* London: Longman.

Hill, A. A. (ed. 1969-78) *Linguistics***,** Washington: Forum.

Hockey, S. and Ide, N. (eds. 1994) *Research in Humanities Computing* Oxford: Clarenden Press.

Hoedt, J. & Turner, R. (eds. 1981) *New Bearings in LSP***,** The Copenhagen School of Economics.

Hoedt, J. & Lundquist, L., Picht, H. & Qvistgaard, J. (eds. 1982)) Proceedings of the 3rd European Symposium on LSP, Copenhagen August 1981, The Copenhagen School of Economics.

Hoey, M. (1991) *Patterns of Lexis in Text***,** Oxford, OUP.

Hoey, M. (ed.1993), *Data, Description, Discourse***,** London, Harper Collins.

Hoffman, L. (1981) *The Linguistic Analysis and Teaching of LSP in the German Democratic Republic,* in Hoedt, J. & Turner, R. (eds. 1981) *New Bearings in LSP,* The Copenhagen School of Economics. Pp. 107-130.

Hofland, K. & Johansson, S. (1982) *Word Frequencies in British and American English.* Bergen: Norwegian Computing Centre for the Humanities/London: Longman.

Holden, S. (ed. 1977) *English for Specific Purposes***.** Oxford: Modern English Publications.

Holliday, A. (1984) Research into classroom culture as necessary input into syllabus design. In Swales, J. & Mustafa, H. (eds.1984) *English for Specific Purposes in the Arab World***,** 29-51. Birmingham: University of Aston.

Holliday, A. & Cooke, T. (1982). An ecological approach to ESP. *Lancaster Practical Papers in English Language Education, 5 (Issues in ESP), 123-43.*

Hopper, P. J. (1997) "Discourse and the category 'Verb' in English" in *Language and Communication*, Vol. 17, Nº 2: Pergamon. Pp.93-102.

Howcroft, S.J. (1989) "CALL in Coimbra" *MUESLI News* Micro Users in English as a Second Language Institutions, The I.A.T.E.F.L. Special Interest Group for Computer Assisted Language Learning December 1989 Pp.6-7.

Howcroft, S.J. (1986) "English for Special Groups: CTT Computer Personnel Course", Report on the British Council Conference on Computer Assisted Language Learning, Lisbon 1986, Lisbon: The British Council.

Huang, J. (1971) *A Chinese Child's Acquisition of English Syntax*. Unpublished Master's Thesis. University of California at Los Angeles quoted in Nattinger and De Carrico (1992) *Lexical Phrases and Language Teaching* Oxford: Oxford University Press.

Huddlestone, R. D. (1971) *The Sentence in Written English*, Cambridge: Cambridge University Press.

Hughes, A. *Achievement and Proficiency: The Missing Link?* in Testing English for University Study, ELT Documents: 127, MEP 1988

Hughes, G. (1997) "Developing a Computing Infrastructure for Corpus-based Teaching" in Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora*, London: Longman. Pp. 292-307.

Hunstan, S. (1994) "Evaluation and organization in a sample of written academic discourse" in Coulthard, M. (ed.1994 - 1998) *Advances in Written Text Analysis"* Pp.191-218 London:Routledge

Hutchinson, T. (1988) "Making Materials Work in the ESP Classroom" in Chamberlain, R. and Baumgardner, R. J. (eds. 1988) *ESP in the Classroom: Practice and Evaluation, ELT Documents 128* London: Modern English Publications. Pp. 71-75.

Hutchinson, T. & Waters, A. (1980). ESP at the crossroads. *ESP Newsletter, 36* (Reprinted in Swales (ed. 1985), Pp 177-85.)

Hutchinson, T. & Waters, A. (1984)"How communicative is ESP?" in the *ELT Journal Volume 38/2 April 1984* Pp.108-113.

Hutchinson, T. & Waters, A. (1987), *English for Specific Purposes*, Cambridge: Cambridge University Press.

Hymes, D. (1971) *On Communicative Competence.* Philadelphia: University of Pennsylvania Press.

Johansson, S. (ed. 1982) *Computer Corpora in English Language Research***.** Bergen: Norwegian Computing Centre for the Humanities.

Johansson, S. (1991) "Times change and so do corpora" in Aijmer, K. & Altenberg, B. (eds. 1991) *English Corpus Linguistics***,** London, Longman.

Johansson, S., Leech, G. N. and Goodluck, H. (1978) *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers***.** Department of English, University of Oslo.

Johns, A.M. (1986) *Coherence and Academic Writing: some Definitions and Suggestions for Teaching,* in TESOL Quarterly Vol. 20 Nº2 1986 Pp247-265).

Johns, T (1994-98) "The text and its message" in Coulthard, M. (ed. 1994 - 1998) *Advances in Written Text Analysis,* London: Routledge. Pp. 102-116.

Johns, T. (1997) "Contexts: the Background, Development and Trialling of a Concordance-based CALL Program" in Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora,* London: Longman. Pp.100-115.

Johns, T. (1989) "Whence and whither classroom concordancing?" in Bongaerts, T. *et al.* (eds. 1989) *Computer Applications in Language Learning.* The Netherlands: Foris Publications. Pp. 9-27.

Jones, C. (1991) An integrated model for ESP syllabus design. *English for specific Purposes,*10**,** 3, Pp.155-72.

Jones, C. and Fortescue, S. (1987) *Using Computers in the Language Classroom,* London: Longman.

Jordan, R. R. (1997) *English for Academic Purposes: A guide and resource book for teachers,* Cambridge: Cambridge University Press.

Jordan, R. & Mackay, R. (1973) A survey of the spoken English problems of overseas postgraduate students at the universities of Manchester and Newcastle-upon-Tyne. *Journal of the Institute of Education of the Universities of Newcastle-upon-Tyne and Durham,* 125.

Kachru, B.B. (1990) "World Englishes and Applied Linguistics" in Halliday, M.A.K., Gibbons, J., Nicholas, H. (eds. 1990) *Learning Keeping and Using Language* Volume II Selected Papers from the 8th World Congress of Applied Linguistics, Sydney, 16-21 August 1987. Amsterdam/ Philadelphia: John Benjamin's Publishing. Pp.203-229.

Kaplan, R. B. (1993) "Conquest of Paradise - Language Planning in New Zealand" in Hoey, M. (ed. 1993) *Data, Description, Discourse,* London: Harper Collins.

Kaplan, R. B. (1966) "Cultural thought patterns in inter-cultural education. *Language Learning* 16, Pp 1-20.

Keatley, A. G. (ed. 1985) *Technological Frontiers and Foreign Relations,* Washington: National Academy Press.

Kelly, L. G. (1969) *25 Centuries of Language Teaching,* Rowley, Mass.: Newbury House

Kennedy, C. (1983) "An ESP Approach to EFL/ESL Teacher Training" in *The ESP Journal Volume 2 1983* Pergamon Press Ltd. Pp.73-85.

Kennedy, C. & Bolitho, R. (1984) *English for Specific Purposes***,** London, Macmillan.

Kennedy, G. (1989) *Collocations: where grammar and vocabulary teaching meet.* Paper presented at the RELC Seminar, Singapore.

Kenning, M. and Kenning, M.M. (1983) *Introduction to computer-assisted language teaching.* Oxford: Oxford University Press.

Kenworthy, J. (1991) *Language in Action***,** London: Longman

Kerridge, R. & Sammells, N. (eds. 1998) *Writing the Environment: Ecocriticism and Literature,* London: Zed Books.

Kinsella, V. (ed. 1978) *Language Teaching and Linguistics: Surveys,* Cambridge: Cambridge University Press.

Krashen, S.D. (1982) *Principles and Practice in Second Language Acquisition,* Oxford: Pergamon.

Kress, G., Leite García, R. and van Leeuwen, T (1997) "Discourse Semiotics" in van Dijk, T. A., (ed. 1997) *Discourse as Structure and Process,* London: SAGE Publications.

Kroch, A.S. and Hindle, D.M. (1982) *A Quantitative Study of the Syntax of Speech and Writing,* Final Report to the National Institute of Education.

Kubanek, A. (1998) "Primary foreign language teaching in Europe - trends and issues" *Language Teaching* 31 October 1998 Pp.193-205, Cambridge University Press.

Labov, W. (1972) *Rules for Ritual Insults*, in Sudnow, D. (ed. 1972) *Studies in Social Interaction*, New York: The Free Press.

Lackstrom, J.E., Selinker, L. and Trimble, P. (1972) "Grammar and Technical English" in *English Teaching Forum,* September-October 1972. Pp. 3- 14.

Langkilde, C. *Typical Aspects of the Reading Comprehension Barrier of Undergraduate Language Students at the Copenhagen School of Economics***.** in Hoedt, J. & Lundquist, L., Picht, H. & Qvistgaard, J. (eds. 1982)) Proceedings of the 3rd European Symposium on LSP, Copenhagen August 1981, The Copenhagen School of Economics. Pp.515-524.

Laurillard, D. (1993-97) *Rethinking University Teaching: a framework for the effective use of educational technology,* London: Routledge.

Leech, G. (1993) 100 million words of English, *English Today, 33* Vol.9 No.1 January 1993, Cambridge University Press. Pp 9-15

Leech, G. and Candlin, C. (eds. 1986) *Computers in English Language Teaching and Research,* London: Longman.

Leech, G. (1997) "Teaching and Language Corpora: a Convergence" in Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora,* London: Longman. Pp. 1-23.

Leech, G. & Fligelstone, S. (1992) Computers and Corpus Analysis, in Butler, C. S. (ed.1992) *Computers and Written Texts***,** P.115-140 Oxford: Basil Blackwell.

Leitner, G. (1993) "Where to Begin or Start?: Aspectual Verbs in Dictionaries," in Hoey, M. (ed. 1993) *Data, Description, Discourse,* London: Harper Collins.

Lemke, J. L. (1990) “Technical Discourse and Technocratic Ideology” in Halliday, M.A.K., Gibbons, J., Nicholas, H. (eds. 1990) *Learning Keeping and Using Language* Volume II Selected Papers from the 8th World Congress of Applied Linguistics, Sydney, 16-21 August 1987. Amsterdam/Philadelphia: John Benjamin’s Publishing. Pp.435-460.

Lemke, J. L. (1998) “Multiplying meaning: Visual and verbal semiotics in scientific text” in Martin, J. R. and Veel, R. (1998) *Reading Science*, London: Routledge. Pp.87-113.

Luke, A. (ed.1992) *Critical Perspectives on Literacy and Education,* writing in the Introduction of Halliday, M.A.K. and Martin, J.R. (1993) *Writing Science. Literacy and Discursive Power* London: The Falmer Press.

Mackay, R. (1978) “Identifying the nature of learners’ needs,” in Mackay & Mountford (eds. 1978), *English for Specific Purposes*, London, Longman. Pp 21-42.

Mackay, R. & Mountford, A.J. (eds.1978), *English for Specific Purposes*, London, Longman.

Mackay, R. & Palmer, J. (eds. 1981) *Languages for Specific Purposes*, Rowley, MA: Newbury House

Martin, J. R. (1992) *English Text: System and Structure,* Philadelphia and Amsterdam: Benjamins.

Martin, J. R. and Veel, R. (1998) *Reading Science,* London: Routledge.

Matthews, M. R. (1994) *Science Teaching,* London: Routledge.

McCarthy, M. (1990), *Vocabulary,* Oxford, Oxford University Press.

McCarthy, M. (1991), *Discourse Analysis for Language Teachers*, Cambridge: Cambridge University Press.

McCarthy, M. & Carter, R. (1994) *Language as Discourse. Perspectives for Language Teaching,* London: Longman.

Meijs, W. (1992) *Computers and Dictionaries* in Butler, C. S. (ed. 1992) *Computers and Written Texts*, Oxford: Basil Blackwell.

Meyers, G. (1994-98) “Narratives of science and nature in popularizing molecular genetics” in Coulthard, M. (ed. 1994 - 1998) *Advances in Written Text Analysis,* London: Routledge. Pp.179-190.

Mettinger, A. (1994) *Aspects of Semantic Opposition in English*, Clarendon Press Oxford, Oxford University Press.

Mindt, D. (1997) Corpora and the Teaching of English in Germany in Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora,* London: Longman. Pp. 40-50

Minugh, D. (1997) "All the Language that's Fit to Print: Using British and American Newspaper CD-ROMs as Corpora," in Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora,* P 67-82 London: Longman.

Moon, R. "The analysis of meaning" in Sinclair, J.M. (ed. 1987) *Looking Up An account of the COBUILD Project in lexical computing*, London, Harper Collins Publishers Pp. 86-103.

Moon, R. (1994) "The analysis of fixed expressions in text" in Coulthard, M. (ed. 1994 - 1998) *Advances in Written Text Analysis,* London: Routledge. Pp. 117-135.

Moreira, A.A. (1996) *Desenvolvimento da Flexibilidade Cognitiva dos Alunos-Futuros professores: uma experiência em Didáctica do Inglês,* Unpublished PhD Thesis, University of Aveiro.

Mountford, A. (1988) "Factors Influencing ESP Materials Production and Use" in Chamberlain, R. and Baumgardner, R. J. (eds. 1988) *ESP in the Classroom: Practice and Evaluation, ELT Documents 128* London: Modern English Publications. Pp.76-84.

Moura Carvalho, J. (1991) "The Minerva Project" in *MUESLI News* October 1991. Pp 3-7.

Munby, J. (1978) *Communicative Syllabus Design*, Cambridge, Cambridge University Press.

Murray, D. E. (1988) "Computer-mediated communication: Implications for ESP", *English for Specific Purposes*, vol.7,1, Pp. 3-18.

Myers, G. (1998) "Narratives of science and nature in popularizing molecular genetics" in Coulthard, M. (ed. 1994 - 1998) *Advances in Written Text Analysis,* London: Routledge. Pp179- 190.

National Education Goals Report: Building a Nation of Learners (1997) Washington: U.S. Department of Education. National Education Goals Panel. (http://www.negp.gov/).

Nattinger, J. R. and DeCarrico, J. S. (1992) *Lexical Phrases and Language Teaching* Oxford: Oxford University Press.

Nelson, M. (1992) *A model for course design in ESP for Business*. Unpublished M.Ed. TESOL dissertation, University of Manchester.

Newmark, L. (1979) "How not to interfere with language learning" in Brumfit, C. J. and Johnson, K. (1979-81) *The Communicative Approach to Language Teaching,* Oxford: Oxford University Press. Pp.160-166

Nunes, M. A. (1999) "Teaching English for Specific Purposes: The Guts to do it" Actas do 4º Encontro Nacional do Ensino das Línguas Vivas no Ensino Superior, Faculadade de Letras da Universidade de Porto. Pp.255-261.

O'Brien, T. & Jordan, R.R. (1985) *Developing Reference Skills***,** London: Collins.

Opitz, K. (1983) *Linguistics between Artificiality and Art: Walking the Tightrope of LSP Research***,** *CILA Bulletin 37*, Neuchatel.

Owen, G. T. (1973) "A Reading/Comprehension Course for Students of Science and Technology", *ELT Documents 73/4,* The British Council.

Palmer, J. (1981) *Register research design***,** in Mackay & Palmer (eds. 1981), *Language for Specific Purposes,* Rowley, MA: Newbury House.

Paltridge, B. (1996) "Genre, text type, and the language learning classroom" *ELT Journal Volume 50/3 July 1996:* Oxford University Press pp 237-243.

Papert, S. (1980) *Mindstorms: Children, Computers and Powerful Ideas,* Brighton:Harvester Press.

Peters, A. (1983). *The Units of Language Acquisition*. Cambridge: Cambridge University Press.

Peters P. (1998) "Langscape: Surveying contemporary English usage" in *English Today: The International review of the English Language* ET **53** Volume 14 Nº 1 January 1998 Pp 3-6.

Phillips, M. (1987) *Communicative Language Learning and the Microcomputer*, London: The British Council.

Pongsurapipat, S. (1975) *Text and Discourse Analysis on some Aspects of EST from Three Scientific Textbooks (Biology, General Chemistry, Physics)*, SEAMEO Regional English Language Centre, Thailand.

Pope, R. (1998) *The English Studies Book*, London: Routledge.

Porter D. (1976) *Scientific English: An Oversight in Stylistics?* Studia Anglica Posnaniensa, no.8, Poznan, Poland

Pilbeam, A. (1979) "The language audit." *Language Training, 1*,2, Pp. 4-5.

Pugh, A. K. & Uljin, J. M. (eds.1984), *Reading for Professional Purposes***,** London, Heinemann.

Quin, A. and Porter, N. (1994) "Investigating English usage with ICECUP" in *English Today, 33* Vol.9 No.1 January 1993, Cambridge University Press. Pp 19-24

Quirk, R. (1995), *Grammatical and Lexical Variance in English***,** London, Longman.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

Quirk, R. and Widdowson, H. G. (eds.1985) *English in the World: Teaching and Learning the Languages and Literatures*, Cambridge: Cambridge University Press and the British Council.

Renouff, A. (1987) "Corpus Development" in Sinclair, J. M. (ed. 1987) *Looking up*. London: Collins. Pp1-22.

Renouff, A. (1987) "Moving On" in Sinclair, J. M. (ed. 1987) *Looking up: An account of the COBUILD Project in lexical computing*. London: Collins. Pp167-178.

Renouff, A. (1997) "Teaching Corpus Linguistics to Teachers of English," in Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora,* P.255-266 London: Longman.

Renouff, A. and Sinclair, J. McH. (1991)"Collocational Frameworks in English", in Sinclair, J. (1991) *Corpus, Concordance, Collocation***.** Oxford University Press Pp. 128-143.

Richterich, R. (1971) Analytical classification of the categories of adults needing to learn foreign languages. Reprinted in Trim *et al.*(1973/1980) Strasbourg: Council of Europe/Oxford: Pergamon.

Richterich, R. (1973/1980) Definition of the language needs and types of adults. In Trim *et al.* (1973/1980) Pp 29-88 Strasbourg: Council of Europe/Oxford: Pergamon.

Riley, P. (1989) "Keeping Secrets: ESP/LSP and the sociology of knowledge" in the *European Journal of Teacher Education, Vol. 12, No. 2, 1989* Pp.69-80.

Riley, R. W. (1997-8) "Foundation of a Nation - Strong and Effective Schools in *Education in the United States: Continuity and Change, U.S. Society & Values Electronic Journals of the U.S. Information Agency*, Vol. 2 **Nº 5** December 1997 5-8 and *American Studies Journal* Number 41 Summer 1998 Pp 55-58

Robinson, M. (1994) "Using Email and the Internet in Science Teaching", *Journal of Information Technology for Teacher Education* Vol. 3, N°2 1994

Robinson, P. (1980), *ESP (English for Specific Purposes)***,** Oxford, Pergamon Press Ltd.

Robinson, P.C. (1991) *ESP TODAY: A Practitioner's Guide***,** London: Prentice Hall International

Rosenthal, J. W. (1996) *Teaching Science to Language Minority Students,* Clevedon: Multilingual Matters.

Rowntree, D. (1992) *Exploring Open and Distance Learning*, London: Koogan Page.

Scollon R. and Scollon, S. W. (1995) *Intercultural Communication: a discourse approach,* Oxford:Blackwell

Sedelow, S., Yeates and Sedelow, W.A. Jr (1994) "A Topologic Model of the English Semantic Code and its Role in Automatic Disambiguation for Discourse Analysis" in Hockey, S. and Ide, N. (eds 1994) *Research in Humanities Computing* Oxford: Clarenden Press.

Selinker, L. (ed. 1981) *English for Academic and Technical Purposes***,** London: Newbury House.

Selinker, L. 81991) *Rediscovering Interlanguage,* London: Longman.

