UNIVERSIDADE FEDERAL DE SANTA CATARINA

PÓS-GRADUAÇÃO EM LETRAS-INGLÊS

**A CORPUS-BASED APPROACH TO ESP: EST VOCABULARY IN
INFORMATION TECHNOLOGY**

SARA GEANE KOBELINSKI

Dissertação submetida à Universidade Federal de Santa Catarina em cumprimento

parcial dos requisitos para a obtenção do grau de

MESTRE EM LETRAS

Opção Língua Inglesa

FLORIANOPÓLIS

Maio de 2005

Esta Dissertação de Sara Geane Kobelinski, intitulada "A corpus-based approach to ESP: EST vocabulary in Information Technology," foi julgada e adequada e aprovada em sua forma final pelo Programa de Pós-graduação em Inglês para obtenção do grau de

MESTRE EM LETRAS

Opção Língua Inglesa e Lingüística Aplicada

....
Dra. Mailce Mota Borges Fortkamp
Nome do coordenador(a)

BANCA EXAMINADORA:

....
Dra. Maria Lúcia de Barbosa Vasconcellos
Orientadora e presidente

....
Dra. Lêda Maria Braga Tomitch
Examinadora

....
Dr. Lincoln Paulo Fernandes
Examinador

Florianópolis, 23 de maio de 2005.

ACKNOWLEDGMENTS

No work is done alone. Sometimes we seem so lucky or maybe we are just blessed, when there seems to be too much to do, along come friends like the ones to whom I have to say my thanks. Their efforts and friendship make a great amount of things possible in my life. Thus, I certainly have reasons to thank two institutions and a lot of people.

Firstly, I would like to thank the institutions which made possible this improvement in my career: Universidade Estadual do Centro-Oeste (UNICENTRO) in Guarapuava, where I teach, and Universidade Federal de Santa Catarina (UFSC), where I took the MA course.

Secondly, I thank my advisor, teacher Maria Lúcia de Barbosa Vasconcellos, who kindly agreed to guide me through this thesis and encouraged me to reflect throughout the process. I also thank her for letting me discover the bright and sensitive human being she is.

I would like to thank teacher Marco Antonio Esteves da Rocha who allowed to use the material which he compiled for his classes: Corpeng. I am also indebted to him for introducing me to Corpus Linguistics research.

I would like to thank the students who agreed to participate in this study. I am especially in indebted to Caroline Hagemeyer, the teacher who allowed me to enter in her classroom.

I would like to thank my daughter, Andréia Lins, for encouraging me, and for sharing with me this important moment in my life. Her willing heart and simple attitudes made me proud of her. My father (in memoriam), my brother Michel and my sister-in-law Andréia for their warm thoughts that are always with me. I would like to

thank my mother for her generosity.

I also would like to thank:

Lidia Stultz, Célia Regina Viganó, José Osvaldo Sampaio Bueno, Debora de Rocco, Neide Garcia, Mônica Godart, for their enlightening words. Specially, Sanae Morisawa, for her sincere friendship and solid support. My colleague Raquel Cristina Mendes de Carvalho for her sympathetic friendship. I also thank them for sharing their views on language learning/teaching during the MA course, and for the suggestions made to this thesis.

Lincoln Fernandes for comments on the method of this thesis and Marcia P. Carazzai for her insights into grammar.

I will be forever indebted to Terezinha Diniz Biasi, Dalila O. L. de Oliveira, and Natália Gomes Thimoteo. They are very special human beings whose kind gestures and attitudes made all the difference.

I am also thankful to Sonia Gotlieb, Josiane Marques and my colleagues, whose simple gestures make it easy for me to say a sincere thank you.

All those, and that for sure means a great amount, who directly or indirectly contributed to the development of this work.

ABSTRACT

**A CORPUS-BASED APPROACH TO ESP: EST VOCABULARY IN
INFORMATION TECHNOLOGY**

SARA GEANE KOBELINSKI

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
2005**

SUPERVISOR: Dra. MARIA LÚCIA DE BARBOSA VASCONCELLOS

The aim of this research is to contribute some ideas to the investigation of vocabulary instruction. In the 90's, Applied Linguistics emphasized the importance of vocabulary teaching since grammar and lexis were understood as inseparable concerns. Nowadays, specialized vocabulary is studied as a "sublanguage" or English for Science or Technology (EST). This sublanguage, which is present in technical contexts, is part of students knowledge, who study English for Specific or Academic purposes (ESP or EAP) while enhancing their skills. In TEFL, the tradition, was to teach grammar and language structures. Typical pedagogical materials for ESP present some limitations due to the following: (i) they use unauthentic texts; (ii) they are not constantly updated, since they are printed and still based on grammar. In an attempt to contribute to remedy this situation, some methodologies are appearing in Corpus Linguistics (CL).Three approaches which make use of CL methodology (Lexical Syllabus, Lexical Approach, and Data-Driven Learning) were applied in one specific unexplored area: EST vocabulary. In this context, the following research questions are addressed: (i) How does EST vocabulary behave in some IT texts of Linux guides? (ii) What are the advantages of using CL methodology to produce ESP materials in Information technology? (iii) How do learners act in response to corpus-based activities designed focusing on vocabulary instruction? The research methodology includes corpus compilation, study of language standards retrieved with WordSmith tools (Scott, 1996), elaboration of pedagogical material supported by corpus-based approaches as well as the application and analysis of the developed material. The most important results pointed out: (i) Lexical combinations are context dependent, (ii) the design of a satisfactory material for teacher and learners, (iii) corpus-based exercises assumed a different position.

Keywords: TEFL, Corpus Linguistics, ESP, EST vocabulary and pedagogical material.
Number of pages: 135

RESUMO

ABORDAGEM BASEADA EM CORPUS PARA ALUNOS QUE ESTUDAM INGLÊS COM PROPÓSITOS ESPECÍFICOS: VOCABULÁRIO DE INGLÊS PARA CIÊNCIA E TECNOLOGIA EM ANÁLISE DE SISTEMAS

SARA GEANE KOBELINSKI
UNIVERSIDADE FEDERAL DE SANTA CATARINA

2005
PROFESSORA ORIENTADORA: Dra. MARIA LÚCIA DE BARBOSA VASCONCELLOS

O objetivo desta pesquisa é contribuir com algumas idéias para a investigação de ensino de vocabulário. Nos anos 90, a Lingüística Aplicada enfatizou a importância do ensino do vocabulário desde que gramática e léxico foram entendidos como conceitos inseparáveis. Atualmente, vocabulário especializado é estudado como uma "sublinguagem" ou Inglês para Ciência e Tecnologia (EST). Esta sublinguagem, que está presente em contextos técnicos, é parte do conhecimento dos alunos que estudam inglês com propósitos específicos ou acadêmicos (ESP ou EAP) enquanto desenvolvendo suas habilidades. Na área de Ensino de Inglês como uma Língua Estrangeira (TEFL), a tradição era ensinar gramática e estruturas lingüísticas. Contudo, material pedagógico típico para ESP apresenta algumas limitações devido aos seguintes fatos: (i) uso de textos não autênticos; (ii) não são atualizados constantemente por serem impressos, e ainda são baseados em gramática. Numa tentativa de contribuir para remediar esta situação, algumas metodologias estão aparecendo em Lingüística de Corpus (CL). Três abordagens que fazem uso desta metodologia (Lexical Syllabus, Lexical Approach, and Data-Driven Learning) foram aplicadas em uma área específica e inexplorada: vocabulário de Inglês para Ciência e Tecnologia. Neste contexto, as seguintes perguntas de pesquisa foram endereçadas: (i) Como vocabulário de Inglês para Ciência e Tecnologia se comporta em alguns textos de Análise de Sistemas dos manuais Linux? (ii) Quais são as vantagens de usar a metodologia de Lingüística de Corpus para produzir material pedagógico em Análise de Sistemas? (iii) Como os aprendizes respondem às atividades elaboradas que enfocam ensino de vocabulário? A metodologia de pesquisa inclui compilação de um corpus de estudo, estudo de padrões de linguagem submetidas à análise do programa WordSmith tools (Scott, 1996),elaboração de material pedagógico baseado em corpus, bem como, a aplicação e análise do material desenvolvido. Os resultados mais importantes indicam (i) a dependência de um contexto para combinações lexicais, (ii) a elaboração de um material satisfatório para professor e alunos, (iii) um posicionamento diferenciado em relação aos exercícios baseados em corpus.