Selinker, L. & Trimble, L. (1976) *Scientific and technological writing: the choice of the tense,* Forum (English Teaching Forum), 14 (4), 22-26.

Serway, R. A. (1992) *Physics for Scientists and Engineers with Modern Physics* Third edition, Updated Version, Saunders Golden Sunburst Series, USA

Sinclair, J.M. (ed. 1987) *Looking Up An account of the COBUILD Project in lexical computing***,** London, Harper Collins Publishers

Sinclair, J. (1991) *Corpus, Concordance, Collocation***.** Oxford University Press

Sinclair, J. McH. (1991) “Shared Knowledge.” In *Georgetown University Round Table on Languages and Linguistics 1991.* Washington, DC: Georgetown University Press, Pp. 489-500.

Sinclair , J. McH. (1992) “Priorities in discourse analysis.” In R.M. Coulthard (ed.1994 - 1998), *Advances in Spoken Discourse Analysis.* London: Routledge, Pp.79-88.

Sinclair, J. McH. (1994-98) “Trust the text” in Coulthard, M. (ed. 1994 - 1998) *Advances in Written Text Analysis,* London: Routledge. Pp.12-25.

Sinclair, J. McH. (1997) “ Corpus Evidence in Language Description” in Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora,* London: Longman. Pp. 27-39.

Sinclair, J. McH. and Brazil, D. (1982) *Teacher Talk***,** Oxford: Oxford University Press

Sinclair, J. McH. and Coulthard, R. M. (1975) *Towards an Analysis of Discourse.* London: Oxford University Press.

Steffensen, M. and Joag-Dev, C. (1984) “Cultural Knowledge and Reading” in Alderson, C. and Urquhart, A. (eds. 1984) *Reading in a Foreign Language,* London: Longman. Pp48-64.

Stern, H. H., (1983), *Fundamental Concepts of Language Teaching***,** Oxford, OUP.

Stern, H.H. (1992) *Issues and Options in Language Teaching,* Oxford: Oxford University Press.

Strevens, P. (1973) *Technical, technological and scientific English***,** English Language Teaching Journal XXVII/3

Strevens, P. (1978) "Special -purpose Language learning: a perspective," in Kinsella, V. (ed. 1978) *Language Teaching and Linguistics: Surveys,* Cambridge: Cambridge University Press.

Strevens, P. (1988) "The learner and teacher of ESP" in Chamberlain, R. and Baumgardner, R. J. (eds. 1988) *ESP in the Classroom: Practice and Evaluation, ELT Documents 128* London: Modern English Publications. Pp.39-44.

Stuart, W. & Lee, E. (1972/1985) *The non-specialist user of foreign languages in industry and commerce.* Sidcup London Chamber of Commerce & Industry Examination Board.

Stubbs, M. (1992) "English Teaching, Information Technology and Critical Language Awareness in Fairclough, N. (ed. 1992) *Critical Language Awareness,* London: Longman. Pp 203-222.

Stubbs M. (1994) *Grammar, Text, and Ideology: Computer-assisted Methods in the Linguistics of Representation* Applied Linguistics, Vol. 15, No.2 Oxford University Press

Stubbs, M. (1996) *Text and Corpus Analysis*, Oxford: Blackwell

Stubbs, M. & Gerbig, A. (1993) Human and inhuman geography: on the computer-assisted analysis of long texts, in Hoey, M. (ed. 1993) *Data, Description, Discourse,* London: Harper Collins. Pp.64-85

Sudhikam, S. (1975) *Lexis,* in *Guidelines Sample Materials for the Teaching of English to First Year Tertiary Level Scientific and Technical Students in Universities in Thailand***,** SEAMEO/RELC (Mimeo)

Sudnow, D. (ed. 1972) *Studies in Social Interaction,* New York: The Free Press.

Svartvik, J. and Quirk, R. (eds.) *A corpus of English conversation***.** Lund: CWK Gleerup.

Swales, J. (1971) *Writing Scientific English***,** London: Nelson

Swales, J. (1973) "Introducing Teachers to English for Science and Technology"**,** *ELT Documents 73/6* The British Council.

Swales, J. (1978) *Writing "Writing Scientific English"***,** in Mackay & Mountford (eds 1978) *English for Specific Purposes***,** London: Longman**.** Pp. 43-55.
Swales, J. (1985), *Episodes in ESP***,** Oxford: Pergamon Press Ltd.

Swales, J. M. (1990) *Genre Analysis***.** Cambridge: Cambridge University Press.

Swales, J. & Mustafa, H. (eds.1984) *English for Specific Purposes in the Arab World***,** Birmingham: University of Aston.

Swales, J.M. and Najjar, H. (1987) "The writing of research article introductions." *Written Communication***,** 9(2), Pp.175-192.

Swift, J.(1976) *Gulliver's Travels*, Oxford University Press.

Tarone, E. & Yule, G. (1989) *Focus on the language learner***.** Oxford: Oxford University Press.

Tavares, C. F., Valente, M. T. and Roldão, M. C. (1996) "Dimensões Formativas de Disciplinas do Ensino Básico Língua Estrangeira." Lisboa: Instituto de Inovação Educacional.

Tavares, J., Santiago, R.A, Lencestre, L., Soares, I. (1996) *Niveis de Sucesso dos Alunos do 1º Ano dos Cursos de Ciências e Engenharia da Universidade de Aveiro*, Universidade de Aveiro.

Tesch, F. (1990) *Die Indefinitpronomina some und any in authentischen englischen Sprachgebrauch und in Lehrwerken:eine empirische Untersuchung.* Narr, Tubingen.

Tickoo, M. (1988) "Michael West in India: a centenary salute." *English Language Teaching Journal, 42,4,* Pp.294-300.

Todaka, Y. (1996) "Between and among: a data-based analysis" in *Journal of the International Linguistic Association WORD*, Volume47, Number 1 (April 1996). Pp. 13-39.

Tribble, C. & Jones, G. (1990) *Concordances in the Classroom***,** London, Longman.

Trim, J. L. M., (1998) "European perspectives on modern language learning" *Language Teaching* 31, October 1998,. Cambridge University Press. Pp.206-217.

Trim, J. L. M, Richterich, R., Van Ek, J. & Wilkins, D. (1973/80). *Systems development in adult language learning***.** Strasbourg: Council of Europe/Oxford: Pergamon.

Trimble, L. (1985), *English for Science and Technology. A discourse approach*, Cambridge, CUP.

Todd Trimble, M & Trimble, L *Rhetorical-Grammatical Features of Scientific and Technical Texts as a major Factor in Written Scientific Communication***,** in Hoedt, J., Lundquist, L., Picht, H. & Qvistgaard, J. (eds. 1982)) Proceedings of the 3rd European Symposium on LSP, Copenhagen August 1981, The Copenhagen School of Economics. Pp 199-218

Turner, , R. (1981) "A note on "special languages" and "specific purposes"", in Hoedt, J. & Turner, R. (eds. 1981) *New Bearings in LSP***,** The Copenhagen School of Economics.

van Dijk, T. A., (ed. 1997) *Discourse as Structure and Process,* London: SAGE Publications.

van Dijk, T. A., (ed. 1997) *Discourse as Social Interaction,* London: SAGE Publications.

Van Ek, J. A., *The Threshold Level for Modern Language Learning in Schools* Council of Europe, Strasbourg 1976, Longman 1977

Van Ek, J. A., Alexander, L. G. & Fitzpatrick, M. A. (1980) *Waystage English,* Oxford: Oxford University Press.

Van Ek, J. A. and Trim, J. (1998) *Threshold 1990,* Cambridge: Cambridge University Press.

Veel, R. (1998) "The greening of school science: Ecogenesis in secondary classrooms" in Martin, J. R. and Veel, R. (1998) *Reading Science,* London: Routledge. Pp. 114-151.

Warschauer, M. (1999) "CALL vs. Electronic Literacy: Reconceiving Technology in the Language Classroom" London: CILT Research Forum.

Waters, M. and Waters, A. (1992) "Study skills and study competence: getting the priorities right" in ELT Journal Volume 46/3 July 1992 Oxford: Oxford University Press.

Weber, H. *Language for specific Purposes, Text Typology, and Text Analysis: Aspects of a Pragmatic-Functional Approach* in Hoedt, J. & Lundquist, L., Picht, H. & Qvistgaard, J. (eds. 1982)) Proceedings of the 3rd European Symposium on LSP, Copenhagen August 1981, The Copenhagen School of Economics. Pp. 219- 234.

Weir, C. (1988) *The Specification, Realization and Validation of an English Language Proficiency Test* in Testing English for University Study, ELT Documents: 127, MEP 1988

West, M. (1953) *Supplementary Scientific and Technical Vocabulary*, in *A General Service List of English Words*, London: Longman

West, R. (1994) "Needs analysis in language teaching: State of the art article," *Language Teaching,* January 1994, Cambridge University Press, Pp 1-19.

White, R. (1988). *The ELT Curriculum: Design, Innovation and Management,* Oxford: Blackwell.

White, P. R. R. (1998) "Extended reality, proto-nouns and the vernacular: Distinguishing the technological from the scientific" in Martin, J. R. and Veel, R. (1998) *Reading Science,* London: Routledge. Pp.266- 296.

Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora,* London: Longman.

Widdowson, H. G. (1974) *The deep structure of discourse and the use of translation,* in Corder, S. P. & Roulet, E. (eds. 1974) *Linguistic Insights in Applied Linguistics,* Second Neuchatel Colloquium in Applied Linguistics, Brussels: AIMAV; Paris: Didier and in Brumfit, C. J. and Johnson, K. (1979-81) *The Communicative Approach to Language Teaching,* Oxford: Oxford University Press. Pp.61-71.

Widdowson, H. G. (1974) A*n approach to the teaching of scientific English discourse,* RELC Journal, 5, Pp 27-40.

Widdowson, H. G. (ed. 1979), *Reading and Thinking in English: Discovering Discourse***,** Oxford, Oxford University Press.

Wilkins, D. A (1976) *Notional Syllabuses*, Oxford: Oxford University Press.

Wilks, Y.A., Slator, B.M.,& Guthrie, L.M. (1996) *Electric Words: Dictionaries Computers and Meanings*, Mass: MIT Press.

Wilson, E. (1997) "The Automatic Generation of CALL Exercises from General Corpora" in Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds. 1997) *Teaching and Language Corpora*, London: Longman. Pp.116-130

Wingard, P. (1981) *Some verb forms and functions in six medical texts*, in Selinker (ed. 1981) London, Newbury House.

Wong-Fillmore, L. (1976) *The Second Time Around: Cognitive and Social Strategies in Second Language Acquisition*. Unpublished doctoral dissertation, Stanford University, quoted in Nattinger, J.R. and DeCarrico, J. S. (1992) *Lexical Phrases and Language Teaching* Oxford: Oxford University Press.

"World Education League: Who's top?" *The Economist* March 29th 1997 Pp.21-25.

Appendices

Appendix A

Biber's Algorithms and Functions

67 linguistic features were counted. These features include all features that: (1) have been assigned distinctive functions by previous research, and (2) can be automatically identified in spoken and written texts. Each of these features is discussed in turn here.

The following notation is used in the descriptions of the algorithms:

+: used to separate constituents
0: marks optional constituents
1: marks disjunctive options
xxx: stands for any word
#: marks a word boundary
T#: marks a 'tone unit' boundary, as defined in Quirk et al. (1972: 937-8) for use in the London-Lund corpus.'
DO: *do, does, did, don't, doesn't, didn't, doing, done*
HAVE: *have, has, had, having, -'ve#, -'d#, haven't, hasn't, hadn't*
BE: *am, is, are, was, were, being, been -'m#, -'re#, isn't, aren't, wasn't, weren't*
MODAL: *can, may, shall, will, -'ll#, could, might, should, would, must, can't, won't, couldn't, mightn't, shouldn't, wouldn't, mustn't*
AUX: MODAL/ DO/ HAVE/ BE/ -'s
SUBJPRO: *I, we, he, she, they* (plus contracted forms)
OBJPRO: *me, us, him, them* (plus contracted forms)
POSSPRO: *my, our, your, his, their, its* (plus contracted forms)
REFLEXPRO: *myself, ourselves, himself, themselves, herself, yourself, yourselves, itself*
PRO:SUBJPRO/OBJPRO/POSSPRO/REFLEXPRO/*you/her/it*
PREP: prepositions (e.g. *at, among* - see no. 39)
CONJ: conjuncts (e.g. *furthermore, therefore* - see no. 45)
ADV: adverbs (see no. 42)
ADJ: adjectives (see nos. 40, 41)
N: nouns (see nos. 14, 15, 16)
VBN: any past tense or irregular past participial verb
VBG: *-ing* form of verb
VB: base form of verb
VBZ: third person, present tense form of verb
PUB: 'public' verbs (see no. 55)
PRV: 'private' verbs (see no. 56)
SUA: 'suasive' verbs (see no. 57)
V: any verb
WHP: WH pronouns - *who, whom, whose, which*
WHO: other WH words - *what, where, when, how, whether, why, whoever, whomever, whichever, wherever, whenever, whatever, however*
ART: articles - *a, an, the, (dhi)*
DEM: demonstratives - *this, that, these, those*
QUAN: quantifiers - *each, all, every, many, much, few, several, some, any*
NUM: numerals - *one… twenty, hundred, thousand*
DET: ART/DEM/QUAN/NUM

ORD: ordinal numerals -*first* ... *tenth*
QUANPRO: quantifier pronouns - everybody, *somebody, anybody, everyone, someone, anyone, everything, something, anything*
TITLE: address titles
CL-P: clause punctuation (‘.’, ‘!’, ‘?’, ‘:’, ‘;’,’-‘)
ALL-P: all punctuation (CL-P plus',')

In the following discussion, the 67 linguistic features have been grouped into sixteen major categories: (A) tense and aspect markers, (B) place and time adverbials, (C) pronouns and pro-verbs, (D) questions, (E) nominal forms, (F) passives, (G) stative forms, (H) subordination features, (I) adjectives and adverbs, (j) lexical specificity, (K) specialized lexical classes, (L) modals, (M) specialized verb classes, (N) reduced or dispreferred forms, (0) coordination, and (P) negation.

(A) TENSE AND ASPECT MARKERS (nos. 1-3)

1. past tense

Any past tense form that occurs in the dictionary, or any word not otherwise identified that is longer than six letters and ends in *ed#*. Past tense forms are usually taken as the primary surface marker of narrative.

2. perfect aspect

(a) HAVE + (ADV) + (ADV) + VBN
(b) HAVE+NIPRO+VBN (questions)
(includes contracted forms of HAVE)

Perfect aspect forms mark actions in past time with ‘current relevance’ (Quirk *et al*. 1985:189ff).

3. present tense

All VB (base form) or VBZ (third person singular present) verb forms in the dictionary, excluding infinitives.
Present tense verbs deal with topics and actions of immediate relevance. They can also be used in academic styles to focus on the information being presented and remove focus from any temporal sequencing.

(B) PLACE AND TIME ADVERBIALS (nos. +-5)
mark direct reference to the physical and temporal context of the text, or in the case of fiction, to the text-internal physical and temporal world.

4. place adverbials *aboard, above, abroad, across, ahead, alongside, around, ashore, astern, away, behind, below, beneath, beside, downhill, downstairs, downstream, east, far, hereabouts, indoors, inland, inshore, inside, locally, near, nearby, north, nowhere, outdoors, outside, overboard, overland, overseas, south, underfoot, underground, underneath, uphill, upstairs, upstream, west*

This list is taken from Quirk *et al.* (1985:514ff). Items with other major functions, for example, *in, on,* which often mark logical relations in a text, have been excluded from the list.

5. time adverbials

afterwards, again, earlier, early, eventually, formerly, immediately, initially, instantly, late, lately, later, momentarily, now, nowadays, once, originally, *presently, previously, recently, shortly, simultaneously, soon, subsequently, today, tomorrow, tonight, yesterday*

This list is taken from Quirk *et al.* (1985:526ff). Items with other major functions, for example, *last, next,* which often mark logical relations within a text, have been excluded from the list.

(C) PRONOUNS AND PRO-VERBS (nos. 6-12)

Some studies have grouped all pronominal forms together as a single category which is interpreted as marking a relatively low informational load, a lesser precision in referential identification, or a less formal style (e.g., Kroch and Hindle 1982; Brainerd 1972). Other studies have grouped all personal pronouns into a single category, and interpret that category as marking interpersonal focus (Carroll 1960; Poole 1973; Poole and Field 1976). In the present analysis, I separate personal and impersonal pronominal forms, as well as each of the persons within the personal pronouns.

1. (C1) PERSONAL PRONOUNS

6. first person pronouns

I, me, we, us, my, our, myself, ourselves (plus contracted forms)

First person pronouns have been treated as markers of ego-involvement in a text.

7. second person pronouns

you, your, yourself, yourselves (plus contracted forms)

Second person pronouns require a specific addressee and indicate a high degree of involvement with that addressee (Chafe 1985). They have been used as a marker of register differences by Hu (1984), Finegan (1982), and Biber (1986a).

8. third person personal pronouns

she, he, they, her, him, them, his, their, himself, herself, themselves (plus contracted forms)

Third person personal pronouns mark relatively inexact reference to persons outside of the immediate interaction.

(C2) IMPERSONAL PRONOUNS

9. pronoun *it*

It is the most generalized pronoun, since it can stand for referents ranging from animate beings to abstract concepts. This pronoun can be substituted for nouns, phrases, or whole clauses. Chafe and Danielewicz (1986) and Biber (1986a) treat a frequent use of this pronoun as marking a relatively inexplicit lexical content due to strict time constraints and a non-informational focus. Kroch and Hindle (1982) associate greater generalized pronoun use with the limited amounts of information that can be produced and comprehended in typical spoken situations.

10.demonstrative pronouns (e.g., *this is ridiculous)*

(a) *that/ this/ these/ those* + V/AUX/ CL-P/T#/WHP/*and*

(where *that* is not a relative pronoun)

(b) *that's*

(c)T# + *that*

(*that* in this last context was edited by hand to distinguish among demonstrative pronouns, relative pronouns, complementizers, etc.)

Demonstrative pronouns can refer to an entity outside the text, an exophoric referent, or to a previous referent in the text itself. In the latter case, it can refer to a specific nominal entity or to an inexplicit, often abstract, concept (e.g., *this shows* ...). Chafe (1985; Chafe and Danielewicz 1986) characterizes those demonstrative pronouns that are used without nominal referents as errors typically found in speech due to faster production and the lack of editing.

11.indefinite pronouns

anybody, anyone, anything, everybody, everyone, everything, nobody, none, nothing, nowhere, somebody, someone, something (Quirk *et al.* 1985:376ff)

These forms have not been used frequently for register comparison. They are included here as markers of generalized pronominal reference, in a similar way to *it* and the demonstrative pronouns.

(C3) PRO-VERBS

12.pro-verb *do* (e.g., *the cat did it)*

DO when NOT in the following constructions:

DO + (ADV) + V (DO as auxiliary)

ALL-P/T#/WHP+DO (DO as question)

This feature was included in Biber (1986a) as a marker of register differences. Do as pro-verb substitutes for an entire clause, reducing the informational density of a text and indicating a lesser informational focus, due to processing constraints or a higher concern with interpersonal matters.

(D) QUESTIONS (no. 13)

Questions, like second person pronouns, indicate a concern with interpersonal functions and involvement with the addressee (Marckworth and Baker 1974; Biber 1986a). *Yes/no* questions were excluded from the present analysis because they could not be accurately identified by automatic analysis in formal spoken genres, where every phrase tends to be a separate intonation unit; that is, many intonation units begin with an auxiliary and therefore are identical in form to direct questions.

13. direct WH-questions

CL-P/T#+WHO+AUX

(where AUX is not part of a contracted form)

(E) NOMINAL FORMS (nos. 14-16)

The overall nominal characterization of a text and the distinction between nominal and verbal styles is identified as one of the most fundamental distinctions among registers by Wells (1960) and Brown and Fraser (1979). A high nominal content in a text indicates a high (abstract) informational focus, as opposed to primarily interpersonal or narrative foci. Nominalizations, including gerunds, have particularly been taken as markers of conceptual abstractness.

14. nominalizations

All words ending in *-tion#, -ment#, -ness#,* or *-ity#* (plus plural forms).

Nominalizations have been used in many register studies. Chafe (1982, 1985, and Danielewicz 1986) focuses on their use to expand idea units and integrate information into fewer words. Biber (1986a) finds that they tend to co-occur with passive constructions and prepositions and thus interprets their function as conveying highly abstract (as opposed to situated) information. Janda (1985) shows that nominalizations are used during note-taking to reduce full sentences to more compact and efficient series of noun phrases.

15. gerunds

All participle forms serving nominal functions - these are edited by hand.

Gerunds (or verbal nouns) are verbal forms serving nominal functions. As such, they are closely related to nominalizations in their functions. Some researchers (e.g., Chafe 1982) do not distinguish among the different participial functions, treating gerunds, participial adjectives (nos. 40-1), and participial clauses (nos. 25-8) as a single feature. In the present study, these functions are treated separately.

16. total other nouns

All nouns included in the dictionary, excluding those forms counted as nominalizations or gerunds.

This count provides an overall nominal assessment of a text. Nominalizations and gerunds are excluded from the total noun count so that the three features will be statistically independent.

(F) PASSIVES (nos. 17-18)

Passives have been taken as one of the most important surface markers of the decontextualized or detached style that stereotypically characterizes writing. In passive constructions, the agent is demoted or dropped altogether, resulting in a static, more abstract presentation of information. Passives are also used for thematic purposes (Thompson 1982; Finegan 1982; Weiner and Labov 1983; Janda 1985). From this perspective, agentless passives are used when the agent does not have a salient role in the discourse; by-passives are used when the patient is more closely related to the discourse theme than the patient.

17. agentless passives 18. by-passives**

(a) BE + (ADV) + (ADV) + VBN + (by)**

(b) BE + N/PRO + VBN + (by)** (question form)

(** no. 18 with the by-phrase)

(G) STATIVE FORMS (nos. 19-20)

Only a few studies have used stative forms for register comparisons. These forms might be considered as markers of the static, informational style common in writing, since they preclude the presence of an active verb. Conversely, they can be considered as non-complex constructions with a reduced informational load, and therefore might be expected to be more characteristic of spoken styles. Kroch and Hindle (1982) analyze existential *there* as being used to introduce a new entity while adding a minimum of other information. Janda (1985) notes that stative or predicative constructions (X *be Y*) are used frequently in note-taking, although the *be* itself is often dropped. Predicative constructions with *be*-ellipsis are also common in sports announcer talk (Ferguson 1983). These predicative constructions might be characterized as fragmented, because they are typically alternatives to more integrated attributive constructions (e.g., *the house is big* versus *the big house*). *Be* as main verb is used for register comparisons by Carroll (1960) and Marckworth and Baker (1974).

19.*be* as main verb

BE + DET/POSSPRO/TITLE/PREP/ADJ

20.existential *there* (e.g., *there are several explanations* ...)

(a) *there* + (xxx) + BE

(b) *there's*

(H) SUBO.RDINATION (nos. 21-38)

Subordination has perhaps been the most discussed linguistic feature used for register comparisons. It has generally been taken as an index of structural complexity and therefore supposed to be more commonly used in typical writing than typical speech. Some researchers, though, have found higher use of subordination in speech than writing (e.g., Poole and Field 1976). Halliday (1979) claims that even conversational speech has more subordination than written styles, because the two modes have different types of complexities: spoken language, because it is created and perceived as an on-going process, is characterized by 'an intricacy of movement [and by] complex sentence structures with low lexical density (more clauses, but fewer high-content words per clause)'; written language, in which the text is created and perceived as an object, is characterized by 'a denseness of matter [and by] simple sentence structures with high lexical density (more high-content words per clause, but fewer clauses)'.

Thompson (1983,1984,1985; Thompson and Longacre 1985; Ford and Thompson to appear) has carried out some of the most careful research into the discourse functions of subordination. She distinguishes between dependent clauses (complementation and relative clauses) and other types of subordination (e.g., adverbial clauses) that function to frame discourse information in different ways. Her studies have focused on the discourse functions of detached participial clauses, adverbial clauses in general, purpose clauses, and conditional clauses (see below). In all of these studies, Thompson emphasizes that subordination is not a unified construct, that different types of structural dependency have different discourse functions, and that particular subordination features are therefore used to different degrees in different types of discourse.

Beaman (1984) and Biber (1986a) also find that different subordination forms are distributed differently. Based on an analysis of spoken and written narratives, Beaman observes that there are more finite nominal clauses (that-clauses and WH-clauses) in speech and more non-finite nominal clauses (infinitives and participial clauses) in writing. She also discusses the distribution of relative and adverbial clauses in these texts (see below). In my own earlier studies, I find that *that*-clauses, WH-clauses, and adverbial subordinators co-occur frequently with interpersonal and reduced-content features such as first and second person pronouns, questions, contractions, hedges, and emphatics. These types of subordination occur frequently in spoken genres, both interactional (conversation) and informational (speeches), but they occur relatively infrequently in informational written genres. Relative clauses and infinitive were found to have a separate distribution from the other types of subordination, but they did not form a strong enough co-occurrence pattern for interpretation. These same features are discussed from the perspective of discourse complexity in Finegan and Biber (1986b).

These studies by Thompson and Beaman, and my own earlier studies, all show that different types of subordination function in different ways. Based on these analyses, I have divided the subordination features used in the present study into four sub-classes: complementation (H1), participial forms (H2), relative clauses (H3), and adverbial clauses (H4). Each of these is now discussed in turn.

(H1) COMPLEMENTATION (nos. 21-4)

21.*that* verb complements (e.g., *I said that he went)*

(a)*and/nor/but/or/also/* ALL-P + *that* +
DET/PRO/*there*/plural noun/proper noun/TITLE
(these are *that*-clauses in clause-initial positions)

(b)PUB/PRV/SUA/SEEM/APPEAR + *that* + xxx
(where xxx is NOT: V/AUX/CL-P/T#/*and)*
(*that*-clauses as complements to verbs which are not included in the above verb classes are not counted - see Quirk *et al.* 1985:1179ff.)

(c)PUB/PRV/SUA + PREP + xxx + N + *that*
(where xxx is any number of words, but NOT = N)
(This algorithm allows an intervening prepositional phrase between a verb and its complement.)

(d)T# + *that*
(This algorithm applies only to spoken texts. Forms in this context are checked by hand, to distinguish among that complements, relatives, demonstrative pronouns and subordinators.)

Chafe (1982, 1985) identifies *that*-complements as one of the indices of integration, used for idea-unit expansion in typical writing. Ochs (1979) describes complementation as a relatively complex construction used to a greater extent in planned than unplanned discourse. In contrast, Beaman (1984) finds more *that* complementation in her spoken than written narratives. Biber (1986a) finds that *that*-complements co-occur frequently With interactive features such as first and second person pronouns and questions, and that all of these features are more common in spoken than written genres. In that paper and in Finegan and Biber (1986b), this is interpreted in a similar way to Halliday's characterization: that this type of structural complexity is used in typical speech, where there is little opportunity for careful production or revision, while other types of linguistic complexity, notably lexical variety and density, are used in typical academic writing, which provides considerable opportunity for production and revision.

Other studies that have used *that*-complements for register comparisons include Carroll (1960), O'Donnell (1974), Frawley (1982), and Weber (1985). Winter (1982) notes that both verb and adjective *that*-complements provide a way to talk about the information in the dependent clause, with the speaker's evaluation (commitment, etc.) being given in the main clause and the propositional information in the *that*-clause.

Some verb complements do not have an overt complementizer (e.g., *I think he went*); these are counted as a separate feature (no. 60).

22.*that* adjective complements (e.g., *I'm glad that you like it*)

ADJ + (T#) + *that*

(complements across intonation boundaries were edited by hand)

Most studies of *that*-clauses consider only verb complements. Winter (1982) points out, however, that verb and adjective complements seem to have similar discourse functions, and so both should be important for register comparisons. Because there is no a priori way to know if *that* verb and adjective complements are distributed in the same way among genres, they are included as separate features here. Householder (1964) has compiled a list of adjectives that occur before *that*-clauses; Quirk *et al*. (1985:1222-5) give a grammatical and discourse description of these constructions.

23.WH-clauses (e.g., *I believed what he told me*)

PUB/PRVISUA + WHP/WHO + xxx

(where xxx is NOT = AUX - this excludes WH questions)

This algorithm captures only those WH clauses that occur as object complements to the restricted verb classes described below in nos. 5 5-7; see Quirk *et al.* 1985:1184-5. Other WH clauses could not he identified reliably by automatic analysis and so were not counted.

Similar to *that*-clauses, WH-clauses are complements to verbs. Chafe (1985) analyzes them as being used for idea unit expansion, and thus they should be more frequent in typical writing. Bearnan (1984) did not find WH-clauses in her written narratives; she writes that they resemble questions and serve interpersonal functions in discourse, accounting for their use in spoken but not written narratives. Winter (1982) notes that WH complements provide a way to talk about questions in the same way that *that*-complements provide a way to talk about statements, that is, with the speaker's evaluation, commitment, etc. provided in the main clause. Biber (1986a) finds WH-clauses to be distributed in a similar pattern to *that*-clauses, both of which co-occur frequently with interpersonal features such as first and second person pronouns and questions.

24. infinitives

to + (ADV) + VB

Infinitives are the final form of complementation to be included in the present study. The algorithm above groups together all infinitival forms: complements to nouns, adjectives, and verbs, as well as ‘purpose’ adverbial clauses (see below). The distribution and discourse functions of infinitives seem to be less marked than that of other types of subordination. Chafe (1982, 1985) includes infinitives as one of the devices used to achieve integration and idea-unit expansion in typical writing. Beaman (1984) finds that infinitives co-occur with other non-finite nominal clauses (especially participial clauses), and that they are more common in written than spoken narratives. Biber (1986a) finds a weak co-occurrence relationship between infinitives and relative clauses. Finally, Thompson (1985) carefully distinguishes between those infinitives functioning as complements and those functioning as adverbial purpose clauses, and she analyzes the thematic discourse functions of the latter in some detail. Although this is an important functional distinction, it is not made here because of the limitations of the automatic analysis.

(H2) PARTICIPIAL FORMS (nos. 25-8)
Participles are among the most difficult forms to analyze. They can function as nouns, adjectives, or verbs, and within their use as verbs, they can function as main verbs (present progressive, perfect, or passive), complement clauses, adjectival clauses, or adverbial clauses. Some studies do not distinguish among these functions, counting all participial forms (excluding main verbs) as a single feature (e.g., Chafe 1982; Beaman 1984). Many studies also do not distinguish between present and past participial clauses, or they count only present participle forms. In the present analysis, each of the different grammatical functions of participles is treated as a separate linguistic feature, since these grammatical functions are likely to be associated with different discourse functions.

Studies that consider participles typically find that they occur more frequently in writing than in speech; the usual interpretation associated with this distribution is that participles are used for integration or structural elaboration.

Thompson (1983) distinguishes syntactically detached participial clauses (e.g., *Stuffing his mouth with cookies, Joe ran out the door)* from other participial functions. She shows how these clauses are used for depictive functions, that is, for discourse that describes by creating an image. No. 25 and no. 26 below are algorithms for detached participial clauses (present and past). These forms were edited by hand to exclude participial forms not having an adverbial function. Participial clauses functioning as reduced relatives, also known as WHIZ deletions, are treated separately (nos. 27 and 28*).* Janda *(1985)* notes the use of these forms in note-taking to replace full relative clauses, apparently because they are more compact and integrated and therefore well-suited to the production of highly informational discourse under severe time constraints. In the present analysis, these forms were also edited by hand to distinguish between subordinate clause functions and other functions; in particular, past participles following a noun can represent either a simple past tense form or the head of a reduced relative clause, and these forms thus needed to be checked by hand. Finally, participles functioning as nouns and adjectives were distinguished (nos. 15 and 40-1 respectively); these forms were also edited by hand to verify their grammatical function.

25.present participial clauses
(e.g., *Stuffing his mouth with cookies, Joe ran out the door)*
T#/ALL-P + VBG + PREP/DET/WHP/WHO/PRO/ADV
(these forms were edited by hand)

26.past participial clauses
(e.g., *Built in a single week, the house would stand for fifty years.)*
T#/ALL-P + VBN + PREP/ADV
(these forms were edited by hand)
27.past participial WHIZ deletion relatives

(e.g., *the solution produced by this process)*
N/QUANPRO + VBN + PREP/BE/ADV
(these forms were edited by hand)

28.present participial WHIZ deletion relatives
(e.g., *the event causing this decline is . . .)*
N + VBG
(these forms were edited by hand)

(H3) RELATIVES (nos. 29-34)
Relative clauses have been used frequently as a marker of register differences. Relatives provide a way to talk about nouns, either for identification or simply to provide additional information (Winter 1982; Beaman 1984). Ochs (1979) notes that referents are marked differently in planned and unplanned discourse: simple determiners are preferred in unplanned discourse while relative clauses are used for more exact and explicit reference in planned discourse. Chafe (1982, 1985) states that relative clauses are also used as a device for integration and idea unit expansion.

In general, these studies find that relative clauses occur more frequently in writing than in speech. Some studies, however, do not treat all relative clauses as a single feature and consequently do not find a uniform distribution. Kroch and Hindle (1982) and Beaman (1984) provide two of the fullest discussions. Beaman analyzes *that* relatives separately from WH relatives and finds more *that* relatives in her spoken narratives but more WH relatives in her written narratives; further, she finds more relativization on subject position in her spoken narratives versus more relativization on object positions in her written narratives. In contrast, Kroch and Hindle find more relativization on subject position in their written texts and more relativization on object position in their spoken texts. They attribute this to a greater use of pronouns in subject position in speech, making this position unavailable for relativization. Both of these studies also analyze pied-piping constructions separately, finding more in written than in spoken texts. In the present analysis, I separate *that* from WH relatives, and relativization on subject position from relativization on object position. Pied-piping constructions are also treated separately.
29.*that* relative clauses on subject position
(e.g., *the dog that bit me)*
N + (T#) + that + *(ADV)* + AUX/V
(that relatives across intonation boundaries are identified by hand.)
30.*that* relative clauses on object position
(e.g., *the dog that I saw)*
N + (T#) + that + *D*ET/ SUBJPRO / POSSPRO / *it/* ADJ / plural noun/ proper noun / possessive noun / TITLE
(This algorithm does not distinguish between *that* complements to nouns and true relative clauses.)
(In spoken texts, *that* relatives sometimes span two intonation units; these are identified by hand.)
31.WH relative clauses on subject position

(e.g., *the man who likes popcorn)*
xxx + yyy + N + WHP + (ADV) + AUX/V
(where xxx is NOT any form of the verbs ASK or TELL; to exclude indirect WH questions like *Tom asked the man who went to the store)*
32.WH relative clauses on object positions
(e.g., *the man who Sally likes)*
xxx + yyy + N + WHP + zzz
(where xxx is NOT any form of the verbs ASK or TELL, to exclude indirect WH questions, and zzz is not ADV, AUX or V, to exclude relativization on subject position)
33.pied-piping relative clauses
(e.g., *the manner in which he was told)*
PREP + WHP
34.sentence relatives
(e.g., *Bob likes fried mangoes, which is the most disgusting thing I've ever heard of)*
T#/, +which
(These forms are edited by hand to exclude non-restrictive relative clauses.)

Sentence relatives do not have a nominal antecedent, referring instead to the entire predication of a clause (Quirk *et al.* 1985:1118-20). They function as a type of comment clause, and they are not used for identificatory functions in the way that other relative clauses are. A preliminary analysis of texts suggested that these constructions were considerably more frequent in certain spoken genres than in typical writing, and they are therefore included here as a separate feature.

(H4) ADVERBIAL CLAUSES (nos. 35-8)
Adverbial clauses appear to be an important device for indicating informational relations in a text. Overall, Thompson (1984) and Biber (1986a) find more adverbial clauses in speech than in writing. Several studies, though, separate preposed from postposed adverbial clauses, and find that these two types have different scopes, functioning to mark global versus local topics, and that they have different distri-butions (Winter 1982; Chafe 1984a; Thompson 1985; Thompson and Longacre 1985; Ford and Thompson 1986).

There are several subclasses of adverbial clauses, including condition, reason/cause, purpose, comparison, and concession (Quirk *et al.* 1985:1077-18; Tottie 1986; Smith and Frawley 1983). The most common types, causative, concessive, and conditional adverbials, can be identified unambiguously by machine (nos. 35-7); the other subordinators are grouped together as a general category (no. 38).
35.causative adverbial subordinators: *because*
Because is the only subordinator to function unambiguously as a causative adverbial. Other forms, such as *as*, *for*, and *since*, can have a range of functions, including causative. Most researchers find more causative adverbials in speech (Beaman 1984; Tottie 1986), although the functional reasons for this distribution are not clear. Tottie (1986) and Altenberg (1984) both provide detailed analyses of these subordination constructions. For example, Tottie notes that while there is

more causative subordination overall in speech, the form *as* is used as a causative subordinator more in writing.

36.concessive adverbial subordinators: *although, though*

Following a general pattern for adverbial clauses, concessive adverbials can also be used for framing purposes or to introduce background information, and they have different functions in pre- and post-posed positions (McClure and Geva 1983; Altenberg 1986). Both Altenberg and Tottie (1986) find more concessive subordination overall in writing.

37.conditional adverbial subordinators: if, *unless*

Conditional clauses are also used for discourse framing and have differing functions when they are in pre- or post-posed position (Ford and Thompson 1986). Finegan (1982) finds a very frequent use of conditional clauses in legal wills, due to the focus on the possible conditions existing when the will is executed. Several researchers have found more conditional clauses in speech than in writing (Beaman 1984; Tottie 1986; Biber 1986a; Ford and Thompson 1986), but the functional reasons for this distribution are not clear.

38.other adverbial subordinators: (having multiple functions)

since, while, whilst, whereupon, whereas, whereby, such that, so that xxx, such that xxx, inasmuch as, forasmuch as, insofar as, insomuch as, as long as, as soon as

(where xxx is NOT: N/ADJ)

(I1) PREPOSITIONAL PHRASES (no. 39)

39.total prepositional phrases

against, amid, amidst, among, amongst, at, besides, between, by, despite, during, except, for, from, in, into, minus, notwithstanding, of, off, *on, onto, opposite, out, per, plus, pro, re, than, through, throughout, thru, to, toward, towards, upon, versus, via, with, within, without*

This list of prepositions is taken from Quirk *et al.* (1985:665-7), excluding those lexical items that have some other primary function, such as place or time adverbial, conjunct, or subordinator (e.g., *down, after, as)*

Prepositions are an important device for packing high amounts of information into academic nominal discourse. Chafe (1982, 1985; and Danielewiez 1986) describes prepositions as a device for integrating information into idea units and expanding the amount of information contained within an idea unit. Biber (1986a) finds that prepositions tend to co-occur frequently with nominalizations and passives in academic prose, official documents, professional letters, and other informational types of written discourse.