Palavras-chave: Ensino de Inglês como língua Estrangeira, Lingüística de Corpus, Inglês para propósitos Específicos, Ensino de vocabulário técnico, desenvolvimento de material pedagógico.
Número de páginas: 135

TABLE OF CONTENTS

Chapter III – Method

Chapter IV – results and Discussion

Chapter V – Final Considerations

Appendices

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACPI.. Advanced Configuration and Power Interface

CG .... Control Group

CL.. Corpus Linguistics

DDL.. Data-driven Learning

EAP..English for Academic Purposes

EBE..English for Business and Economics

EG...Experimental Group

ELT... English Language Teaching

ESL.. English as a second Language

ESS.. English for Social Science

ESP.. English for Specific Purposes

EST...English for Science and Technology

EOP...English for Occupational Purposes

IT... Information Technology

LA... Lexical Approach

LS... Lexical Syllabus

SLA... Second Language Acquisition

TEFL.. Teaching English as a Foreign language

SEU... Survey of English Usage

TTT... Teacher Talking Time

NLP... Natural Language Processing

Chapter One

Introduction

> **Here is a work that (sic) is based not just on real language data rather than on intuitions, but on vast amounts of real language data. Stubbs, 1996**.

1.1. **Initial Remarks**

This investigation is carried out in the context of the research area of *"Aprendizagem e Ensino"* (Teaching and Learning) of the Post Graduate Program in English [PGI Programa de Pós-Graduação em Letras / Inglês], more specifically in the field of Teaching English as a Foreign Language (TEFL). In this context, it explores the area of English for Specific Purposes (ESP), with a focus on English for Science and Technology (EST) vocabulary. In doing so, this study is developed at the interface between EST and Corpus Linguistics (CL), which provides the methodological tools for corpus design, building, processing and analysis. Figure 1 below presents a general view of the composite theoretical and methodological framework informing the study.
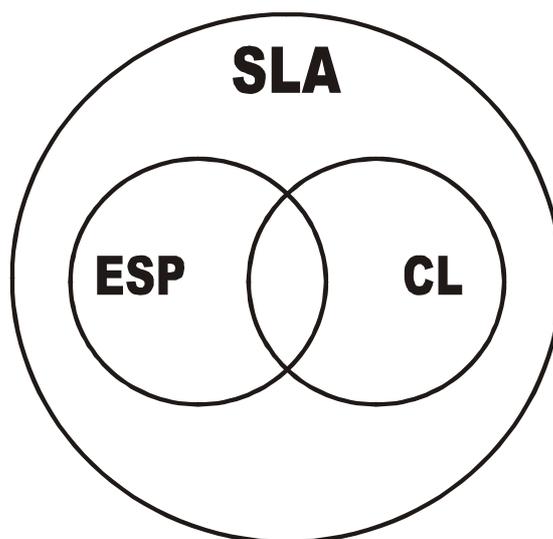

Figure 1: the composite theoretical and methodological framework for the

research

As the acronym indicates, EST refers to the general area of English acquisition for Science and Technology. The focus the present study is on vocabulary instruction in Information Technology (IT), as related to information disseminated via on-line resources. In order to explain how these different dimensions are integrated in the current research, the following section contextualizes the study and shows the relevance of each aspect of the composite framework within which the investigation is carried out.

1.2. Context of investigation

Investigation of vocabulary acquisition has been carried out for some time. Typically, studies in the vocabulary area concentrate on acquisition by means of communicative activities, the perspective on teaching vocabulary being "greatly influenced by the top-down naturalistic and communicative approaches of the 1970's and 1980's" (Söken, 1997, p. 20). Teaching activities typically include direct acquisition of vocabulary, use of dictionaries as well as the exploration of the four skills (e.g., speaking, reading, listening and writing).

As the English language dominates many scientific areas of our global knowledge, teachers of English as a Foreign Language (EFL) have come to realize that they need to know not only about the language itself, but also about the different domains in which this language serves as the means of communication and dissemination of information. This fact has led them to enter different areas of the teaching of English related to specific domains in which they are not necessarily experts that is, to English for Specific Purposes (ESP). During an ESP class, teachers are not in a position to provide an immediate answer to many of the questions put by students, as regards the

specialized field at stake. Teachers are then led to do research in grammars, glossaries, and pedagogical materials or in other sources, in their search for satisfactory answers. This effort sometimes seems to be insufficient, resulting in unconvincing answers and in frustration**:** such sources tend not to explore the topic thoroughly and, many times, they neglect both special difficulties students might have and the linguistic complexity of the texts themselves.

Glossaries, specialized dictionaries, textbook materials are some of the resources that scaffold the acquisition of ESP students' skills. Nevertheless, they suffer from limitations for a number of reasons: (i) glossaries and specialized dictionaries provide vocabulary for English Science and Technology (EST), but the definitions are either only translated to the students' first language or defined in the target language; (ii) the use of specialized dictionaries in ESP classes ends up as the only resource to deal with unfamiliar vocabulary in specialized fields (while it cannot be denied that dictionaries are useful tools, they are not the only source of information to deal with meaning, particularly in authentic texts, in which language is not facilitated for pedagogical purposes); (iii) text book materials are generally designed under artificial language samples, which it is not used for learners out of the classroom (See Chapter 5, section 5.4). Thus, such an artificial language can cause misunderstandings to ESP students. For instance, ESP learners are usually confused about words that have more than one meaning. In fact, it is a consensus in language teaching that the use of authentic material is beneficial to the learning process (Guariento & Morley, 2000).

The methodologies for vocabulary learning applied until now have some relevant drawbacks: (i) the slow process of acquiring vocabulary; (ii) learners' styles; (iii) the inference of wrong meaning to words, (iv) lack of results in long term vocabulary's retention; (v) problems in guessing words in context as a slow process.

Important as they are, these materials and methodologies fail to address *directly* the issue of how to deal with specialized sorts of vocabulary, in which not even native speakers are experts. For the purpose of teaching ESP students in the area of Information Technology (IT) research on specialized vocabulary must be further explored.

According to Sardinha (2004), today a new dimension added to the field of vocabulary acquisition is that based on electronic corpora and other contributions from Corpus Linguistics (CL). This new methodology offers many advantages to explore word meanings by means of empirical evidence, e.g. authentic texts. For instance, (i) it makes extensive use of computers in the analysis, employing automatic and interactive techniques; (ii) it is dependent both on qualitative and quantitative techniques (p. 38, my translation). CL allows for the emergence of a new concern with a well-organized terminology, as specialized languages are coming to be seen as a sort of "sublanguage"[1], which can be explored by CL methodologies. The term sublanguage can be explained as the specific language a community of specialists uses in discussing a restricted domain as a form a "natural" language. Sublanguages differ from each other, and from the standard language, in terms of syntactic, semantic and discourse properties (Grishman & Kittredge , 1985,p.2)

For instance, the words "mouse" and "keyboard" are common words that acquired the status of a "natural" language in the context of IT, as well as the acronym "BIOS", which, for the specialists using it, is both technical and "natural". The behavior of words which formerly act as language and acquire the status of a sublanguage can be thus spelled out: (i) they have more than one meaning, thus their lexical combination is

---

[1] A sublanguage is a version of a natural language, which does not display all of the creativity of the latter. One essential feature that has been hypothesized for a sublanguage is that it will show a high degree of closure at various levels. Closure means that it is tending towards being finite, for example, it is possible to develop a simple list of sentence type rules for the language (McEnery & Wilson, p.148, 1997).

context-dependent; (ii) they do not behave as a sublanguage item, since they can be found in more than one context: sublanguage items are isolated in a specific context; (iii) they show degree of closure at various levels. Issues of words behavior in specialized or technical form can be studied by means of their occurrence in a specific context of use (See chapter 2, section 2.3.1)

1.3. Objective and Research Questions

The discussion carried out so far has pointed out that vocabulary acquisition is an important aspect of Second Language Acquisition (SLA) and word behavior needs to be placed in a context that considers real needs for ESP learners. Typically, textbook materials are not based on empirical evidence, as Mindt (1996) demonstrated. Thus, it seems to be necessary to develop corpus-based materials to stimulate ESP students' acquisition of technical lexis.

Valuable to note that an ESP teacher has to be aware that English in authentic contexts is a real need for ESP students, specially if the ability to read authentic second language texts is desired: a large vocabulary size then becomes necessary which can be acquired through corpus-based approaches. Furthermore, the main point is that some scholars as Swan (1985), Guariento and Morley (2001), Sinclair (1997) view text in language use (authenticity) as the most essential component of any course.

The current research highlights the fact that CL can provide samples of collected language use and empirical data for the production of pedagogical activities based on technical lexis for ESP students. In fact, CL is a linguistic field devoted to research by the observation of a large quantity of authentic retrieved texts, which are analyzed by means of software programs that can calculate, select, count and order them.

Authenticity - together with issues of collocation and the value of corpora - seems to permeate every discussion of corpora. Despite all the discussion, the term merits a more careful definition, which is discussed in the following statements.

A General consensus is that an authentic text is one created to fulfill some social purpose in the language community in which it was produced. According to Johns (1988 as cited in Petitclerc 1998, 8) authenticity has three attributes: (i) that of Script: speaking scripts or texts, which are written by and for native speakers and are also context dependent; (ii) that of Purpose: the text or corpus used by students should be of value to the students outside the classroom; (iii) that of Activity: what a student does with the text or corpus gives him/her language learning skills that are transferable outside the classroom. Although I acknowledge the value of the attempts above to define "authenticity", I share Taylor's (1994 as cited in Petitclerc 1998, 12) ideas that the concept is of an abstract nature, "authentic" being an abstract quality that depends on too many variables to be defined in clear cut terms. Thus, in this study I consider the three attributes of authenticity due to the fact that authenticity and corpora issues are intertwined concerns.

As the relation between CL and SLA is considered a new one and the application of CL for pedagogical purpose is in its infancy, the results of such an interface can be difficult to measure without a focused investigation on this issue. Within this context, the main objective of the current study is to observe whether samples of English language (authentic material) collected and analyzed by means of a corpus-based approach is an adequate methodology for the production of pedagogical materials. Corpus-based analysis provides data about EST vocabulary behavior in IT texts, which is the subject under study. In addition, it investigates learners acting in response to this empirical approach.

The Research Questions (RQ) are put as follows:

RQ 1 : How does EST vocabulary behave in some IT texts of Linux guides?

RQ 2 : What are the advantages of using CL methodology to produce ESP

materials in Information Technology?

RQ 3 : How do learners act in response to corpus-based activities designed

focusing on vocabulary instruction?

1.4. Significance of the research.

Despite the fact that there is a great amount of publications in CL and ESP studies, little research exploring teaching and learning technical lexis by means of CL methodology (such as concordances and collocations) is carried out. To help remedy this situation and offer some relevant contribution to the area, this study explores the possibilities of the use of Corpus in the design of materials for vocabulary learning in undergraduate courses for students of English as a foreign language.

Thus it intends to shed light on this area of research, particularly on the under/unexplored interface between Corpus Linguistics (CL) and English for Science and Technology (EST)/Information Technology (IT), by taking into consideration the necessity to offer a different methodology to Brazilian students so as to facilitate their learning of technical vocabulary, particular IT vocabulary. Assuming that a corpus-based approach can provide activities that enhance students' acquisition of a technical lexis, this study will hopefully contribute with relevant information to further studies in this area.

1.5. Organization of the thesis

This thesis is organized into 5 chapters. This first chapter introduced the research and its objectives. Chapter 2 discusses the theoretical framework as follows: firstly, TEFL and its relation to the ESP context and the role of pedagogical material are addressed; secondly, the issue of vocabulary as a sublanguage is explored in the EST scenario; finally, the main concepts and relevant studies carried out in CL are discussed, focusing on those devoted to language description and language teaching and learning, specially on corpus-based methodologies and the development of pedagogical material.

Chapter 3 discusses the research method: corpora, tools for data analysis and relevant information about the design of the study, namely procedures to select EST vocabulary, procedures to design corpus based exercises, needs analysis and participants. Chapter 4 focuses on the description and data analysis as well as in the research questions. Chapter 5 presents the discussion of the results, some final remarks and suggestions for further investigation.

Chapter Two

Theoretical Framework

> **Vocabulary instruction is fascinating, it can be done systematically, and its results are predictable.**
>
> **Cobb, 2002**

2.1. Initial Remarks

As this study interfaces with TEFL, ESP, and CL, this chapter is devoted to a brief overview of (i) TEFL; (ii) the ESP context and the related EST field; (iii) reading; (iv) Corpus Linguistics and Lexical Approaches. It also includes a brief review of the literature on issues that permeate corpus-based methodologies and that are intrinsically related to it.

2.2. TEFL

Focusing on Teaching English as a Foreign Language (TEFL), the beginning of this theoretical framework investigates Applied Linguistics in the development of TEFL.

According to Nunan (1989) the discipline known as Second Language Acquisition (SLA) emerged from comparative studies of similarities and differences between languages. These studies were conducted in the belief that a learner's first language (L1) has an important influence on the acquisition of a second (L2) (p. 87). Adult foreign language acquisition has been addressed in several studies in the last few years. An example is Cyrino's (1999) research, which addresses the issue of the difference between child and adult learning in the following terms: "Firstly, everybody knows that it is more difficult to learn an L2 in the adult age than in childhood years. Secondly, the

majority of adults never dominate completely the L2" (p.178). The author (quoted in Zolb, 1986, p.181) goes on to say that "[i] t is clear that the acquired L2, that is the interlanguage[2], reflects the acquisition of L2 grammar, which occurs before the acquisition of marked pragmatic-semantic aspects of L2"(p.8).

Studies carried out at the interface between TEFL (Teaching English as a Foreign Language) and Second Language Acquisition (SLA) have yielded interesting findings about acquisition and learning, particularly those related to the possibility of learners becoming as proficient in a second language as they are in their first. A now well-established distinction deriving from such studies is that between acquisition and learning. Discussing the distinction, Yule (1996,p.191) says that acquisition refers to the gradual development of the ability in a language, using it naturally in communicative situations, while learning applies to a conscious process of accumulating knowledge of a vocabulary and grammar language. Current SLA research orientations can be captured by the meaning of complexity. The dimensions of acquisition cover issues of organic rather than linear process. In time, for the purpose of the current research the terms learning and acquisition are used interchangeably.