(I2) ADJECTIVES AND ADVERBS (nos. 40-2)

Adjectives and adverbs also seem to expand and elaborate the information presented in a text. Chafe (1982, 1985; and Danielewicz 1986) groups adjectives together with prepositional phrases and subordination constructions as devices used for idea unit integration and expansion. However, the descriptive kinds of information presented by adjectives and adverbs do not seem equivalent to the logical, nominal kinds of information often presented in prepositional phrases. In my earlier work (e.g., Biber 1986a), I find that prepositions, subordination features, adjectives, and adverbs are all distributed differently; for example, prepositional phrases occur frequently in formal, abstract styles, while many types of subordination occur frequently in highly interactive, unplanned discourse; adjectives and adverbs are distributed in yet other ways. All of these features elaborate information in one way or another, but the type of information being elaborated is apparently different in each case.

Some studies distinguish between attributive and predicative adjectives (e.g., Drieman, O'Donnell, and Chafe). Attributive adjectives are highly integrative in their function, while predicative adjectives might be considered more fragmented. In addition, predicative adjectives are frequently used for marking stance (as heads of *that* or *to* complements; see Winter 1982). The present analysis distinguishes between attributive and predicative adjectives, including both participial and non-participial forms.

40.attributive adjectives (e.g., *the big horse)*

ADJ + ADJ/N

(+ any ADJ not identified as predicative - no. 41)

41.predicative adjectives (e.g., *the horse is big)*

(a) BE + ADJ + xxx

(where xxx is NOT ADJ, ADV, or N)

(b) BE + ADJ + ADV + xxx

(where xxx is NOT ADJ or N)

42.total adverbs

Any adverb form occurring in the dictionary, or any form that is longer than five letters and ends in -ly. The count for total adverbs excludes those adverbs counted as instances of hedges, amplifiers, downtoners, amplifiers, place adverbials, and time adverbials.

(J) LEXICAL SPECIFICITY (nos. 43-4)

Two measures of lexical specificity or diversity are commonly used: type/token ratio and word length. Unlike structural elaboration, differences in lexical specificity seem to truly correlate with the production differences between speaking and writing; the high levels of lexical diversity and specificity that are found in formal academic writing are apparently not possible in spoken texts due to the restrictions of on-line production (Chafe and Danielewicz 1986; Biber 1986a). Type/token ratio (the number of different words per text) was a favorite measure of psychologists and researchers in communication studying linguistic differences between speech and writing (Osgood 1960; Drieman 1962; Horowitz and Newman 1964; Gibson *et al.* 1966; Preston and Gardner 1967; Blankenship 1974). Longer words also convey more specific, specialized meanings than shorter ones; Zipf (1949) has shown that words become shorter as they are more frequently used and more general in meaning. Osgood, Drieman and Blankenship include measures of word length in their studies. These two features are found to co-occur frequently in planned written genres by Biber (1986a), and this distributional pattern is interpreted as marking a highly exact presentation of information, conveying maximum content in the fewest words.

43.type/token ratio

the number of different lexical items in a text, as a percentage

This feature is computed by counting the number of different lexical items that occur in the first 400 words of each text, and then dividing by four; texts shorter than 400 words are not included in the present analysis. In a preliminary version of the computer programs used here, I computed this feature by counting the number of different lexical items in a text, dividing by the total number of words in the text, and then multiplying by 100. If the texts in the analysis were all nearly the same length, these two methods of computing type/token ratio would give nearly equivalent results. If text length varies widely, however, these two methods will give quite different results, because the relation between the number of 'types' (different lexical items) and the total number of words in a text is not linear. That is, a large number of the different words used in the first 100 words of a text will be repeated in each successive 100-word chunk of text. The result is that each additional 100 words of text adds fewer and fewer additional types. In a comparison of very short texts and very long texts, the type/token ratio computed over the entire text will thus appear to be much higher in the short texts than in the long texts. To avoid this skewing, the present study computes the number of types in the first 400 words of each text, regardless of the total text length.

44.word length

mean length of the words in a text, in orthographic letters

(K)LEXICAL CLASSES (nos. 45-51)

45.conjuncts

alternatively, altogether, consequently, conversely, eg, e.g., else, furthermore, hence, however, i.e., instead, likewise, moreover, namely, nevertheless, nonetheless, notwithstanding, otherwise, rather, similarly, therefore, thus, viz.

in+comparison/contrast/particular/addition/conclusion/consequence/sum /summary/any event/any case/other words

for + example/instance

by + contrast/comparison

as a + result/consequence

on the + contrary/other hand

ALL-P/T# + that *is/else/altogether + T#/,*

ALL-P/T# + *rather + T#/,/xxx*

(where xxx is NOT: ADJ/ADV)

Conjuncts explicitly mark logical relations between clauses, and as such they are important in discourse with a highly informational focus. Quirk *et al.* (1985:634-6) list the following functional classes of conjuncts: listing, summative, appositive, resultive, inferential, contrastive, and transitional. Despite their importance in marking logical relations, few register comparisons have analysed the distribution of conjuncts. Ochs (1979) notes that they are more formal and therefore more common in planned discourse than unplanned. Biber (1986a) finds that they co-occur frequently with prepositions, passives, and nominalizations in highly informational genres such as academic prose, official documents, and professional letters. Altenberg (1986) looks at concessive and antithetic conjuncts and finds that they are generally more common in writing than speech.

46.downtoners

almost, barely, hardly, merely, mildly, nearly, only, partially, partly, practically, scarcely, slightly, somewhat

Downtoners 'have a general lowering effect on the force of the verb' (Quirk *et al.*, 1985:597-602). Chafe and Danielewicz (1986) characterize these forms as 'academic hedges', since they are commonly used in academic writing to indicate probability. Chafe (1985) notes that downtoners are among those evidentials used to indicate reliability. Holmes (1984) notes that these forms can mark politeness or deference towards the addressee in addition to marking uncertainty towards a proposition.

47.hedges

at about/something like/more or less/almost/maybe/xxx sort of/xxx kind of

(where xxx is NOT: DET/ADJ/POSSPRO/WHO - excludes *sort* and *kind* as true nouns)

Hedges are informal, less specific markers of probability or uncertainty. Downtoners give some indication of the degree of uncertainty; hedges simply mark a proposition as uncertain. Chafe (1982) discusses the use of these forms to mark fuzziness in involved discourse, and Chafe and Danielewicz (1986) state that the use of hedges in conversational discourse indicates an awareness of the limited word choice that is possible under the production restrictions of speech. Biber (1986a) finds hedges co-occurring with interactive features (e.g., first and second person pronouns and questions) and with other features marking reduced or generalized lexical content (e.g., general emphatics, pronoun *it*, contractions).

48.amplifiers

absolutely, altogether, completely, enormously, entirely, extremely, fully, greatly, highly, intensely, perfectly, strongly, thoroughly, totally, utterly, very

Amplifiers have the opposite effect of downtoners, boosting the force of the verb (Quirk *et al.* 1985:590-7). They are used to indicate, in positive terms, the reliability of propositions (Chafe 1985). Holmes (1984) notes that, similar to downtoners, amplifiers can be used for non-propositional functions; in particular, they can signal solidarity with the listener in addition to marking certainty or conviction towards the proposition.

49.emphatics

for sure/a lot/such a/real + ADJ/*so* + ADJ/DO + V/*just/really/most/more*

The relation between emphatics and amplifiers is similar to that between hedges and downtoners: emphatics simply mark the presence (versus absence) of certainty while amplifiers indicate the degree of certainty towards a proposition. Emphatics are characteristic of informal, colloquial discourse, marking involvement with the topic (Chafe 1982,1985). As noted above, Biber(1986a) finds emphatics and hedges co-occurring frequently in the conversational genres. Labov (1984) discusses forms of this type under the label of 'intensity': the ‘emotional expression of social orientation toward the linguistic proposition'. Other studies of emphatics include Stenstrom's (1986) analysis of *really* and Aijmer's (1985) analysis of *just.*

50.discourse particles

CL-P/T# + w*ell/now/anyway/anyhow/anyways*

Discourse particles are used to maintain conversational coherence (Schiffrin 1982, 1985a). Chafe (1982, 1985) describes their role as 'monitoring the information flow' in involved discourse. They are very generalized in their functions and rare outside of the conversational genres.

51.demonstratives *that/this/these/those*

(This count excludes demonstrative pronouns (no. 10) and *that* as relative, complementizer, or subordinator.)

Demonstratives are used for both text-internal deixis (Kurzon 1985) and for exophoric, text-external, reference. They are an important device for marking referential cohesion in a text (Halliday and Hasan 1976). Ochs (1979) notes that demonstratives are preferred to articles in unplanned discourse.

(L)MODALS (nos. 52-4)

It is possible to distinguish three functional classes of modals: (1) those marking permission, possibility, or ability; (2) those marking obligation or necessity; and (3) those marking volition or prediction (Quirk *et al.* 1985:219-3 6; Coates 1983; Hermeren 1986). Tottie (1985; Tottie and Overgaard 1984) discusses particular aspects of modal usage, including the negation of necessity modals and the use of would. Chafe (1985) includes possibility modals among the evidentials that mark reliability, and necessity modals among those evidentials that mark some aspect of the reasoning process.

52.possibility modals

can/may/might/could (+ contractions)

53.necessity modals

ought/should/must (+ contractions)

54.predictive modals

will/would/shall (+ contractions)

(M) SPECIALIZED VERB CLASSES (nos. 55-8*)*

Certain restricted classes of verbs can be identified as having specific functions. Several researchers refer to 'verbs of cognition', those verbs that refer to mental activities (Carroll *1960;* Weber *1985).* Chafe *(1985)* discusses the use of 'sensory' verbs (e.g., *see, hear, feel)* to mark knowledge from a particular kind of evidence. In the present analysis, I distinguish four specialized classes of verbs: public, private, suasive, and *seem/appear.* Public verbs involve actions that can be observed publicly; they are primarily speech act verbs, such as *say* and *explain,* and they are commonly used to introduce indirect statements. Private verbs express intellectual states (e.g., *believe)* or nonobservable intellectual acts (e.g., *discover);* this class corresponds to the 'verbs of cognition' used in other studies. Suasive verbs imply intentions to bring about some change in the future (e.g., *command, stipulate).* All present and past tense forms of these verbs are included in the counts.

55.public verbs

(e.g., *acknowledge, admit, agree, assert, claim, complain, declare, deny, explain, hint, insist, mention, proclaim, promise, protest, remark, reply, report, say, suggest, swear, write)*

This class of verbs is taken from Quirk *et al. (1985:1180-1).*

56.private verbs
(e.g., *anticipate, assume, believe, conclude, decide, demonstrate, determine, discover, doubt, estimate, fear, feel, find, forget, guess, hear, hope, imagine, imply, indicate, infer, know, learn, mean, notice, prove, realize, recognize, remember, reveal, see, show, suppose, think, understand)*
This class of verbs is taken from Quirk *et al.* (1985:1181-2*).*
57.suasive verbs
(e.g., *agree, arrange, ask, beg, command, decide, demand, grant, insist, instruct, ordain, pledge, pronounce, propose, recommend, request, stipulate, suggest, urge)*
This class of verbs is taken from Quirk *et al.* (1985:1182-3*).*
58.*seem/appear*
These are 'perception' verbs (Quirk *et al.* 1985:1033, 1183*).* They can be used to mark evidentiality with respect to the reasoning process (Chafe 1985), and they represent another strategy used for academic hedging (see the discussion of downtoners - no. 46).

(N) REDUCED FORMS AND DISPREFERRED STRUCTURES (nos.59-63)
Several linguistic constructions, such as contractions, stranded prepositions, and split infinitives, are dispreferred in edited writing. Linguists typically disregard the prescriptions against these constructions as arbitrary. Finegan (1980, 1987; Finegan and Biber 1986a), however, shows that grammatical prescriptions tend to be systematic if considered from a strictly linguistic point of view: they tend to disprefer those constructions that involve a mismatch between surface form and underlying representation, resulting in either a reduced surface form (due to contraction or deletion) or a weakened isomorphism between form and meaning (e.g., split infinitives). Biber (1986a) finds that these features tend to co-occur frequently with interactive features (e.g., first and second person pronouns) and with certain types of subordination. Chafe (1984b) discusses the linguistic form of grammatical prescriptions and analyzes the historical evolution of certain prescriptions in speech and writing. Features 59-63 are all dispreferred in edited writing; nos. 59-60 involve surface reduction of form and nos. 61-3 involve a weakened isomorphism between form and meaning.
59.contractions
(1) all contractions on pronouns
(2) all contractions on auxiliary forms (negation)
(3) *'s* suffixed on nouns is analyzed separately (to exclude possessive forms):
N's + V/AUX/ADV+V/ADV+AUX/DET/POSSPRO/PREP/ADJ+CL-P/ADJ+T#

Contractions are the most frequently cited example of reduced surface form. Except for certain types of fiction, they are dispreferred in formal, edited writing; linguists have traditionally explained their frequent use in conversation as being a consequence of fast and easy production. Finegan and Biber (1986a), however, find that contractions are distributed as a cline: used most frequently in conversation; least frequently in academic prose; and with intermediate frequencies in broadcast, public speeches, and press reportage. Biber (1987) finds that contractions are more frequent in American writing than in British writing, apparently because of greater attention to grammatical prescriptions by British writers. Chafe and Danielewicz (1986) also find that there is no absolute difference between speech and writing in the use of contractions. Thus, the use of contractions seems to be tied to appropriateness considerations as much as to the differing production circumstances of speech and writing.

60.subordinator-*that* deletion

(e.g., *I think [that] he went to ...*)

(1) PUB/PRV/SUA+ (T#) +demonstrative pro/SUBJPRO

(2) PUB/PRV/SUA+PRO/N+AUX/V

(3) PUB/PRV/SUA + ADJ/ADV/DET/POSSPRO + (ADJ) + N +AUX/V

While contractions are a form of phonological (or orthographic) reduction, subordinator- *that* deletion is a form of syntactic reduction. There are very few of these deletions in edited writing, even though few explicit prescriptions prohibit this form. Apparently the concern for elaborated and explicit expression in typical edited writing is the driving force preventing this reduction.

61.stranded prepositions

(e.g. *the candidate that I was thinking of*)

PREP + ALL-P/T#

Stranded prepositions represent a mismatch between surface and underlying representations, since the relative pronoun and the preposition belong to the same phrase in underlying structure. Chafe (1985) cites these forms as an example of spoken 'errors' due to the production constraints of speech.

62.split infinitives

(e.g., *he wants to convincingly prove that* ...)

to + ADV + (ADV) + VB

Split infinitives are the most widely cited prescription against surface-underlying mismatches. This notoriety suggests that writers *would* use split infinitives if it were not for the prescriptions against them, but these forms in fact seem to be equally uncommon in spoken and written genres (Biber 1986a; Chafe 1984b). This feature did not co-occur meaningfully with the other features included in the present study, and it was therefore dropped from the factor analysis(Chapter 5).

63.split auxiliaries

(e.g., *they are objectively shown to* ...)

AUX + ADV + (ADV) + VB

Split auxiliaries are analogous to split infinitives, but they have not received much attention from prescriptive grammarians. They are actually more common in certain written genres than in typical conversation; Biber (1986a) finds that they frequently co-occur with passives, prepositions, and nominalizations.

(O) COORDINATION (nos. 64-5)

Phrase and clause coordination have complementary functions, so that any overall count of coordinators would be hopelessly confounded. And as a clause coordinator is a general purpose connective that can mark many different logical relations between two clauses. Chafe (1982, 1985) relates the fragmented style resulting from this simple chaining of ideas to the production constraints of speech. And as a phrase coordinator, on the other hand, has an integrative function and is used for idea unit expansion (Chafe 1982, 1985; Chafe and Danielewicz 1986). Other studies that analyse the distribution and uses of *and* include Marckworth and Baker (1974), Schiffrin (1982), and Young (1985). The algorithms used in the present study identify only those uses of and that are clearly phrasal or clausal connectives.

64.phrasal coordination

xxxxl + and + xxxx2

(where xxxl and xxx2 are both: ADV/ADJ/V/N)

65.independent clause coordination

(a) T#/, + *and* + *it/so/then/you/there* + BE/demonstrative pronoun/SUBJPRO

(b) CL-P+ *and*

(c) *and* + WHP/WHO /adverbial subordinator (nos. 35-8)/discourse particle (no. 50)/conjunct (no. 45)

(P) NEGATION (nos. 66-7)

There is twice as much negation overall in speech as in writing, a distribution that Tottie (1981, 1982, 1983b) attributes to the greater frequency of repetitions, denials, rejections, questions, and mental verbs in speech. Tottie (1983a) distinguishes between synthetic and analytic negation. Synthetic negation is more literary, and seemingly more integrated; analytic negation is more colloquial and seems to be more fragmented.

66.synthetic negation

(a) *no* + QUANT/ADJ/N

(b) *neither, nor*

(excludes *no* as a response)

67. analytic negation: *not*

(also contracted forms)

Appendix B

Alphabetical Frequency List

b	12320/15281
bachelor	110/145
back	1369/1910 P
backed	205/241
background	347/441
backward	108/122 P
bad	145/159 P
bah	132/135
balance	406/582 CP
balanced	141/161 CP
ban	101/115
banned	167/187
bark	199/348
barley	130/174
basin	343/668 C
basins	132/285
basis	968/1216 CP
bass	139/321
BC	1692/4389
beach	176/289
beaches	116/161
beam	173/393
beams	101/194
beans	101/194
bear	428/589
bearing	312/405 C
bears	341/536 CP
beat	101/140
beautiful	305/358
becomes	566/733 CP
becoming	655/703
bed	141/195
beds	301/363
bee	175/269 P
beer	106/156
began	3077/4682 CP
begin	340/426
beginning	1225/1582 P
beginnings	191/206
begins	319/383 CP
behalf	118/127
behavior	640/1441 CP
behavioral	102/167
behind	540/639 P
beings	237/336 C
belief	408/505 C
beliefs	270/340
believe	346/409 CP
believed	009/1222 C
bell	313/491 P
belong	295/321
belonged	111/112
belonging	305/317
belongs	147/163
below	764/997 CP
belt	183/305
beneath	257/331 P
benefit	143/178 P
benefits	172/256
bent	100/111
berries	109/165
besides	159/176
best	2786/3536 CP
better	570/687 C
beyond	644/785
big	358/517 P
bill	500/825
billion	534/1319 CP
bills	154/224
binding	106/145
bird	417/729
birds	605/1225
birth	547/838
births	217/256
bitter	234/268 C
black	1921/3783 C
blacks	273/600
blend	122/135
blind	137/206 C
block	257/341 C
blocks	177/222 P
blood	732/2125 C
bloody	112/125
bloom	105/155
blow	115/136
blue	668/1101 CP
blues	100/243
board	422/585 P
boat	129/248 P
bodied	108/124
bodies	543/764 CP
body	1807/3647 CP
boh	122/123
boiling	145/195 C
bold	194/228
bomb	109/252 C
bond	164/321
bonds	157/367
bone	286/613 P
bones	239/516 P
book	1558/2287 CP
books	1088/1517 P
boom	106/140
border	563/819
bordered	202/209
borders	222/270
bore	174/209
born	1391/1622
borne	148/165
borrowed	125/149
both	4356/7133 CP
bottom	414/749 CP
bought	134/147
bound	231/273
boundaries	235/300
boundary	232/316
bounded	187/199
bow	120/173
box	189/262 CP
boy	193/212 P
brain	347/975 C
branch	499/662 C
branches	418/611 P
brass	108/170 P
break	388/440 CP
breakdown	121/141 C
breaking	208/233 P
breathing	131/196 CP
bred	136/235
breed	245/522
breeding	227/380
breeds	124/235
brick	143/222
bridge	305/555 P

bridges	148/231 P
brief	386/400
bright	352/447
brightly	107/130 C
bring	428/472 CP
bringing	208/222
broad	651/802 C
broadcasting	127/292
broader	104/112 C
broadly	121/130
broke	392/469 C
broken	338/395 CP
brother	700/852
brothers	318/453
brought	1462/1939 CP
brown	1062/1544 C
budget	149/253
buffalo	191/275
build	385/471 CP
builders	112/150 C
building	1099/1782 P
buildings	760/1219 P
built	1675/2645 CP
bulk	171/208 C
bull	171/306
bur	113/116
bureau	154/239 P
burial	160/270
buried	255/287
burn	116/144 C
burned	227/261 C
burning	229/313 C
burns	122/190 C
bush	193/339
business	769/1230
businesses	118/157
buy	122/149 CP

unlisted (too common)
be
became
because
become
been
before C
begun
being
between C
bibliography
but C
by CP

c	6214/11873
cabinet	335/473
cable	124/307 P
call	304/354 CP
calling	150/161
calls	152/171
came	1589/2164 P
camp	172/224 P
campaign	421/690
campaigns	209/278
camps	127/180
canal	382/788
canals	162/242
cannot	629/814 CP
capable	470/598 CP
capital	1798/3099 P
car	143/290 P
care	402/828
careful	156/164 P
carefully	272/300 C
carried	873/1071 CP
carries	237/288 C
carry	499/654 CP
carrying	321/375 CP
cars	106/237 P
carved	222/364
cash	156/223 C
cast	331/440 C
castle	290/392
cat	143/373
catch	150/189 P
cattle	467/818 P
caught	181/206 P
cave	149/336 P
caves	100/189
ceased	129/137
ceded	143/155
cell	399/1372 P
cells	591/2105 P
cement	193/240
census	500/633
centuries	1404/2018 CP
century	5988/14669 CP
certain	1512/2184 CP
certainly	101/110 P
chain	330/551
chains	137/235
chairman	272/331
challenge	224/261
challenged	203/218
chamber	421/631 CP
chambers	167/236
chancellor	159/263
change	1067/1624 CP
changed	552/628 CP
changes	938/1508 CP
changing	427/497 CP
channel	288/503
channels	158/256
chapel	184/288
charge	404/733 C
charged	402/551 P
charges	252/369 C
charter	161/237
chartered	107/118
check	112/140 CP
chemistry	547/1054 CP
chest	130/181
chief	1384/1937
chiefly	334/360 C
chiefs	126/164
child	576/936 P
childhood	237/307
children	1095/2092
chloride	101/208 C
choice	255/295 CP
choose	130/148 CP
choral	138/211
chose	193/209 C
chosen	379/442 CP
Christmas	119/161
chronic	178/262
church	1690/3823
churches	505/901
circle	384/562 CP

circles 181/217 CP
circular 271/373 P
cities 1012/1927
citizen 197/248 P
citizens 319/496
citrus 106/193 C
city 4147/10255 P
civic 114/131
civilian 205/293
claim 338/417 P
claimed 437/508 P
claims 307/406 P
clarity 119/127
classes 393/522 C
clay 346/583
clean 119/142 C
clear 553/659 P
clearly 352/409 P
clergy 155/194
climate 778/1307
climates 212/342
climatic 152/254
clinical 127/216
close 975/1162 CP
closed 386/477 C
closely 763/384 CP
closer 193/211 CP
closest 110/137
cloth 149/212
clothing 267/344 CP
cloud 171/317 CP
clouds 143/309 C
cluster 102/165
clusters 275/483
cm 1467/2372 P
co 211/258
coal 584/1341
coalition 275/417
coarse 153/170
coast 1183/2260 P
coastal 522/936
coastline 105/123
coasts 176/242
coat 292/467 C
coated 126/176
code 262/567
codes 111/154
coffee 193/289
coined 131/141 C
cold 580/915 CP
collected 500/556 C
collecting 109/142 C
collection 908/1281
collections 392/486
collective 164/224
collectively 137/142
color 1153/2019 C
colored 537/686 C
colorful 219/262
colorless 107/125 C
colors 514/770 C
column 185/310 CP
columns 165/265 CP
combat 189/263
come 727/854 CP
comes 381/433 CP
coming 339/399 P
commitment 154/175
committed 292/326 C
committee 375/573 P
commodities 104/130 C
common 2536/4170 CP
commonly 1050/1298 CP
commonwealth 171/265
compact 212/271 P
companies 400/640
companion 162/179
company 1047/1913 P
comparable 182/200 CP
compared 311/358 CP
comparison 129/148 CP
compound 270/446 CP
compounds 433/1089 C
comprise 231/262
comprised 106/113
comprises 330/381
comprising 183/196
compromise 191/263
compulsory 142/175
conceived 196/209
concept 670/934 CP
conception 145/179
concepts 305/394 P
concern 489/582 C
concerned 505/653 CP
concerning 317/373 P
concerns 365/408 C
concrete 229/419 P
condemned 203/227
conduct 274/341 C
conducted 425/489 P
conducting 156/197 C
conductor 211/417 C
confidence 115/135
confined 270/307
connected 439/506 P
connecting 135/152 P
connection 206/230
connections 127/146
connects 103/114
conquered 411/535
conquest 403/571
conscious 134/178
consciousness 201/318 C
contain 804/1119 CP
contained 494/594 CP
containing 697/898 CP
contains 1076/1413 CP
content 437/611 C
convicted 121/139
conviction 119/144
convinced 165/178
cook 150/237
cooking 103/176 C
cool 315/396
cooled 127/173 C
cooler 112/139
cooling 151/278
copper 465/962 CP
core 279/457
corn 285/488
corner 131/150 P
costly 166/205 C
cotton 483/784

could 1680/3136 CP
council 761/1364
councils 137/186
counterpart 108/111
counties 154/297
countries 1182/2417
country 1685/3428 P
county 878/1293
coupled 123/143 C
course 610/802 CP
courses 179/257 P
court 1367/2955
courts 336/596
cousin 127/140
cover 434/578 P
coverage 102/138
covered 732/947 CP
covering 333/368
covers 432/487 P
craft 234/406
crafts 103/157
creek 144/215
crew 141/310 CP
crisis 427/634
crop 304/539 C
crops 507/822 C
cross 627/891 CP
crossed 204/245 P
crosses 101/119 P
crossing 154/174 P
crown 493/641
crowned 164/206
crude 152/191 P
crushed 143/158 P
crust 165/338
cup 128/206 C
currency 273/362
custom 116/133
customs 246/319
cut 606/862
cuts 107/145 P
cutting 225/354
cylinder 131/309 CP
cylindrical 141/167

unlisted (too common)
called P
can P
caused

d 10966/14425
dah 108/110
daily 444/591 CP
dairy 161/231
dam 108/287
damage 456/628 C
damaged 234/258 C
dame 126/199
dams 113/217
dancer 157/266
dancers 112/195
danger 147/187 C
dark 630/794 C
darkness 105/132
data 422/859 CP
date 510/620
dated 140/165
dates 389/455
dating 379/510
daughter 533/621
daughters 106/125
day 2019/2908 CP
days 932/1297 P
dead 429/621 CP
deal 468/613 CP
dealing 223/246 CP
deals 207/248 CP
dealt 189/215
death 2634/3643
deaths 154/197
debt 208/303
Dec 1653/1798
decay 153/305
December 499/649
deciduous 182/246
decline 510/648
declined 511/585
declining 106/118
decrease 186/247 CP
decreased 132/157 CP
decreases 49/175 CP
deep 829/1177 C
deeper 127/141 P
deeply 298/326
defeat 476/617
defeated 720/967
defeating 102/112
defense 721/1049
defensive 116/169
define 164/184 CP
defined 608/772 CP
definition 240/315 CP
deg 1039/4147
degree 714/967 C
degrees 392/625 C
delayed 114/122 P
delivered 121/133
delta 180/322 P
demand 372/546
demands 284/346
demonstrate 147/151 CP
demonstrated 429/509 C
denied 191/218
dense 348/410 C
densely
145/167
density 535/808 CP
depict 111/128
depicted 252/315
depicting 151/164
depicts 119/132
deposed 162/188
deposit 133/190 C
deposited 190/252
deposition 106/151
deposits 602/1141 C
depression 364/537 C
depth 350/496 CP
depths 178/249 C
deputy 123/387
derive 101/108 CP
derived 1246/1545 CP
derives 142/144
descendants 250/287

descended 171/195
descent 287/348
describe 382/439 CP
described 676/797 CP
describes 254/273 CP
describing 146/153 CP
description 240/290 CP
descriptions 106/115 P
descriptive 106/126
desert 386/759
deserts 119/204
design 859/1592 P
designated 272/309
designed 1201/1687
designer 189/241
designs 413/605
desire 241/293 C
desired 203/260 C
despite 931/1171 C
destroy 173/215 C
destroyed 670/823 C
destruction 327/399 C
destructive 113/130
detail 318/385 C
detailed 266/291 CP
details 235/274 CP
detect 160/207 P
detected 157/212
detection 116/156
determination 187/233
determine 462/593 CP
determined 682/896 CP
determines 151/172
determining 223/262 CP
devastating 130/133
develop 829/1083 CP
developing 579/723 C
developments 337/418CP
develops 210/256
device 420/595 CP
devices 518/837
devil 128/146
devised 268/300 CP
devoted 482/525
diagnosis 107/160
dialect 129/164
dialogue 173/225
diameter 410/592 P
diamond 115/189
dictator 107/152
did 1632/2368 CP
die 294/408
died 1283/1561
dies 109/139
diet 242/354
differ 297/342 CP
difference 299/401 CP
differences 462/644 CP
differing 119/126
differs 186/204 CP
difficult 689/806 CP
difficulties 225/251
difficulty 198/219
digestive 116/201
dimensional 192/286 CP
diminished 111/117 P
dioxide 219/421 C
direct 823/1094 CP
directed 570/708
disappeared 196/219
disaster 116/144
disastrous 119/127
discharge 124/234 C
discovered1207/1599 CP
discoveries 218/270 P
discovery 661/894 P
discussed 250/168 CP
discussion 142/179 CP
disease 704/1738
diseases 512/1201
disk 121/201 P
disorder 188/347
disorders 240/517
dispersed 130/152
display 296/367 C
displayed 210/229
displays 185/220
dispute 221/258
disputed 117/134
disputes 203/254
dissolved 262/350 C
distinguished 781/866 CP
disturbances 111/141
divine 345/495
DNA 106/479
do 1274/1843 CP
documentary 124/156
does 912/1167 CP
dog 274/737 P
dogs 195/374 P
doh (pronun) 145/147
doing 151/168 P
dollar 102/148 C
domain 116/161 P
dome 133/248
domestic 573/899 C
dominance 135/161
done 467/569 C
door 102/144
doors 119/149
double 487/736 CP
doubt 117/132 P
down 1287/1741 CP
downward 149/166 P
dozen 147/161 P
Dr 180/234
draft 150/248
drainage 234/373
drained 177/208
drains 128/151 C
draw 162/183 CP
drawing 283/394 CP
drawings 287/425
drawn 503/601 P
dream 209/264
dreams 120/188
drew 250/287 P
dried 220/338 C
drift 114/206
drinking 112/175 C
drive 251/352 CP
driven 293/368 C
driving 145/180 C

drop	187/224 CP
dropped	177/214 P
drought	122/175
drove	132/152
drug	337/868
drugs	342/890 C
dry	613/945
duchy	109/144
due	606/806 CP
duh (pronun)	263/276
dur (pronun)	101/103
durable	102/112
duration	135/176 P
during	6721/12543 CP
dust	180/334 P
duties	172/246
duty	164/204
dwarf	102/136
dwellers	107/125
dwelling	154/196
dying	109/141
dynamic	174/202
dynamics	121/164 P

unlisted (too common)
developed P
development P
different P

e	3676/5155
ear	127/279
earl	286/460
earlier	1046/1280 CP
earliest	886/1132 C
earned	414/450
ears	223/304
earth	1209/3113 CP
earthquake	108/165
earthquakes	107/157 P
ease	124/132 C
easier	114/124 C
easily	530/629 CP
east	2106/4017 P
eastern	1651/2954
eastward	157/194 P
easy	222/254 CP
eat	213/274 CP
eating	192/243 P
ecology	134/216
economic	1520/3406
economically	239/278
economics	301/440
economy	1053/2114
edge	323/420 P
edges	137/160
edible	148/265
edited	229/255 P
eds	1545/1859
ee (pronun)	846/861
effort	555/673 C
efforts	829/1086 CP
egg	225/496
eggs	414/795
eight	496/567 CP
either	1457/1961 CP
elaborate	436/530
elder	205/248
electric	527/1147 CP
electricity	281/519 CP
elongated	186/222
else	120/128 P
elsewhere	356/414
emerge	157/182
emerged	471/604
emergence	195/230
emergency	151/210 P
emerging	140/151
emperor	839/1610
emperors	137/199
emphasis	472/590
emphasize	111/140
emphasized	288/320 P
empire	1098/2330 P
employ	187/209 CP
employed	580/702
employees	129/213
employing	103/114
employment	243/406
employs	141/155
empty	141/153 C
enable	157/171 CP
enabled	267/299
enables	107/116 CP
enabling	109/118
enacted	157/187
enclosed	218/244
encompasses	119/121
encountered	148/160 C
encourage	159/172
encouraged	470/524
end	1961/3027 CP
ended	634/791
ending	222/238
ends	284/333 CP
enduring	106/117
enemies	229/282
enemy	240/360
energy	1011/3301 CP
engage	108/119
engaged	306/330
engine	251/823 CP
engines	181/395 CP
enhance	105/111
enhanced	150/155 P
enjoyed	282/304
enlarged	216/249 P
enlightenment	134/207
enormous	494/581 CP
enormously	105/107
enough	561/691 CP
enrollment	363/904
ensuing	128/134
ensure	189/207
enter	314/355 CP
entered	770/859
entering	216/230
enterprise	174/206
enterprises	114/131
enters	192/225 P
entertainment	105/154
entrance	201/233

entry	198/237
environment	607/945 CP
environmental	448/759
environments	180/248
epic	306/469
equal	621/908 CP
equality	138/199 P
equally	351/398
equals	103/124 CP
era	786/1163 P
erect	117/129
erected	173/219
error	105/170 CP
erupted	113/132
essay	186/238 P
essayist	119/152
essays	630/836
essence	115/129
estate	258/334
estimated	475/588
evening	143/178
event	326/402
events	667/982
eventual	120/130
eventually	1091/1371 CP
ever	562/679 CP
evergreen	193/327
everyday	170/197 P
everything	129/152
everywhere	112/127
evidence	687/1079 C
evil	186/240
excavated	149/183
excavations	155/205
excelled	111/116
exchange	406/574
exerted	172/199 C
expelled	212/246
expert	102/127 P
experts	157/169
explain	225/272 CP
explained	189/235 CP
explanation	146/175 P
exploitation	133/149
exploration	340/484
exposure	229/383 CP
express	228/293 CP
expressed	441/526 CP
eye	397/660 P
eyes	385/586

unlisted (too common)
each
early
ed
est
established CP
even P
every P

f	3325/5440
fabric	129/221
fabrics	104/167
face	518/672 P
faced	294/343 P
faces	160/208 P
facilities	445/663
facing	158/176
factories	196/247
factory	218/333
fail	116/148 P
failed	568/698 P
failure	413/522
fair	257/323 C
fairly	257/287 C
faith	373/519
fall	863/1116 P
fallen	108/114 P
falling	179/208 P
falls	376/564 CP
familiar	368/420 CP
far	1259/1738 CP
farm	386/645
farmer	142/182 P
farmers	349/550 C
farming	380/575
farms	199/360 C
farther	172/223
fashion	251/379
fast	320/407 CP
faster	165/198 CP
fat	174/359
fate	112/145 P
favor	455/512
favorable	157/170
favored	342/384
favorite	261/285
fear	219/274
feared	132/148
feast	260/283
feathers	128/264
feature	458/524
featured	116/127
features	659/902 C
Feb	1657/1782
February	377/467
fed	215/266 P
fee	138/160
feed	494/678 C
feeding	223/306 C
feeds	137/179
feel	118/145 CP
feeling	190/223
feelings	143/181
feet	324/444 CP
fell	429/503 CP
fellow	201/214
felt	303/365
female	673/1254 C
females	243/375
fever	219/387
few	1910/2757 CP
fewer	219/253
fiber	172/370
fibers	230/528 P
field	1231/2356 CP
fields	672/972 P
fifth	300/377
fight	224/266
fighting	424/606
figure	815/1066 CP
figures	796/1275 CP
fill	122/138 P
filled	418/497 CP

fin	112/201
find	439/513 CP
finding	180/212 P
findings	107/119 C
finds	132/159 CP
fine	827/1064
finely	110/125
finest	493/550
finished	204/238
fins	118/209
fire	619/989
fired	151/235
firm	419/535
firmly	163/173
firms	146/246
first	8401/15779 CP
fish	705/1720 CP
fishes	186/467
fishing	621/953
fit	130/148 CP
fitted	124/152
five	1417/1849 CP
flag	106/230
flat	532/692 P
flattened	161/175
flavor	165/302 C
fled	333/368
fleet	230/446
flesh	150/187 P
fleshy	100/137
flew	102/145
flight	398/930 CP
floating	135/163 P
flood	149/240
floor	268/408 CP
flour	108/170
flourished	386/434
flow	548/1012
flower	333/629
flowering	222/341
flowers	706/1580
flowing	228/290
flows	420/576
fly	164/292 P
flying	217/352 CP
focus	381/480 C

focused	232/274
folk	374/781
folklore	142/179
follow	359/393 CP
followed	1346/1687 CP
followers	282/329
following	1754/2200 CP
follows	275/298 CP
food	1435/3002 CP
foods	280/505
foodstuffs	122/130
foot	276/427 CP
forbidden	102/127
foreign	1169/2185
foremost	292/327
form	3520/6359 CP
former	986/1291
formerly	553/624
forms	2034/3334 P
fort	444/767
forth	276/328 CP
fortress	149/197
forward	291/393 P
fostered	113/115
fought	545/653
found	3353/5597 P
four	2109/3050 P
fourth	505/601 P
fox	226/357
fragrant	107/173
frame	181/314 P
framework	193/223 CP
free	1493/2205 CP
freed	155/164
freedom	603/961
freely	179/192 CP
freezing	123/177 C
fresh	308/395 C
freshwater	216/364
friend	358/379 P
friends	283/346
friendship	134/148
front	554/965 CP
frost	104/183
frozen	146/204 P
ft	2511/5362 CP
fuel	285/696 CP

fuels	107/213 C
full	948/1175 C
fully	458/513 CP
fur	309/558
furniture	298/580
further	1107/1424 CP
furthermore	118/135 CP
fused	115/150

unlisted (too common)
for P
from CP

g	2864/3937 C
gain	317/370 C
gained	814/978 C
gaining	147/154
gains	120/139
game	438/1078 P
games	316/834
gamma	118/224
gap	100/145
garden	391/661
gardens	277/436
gasoline	100/193 CP
gate	110/176
gates	125/171
gathered	167/181 C
gathering	153/188
gave	1039/1273 C
GDP	155/171
gene	168/465
genera	223/301
general	2464/4043 CP
generate	139/164
generated	293/349 C
generating	117/153
genus	800/1050
get	187/212 CP