Adult foreign language acquisition has been addressed in several studies due to the growing interest in adults becoming proficient in a second language, especially English. This is one of the reasons why many things have changed in recent years in the field of applied linguistics. Some research has been devoted to language acquisition and to approaches to language teaching. Such an interest is also an attempt to improve a field of studies known as English for Specific or Special Purposes (ESP). As regards this issue, Graves (1996) suggests that teachers need to be informed by research about their learners' objectives if the main goal is to turn learners proficient in English.

---

[2]The term "Interlanguage" refers to the language produced by learners at initial stages of learning, containing a large number of 'errors' which seem to have no connection with the forms of either L1 or

Graves (1996), tracing the evolution of ESP, highlights the so called "proficiency movement," which he puts in the following terms:

> "The proficiency movement, the concept and various models of communicative competence, the advent of ESP (English for Specific Purpose), the proliferation of methods of language teaching, and the diversification of the population of English learners have all provided the teacher with many more options to consider in deciding what will be the backbone of her course" (p.20)

In fact, technology and commerce increased the number of people who want to learn English since being fluent in this language is the key to be successful. Nevertheless, as technology and commerce are different poles, a clear cut distinction about specific reasons explaining learners needs are necessary to be made. To this end, the next section deals with issues related to the teaching of English for specific purposes - ESP, for short.

2.3. The ESP context

As the acronym indicates, ESP means English for Special or Specific Purposes. Other acronyms related to English for other purposes are found in: English for Academic Purposes (EAP), English for Occupational Purposes (EOP), English for Social Science (ESS), and English for Business and Economics (EBE). The acronyms are used to define studies in which the English language is used as a foreign language in specific environments. ESP is differentiated according to whether the learner requires English for academic study or for work training. Nevertheless, there is not a clear-cut distinction between the two poles. Hutchinson and Waters (1987) build the tree of ELT

L2. It is an in-between system used in L2 acquisition.

(see Appendix Three) in an attempt to exemplify differences that could account for the differentiated branches of ESP (p. 17).

According to Hutchinson and Waters (1987, p. 8) three reasons can be identified for the emergence of all kinds of ESP. First, the demands of "a brave new world after the Second World War," that is, the expansion in scientific, technical and economic activity on an international scale. Second, a revolution in linguistics: it shifted the focus of attention to the way in which language is actually used instead of describing rules of English usage (grammar). Third, focus on the learner, that is the educational psychology emphasis on the centrality of the learners and their attitudes to achieve motivation for better and faster learning.

For the purposes of the current study, the focus of attention is drawn from the statement above. In fact, the reasons mentioned are intertwined concerns in the sense that the way in which English language is used in the ESP field and the reason why focus should be on learners are issues that pervade all these related areas of study.

The discussion around the role of ESP started in the 1960's. Barber (1962, as cited in Hutchinson & Waters, 1987 p.8) wrote an article on the nature of Scientific English, published in early 1962. Swales (1985, as cited in Hutchinson & Waters, 1987 p.8) reviewed the article to state that: in the late 1960's and early 1970's expansion of research into the nature of particular varieties of English marked the beginning of ESP advent. According to the latter author, English for Science and Technology (EST) was a synonym for ESP, due to the nature of studies developed on that time, which did not distinguished between General English and ESP. The next authors' statements trace the distinction among all these studies.

Smöak (2001, as cited in Kavaliauskienë & Janulevièienë 2001) subscribes to Swales (1990) concluding that the way we view the field of ESP today is far different

from the way we viewed it in the 1960's. The fact is that in the 1960's ESP practitioners believed their main job was to teach the technical vocabulary of a given field or profession. Teachers followed a general English syllabus adding technical vocabulary to it.

From the conception that there is a distinction of language use between ESP and General English, authors attempt to clarify obscure points of language use in the ESP field. Kennedy and Bolitho (1984) notice that ESP teachers began to recognize the sub technical vocabulary the words and phrase that surround the technical words. Some studies highlighted ESP issues related to vocabulary namely, Applied Linguistics, studies in learner-centered teaching, discourse and genre analysis and linguistic corpora (see Widdowson 1981;Trimble 1985; Swales 1990, Dudley-Evans & Johns 1993 as cited in Kavaliauskienë and Janulevièienë 2001).

Evidence point out that students' needs gave rise to the distinction between General English and ESP. Robinson (1980) states that an ESP syllabus is based on a rigorous analysis of students' needs and should be *"tailor-made."* In addition, Willmot (1986 as cited in Hutchinson and Waters 1987) argues that an ESP or EAP course is characterized by a number of factors, such as students' needs, strict objectives, specific content and authenticity of material as discussed in Kennedy (1986 as cited in Hutchinson and Waters 1987), who carried out research by means of corpus analysis in the ESP context.

Taking part in this discussion, Vian (1999) concludes: "ESP should be seen as an approach not as a product. ESP is not a particular kind of language or methodology, nor does it consist of a particular type of teaching material. Understood properly it is an approach to language learning, which is based on (sic) learner (sic) need." (p.2).

As a matter of fact, it is difficult to come to an understanding of the nature and benefits of teaching ESP due to the fact that ESP involves many of the complexities of language, and hence of language learning. Nevertheless, what turns evident is that TEFL, ESP, and even EAP merit some special attention since they are different from teaching General English.

Smöak (2001) emphasizes four important aspects as regards ESP: (i) ESP is not teaching lists of technical vocabulary; (ii) assumptions and intuition about language use in ESP situations are probably inaccurate, (iii) needs analysis should include observation of the language use in context; (iv) materials should be appropriate and authentic

The aspects above as regards ESP are main drawbacks that do not obtain a consensus. In order to remedy this situation, the integration of corpus analysis in the ESP context is pointed out as a possible solution to achieve a desired consensus. Corpus analysis is related to the specific need for authentic material (Robinson & Willmot, 1984). As Kavaliauskienë and Januleviêienë (2001) highlight, we must be prepared to find out how a language is used in real world situations to teach that language. The authors further say that "[we] must be ready to develop courses that teach authentic language from many different fields, based on accurate needs analysis and appropriate materials and methodologies. [We] need to acknowledge the fact that much of the language that our students need will not be found in any course books or pre-packaged materials." (p.6)

Kavaliauskienë and Januleviêienë (2001) state that the scope of specialized vocabulary in teaching ESP is a primary goal, but this does not mean just adding lists of technical vocabulary to the syllabus. In fact, corpus analysis can explore language use in specific situations, as the teaching of ESP requires. Lewis (1993) highlights the

importance of students becoming familiar with the notion of "collocation" and concludes that there are specific types of collocations in ESP, which cause students' errors due to a lack of translational equivalence between the first language (L1) and the second language (L2). Teachers must help the learner become familiar with ESP collocations, and such familiarity will develop best when the learner is consciously aware of this tendency of words to go together.

According to Lewis (1993), there is a tendency among learners to translate any professional text word-for-word, e.g. they usually try to simplify most lexical phrases and to separate words. Thus, the role of teachers is to arouse students' awareness of the existence of lexical collocations. Both authors' conclusions are intertwined in the same point as the quote below shows:

> "Quite recently, integrating corpus analysis in the ESP context, computer analysis of the English language has revealed a widespread occurrence of lexical patterns in language use. Some researchers call them "lexical phrases or lexical items"; others prefer the term "multi-word chunks of just chunks" of language. Whatever the term, they are an important feature both in language use and language acquisition and offer advantages for language teaching, particularly for teaching ESP" (Kavaliauskienë and Janulevièienë, 2001, p. 5).

In fact, various studies, such as those by Ghadhessy and Resebery (2001), discuss the importance of integrating corpus analysis in the ESP context. Based on such evidence many scholars carried out studies in an attempt to provide this integration. An example is Mindt (1996) and others that have built corpora of particular varieties of English with the aim of exploring these corpora in language teaching for specific purposes.

According to Mindt (1996), "such corpora can be used to provide many kinds of domain specific material for language learning, including quantitative accounts of vocabulary and usage which address the specific needs of students in a particular domain more directly than those taken from more general language corpora". (p.15)

In fact, Kavaliauskienë and Janulevièienë (2001) and Smöak (2001) state that there are other dimensions to vocabulary knowledge that should be considered mainly as regards the extent to which a learner knows the combinator possibilities of a word. For any given word, a native speaker also knows a range of other words, which can occur or collocate with it. According to the authors, this is an aspect of vocabulary knowledge that has until recently been largely ignored.

Thus, drawing upon the theories based on ESP it is important to realize that corpus-based research presents a challenge to the traditional approaches to teaching vocabulary to ESP learners in the sense that it confers another status to specialized vocabulary, namely that of sublanguage. This topic is discussed in the next section.

2.3.1. Vocabulary as a sublanguage and EST vocabulary

The aim in this section is to address the notion of sublanguage by focusing on insights into the nature of restricted language.

In his study on students' reading difficulties, Griva (2002) reported that a number of students ranked specialist vocabulary in the text as a problematic aspect of it. New discipline specific terminology was a major cause of difficulty for them. The research participants reasoned that unfamiliar terminology made them read the text considerably slowly. They declared that materials which are subject-specific in content met their needs although they found them difficulty and hardly accessible. The sub-technical vocabulary lies between General English and technical vocabulary of a particular

specialism and it is of great importance to ESP learners, as it is precisely this language that they need to communicate about their specialism Powel (as cited in Lewis, 2000).

According to Oliveira and Isquerdo (2001), in order to define a specialized vocabulary in the academic area, the lexis has to be simplified to a basis, which can articulate conceptual definitions. Nevertheless, this simplification is not a real practice in English Science and Technology (EST). For Bell (2003), "Usually EST (English for Science and Technology) vocabulary instruction is carried out through the use of dictionaries, glossaries and words lists, where EST words are merely either translated into the students first language (L 1) or defined in the target language" (p. 25).

Sublanguages have also been discussed under the term "EST vocabulary" as there is a subset of words in any language that is less frequent than ordinary words: They occur only into a specific domain. Such subsets are known as a sublanguage, a constrained variety of a language (McEnery & Wilson, 1996, p.179), which may be naturally occurring, although its essential feature is that it lacks the productivity generally associated with language.

McEnery and Wilson (1996) carried out a study to explore the possibility of the existence of a specific sublanguage. In order to do that, they explored the issue of sublanguage in IBM manuals. The authors stated that they built corpora with these manuals and contrasted these corpora with two other corpora available. They hypothesized that the IBM manual corpora show a potential sublanguage and the two other corpora do not due to the fact the latter corpora deals with different contexts. Their findings have practical benefits for Computational Linguistics and Lexicographers.

McEnery and Wilson (1996) state those sublanguages are also known as the languages of science. For this reason they have some specific characteristics: (i) they

represent a constrained subset of natural language; (ii) it is hypothesized that they are present in certain genres of writing; (iii) they do not display the creativity of natural language; (iv) it is hypothesized that they show a high degree of closure at various levels of description, e.g., it is tending towards to be finite. It is possible to develop a list of sentence type rules for the language (p.148)**.**

The discussion about sublanguages arouses in the EST vocabulary contexts, in which there does not seem to exist a clear and established position among scholars. Nation and Kyongho (1995, as cited in Bell, 2003), based on the criteria of frequency, coverage and range, state that words can be classified as general service, academic, technical and low frequency words. Agreements about a specific definition are not achieved properly.

While drawing upon these scholars, in the current study I use the term EST vocabulary to mean vocabulary with different meanings in a particular field (IT) be they highly frequent or not. This definition is close to Nation's (2001, as cited in Bell, 2003) categorization. I have adopted this position being aware that a corpus should be a resource for identifying and describing a sublanguage because it has one important feature – its finite nature.

In order to clarify different scholars' positions, the following table presents how issues of sublanguages as EST vocabulary are categorized by some authors (as cited in Bell, 2003):

Table 1

Definition of words categorized by different authors

Authors	Category one	Category two	Category three	Category four
Trimble (1985)	TECHNICAL: words with particular sense dictated by the subject matter.	SUB-TECHNICAL: high frequency words that occur across various disciplines and words that co-occur with special meaning in specific fields.	NOUN COMPOUNDS: Lexical units composed of two or more words that together form single meaning. They can be considered simple or complex	
Jordan (1997)	CORE VOCABULARY: neutral words that do not indicate degrees of intensity or formality and words that will only be expressive of a particular field.	ACADEMIC VOCABULARY: Words that occur across a number of disciplines.	TECHNICAL VOCABULARY: Words that are closely connected with learning the subject matter.	
Dudley-Evans and John (1998)	TECHNICAL VOCABULARY: general service words that have specific meanings in certain disciplines	GENERAL SERVICE WORDS: Those having higher frequency in a specific field.		
Nation (2001)	ONE words that rarely appear outside their field	TWO Words that appear inside and outside their field of study but with different meanings	THREE EST words that are accessible through their meanings outside the field but that are used the most inside the field.	FOUR words that can be easily found outside the field but are better understood by people who know the field.

As pointed out above, there are differences in the categorizations as regards a specific definition for EST vocabulary. Such differences are reflected in the language which is typical of the particular discipline. These differences are important and can be

investigated empirically by means of corpus-based studies. According to Lewis (2000) The solutions to the ESP teachers' problems are relatively easy (p.195) : The author suggests that a comparatively small corpus can be gathered in an attempt to address EST vocabulary instruction. It is also important to be aware of how such a vocabulary is learned. Thus, the next section deals with vocabulary acquisition.

2.3.2. Vocabulary teaching and learning

Research concerning vocabulary acquisition in second/foreign language and teaching methodologies for vocabulary acquisition has merited attention. Among main scholars, Meara (1995), Ellis(1985), O'Dell (1995) (as cited in Schmitt and McArthy,1997), Laufer (1997), Lewis (2000), Nation (2001), Cobb (2002), Bell (2003) are researchers that contribute relevant reflections on vocabulary learning. Their works are drawn upon in the present study to show how the process of vocabulary instruction can be carried out. For instance, the authors above discuss issues that go from main drawbacks of vocabulary acquisition to vocabulary collocations.

To start the discussion Cobb (2002) states that vocabulary learning has often been seen as peripheral to language acquisition, both theoretically and practically. The assumption was that vocabulary is too vast a quantity for direct instruction. In the beginning of the 21ST century, vocabulary acquisition has acquired a more important position in second language learning (Lewis, 1997, p.236). Thus, studies (for instance, Meara 1995, Ellis 1985, as cited in Schmitt & McArthy 1997), state that some of the major issues are related to (i) the slow process of students vocabulary acquisition; (ii) learners' styles; (iii) inferring wrong word meaning; (iv) no results in long -term vocabulary retention.

Nation (2003) - a researcher who has studied the incorporation of vocabulary learning into the EFL and ESL syllabus - and his colleagues carried out research from the 1970's until today in an attempt to solve problems in vocabulary acquisition. Their point is to show that vocabulary acquisition does not happen by itself to any satisfactory degree, particularly as needed for the first language literacy or second language generally (Cobb, 2002). Nation (2003) also explored computer driven lexical analysis and description. The author has provided relevant basis for several theoretical and practical developments (p.20).