giant	325/447 P
gift	134/150
gifted	109/119 C
gifts	101/113
girl	161/192
girls	106/147
give	690/821 CP
given	1607/2132 CP
gives	373/427 CP
giving	450/517 C
gland	169/330
glass	592/1282 C
go	369/449 CP
goats	112/138
goddess	163/228
gods	281/471
goes	193/239 CP
goh (pronun)	107/111
going	180/188 P
gold	779/1502 CP
golden	483/648
gone	125/135
good	961/1221 CP
goods	565/885
grade	149/224
graduate	378/479
graduated	277/305
graduating	141/143
grain	344/577 C
grains	262/424 C
grammar	113/193
grandson	144/165
grant	410/703
granted	427/510
grants	142/186
grapes	122/170
graphic	129/188
grass	213/362 P
grasses	116/196
grasslands	109/168
grave	113/139
gravel	121/172
gray	565/737
great	3871/6979 CP
greater	1093/1475 CP
greatest	1195/1575 CP
greatly	825/967 C
green	873/1399 CP
grew	763/897
gross	159/212
ground	916/1428
grounds	208/250
grow	614/914
growing	912/1278
grown	768/1156
grows	439/698
growth	1363/2541 CP
guh (pronun)	109/109
guidance	117/170
gun	124/281 P
guns	104/228

unlisted (too common)
generally P
got

h	3881/5697
ha	198/290 C
habits	120/141
hair	278/485
half	1369/2036 CP
hall	622/855 P
hand	906/1322 CP
handbook	430/467 C
handle	135/159
handled	122/127
handling	189/218 C
hands	408/535 CP
harbor	336/491
hard	616/809 CP
hardness	160/194
harmful	103/123
harsh	193/209 C
harvest	111/138
harvested	116/177
hatch	111/134
haven	108/140
having	1060/1283 CP
hay	151/196
head	1300/1864 P
headed	458/539 P
headquarters	349/386
heads	308/407 P
health	870/1694 C
heard	190/214 P
hearing	127/203
heart	614/1311 P
heat	581/1564 CP
heated	221/307 C
heating	188/297 C
heaven	143/186
heavier	181/224 C
heavily	521/597
heavy	899/1213 CP
height	563/745 CP
heights	202/247 CP
heir	143/171
held	1788/2352 CP
helium	108/258 CP
help	673/807 CP
helped	1160/1378
helping	140/150
helps	126/141 C
hence	322/372
herb	141/274
herbs	150/207
herds	111/154
here	583/730
hereditary	189/262
heritage	318/357
herself	145/160
hidden	147/158
high	3145/5679
higher	1066/1890
highest	811/1293
highlands	188/376
highly	1494/1919
highway	171/243
highways	126/160
hill	401/540
hills	336/462
hind	142/214
hit	170/229
hold	447/529
holding	289/333
holdings	122/155
holds	257/284 P

hole 144/236 CP
holes 152/251 C
hollow 167/201 P
home 1162/1671
homeland 140/170
homes 212/277
honor 340/387
hope 286/372 C
hoped 149/179
hopes 101/111
horn 151/247
horse 339/760 P
horses 245/412
host 210/366
hot 542/806 CP
hour 228/299 CP
hours 426/590 CP
house 1452/2544
housed 142/155 P
household 171/218 C
houses 578/843
housing 265/516
how 894/1310 CP
however 4654/8416 CP
huge 537/663 C
hundred 424/489 C
hundreds 373/430 C
hunt 232/245
hunters 164/203
hunting 398/616
husband 348/435
hy (pronun) 102/103
hydroelectric 243/354
hydrogen 356/918 CP

unlisted (too common)
had CP
has P
have CP
he CP
her P
him P
himself
his CP

I 3595/5925 C
ice 426/1086 CP
ill 391/447
illness 184/250
improve 307/359 C
improved 471/589
improvement 177/207 C
improvements 196/234
improving 125/136 P
incident 131/174
income 395/1041 C
increase 748/1122 CP
increased 1074/1658 CP
increases 456/586 CP
increasing 772/1027 CP
increasingly 785/985
indeed 187/213 CP
indigenous 236/297
inexpensive 117/234
inherent 106/117
inheritance 125/182
inherited 237/276
injury 170/254
inland 309/411
inner 466/655 P
input 102/226 P
inquiry 103/131
inside 448/546 CP
insight 102/113 C
instance 299/382 C
instances 125/141
instead 689/839 C
insurance 256/588
intended 422/493
interest 1192/1660 CP
interested 289/321 C
interesting 103/111 CP
interests 421/551

interrupted 113/116
intricate 120/125
invited 114/114 C
involve 318/391 C
involved 919/1160 CP
involvement 194/227
involves 428/549 CP
involving 391/444 CP
iron 969/2001 CP
irrigated 105/152
irrigation 229/436
island 1164/2570 P
islands 794/2147
isotopes 111/207 C
issue 372/499
issued 229/285
issues 445/596
itself 1104/1434 CP
ivory 161/317

unlisted (too common)
if P
in P
include CP
included P
includes CP
including CP
into P
is CP
it CP
its P

j 5518/8052
January 619/744 CP
Jan 1963/2274
jaws 108/157

jay	191/252
jee (pronun)	103/105
jet	145/286 CP
job	193/283 P
jobs	211/297
join	272/301
joined	1298/1486
joining	188/195 P
joint	302/454
journal	304/381
journals	120/141
journey	229/254
judge	204/291
judges	120/188
judgement	185/250
July	2322/2848
June	2035/2473
just	765/934 CP

k	1780/2259
kah (pronun)	244/252
kee (pronun)	162/166
keep	387/455 CP
keeping	177/198 P
kept	433/502 C
key	505/649 C
keyboard	106/226
kg	658/1083 CP
kidney	121/252 P
kill	165/207
killed	613/748 C
killing	183/206
kind	519/624 CP
kinds	509/625
king	2070/4107
kingdom	891/1524 P
kingdoms	177/281 P
kings	474/697
kinship	101/157
km	2847/6227 CP
knew	116/133 P
knight	135/188
knighted	154/159
knights	105/189
know	225/281 CP
knowledge	698/1157 CP
koh (pronun)	251/256
kuh (pronun)	198/198

unlisted (too common)
known CP

l	3450/4540 C
labor	897/1822
laboratories	145/169 P
laboratory	417/613 CP
lack	553/700 P
lacked	134/140 P
lacking	153/169
lady	274/356
lah (pronun)	189/194
laid	460/523 CP
lake	885/2526 CP
lakes	454/932 C
land	1711/4004 P
landed	162/209
landing	162/313 P
landmark	129/130
landmarks	213/227
lands	446/689 P
landscape	416/727
landscapes	232/296
lane	102/125
language	1335/3102 P
languages	592/1667
large	3715/6916 CP
largely	1083/1357 CP
larger	1067/1460 CP
largest	1926/3328 CP
larvae	119/303
last	1791/2323 C
lasted	221/239
lasting	243/261
late	2417/3461 C
latter	930/1073 CP
launch	125/288 P
launched	437/799 CP
law	1918/4274 CP
laws	801/1488 CP
lawyer	302/349
lay	534/610 P
layer	334/627 CP
layers	318/469 C
lb	673/1107 P
lead	789/1115 CP
leader	1371/1940
leaders	673/994
leadership	516/748
leading	1766/2329 C
leads	258/309 C
leaf	202/366
league	535/1071 P
learn	162/214 C
learned	401/481 C
learning	440/823
least	943/1236 CP
leather	236/312
leave	270/292 C
leaves	815/1697 P
leaving	357/389 P
lectures	128/141
left	1590/2336 CP
leg	109/181 P
legend	351/433
legends	138/165
legs	388/632
length	1133/1573 P
lengths	118/132 P
lengthy	108/117 C
less	1976/2963 CP

lesser	263/311
let	113/123 CP
letter	270/472 P
letters	628/873 P
level	1202/1911 CP
levels	558/830 C
libraries	101/168
library	828/1929 P
lie	304/371
lies	795/910 CP
life	4668/8121 CP
lifelong	113/117
lifetime	266/309 CP
light	1598/3698 CP
lighter	218/257 C
lighting	137/195
like	1816/2672 CP
likely	308/393 CP
limbs	117/211 P
limestone	268/382 C
lined	134/158
lines	756/1282 CP
lining	106/140
link	274/303
linked	432/503
linking	108/121
links	166/189
literacy	237/328
little	1536/2087 CP
live	1141/1679 CP
lived	1062/1290 P
lively	146/153
liver	158/305
lives	680/925
livestock	323/426
living	1168/1654 P
load	103/177 P
located	1930/2735 P
loh (pronun)	152/155
long	4041/6701 CP
longer	737/920 CP
longest	231/262
loo(pronun)	107/111
look	284/320 CP
looked	106/119
looking	151/164 P

loose	181/202
loosely	129/135
lose	158/175 C
losing	142/151
loss	558/804 CP
losses	195/280 P
lost	1149/1425 CP
love	771/1076
low	1488/2344 CP
lower	1247/1818 C
lowered	107/131
lowest	406/522 CP
lowland	128/194
lowlands	155/267
loyal	101/108
loyalty	125/133
luh (pronun)	163/165
lumber	203/309
lung	110/230
lungs	146/318 C
luster	114/129
lying	242/268 P

unlisted (too common)
later CP
led P

m	6223/11357 CP
machine	493/984
machinery	496/607
machines	263/516
magazine	316/448
magazines	117/199
mah (pronun)	204/212
mahn (pronun)	142/142
main	1385/1878 CP
mainland	212/339
mainly	967/1256 CP
maintain	536/666 CP
maintained	544/631 CP
maintaining	217/241 P
maintains	131/144 P
maintenance	165/213
maize	129/161
major	4083/7478 CP
majority	661/1039 C
makers	109/137

makes	581/675 CP
making	1400/1753 CP
male	709/1446 C
males	306/455
mammals	306/607 P
man	1771/2436 P
managed	215/246
management	369/635
manager	119/159
manganese	136/233 C
Mar	1761/1929
marble	202/305 P
March	607/852 P
market	572/906
marketing	127/197
markets	218/332
marks	218/250 C
marriage	563/856
married	687/802
marry	112/122
marshes	105/171
masks	103/158
massive	531/668 P
master	680/885 CP
masters	311/387
mastery	118/133
match	115/140 C
mate	127/160
mating	118/172
matter	743/1250 CP
matters	259/317 P
mature	313/389
maturity	148/165
May	6133/12728 C
mayor	161/228
me	177/212
mean	423/574 CP
meaning	751/1027 CP
means	1317/1796 CP
meant	178/210 P
meanwhile	164/232
measure	493/646 CP
measured	334/503CP
measurement	222/387 CP
measurements	186/316C P

measures 431/548 CP
measuring 242/322 CP
meat 265/457
medium 455/693 CP
mee (pronun) 191/192
meet 356/402 P
meeting 236/270
melting 184/271 C
men 1410/2000 CP
merged 179/188
message 152/217
met 541/622
meters 152/212 C
mi 2747/5872 P
mid 1185/1612 P
middle 1644/2680 P
might 603/842 CP
mild 316/357 C
miles 193/237 CP
milk 174/504
mill 149/250 P
mills 195/289
mind 434/636 CP
mirror 161/286
mixed 442/544 C
mixture 338/463 C
mode 246/322
modes 123/165
moh (pronun) 147/147
moist 189/216 C
moisture 192/297
molten 111/181 C
money 433/825
month 254/326 CP
monthly 127/145
months 727/923 CP
mood 114/132
moon 329/765 CP
more 6171/14182 CP
moreover 219/255
morning 160/193
mostly 789/991 CP
mother 651/924
motifs 137/203
motion 88/1405 CP
motions 114/256 CP
mountain 767/1438 P
mountainous 300/346 P
mountains 889/1707 P
mounted 279/422
mouth 472/689
movies 120/173
mph 184/352 P
mud 158/248
muh(pronun) 174/175
murder 229/284
murdered 151/167
my 559/685 P

unlisted (too common)
made C
make CP
many CP
most CP
much CP
must CP

n 2447/3142
nah (pronun) 152/152
name 3023/3687 CP
named 1809/2111 P
names 314/423
narrow 587/713 C
nay (pronun) 124/126
nearly 1121/1455 CP
neck 205/262
nee (pronun) 337/348
need 664/821 CP
needed 540/695 CP
needs 422/557 P
neighboring 247/270
neighbors 147/173 C
neither 323/381 CP
nephew 135/141
nerve 180/453
nerves 100/167
nervous 270/491
nest 165/351
nests 123/191
net 161/243 CP
network 345/499 C
never 926/1089 CP
nevertheless 579/581 CP
new 6738/15987 CP
newer 114/117
newly 394/437
news 285/608
newspaper 303/452
newspapers 216/341
next 1011/1258 CP
nickel 147/226 C
nickname 146/159
night 623/815 CP
nine 387/453 C
nineteenth 126/134 C
nitrogen 202/436 C
noh (pronun) 186/192
non 442/599
none 324/392
nontheless 192/214
nor 373/461 CP
north 3299/7340 P
northeast 453/597 P
northeastern 453/514
northern 1916/3324
northward 149/183 P
northwest 636/908 P
northwestern 365/427 P
nose 140/192
not 5323/11796 CP
nothing 224/265 C
Nov 1641/1784
November 445/612
now 3031/4448 CP
nuclei 141/345
nucleus 216/496 P
nuh (pronun) 139/141
number 2584/4587 CP
numbered 218/248
numbers 844/1401 P

numerous	1390/1707 C
nuts	108/226

unlisted (too common)
near CP
nearby
no CP

o	1256/1816
oak	158/261 C
obtain	332/373 CP
obtained	618/772 CP
obtaining	130/136 P
occur	1087/1694 CP
occurred	633/845 C
occurrence	152/164
occurring	249/283 P
occurs	935/1300 CP
Oct	1685/1835
October	484/636 P
off	1200/1641 CP
offer	274/349
offered	352/417
offering	160/180
offers	241/257
office	872/1381
offices	221/255
offshore	138/179
offspring	108/171
often	3529/5736 CP
oil	858/2052 CP
oils	110/211
old	1947/2894 CP
older	503/634 C
oldest	577/652
once	1359/1708 CP
one	8070/16409 CP
ones	367/418 P
onset	111/137
onto	285/335 CP
open	1047/1565 C
opened	649/804 CP
opening	408/509 C
orange	295/478 CP
order	2031/2978 CP
ordered	262/323
orders	246/372 P
ordinary	350/423 CP
ore	347/654 C
ores	116/202 C
oriented	223/269 CP
others	1697/2270 CP
our	430/595CP
out	2639/3957 CP
outbreak	190/209
outdoor	112/133
outer	405/593 CP
output	310/542
outside	739/899 C
outstanding	476/548
outward	114/140 P
overall	252/293 CP
overcome	153/163 C
overseas	169/238
overthrow	177/206
overthrown	121/150
own	2301/3177 CP
owned	313/425 P
owner	129/182
owners	126/170
ownership	153/235
oz	102/146

unlisted (too common)
of CP
on CP
only CP
or CP
other CP
over CP

p	2357/3404
pace	107/117
pah (pronun)	100/101
paid	240/313
pain	231/434
painful	135/151 C
pair	279/403 C
pairs	224/354 C
pale	157/169
pan	152/218
panels	118/165
paper	669/1085 CP
papers	222/288 C
parent	149/250
parents	356/517
park	443/783 P
parks	223/361
part	3433/5171 CP
parties	324/605
partly	303/375
partner	146/162
partners	258/302 C
parts	1479/2197 C
party	1272/3662 P
past	831/1108
path	215/296 CP
patient	193/450
patients	227/535
pattern	560/801 P
patterns	631/925
paved	104/109 P
pay	356/486
payment	114/137
payments	109/176
peace	791/1420
peaceful	124/139
peak	433/536 CP
pearl	114/204
peasant	207/285
peasants	183/302
pee (pronun)	115/116

pen	138/169
people	2484/5106 CP
peoples	545/990
per	901/2150 CP
perceived	133/168
percent	669/1128
perennial	266/364
perfected	120/130
perform	288/350 CP
performance	485/698 P
performances	231/285
performed	651/868 P
performer	107/131
performers	109/156
performing	230/258 P
perhaps	898/1119 CP
person	890/1525 CP
personal	949/1284
persons	835/1257
petroleum	509/1033 P
pharmaceuticals	107/119
phase	361/627 CP
phases	138/183 P
phenomena	273/434 CP
phenomenon	309/385 C
phosphate	104/201 C
photo	330/684 CP
photograph	118/182 CP
photographic	198/311 C
photographs	172/261
photography	184/393
phrase	107/136
physical	1054/1730 CP
physically	104/118 P
physician	268/394
physicians	304/413
physicist	344/432 CP
physics	584/1150 CP
picture	365/650
pictures	281/430
piece	239/310 P
pieces	449/593 C
pine	205/414 P
pink	204/256 C
pipe	104/211 C
pitch	124/257 P
place	1566/2140 CP
placed	884/1107 CP
places	435/529CP
placing	128/135 C
plain	409/666
plains	396/746
plan	540/834
plane	302/630 CP
planes	153/255
planned	301/365
planning	327/511
plans	302/422
plant	988/2297 CP
planted	171/237
plants	1105/2581 C
plate	227/555 P
plateau	298/575
plates	207/386
play	1011/1599
played	972/1323 C
player	325/841 P
players	199/426
playing	324/401
pleasure	139/176
plus	252/317 C
point	1527/2388 CP
pointed	227/260 C
points	572/910 CP
poison	106/204 C
poisonous	137/214
policies	534/786
policy	794/1447
pollution	205/441 C
poor	675/922 CP
poorly	149/163
possess	221/266 C
possessed	141/172
possesses	102/110
possession	165/199
possessions	107/126
potatoes	177/217
pottery	199/361
pound	102/215
pounds	111/129 CP
poverty	230/347
powder	128/217 C
power	2435/5436 CP
powered	241/394
powerful	985/1293 CP
powers	661/1083 P
prairie	113/218
premier	297/430
press	375/623 P
pressed	164/187
pressure	825/1713 CP
pressures	205/267 C
prevailing	122/136
prevalent	134/147
prey	221/500
price	337/577 C
prices	246/437
primarily	1453/1786 P
prime	669/1344
print	121/225 C
printed	229/345
printing	270/592
prints	149/238
prior	359/421 P
private	1067/1617
privately	118/130
prize	942/1239
prized	134/160
procedure	217/302 CP
procedures	267/376
profits	117/162
prominence	225/243
prominent	752/849
promise	162/187
promised	142/170
promote	268/302
promoted	292/310
promoting	114/124
prompted	134/141
proof	117/153 P
proper	400/500 P
properly	162/170 C
properties	571/1102 CP
property	540/1007 CP
proposed	574/731 P
protect	404/491 CP
protected	311/374
protective	179/221 C
prove	147/173 P
proved	606/733 P
provide	1090/1506 CP

provided 1037/1398 P
provides 601/737 CP
providing 365/430
published 1697/2229 CP
publisher 111/128
publishing 251/344
pulled 104/122 P
pulp 120/233
punishment 149/240
pupil 170/181
pupils 108/118
purchase 177/229
purchased 165/179
purity 110/125 C
purple 212/288 C
purpose 511/609 CP
purposes 441/528
pursue 110/128
pursued 222/247
pursuit 122/131
pushed 156/194 C
put 476/566 CP

unlisted (too common)
probably P

qantum
130/341 P
quarter 259/328
quarters 131/143
queen 530/844
quest 162/179
question 396/551 CP
questions 290/397 CP
quick 156/175
quickly 597/697
CP
quiet 111/130
quite 299/341 CP