Laufer (1997) highlights the fact that knowing a word implies knowing its (i) form (ii) structure,(iii) syntactic pattern,(iv) meaning, (v) lexical relation and (vi) common collocation. However, the general background knowledge of related acquisition vocabulary theories and practice is, somehow, incomplete as Bell (2003) puts it:

> "Over the past two decades, the teaching of vocabulary has gained greater importance in second language research. Recent research has benefit from the development of corpora of spoken and written language.... Yet, many questions remain unanswered, especially where English for Science and Technology (EST) is concerned. Approaches to EST have focused on different perspective" (p. 23)."

Among the different perspectives pointed out by Bell in the quote above, Söken (1997, p.321) focuses on how the learners build vocabulary and establishes some principles about the notion of learner independence as regards vocabulary acquisition. Along similar lines, O'Dell (1995, as cited in Schmitt & McCarthy, 1997) claims that the final goal in vocabulary teaching and learning is to allow students make use of their second language in any situation they wish.

However, little research on EST vocabulary instruction has been carried out. There are few studies focusing on non-native learners of EST attending scientific and

technical courses in their L1 (Bell, 2003. p. 31). How EST vocabulary can be a subject for instruction is regarded as a continuing investment of time and energy. This is the point where this research intends to give a contribution, in the following manner: (i) in the present study, EST vocabulary is retrieved from electronic corpora aided by a suite of software programs (see Chapter Three, section 3.3); (ii) the notion of sub-languages is drawn upon and applied for pedagogical purposes.

The next section provides some background information on how pedagogical materials address vocabulary learning issues. During the past 20 years, there has been a revolution in the way in which language can be studied. Because of the rapid development of computers to handle great large amounts of language data it is now possible to build language corpora which allow researchers to analyze how the language is being used at the present time, or indeed how it was used at particular times in the past. These new perspectives have a bearing on pedagogical material development, which is explored next.

2.3.3. Pedagogical Materials

This subsection aims to review some of existing pedagogical materials for vocabulary instruction. When it comes to pedagogical material for vocabulary instruction, two opposing situations occur: (i) in one of them, grammars, glossaries, dictionaries are used as ready made tools, and in many cases, disassociated from the real needs of individual ESP students and presenting either decontextualized or even invented language segments; (ii) in the other, authentic material is used, selected from electronic corpora related to the specialized vocabulary at stake.

In this subsection, issues of pedagogical materials used as ready-made tools are

addressed and then the use of authentic material is dealt with. As Sinclair (1991) pointed out (see Chapter 1, section 1.4), grammars are typically constructed under the open-choice principle. Nevertheless, there is a growing interest in changing this sort of approach to grammar. Throughout history the term "grammar" has received different definitions in the area of Linguistics (Carrazzai, 2002, p. 6). From the perspective that grammar can be understood in diverse ways, each scholar follows a theory or tendency. For example, there are descriptive grammars, prescriptive grammars, learner's grammar or interlanguage grammars and Pedagogical grammars. Due to space and time constraints, this section can not cover all the issues related to grammar. The focus of attention will then be pedagogical grammars. In this research, the view of pedagogical grammar interfaces vocabulary instruction, for the following reasons: (i) besides their use in language analysis, pedagogical grammars also contain metalingüistic terms and concepts, simplified in order to facilitate usage while helping learners achieve linguistic competence (Mohammed, 1996, as cited in Carazzai, 2002 p.7); (ii) this sort of grammar aims to provide teachers with "facts" of the language in a way that can be directly presented to students. In this research, "facts" of grammar are addressed as collocations that co-occur with EST vocabulary, while, "directly presented" means that lexical chunks (see, Chapter 3, section 3.4.2) are introduced to learners in an explicit fashion.

Having outlined this discussion about grammar, I will make some remarks concerning how grammar and vocabulary instruction are intertwined. At this interface, Willis (1993) and Tomlinson (1998) are the main authors drawn upon.

In Willis' view (1993), we should look at a number of different perspectives on grammar as different perspectives explore a specific interrelation between lexis and grammar. When it comes to ESP, a special situation happens: as ESP teachers cannot

teach all vocabulary learners need to know, for the simple fact that these teachers are not necessarily experts in the different areas they deal use in their teaching practice - learners go for dictionaries or glossaries to solve exercises, but this proves to be insufficient. Thus teachers are faced with the problem of teaching learners how to deal with new lexis or words with specific or specialized meanings. In this context, Willis (1993) argues that the teaching of vocabulary depends on the integration of lexis, grammar, and discourse, and that this can be achieved if lexis is taught through reading. It is suggested that traditional criteria for item selection might have to give way to new ones that allows specific learner needs to be more directly attended to. According to him, "We are not as helpful as we might be in making the bridge between the grammar and the lexicon" (p. 5) The author points out two reasons, at least, why the starting point should be on lexis: (i) grammar is likely to be more complete if we take words as the starting point: take care of the words and the structures will take care of themselves; (ii) the word provides the learner with tangible and recognizable starting point for the exploration of language. In the earlier stages learners can retrieve and organize a good deal of knowledge from a study occurrence of the word in, for example, in context, which are familiar to them (p. 56).

According to Willis (1993), lexical competence implies more than just knowing what a word means. It subsumes a number of other kinds of knowledge related to word meaning, for instance, word's combinations (collocations). Dictionaries are only one source of information that can be important in lexical acquisition, text corpora being another under-explored source. Other sources are encyclopedias, thesauri, gazetteers, collection of technical vocabulary and any other reference work or data base that is likely to contribute to a characterization of the syntactic and semantic properties of uncommon words and names (p. 88).

Tomlinson (1998) shares Willis' (1993) view in that both agree with the claim that to focus on the learner's language, more specifically on the lexicon, and on the grammar could provide valuable information to English Language Teaching (ELT) materials writers in the shaping of inventory resources available to learners.

Teaching materials and specially textbooks for EFL are unaffected by the findings of corpus-based approaches, as Mindt (1996) points out. However, learners should be given opportunities to use language for communication rather than just to practice it in situations controlled by the teacher and the materials. Using language for communication involves attempts to achieve a purpose in a situation in which the learners determine the content, strategies, and expressions of the interaction.

In section 2.2, exploring the revolution in Linguistics and focusing on learners and their acquisition of a specialized vocabulary for ESP, a new tendency was discussed: the tendency to use authentic material based on corpora collection to explore vocabulary learning and the application of a method which helps learners deal with vocabulary. This new tendency is extended to the reading context as the need to read authentic material is also brought to the fore. In this connection, the next section is devoted to a general overview of reading studies.

2.4. Reading

Research carried out by Goodman (1967) and Smith (1971, as cited in Xatara, 2001), based on reading in the mother tongue, was the basis of theoretical reflections from the 60's to the 70's. The authors claim that reading involves an interaction between previous knowledge and the information obtained from texts. In fact, they criticized the concept of reading as a simple game of decodification.

At the end of the 70's and in the beginning of the 80's, cognitive studies gave emphasis to the individual's knowledge in the process of understanding and learning. These studies were carried by some authors, such as Rumelhart and Ortony, (1977), Adams and Collins (1979), and Rumelhart (1980) (as cited in Xatara, 2001).

According to Braga (1997), those theoretical tendencies and the methodology applied to the teaching of a foreign language based on Krashen (1982, as cited in Xatara ,2001) advocated the importance of a learner centered methodology. As regards reading, this tendency had two consequences. Firstly, it made evident the active role of the reader during the reading process and secondly, it relegated to a secondary plan the formal aspect of target language (Braga & Busnardo, 1993).

Eskey (1988) demonstrates that effective reading is based much more on linguistic knowledge than on the guessing game proposed by Goodman (1967). According to Eskey (1988), learners whose attention is not drawn to the typical behavior of words are in danger of overlooking important meaning resources and failing to recognize markers in the text which enable faster and more efficient reading.

Carrell (1984) and Kern (1989, both as cited in Xatara 2001) claim that, in order to develop reading skills, the approach used needs to be centered in complex strategies. Such strategies emphasize the importance of text comprehension, previous knowledge, textual structure, or connectors named as cohesion and coherence. This set of strategies is called metacognition. They posit that metacognition plays an important role in reading as it involves both the conscious awareness and the conscious control of one's learning.

 From the conception of text as a kind of social cognition that organizes and constitutes the knowledge and with the development of new digital technologies, giving access to verbal interaction between two social actors, reducing distances and time, texts

- once digitized - seem to open space to the hypertext (Xavier, 2001), that is the text produced on a computer screen. Unfortunately, it was not possible to read hypertexts due to technical problems.

This brief overview of reading is necessary since participants in this research acquire vocabulary by reading processes in printed materials or on the computer screen as Data-Driven Learning (DDL) suggests (see section 2.5.3). In this scenario, this study aims to: (i) engage Brazilian ESP students in the complex process or significant interaction with authentic texts; (ii) establish contact with linguistic complexities in use.

The first aim above highlights the need for ESP students to be guided as regards to towards the acquisition of reading strategies devoted to academic texts and to be exposed to different text genres so as to overcome their difficulties in the acquisition of EST vocabulary. Braga and Busnardo (1993) claim that the teaching of foreign language reading needs to include comprehension and acquisition strategies, since it occurs in a precarious environment of acquisition, "precarious" meaning lacking basic pedagogical material, technological services or suffering from the absence of a fast paced environment.

The problem for successful learners is not that they make mistakes but that they are inefficient in their use of language as a result of the fact that their reading is imprecise. Nowadays it is possible to apply methods such as Data-driven Learning (DDL, see section 2.3.3) in order to read a hypertext. Thus, the issue of precarious environment needs to be explored in order to turn students into independent learners. To clarify issues of this new methodology, the next section provides some information related to CL and lexical approaches.

2.5. Corpus Linguistics and Lexical Approaches:

There are three relevant sorts of approaches to facilitate vocabulary acquisition, to provide authentic materials and develop to learner's independence. They are all related to the idea that many textbook materials have gaps in their linguistic description. These approaches are known as the Lexical Syllabus (LS); the Lexical Approach (LA) and Data-Driven Learning (DDL). In the following sub-section, CL is discussed before introducing corpus-based approaches.

2.5.1 Corpus Linguistics

In the context of the present study, Corpus Linguistics is seen as a methodology to teach Linguistics and foreign languages. It is a well-known fact that lexicographers have always based on Corpus in order to build the first dictionaries (McEnery & Wilson, 1996). The innovation is that by means of computing tools dictionary writing improved in quality as regards real life language, since they are supported by an accurate analysis of real data.

Nevertheless, the importance which Corpus Linguistics has been acquiring was criticized before the computing revolution. In the 1980's, due to the difficulties researchers had in analyzing amounts of text by ocular scan. The criticisms made to Corpus Linguistics were right, if one considers the time when they were made: exploration of corpora was tiresome, imprecise and time and effort consuming without the aid of technology. Despite the criticism, some linguists carried out studies dealing with corpus and tried to establish a balance between the use of the corpus and intuition (see McEnery & Wilson, 1996).

As Sardinha (2004) points out, although not particularly interested in corpus studies, some linguists have pointed out the probabilistic nature of language, which opens up room for the development of corpus-based investigations. For example, Halliday's theory highlights the probabilistic nature of language - as opposed to Chomsky´s theory of language is a possibility (Sardinha, 2004). In fact, Halliday is not a CL's researcher but his theory gave more credits for CL studies than Chomsky's did, as the latter author discusses probability in relation to linguistic systems as used in contexts, by real speakers. Probability is at the core of CL (see chapter 5, section 5.4.).

Despite the criticisms leveled against CL, some linguists insisted on corpus-based methodologies. For instance, Quirk (1960, as cited in Sardinha, 2004) and his followers planned and executed the construction of their corpus, known as the Survey of English Usage (SEU). Francis and Kucera (1961) began to work on the now famous Brown corpus, which took almost two decades to be completed, (Sardinha, 2004).

In the meantime, Svartvik, Quirk, Greenbaum, Leech (as cited in McEnery and Wilson, 1996) designed the Comprehensive grammar of the English language, based on SEU corpus and on the London-Lund corpus. The latter is a spoken part of computerized SEU, e.g., "an unmatched recourse for studying spoken English" (McEnery & Wilson, 1996, p. 95) .

From the 1980's, corpus studies - with the advances in technology - received new positive criticisms, especially as regards its empirical nature. The definition explaining the meaning of the term corpus was expanded, since it acquires extra meanings.

For instance, corpus can be (loosely) defined as *any body of text* or (most commonly) *a body of machine-readable text*. Nowadays, from the perspective of new advances in technology the definition is refined as *a finite collection of machine-*

readable text, sampled to be maximally representative of a language or variety"
(McEnery, & Wilson 1996 p. 177).

In accordance with the authors cited, Ball (1995) uses the term corpus in the
context of modern linguistics:

"In linguistics and lexicography, a body of texts, utterances, or other specimens
considered more or less representative of a language, and usually stored as an
electronic database. Currently, computer corpora may store many millions of
running words, whose features can be analyzed by means of *tagging* (the addition
of identifying and classifying tags to words and other formations) and the use of
concordancing programs (Chapter 3, section, 3.3.). Corpus Linguistics studies data
in any such corpus" (p. 5).

McEnery and Wilson (1996) point out that examples analyzed by means of corpus
methodologies are important in language learning, since they can provide empirical data
for learners to be exposed to. Corpus-based studies become an important source of data
for empirical linguistic research. In sum, corpus Linguistics deals with corpora
collection and exploration, which is collection of linguistic textual data that were strictly
retrieved with the purpose to carry out research of a language or a linguistic variety.
Thus, CL is devoted to language exploration by means of empirical evidence retrieved
through computers resources (Sardinha 2000, p. 46, my translation.).

CL collects and analyzes samples of language use. Thus the term authenticity
related to Corpus can be understood as sampling frame. Despite this fact, a corpus is
artificial, due to the fact that it is created for research purposes, e.g., it represents a
sample of language (Sardinha 2000, my translation). A new consensus that arises is that
Corpus Linguistics is its importance in teaching languages, since it can provide
empirical evidence of language use. To achieve the goals of each specific kind of corpus

exploration in language teaching, the corpus has to be planned and built on more specific basis of linguistic features and for specific purposes. The next section deals with the issue of corpora construction.

2.5.2. Building corpora

The use of corpora in second language teaching was investigated by Aston (1995-97), Barlow (1992-95), Burnard and McEnery (1999), Celce-Murcia (1990), Collins (1999) Flowerdew (1999), Gavioli (1997), Higgins (1991), Johns (1986 to 1993), Johns and King (1991), Leech and Candlin (1986), as cited in McEnery and Wilson (1996).

According to Rocha (2001) for the teacher who intends to take advantage of this methodology, a possible difficulty is the decision concerning the design of the corpus to be used to cover the teaching purposes in question. Corpus design implies the central need for the corpus to be adequate to answer the research question in the study at stake or to address the teaching goals to be achieved. The author highlights the fact that it is important to remember that the use of corpora in the context of teaching has no priority in linguistic research; priority is based on the target objectives designed for learning (p. 149, my translation). As Leech (1992, as cited in McEnery & Wilson, 1996, p.149) puts it: "We must understand what we are doing when we are looking in a corpus and building one. Corpus linguistics is, and should be, a synthesis of introspective and observational techniques relying on a mix of artificial and natural observation"(p.149). In this study, the focus of attention is the development of pedagogical activities by means of lexicographic and inductive grammar.