r 4064/885
race 356/549 P
races 158/302
racing 107/265 P
radius 103/178 P
rah (pronun) 184/186
rail 312/442
railroad 562/870 P
railroads 395/566
railway 160/223
rain 327/519 C
rainfall 473/657 P
raise 237/264 C
raised 668/818 CP
raising 386/423 C
ran 258/301
random 112/184 C
range 1767/2641 CP
ranged 114/118
ranges 522/738 C
ranging 493/557
rank 307/554 C
ranked 118/147 C
ranks 245/313
rate 693/1313 CP
rates 333/554 C
rather 1352/1793 CP
ratio 198/280 CP
raw 277/380 C
ray 490/840 P
rays 248/589
re 159/205
reach 666/817 CP
reached 880/1156 CP
reaches 394/451 CP
reaching 403/454 CP
read 374/502 CP
readily 253/298 C
reading 261/368 CP
ready 160/198 C

real 601/823 CP
rear 183/278
reason 466/633 CP
reasons 292/327 CP
rebuilt 234/259
receive 383/491
received 1427/1715 CP
receives 221/266
receiving 264/305
recognize 228/249 CP
recognized 972/1214 C
record 723/1137 CP
recorded 465/636 C
recording 198/434 C
recordings 124/177
records 447/646
recover109/109 C
recovered 186/212
recovery 212/281 C
red 1341/2432 CP
reddish 179/195 C
reduce 441/543 CP
reduced 692/877 CP
reduces 124/147 CP
reducing 240/290 C
reduction 258/336 C
ree (pronun) 272/275
reed 115/201
refer 222/239 CP
referred 441/495 CP
refers 655/723 C
refused 367/426
regained 148/157
regard 191/209 P
regarded 723/866 C
regarding 129/136 C
regardless 125/128 CP
reinforced 146/202
related 1329/1756 CP
relating 107/109 CP
relation 250/319 CP
relations 558/822 CP
relationship 506/646 CP
relationships 284/361 CP
relative
470/605 P
relatively 1029/1350 CP

relatives	163/193
release	340/439 P
released	410/508 CP
reliable	120/139 P
relied	127/137
relief	339/522
rely	114/115
remain	779/908 CP
remainder	220/240
remained	1338/1696 P
remaining	465/533 CP
remains	1080/1378 CP
remarkable	361/399
remarkably	111/114 C
remembered	364/396
removal	235/294C
remove	234/287 CP
removed	442/553 C
removing	120/137 C
renamed	228/240
rendered	148/165
renewed	218/240
renowned	280/296
replace	247/275 CP
replaced	846/1080 CP
replacement	111/124
replacing	129/136
report	220/280 P
reports	203/260 P
require	470/594 CP
required	868/1223 CP
requirements	264/332 P
requires	431/549 P
requiring	199/218
research	1279/2111 CP
researchers	191/249
resemble	286/328 C
resembles	244/292
resembling	161/170
resolution	141/229
resort	233/263
resource	122/173
resources	663/1379
response	507/712
responses	116/169
rest	705/846 CP
restore	205/225
restored	448/507 C
resumed	165/174
retail	117/148
retain	199/217 C
retained	366/400
retirement	268/313
revenue	137/206
revenues	116/161
reverse	130/153 C
reversed	129/142 C
review	207/263 CP
reviewed	145/145
rice	354/583
rich	893/1147 C
richest	110/118 C
richly	116/125
ridge	167/314 C
ridges	132/197
right	1212/2244 CP
rights	807/1790
ring	330/606
rings	152/257
rise	1064/1416 CP
rises	360/413 CP
rising	419/486 C
risk	158/249
river	2273/5059 P
rivers	813/1389
road	415/644 P
roads	372/548 CP
rock	792/1615
rocket	137/387
rocks	477/1210
rocky	277/388
rod	102/163 P
rodents	103/171
roh (pronun)	203/215
role	1406/236 C
room	291/419 CP
rooms	131/188
root	289/493 CP
rooted	105/112
roots	480/673 C
rose	601/793 C
rough	215/258 CP
roughly	257/294 CP
round	391/482 CP
rounded	212/236
route	334/437 P
routes	202/274
row	107/138
rows	115/136
rubber	278/457
rugged	163/187
ruh (pronun)	118/118
rule	1066/1707 CP
ruled	693/873
ruler	355/454
rulers	330/441
rules	391/653 P
ruling	280/328
run	524/701 CP
running	386/524 P
runs	249/342 CP
rush	101/154

unlisted (too common)
repr CP

s	17035/58459
safe	145/178
safety	256/411
sah (pronun)	121/124
said	610/735 CP
sailed	142/220
sale	204/268
sales	184/295
salmon	134/243
salt	461/813 C
salts	153/263 C
same	2217/3352 CP

sample	100/218 C
sand	366/684
sandstone	102/144
sandy	161/192
save	126/138
saved	127/137 P
saw	480/602 CP
say	231/282 CP
scale	892/1333 CP
scales	196/329 CP
scandal	127/155
scattered	204/233
scene	366/517
scenes	410/608
scenic	105/128
scheme	154/190
scholar	255/283
scholarly	113/128
scholars	359/465
scholarship	133/150
school	1832/3129
schools	797/1629
science	1251/2173 CP
sciences	471/693 P
scientific	899/1382 CP
scientist	255/305 C
scientists	504/714
scope	187/212
score	117/176
scored	107/141
scores	132/162
screen	229/414
sea	2057/4549 C
seaport	122/134
search	531/646
seas	258/396
season	489/728 P
seasonal	151/189
seasons	216/253
seat	739/855
seats	127/189
sec	134/322 P
second	2411/3596 CP
secondary	440/766
seconds	134/167 CP
section	465/688 CP
sections	319/375 CP
secure	251/284
secured	201/226
security	330/584
seed	257/491
seeds	333/648
seek	261/293
seeking	239/262
seeks	100/110
seem	324/377 CP
seemed	220/297
seems	308/348
seen	894/1145 CP
seized	260/308
seldom	141/152 P
self	1112/1543 C
sell	151/193
selling	190/226
send	122/138 C
sense	799/1107 CP
sensitive	307/391
sensitivity	113/133
sensory	112/249
sent	614/767 P
sentence	102/223
sentenced	103/111
Sept	1655/1805
September	442/613
series	1782/2520 CP
serious	684/856 CP
seriously	146/155
services	726/1254
set	1869/2680 CP
sets	334/440 CP
setting	394/450 P
settings	181/235
settle	179/198
settled 1	123/1278
settlement	697/998
settlements	246/338
settlers	385/567
settling	119/127
seven	767/964 CP
seventh	118/133
shadow	132/159
shallow	399/516
shape	680/902 CP
shaped	847/1057 P
shapes	267/336 P
share	314/349 C
shared	486/540
shares	122/164
sharing	106/137
sharp	343/391 C
sharply	218/238
sheep	327/502
sheet	142/237 C
sheets	135/207
shell	252/521 P
shells	177/315
shelter	105/137
shield	126/213
shift	186/268 C
shifted	172/205 CP
shifting	111/122
ship	339/699
shipbuilding	170/191
shipping	272/343
ships	438/830 P
shock	137/231
shore	304/373
shores	142/172
short	1967/2756 CP
shorter	223/254
shortly	377/417 P
shot	184/255 P
should	777/1180 CP
shoulder	186/230
shoulders	134/162
show	860/1219 CP
showed	442/546 C
showing	196/214 CP
shown	622/945 CP
shows	440/549 CP
shrub	124/234
shrubs	222/306
side	968/1503 CP
sided	126/134
sides	437/586 CP
siege	141/187

sight	128/152
sign	230/303 CP
signal	199/558 P
signals	190/492
signed	388/499
signs	207/307 P
silent	140/184
silicon	117/265 C
silk	175/343
silver	536/961 C
similarly	289/318 CP
simpler	107/121 C
simplest	160/184 C
simply	430/494
sin	140/219
since	2992/4519 CP
singer	273/453
singing	162/243
single	1506/2201 CP
sir	1128/1929
sister	215/250 P
site	1263/1718
sites	490/775
six	1016/1228 CP
sixth	145/169
size	1188/1797 CP
sized	171/205
sizes	154/185
skeleton	106/191
skill	235/284
skilled	158/186
skillful	114/126
skills	266/367
skin	522/1061 P
skull	134/249
sky	260/406
slave	265/456
slavery	233/564
slaves	238/434
slender	245/276
slight	122/128 CP
slightly	492/542 CP
slope	101/158 P
slopes	179/239 P
slow	410/516 P
slowly	360/419 C
small	3443/6011 CP
smaller	959/1251 CP
smallest	196/230 CP
smoke	111/179
smooth	293/352 P
snake	138/327
snow	275/428
sodium	215/528 C
soft	449/578 CP
soil	562/1168 C
soils	277/703
sold	386/466 CP
soldier	241/271
soldiers	249/358
sole	212/245
solely	138/147
solid	559/897 P
soluble	111/163 C
solution	386/681 CP
solutions	198/295 P
solve	108/134 CP
solved	101/119
something	162/201 P
sometimes	1696/2222 CP
somewhat	462/502 C
son	1708/2328
song	436/692
songs	430/713
sons	310/373
sort	106/115
sought	745/934 C
sound	617/1420 P
sounds	240/470 C
source	1247/1750 C
sources	722/1013 C
south	2872/6531 P
southeast	620/923 P
southeastern	498/605
southern	1887/3226
southward	151/192
southwest	566/766
southwestern	483/556
sovereign	113/147
sovereignty	151/237
space	947/2566 CP
spacecraft	142/426
spaces	154/203
span	170/247
sparsely	114/126
spatial	106/145
speak	323/370 C
speaker	110/194
speakers	136/269
speaking	552/685 CP
special	1157/1687 CP
specialized	506/728
specially	123/134
species	1778/4735 C
specific	832/1203 CP
specifically	264/276
specified	191/236 P
specimens	112/144
spectrum	204/380
speech	358/746
speed	604/1120 CP
speeds	184/274 CP
spend	139/158 C
spending	133/188
spent	705/771 P
sperm	121/310
sphere	151/236 P
spiral	129/193
spite	109/122
split	241/283
spoke	139/150
spoken	252/490
spokesman	126/133
sponsored	242/266
spontaneous	111/153
sport	201/398
sports	223/384 P
spot	106/153
spots	127/151 P
spread	769/1015 C
spreading	173/233
spring	560/779
springs	199/307
spurred	145/156
sq	1182/3547
square	408/652 CP
stability	237/282
stable	318/447 CP
staff	392/586
stage	948/1538 CP

stages	341/478 C
stand	302/346 P
standard	746/1005 CP
standards	408/574 P
standing	325/406 CP
stands	317/362 P
star	622/1323 P
stars	420/1263 P
start	234/287 CP
started	325/357 CP
starting	234/272 CP
state	3263/9450 CP
stated	166/214 CP
statement	201/261 CP
statements	105/157 C
states	4887/11320 CP
statesman	280/308
station	317/494
stations	277/476 CP
statistical	152/247 C
stay	159/167
steadily	157/170
steady	157/179
steam	254/611 C
steel	697/1268 C
steep	132/151
stellar	103/193
stem	185/313
stems	206/297
step	289/387 CP
steps	232/307 C
still	2009/2844 CP
stimulate	107/129
stimulated	225/280
stock	366/607
stomach	115/211
stone	798/1520 P
stones	215/351 P
stood	144/162
stop	204/248 P
stopped	130/146
storage	257/417 C
store	178/252 P
stored	218/349 C
stores	118/157
stories	662/1040
storm	159/219
storms	127/187
story	1235/1592
straight	321/396 CP
strait	240/352
strange	109/119 C
strategic	225/337
strategy	170/313 P
stream	290/436
streams	287/426
street	337/458
streets	171/210 P
strength	528/723 P
strengthen	104/108
strengthened	145/160
stress	275/427
stressed	198/226
stresses	106/125
stretched	105/114
strict	189/207
strictly	128/131 CP
strike	227/355 CP
strikes	138/177 P
striking	293/336
string	206/314 P
strings	120/262 P
strip	204/289 C
stripes	109/157
stroke	145/190
strong	1407/1836 CP
stronger	172/194 CP
strongest	124/136 C
strongly	488/543 C
struck	159/225 P
structural	333/505
structure	1361/2326 P
structures	732/1219
struggle	466/588
struggles	109/115
student	415/578 CP
students	472/944 P
studied	1243/1394 CP
studies	1321/1807 C
study	2011/3142 CP
studying	408/426 CP
style	1924/3617
styles	391/642
stylistic	120/131
subdivisions	220/229
subject	1037/1366 CP
subjected	142/151 C
subjects	602/782
submarine	128/239 P
subtle	166/176
subtropical	198/254
succeed	143/149
succeeded	987/1281 C
succeeding	220/227
successes	143/156
successful1	287/1543 P
successfully	414/466 P
succession	389/498
sudden	147/167 C
suddenly	111/120 P
suffered	413/472 P
suffering	193/221 C
sugar	399/741 P
sugarcane	132/163
suggest	254/278 C
suggested	349/392 P
suggesting	101/105
suggests	243/269
suh (pronun)	122/122
suit	128/209 P
suitable	257/299 CP
suited	155/171
sulfur	163/345 C
sum	139/205 CP
summer	721/1063
summers	191/234
sun	679/1443 CP
sung	124/180
sunlight	134/177
superb	119/125
supplied	292/341
supplies	320/461
supply	665/1033 CP
surface	1233/2798 CP
surfaces	345/453 P
surgery	165/419
surgical	128/218
surpassed	107/117
surrender	159/211
surrendered	106/124
surrounded	383/434 C

surrounding 576/652
survey 262/323 C
survival 211/249 C
survive 351/415
survived 355/407
survives 109/113
surviving 290/320
sustained 152/165
sweet 234/417
swept 109/134
swift 148/190
swimming 143/220 P
symbol 384/442 CP
symbolic 237/295 C
symbolism 147/199
symbols 266/377 CP
sympathetic 112/124
symptoms 263/477
syndrome 112/192
synthesis 253/402 C
synthesized 122/146 C
synthetic 283/460
system 2740/6630 CP
systematic 231/280 P
systems 1282/2840 CP

unlisted (too common)
see CP
served
several CP
she CP
so CP
some CP
soon CP
such CP

t 1916/2645
table 290/469 CP
tactics 112/171
tah (pronun) 142/146
tail 531/876 P
tails 129/153 P
take 942/1195 CP
taken 1051/1311 CP
takes 524/599 CP
taking 470/528 CP
talent 154/167
tall 365/485 P
tan 117/167 P
target 152/236 P
task 222/271 CP
tasks 132/170
taste 250/324 C
taught 866/969
tax 337/848
taxation 115/154
taxes 200/349
tea 129/282 C
teach 149/178
teacher 455/572
teachers 197/353
teaching 486/716
teachings 147/176
team 304/535
teams 148/249
technical 476/625
technique 658/932 CP
techniques 909/1391 CP
technological 216/291 P
technologies 136/208
technology 843/1346 P
tee (pronun) 222/231 P
teeth 297/655
telegraph 128/204
telephone 148/296
tell 148/176 CP
temperate 371/506
ten 590/698 P
tend 454/580 CP
tended 219/267
tends 187/213
tenure 124/135
term 1892/2666 CP
termed 225/280
terms 1018/1481
terrain 175/209
terrestial 139/225
test 382/637 C
tested 147/165
testing 195/307 P
tests 224/453 C
theme 334/398
themes 439/538
themselves 831/1101 P
theory 1405/3209 CP
thereafter 542/584
therefore 834/1047 C
thick 433/558 CP
thickness 109/143 P
thin 452/633 CP
things 317/425 CP
think 120/149 CP
thinkers 104/139
thinking 174/212 C
third 1424/1949 CP
thirds 233/308
thirty 132/175
thought 1177/1686 C
thousand 287/332 P
thousands 576/721 C
threat 242/300
threatened 279/335
three 3440/5310 CP
throat 128/182
throughout1810/2517 CP
thrust 127/244 P
tidal 108/172
tied 147/169 P
ties 232/295
tightly 112/119
timber 232/325
time 4090/7407 CP
times 2322/3235 CP
tin 203/338 CP
tiny 360/464 CP
tip 263/330 P
tissue 331/720 C
tissues 254/489
today 1728/2452 CP
toes 111/167
together 1225/1605 CP
toh (pronun) 199/200
told 122/132 P
tone 250/375
tongue 165/262
tons 230/490 CP

too	736/948 CP
tool	227/318 P
tools	375/633 P
top	673/1043 CP
touch	156/176
tough	136/157
tour	204/263
toured	124/135
tourism	383/505
tourist	252/277
tourists	152/166
toward	1300/1799 P
tower	227/363 P
towers	127/175
town	996/1343
towns	402/609
trace	151/179 C
traced	229/245
traces	156/164
track	172/328 CP
tract	171/261
trade	1391/2882
traders	167/205
trading	350/486
traffic	196/319
trail	100/182
train	179/232 P
trained	532/598
training	621/893
traits	142/220
travel	439/590
traveled	378/410 C
traveling	201/220 C
travels	196/238
treason	116/144
treasury	207/276
treat	158/200 CP
treated	360/439 CP
treaties	164/240
treating	143/170 C
treatise	203/247
treatment	688/1169 C
treaty	590/1128
tree	666/1657 P
trees	669/1299 P
tremendous	121/137 P
trend	236/272
trends	168/196
trial	334/541 C
trials	139/181
tried	550/654
trip	159/171 P
triple	141/205
troops	594/1037
troubled	100/105
true	851/1131 CP
truly	124/136 P
trust	148/268
truth	242/338
try	149/169 CP
trying	166/177 C
tube	274/603 CP
tubes	205/351
tuh (pronun)	145/146
tur (pronun)	159/163
turn	849/1065 CP
turned	827/981 C
turning	274/312 P
turns	182/218 CP
twelve	101/130 C
twentieth	220/238
twenty	120/151
two	5736/10916 CP
type	1429/2313 CP
types	1286/2050 CP

unlisted (too common)
than CP
that CP
the CP
their CP
them CP
then CP
there CP
thereby
these CP
they CP
this CP
those CP
though CP
through CP
thus CP
to CP
took CP

u	3361/7428
uh (pronun)	1412/1526
uhm (pronun)	137/139
uhn (pronun)	248/251
uhs (pronun)	235/244
ultimate	210/249
ultimately	338/387 C
unable	326/382 P
uncertain	154/161
uncle	146/190
undergo	127/164 CP
undergraduate	132/183
underground	258/346 C
underlying	214/249
understand	173/209 CP
understanding	502/665CP
understood	309/369 CP
undertaken	104/113
undertook	126/135
underwater	104/173 C
underwent	104/108
unemployment	165/283
unfinished	130/139
unified	227/271 P
unique	516/611
unit	548/842 CP
unite	107/114
united	4674/9912 CP
units	521/902 CP
unity	313/431 CP
unknown	419/487 CP
unless	196/236 P
unlike	555/633 C
unprecedented	129/140
unrest	105/126
unstable	129/157 C
unsuccessful	335/367

unsuccessfully 137/139
unusual 320/363 CP
unusually 103/106 CP
upper 964/1401 CP
upright 110/128
uprising 168/208
upward 185/219 P
uranium 134/367 CP
urged 163/178
us 130/170 CP
usage 131/147
useful 471/593 CP
uses 761/1062 CP

unlisted (too common)
under P
until CP
up CP
upon CP
use CP
used CP
using CP
usually CP

v 1455/2305
vacuum 140/220 CP
valley 706/1306 P
valleys 270/449
valuable 442/536 P
value 727/1325 CP
valued 199/233
values 467/735 CP
vapor 115/229 C
variable 264/406
varied 374/432 C
varies 393/447 CP
varieties 341/568
variety 1323/1777 C
variously 105/106
vary 547/623 CP
varying
319/355
vast 622/796 C
vee (pronun) 139/139
vegetable 179/289
vegetables 211/285 C
vegetation 386/616
vehicle 217/393 P
vehicles 170/295 C
veins 111/178
vertebrates 121/216
very 1319/2048 CP
vessel 160/274 C
vessels 373/648 C
via 199/250
victim 129/179 C
victims 166/196
view 771/1072 CP
viewed 302/345 CP
views 471/571
village 502/603
villages
283/388
vivid 156/163
voice 341/451
volcanic 315/502 P
voyage 154/235
voyages 110/168

unlisted (too common)
various CP

w 3408/4398
wage 111/165
wake 105/142
walk 119/171 P
wall 513/855 CP
walled 108/125
walls 503/759 CP
wanted 139/176 CP
war 4534/11217 C
warfare 258/481
warm 440/608 P
warmer 130/151
warning 100/131
warrior 136/168
warriors 128/167
wars 778/1185
wartime 139/154
waste 191/381
water 2069/6005 CP
waters 507/816
wave 335/717
wavelength 119/269
wavelengths 101/228 P
waves 298/821 P
way 1731/2477 CP
ways 654/837 CP
we 215/291 CP
weak 269/340 CP
weakened 177/199 C
weakness 121/145
wealth 400/510
wealthy 309/342
weapon 104/217
weapons 272/552
wear 130/182 C
weather 334/620 C
weaving 100/161
week 186/227 P
weekly 153/180
weeks 332/437 C
weigh 205/250 CP
weighing 119/149 C
weighs 167/189 C
weight 763/1203 CP
weights 102/197 P
welfare 220/364
well 3766/5839 CP
wells 152/259
west 2616/4841 P
western 2343/4260
westward 191/250 P

wet	169/230 CP
whatever	135/143 C
wheat	343/645
wheel	142/336
whereas	753/920 CP
whereby	123/132
whether	647/866 CP
white	2138/3721 CP
whites	139/264
whole	610/801 C
wide	1515/1876 P
wider	162/170 P
widespread	632/738 C
widow	117/141
width	175/216 P
wife	691/838
wild	527/839
wilderness	106/151
wildlife	141/237
will	1606/3300 CP
win	324/411
wind	448/911 P
window	110/139 P
windows	121/206
winds	226/402
wine	239/358 C
wing	373/561
wings	266/455
winner	131/165 P
winning	376/459
winter	725/1058 P
winters	194/245
wire	186/329 CP
wisdom	127/177
withdraw	106/121 C
withdrawal	140/188
withdrew	198/225
without	1515/1963 CP
wolf	144/224
woman	630/922 P
women	1842/2727 C
wood	843/1481 CP
wooden	266/369
woods	157/197
woody	104/163
wool	173/294 C
word	812/1226 P
words	512/877 CP
work	4163/7036 CP
worked	1260/1450 C
worker	146/212
workers	557/1069
working	894/1134 CP
works	2887/4872 C
workshop	105/121
world	5787/12090 CP
worlds	138/154 C
worldwide	586/771
worms	128/220
worn	103/163
worth	159/219 P
would	1664/3303 CP
wounded	153/195
write	322/375 CP
writing	818/1311 C
wrote	1671/2168 P

unlisted (too common)

want CP
was CP
were CP
what CP
when CP
where CP
which CP
while CP
who CP
whom
whose CP
why CP
widely
with CP
within CP

x	567/1644
XIV	186/264
y	939/1234
yah (pronun)	125/126
year	2689/4357 CP
yearly	131/139 P
years	4912/8977 CP
yellow	683/1036 C
yellowish	111/122
yet	728/895 CP
yield	260/327 CP
yielded	151/162 C
yields	183/240 C
you	216/292 CP
young	1606/2300 CP
younger	421/504
youngest	118/126
your	160/189 CP
youth	344/393
z	163/240
zero	191/353 CP
zinc	199/340 C
zone	335/664
zones	225/334

ESSAY

Scaling—the Physics of Lilliput*

Philip Morrison
Massachusetts Institute of Technology

The fictional traveler Lemuel Gulliver spent a busy time in a kingdom called Lilliput, where all living things—men, cattle, trees, grass—were exactly similar to our world, except that they were all built on the scale of one inch to the foot. Lilliputians were a little under 6 inches high, on the average, and built proportionately just as we are. Gulliver also visited Brobdingnag, the country of the giants, who were exactly like men but 12 times as tall. As Swift described it, daily life in both kingdoms was about like ours (in the 18th century). His commentary on human behavior is still worth reading, but we shall see that people of such sizes just could not have been as he described them.

Long before Swift lived, Galileo understood why very small or very large models of man could not be like us, but apparently Dean Swift had never read what Galileo wrote. One character in Galileo's "Two New Sciences" says, "Now since . . . in geometry, . . . mere size cuts no figure, I do not see that the properties of circles, triangles, cylinders, cones, and other solid figures will change with their size. . . ." But his physicist friend replies, "The common opinion is here absolutely wrong." Let us see why.

We start with the strength of a rope. It is easy to see that if one man who pulls with a certain strength can almost break a certain rope, two such ropes will just withstand the pull of two men. A single large rope with the same total area of cross-section as the two smaller ropes combined will contain just double the number of fibers of one of the small ropes, and it will also do the job. In other words, the breaking strength of a wire or rope is proportional to its area of cross-section, or to the square of its diameter. Experience and theory agree in this conclusion. Furthermore, the same relation holds, not only for ropes or cables supporting a pull, but also for columns or struts supporting a thrust. The thrust which a column will support, comparing only those of a given material, is also proportional to the cross-sectional area of the column.

Now the body of a man or an animal is held up by a set of columns or struts—the skeleton—supported by various braces and cables, which are muscles and tendons. But the weight of the body which must be supported is proportional to the amount of flesh and bone present, that is, to the volume.

Let us now compare Gulliver with the Brobdingnagian giant, 12 times his height. Since the giant is exactly like Gulliver in construction, every one of his linear dimensions is 12 times the corresponding one of Gulliver's. Because the strength of his columns and braces is proportional to their cross-sectional area and thus to the square of their linear dimension (strength $\propto L^2$), his bones will be 12^2 or 144 times as strong as Gulliver's. Because his weight is proportional to his volume and thus to L^3, it will be 12^3 or 1728 times as great as Gulliver's. So the giant will have a strength-to-weight ratio a dozen times smaller than ours. Just to support his own weight, he would have as much trouble as we should have in carrying 11 men on our back.

In reality, of course, Lilliput and Brobdingnag do not exist. But we can see real effects of a difference in scale if we compare similar animals of very different sizes. The smaller ones are not scale models of the larger ones. Figure 1 shows the corresponding leg bones of two closely related animals of the deer family: one a tiny gazelle, the other a bison. Notice that the bone of the large animal is not at all similar geometrically to that of the smaller. It is much thicker for its length, thus counteracting the scale change, which would make a strictly similar bone too weak.

Galileo wrote very clearly on this very point, disproving the possibility of Brobdingnag, or of any normal-looking giants: ". . . if one wishes to maintain in a great giant the same proportion of limb as that found in an ordinary man he must either use a harder and stronger material for making the bones, or he must admit a diminution of strength in comparison with men of medium stature; for if his height be increased

(Continued on next page)

*Adapted from PSSC PHYSICS, 2nd edition, 1965; D.C. Heath and Company with Education Development Center, Inc., Newton, MA.

Figure 1a The front leg bones of a bison and a gazelle. The animals are related, but the gazelle is much smaller. The photos show the approximate relative sizes of the bones.

Figure 1b The leg bone of a gazelle enlarged to the same length as the bison bone. Note that the bone of the larger animal is much thicker in comparison to its length than that of the gazelle. The small deer is generally more lightly and gracefully built. Can you visualize how much different Lilliputians must have been from men of normal size?

inordinately he will fall and be crushed under his own weight. Whereas, if the size of a body be diminished, the strength of that body is not diminished in the same proportion; indeed, the smaller the body the greater its relative strength. Thus a small dog could probably carry on his back two or three dogs of his own size; but I believe that a horse could not carry even one of his own size." The sketch of Figure 2 is taken from Galileo, who drew it to illustrate the paragraph just quoted.

An elephant is already so large that his limbs are clumsily thickened. However, a whale, the largest of all animals, may weigh 40 times as much as an elephant; yet the whale's bones are not proportionately thickened. They are strong enough because the whale is supported by water. What is the fate of a stranded whale? His ribs break. Some of the dinosaurs of old were animals of whalelike size; how did *they* get along?

Following Galileo, we have investigated the problems of scaling up to giants. Now let's take a look at some of the problems that arise when we scale down.

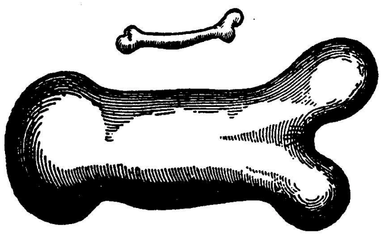


Figure 2 Galileo's drawing illustrating scaling. Over 300 years ago, Galileo wrote concerning the fact that a bone of greater length must be increased in thickness in greater proportion in order to be comparably strong.

When you climb dripping wet out of a pool there is a thin film of water on your skin. Your fingers are no less wet than your forearm; the thickness of the water film is much the same over most of your body. Roughly, at least, the amount of water you bring out is proportional to the surface area of your body. You can express this by the relation

$$\text{amount of water} \propto L^2,$$

where L is your height. The original load on your frame is as before, proportional to your volume. So, the ratio *extra load/original load* is proportional to L^2/L^3, or to $1/L$. Perhaps you carry out of the pool a glassful or so, which amounts to about a 1% increase in what you have to move about. But a Lilliputian will bring out about 12% of his weight, which would be equivalent to a heavy winter suit of clothing with an overcoat. Getting out of the pool would be no fun! If a fly gets wet, his body load doubles, and he is all but imprisoned by the drop of water.

There is a still more important effect of the scale of a living body. Your body loses heat mainly through the skin (and some through breathing out warm air). It is easy to believe — and it can be checked by experiment — that the heat loss is proportional to the surface area, so that

$$\text{heat loss} \propto L^2,$$

keeping other factors, like the temperature, nature of skin, and so on, constant. The food taken in must supply this heat, as well as the surplus energy we use in moving about. So minimum food needs go as L^2. If a man like Gulliver can live off a leg of lamb and a loaf of bread for a day or two, a Lilliputian with the same body temperature will require a volume of food only $(\frac{1}{12})^2$ as large. But his leg of lamb, scaled down to his world, will be smaller in volume by a factor of $(\frac{1}{12})^3$. Therefore, he would need a dozen of his roasts and loaves to feel as well fed as Gulliver did after one. Lilliputians must be a hungry lot, restless, active, graceful, but easily waterlogged. You can recognize these qualities in many small mammals, like a mouse.

We can see why there are no warm-blooded animals much smaller than the mouse. Fish and frogs and insects can be very much smaller because their temperature is not higher than their surroundings. In accord with the scaling laws of area and volume, small, warm-blooded animals need relatively a great deal of food; really small ones could not gather or even digest such an enormous amount. Certainly the agriculture of the Lilliputians could not have supported a kingdom like the one Gulliver described.

(Continued on next page)

Now we see that neither Brobdingnag nor Lilliput can really be a scale model of our world. But what have these conclusions to do with physics?

Let's start again with the very large. As we scale up any system, the load will eventually be greater than the strength of the structure. This effect applies to every physical system, not just to animals, of course. Buildings can be very large because their materials are stronger than bone, their shapes are different, and they do not move. These facts determine the constants like K in the equation

$$\text{strength} = KL^2$$

but the same laws hold. No building can be made which will look like the Empire State but be as high as a mountain, say 10 000 m. Mountains are solid structures, for the most part, without interior cavities. Just as the bones of a giant must be thick, an object of mountainous size on the earth must be all but solid, or else built of new materials yet unknown.

Our arguments are not restricted to the surface of the earth. We can imagine building a tremendous structure far out in space away from the gravitational pull of the earth. The load then is not given by the earth's gravitational pull, but as the structure is built larger and larger each part pulls gravitationally on every other and soon the outside of the structure is pulled in with great force. The inside, built of ordinary materials, is crushed, and large protuberances on the surface break off or sink in. As a result any large structure like a planet has a simple shape, and if it is large enough, the shape is close to a sphere. Any other shape will be unable to support itself. Here is the essential reason why the planets and the sun tend to be spherical. The pull of gravity is important for us on earth, but as we extend the range of dimensions which we study, it becomes absolutely dominant in the very large. Only motion can change this result. The great masses of gas which are nebulae, for example, are changing in time, and hence the law that large objects must be simple in shape is modified.

When we go from our size to the very small, gravitational effects cease to be important. But as we saw in investigating Lilliput, surface effects become significant. If we go far enough toward the very small, surfaces no longer appear smooth, but are so rough that we have difficulty in defining a surface. Other descriptions must be used. In any case, it will not come as a complete surprise that in the domain of the atom, the very small, scale factors demonstrate that the dominant pull is one which is not easily observed in everyday experience.

Such arguments as these run through all of physics. Like order-of-magnitude measurements, they are extremely valuable when we begin the study of any physical system. How the behavior of a system will change with changes in the scale of its dimensions, its motion, and so on, is often the best guide to a detailed analysis.

Even more, it is by the study of systems built on many unusual scales that physicists have been able to uncover unsuspected physical relations. When changing scale, one aspect of the physical world may be much emphasized and another one may be minimized. In this way we may discover, or at least get a clearer view of, things which are less obvious on our normal scale of experience. It is largely for this reason that physicists examine, in and out of their laboratories, the very large and the very small, the slow and the rapid, the hot and the cold, and all the other unusual circumstances they can contrive. In examining what happens in these circumstances we use instruments both to produce the unusual circumstances and to extend our senses in making measurements.

It is hard to resist pointing out how much the scale of man's own size affects the way he sees the world. It has been largely the task of physics to try to form a picture of the world which does not depend upon the way we happen to be built. But it is hard to get rid of these effects of our own scale. We can build big roads and bridges which are long and thin, but are essentially not three-dimensional, complex structures. The very biggest things we can make which have some roundness, which are fully three-dimensional, are buildings and great ships. These lack a good deal of being a thousand times larger than men in their linear dimensions.

Within our present technology our scaling arguments are important. If we design a new large object on the basis of a small one, we are warned that new effects too small to detect on our scale may enter and even become the most important things to consider. We cannot just scale up and down blindly, geometrically, but by scaling in the light of physical reasoning, we can sometimes foresee what changes will occur. In this way we can employ scaling in intelligent airplane design, for example, and not arrive at a jet transport that looks like a bee—and won't fly.

Suggested Readings

Galilei, Galileo, "Dialogues Concerning Two New Sciences," trans. by Henry Crew and Alphonso De Salvio, Evanston, Northwestern University Press, 1946, pp. 1–6, 125–128.

Haldane, J. B. S., "On Being the Right Size." *World of Mathematics*, Vol. II, edited by James R. Newman. New York, Simon & Schuster, 1956.

Holcomb, Donald F. and Philip Morrison, *My Father's Watch—Aspects of the Physical World*, Englewood Cliffs, NJ, Prentice Hall, 1974, pp. 68–83.

Smith, Cyril S., "The Shape of Things," *Scientific American*, January, 1954, p. 58.

Thompson, D'Arcy W., "On Magnitude," in *On Growth and Form*, Cambridge University Press, 1952 and 1961.

Essay Questions

1. The leg bones of one animal are twice as strong as those of another closely related animal of similar shape. (a) What would you expect to be the ratio of these animals' heights? (b) What would you expect to be the ratio of their weights?
2. A hummingbird must eat very frequently and even then must have a highly concentrated form of food such as sugar. What does the concept of scaling tell you about the size of a hummingbird?
3. About how many Lilliputians would it take to equal the mass of one citizen of Brobdingnag?

Essay Problems

1. The total surface area of a rectangular solid is the sum of the areas of the six faces. If each dimension of a given rectangular solid is doubled, what effect does this have on the total surface area?
2. A hollow metal sphere has a wall thickness of 2 cm. If you increase both the diameter and thickness of this sphere so that the overall volume is three times the original overall volume, how thick will the shell of the new sphere be?
3. If your height and all your other dimensions were doubled, by what factor would this change (a) your weight? (b) the ability of your leg bones to support your weight?
4. According to the zoo, an elephant of mass 4.0×10^3 kg consumes 3.4×10^2 times as much food as a guinea pig of mass 0.70 kg. They are both warm-blooded, plant-eating, similarly shaped animals. Find the ratio of their surface areas, which is approximately the ratio of their heat losses, and compare it with the known ratio of food consumed.
5. A rectangular water tank is supported above the ground by four pillars 5 m long whose diameters are 20 cm. If the tank were made 10 times longer, wider, and deeper, what diameter pillars would be needed? How much more water would the tank hold?
6. How many state maps of scale 1 : 1·000 000 would you need to cover the state with those maps?

CHEMISTRY IN ACTION

Salvaging the Recorder Tape from the *Challenger*

When the space shuttle *Challenger* exploded in flight on January 28, 1986, the crew cabin separated from the rest of the orbiter and broke up when it hit the water. The cabin was equipped with tape recorders to collect shuttle data and record conversations among the crew. However, there was no "black box" to protect the tapes as is used in airplanes. Thus, when the tapes were found six weeks later in 90 feet of water, they were considerably damaged by exposure to seawater and resultant chemical reactions. The tapes were described as "a foaming, concretelike mess, all glued together."

The major problem was the formation of magnesium hydroxide [$Mg(OH)_2$] by reaction of seawater with magnesium used in the tape reel:

$$Mg^{2+}(aq) + 2OH^-(aq) \longrightarrow Mg(OH)_2(s)$$

(Seawater is somewhat basic and therefore contains enough hydroxide ions to react with the Mg^{2+} ions formed when Mg metal comes in contact with ions of less active metals.) The magnesium hydroxide gradually covered the tape layers and glued them together. In addition, binders holding the iron(II) oxide (the magnetic material used in tapes) to the plastic backing were weakened, exposing bare tape in some places. After experimenting with the recovery process using less important retrieved tapes, a team of scientists prepared to salvage the central tape—the one that recorded the crew conversations. In a very tedious and lengthy process, they carefully neutralized the magnesium hydroxide, removing it from the tape, and stabilized the iron oxide layer. All the work had to be done with the tape still coiled. The tape was alternately treated with nitric acid and distilled water (Figure 3.8). The acid-base neutralization reaction is

$$Mg(OH)_2(s) + 2HNO_3(aq) \longrightarrow Mg(NO_3)_2(aq) + 2H_2O(l)$$

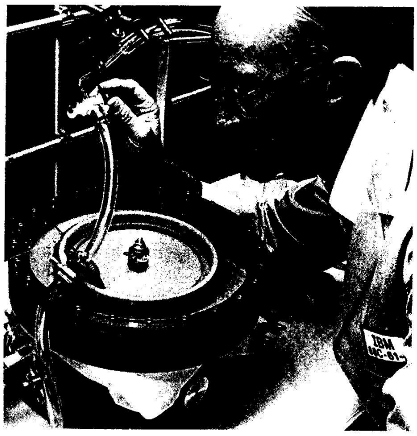


Figure 3.8 *A scientist examines the magnetic tape from the space shuttle* Challenger *as the tape soaks in a solution.*

The purpose of the distilled water was to slowly rinse the tape as the magnesium hydroxide was removed. The tape was then rinsed with methanol to remove the water and treated with the lubricant methyl silicone to protect the tape layers. Finally, the 350-foot-long tape was unwound, transferred to a new reel, and rerecorded on a fresh tape.

The recording showed that at least some of the crew members were aware in the final seconds that the shuttle was in trouble. The impressive fact about this tape-salvaging project is that the principle involved is no more complex than what you would encounter in an introductory chemistry experiment!

These processes fit into a broad class of reactions called ***oxidation-reduction*** (or ***redox***) ***reactions***. (The term "oxidation" was originally used by chemists to denote