There are some basic requirements to build an electronic corpus. Firstly, the texts have to be written in authentic and natural language (not in artificial language as

mathematics annotation, for instance). Secondly, the texts produced for pedagogical or linguistic purposes are excluded, as they are not authentic or spontaneous texts. Thirdly, in order to build a corpus the main objectives have to be pre-selected and it will be related to the goals that the researcher needs to achieve. Finally, the corpus needs to be representative of a language, of a linguistic variety or of an idiom (Sardinha, 2000, my translation).

Corpora and corpus-based exercises are useful (Tribble, 1997), since they propose learning by discovery. On the other hand, the author states that some points can be considered. In fact, grammar, vocabulary, structure of discourse can be studied without a corpus. In the same vein, it is necessary to emphasize that language in use alone (authentic material) will not necessarily motivate learners to learn. It is necessary to teach taking into consideration an approach that explores the use of such texts. The next section provides information about corpus-based approaches relying on CL.

2.5.3. Corpus-based approaches:

Corpus linguistics interprets language as a probabilistic system, in which linguistic traces occurs with lower or higher frequency based on the use principle (Halliday, 1997, as cited in McEnery and Wilson, 1996). The discovery that language is standardized lexically and grammatically is based on statistical concordances (Jacobi, 2001,p.11).

Based on class evidence, Dave Willis (1990, 1993), Michael Willis (1993, 1997) and Johns (1986, as cited in McEnery and Wilson, 1996) explore, in the context of language teaching, the Lexical Syllabus (LS), the Lexical Approach (LA) and Data-Driven Leaning (DDL), respectively.

Table 2 shows the contribution of CL to vocabulary acquisition, though the compilation does not reach the materials writers and teachers working in the field of EST. The table is inspired by Nunan's proposal (1989). The converging points in the 3 (three) approaches are: (i) the theory of language is related to frequency of use based on linguistic research; (ii) the three approaches consider lexis as raw material to be researched by learners by means of inductive acquisition; (iii) the teachers' role is the same in all approaches. Although there is less agreement about theory of learning and the sort of activities - as each approach follows a divergent path - it is important to remember that there is not a clear cut distinction among them.

Table 2

Corpus-based approaches: Lexical Syllabus, Lexical Approach and Data-driven learning:

Theory of language	theory of learning	Objective	Syllabus	activities	learners roles	teachers roles	Materials Roles
Language is frequency of use.	Language use	Focus on tasks; fluency	Considers lexical issues as syllabus, Authentic texts	based on collocation	learner-centered, discovery, cooperative work	advisor	based on lexis, avoid invented sentences
Language is frequency of use.	Pre fabricated lexis reduces difficulties in learning.	Focus on, basically, form;	Considers lexical issues	"lexical chunks" (See 3.3.5, chapter 3)	Learner centered, discovery, cooperative work	advisor	based on inputs quality
Effective language learning is itself a form of linguistic research	Stimulation of inductive learning strategies	Grammar Teaching: By learners' discovery.	Inductive acquisition	discovery and analysis activities	driven by data, observation and analysis, discovery,	advisor,	taken by computer's resources or printed materials

2.6. Final Remarks

To summarize, Applied Linguistics, learner centered approaches and technologies are factors which changed TEFL perspectives. The key of perspective changing is made clear in the open-choice principle vs empirical data distinction, as pointed out above. The need to come up with and use interesting didactic materials has moved both researchers and teachers to the exploration of new prospects. These prospects are away from pedagogical materials based on Chomsky's view, as regards the use of invented sentences presents language in a facilitated form and exercises based on grammar alone.

The new perspectives have placed lexis and grammar issues in central stage. Here, technology plays a central role, as regards CL methodologies. Chapter 3, Method, will be devoted to the use the present research makes of the ideas discussed so far and method used in data investigation.

Chapter Three

Method

> **Few teachers have time or opportunity to design their own courses. In the majority of cases, a basic textbook is chosen and it is this which provides the practical classroom syllabus. Inevitably, teachers tend to follow the book, deciding in advance how long they can spend on each unit so that they will finish the book in a certain time. But the object of the course is to teach the students, not to finish the book.**
>
> **Lewis and Hill, 1999.**

3.1 Initial remarks

This chapter outlines the corpus-based methodology for the investigation of standards of language use in authentic texts and for the development of corpus-based activities for IT vocabulary teaching. For the purposes of this study, the researcher compiled a specialized corpus in order to investigate the behavior of EST vocabulary. "Specialized," in this context, means that the researcher combined, e.g., put together IT texts. The prediction is that compiled corpora are representative of EST lexical items used in the IT context. The purpose of corpus creation and method applied are all described in this Chapter. The following analytical procedures are used in the investigation carried out in the present study:

5) Selection of authentic texts from Internet open sources to build the corpus for the research;

6) Contrast of the corpus under study with a reference corpus - a corpus compiled by a group of researchers and used as examples in courses;

7) Analysis of the occurrence of EST vocabulary and their behavior with the aid of suit Software WordSmith (Scott, 1996), a computer program used as an electronic search tool;

8) Preparation of activities supported by corpus-based approaches;

9) Analysis of the results obtained from the application of exercises supported by

the corpus- based approach to English for Specific Purposes (ESP).

The activities supported by corpus-based approaches focus on the acquisition of technical lexis and are designed for the study of concordances and collocations retrieved by WordSmith (Scott, 1996).

To this end, the chapter is subdivided into sections devoted to specific aspects, namely: Corpus and method; Corpus creation; the use of the suite of software WordSmith Tools; procedures to select EST vocabulary and to design corpus-based activities; description of participants in the study; the application of pre- and post tests. Finally, a concluding section rounds off the discussion. These aspects are detailed below.

3.2. Corpus and Method

In the 50′s, CL criticisms were leveled against CL as at that time language study by introspection was the focus of attention. It was the open-choice principle at work (see chapter 5, section 5.4.). Here, Noam Chomsky's contribution (1957, as cited in McEnery & Wilson 1996, p.2) undermined the potentials of CL, since the author based linguistic studies on speaker′s competence instead of performance. The author′s theory was well accepted by many scholars. Thus, introspection instead of empirical methods was considered relevant.

It must be made clear, however, that the use of a collection of texts (corpora) in language study is not a new idea. In fact, language study by empirical observation has a long tradition in the linguistic field. McEnery and Wilson (1996, p.2) refer to studies before the 1950´s to emphasize the importance of non-electronic corpora in linguistic research. For instance, corpora of children′s language acquisition by their parents′ registers were an issue investigated by Stern (1924) and Thorndike (1921). They used a corpus of 4,5 billion words and published the results of their research as a study of word frequency.

In fact, present studies follow the process of classification used by the *Survey of English Language* (SEU(sic)). A result of research developed by Randolf Quirk (1959, as cited in Sardinha 2004) and his followers. The data consisted of compiled work organized in paper cards, containing analyzed and contextualized words from some texts, which were annotated under a grammatical order. Such an organization is a standard reference to computerized corpora.

However, a difference can be pointed out as regards the use of computer technology to explore specialized corpora, which allows for the development of relevant descriptions of collocations and lexical phrases, fundamental for pedagogical material. Here, studies based on electronic corpora show that electronic identification of the text is possible. Before the 1980's the results obtained from corpus studies were not completely reliable, since they were done by ocular scan and this process did not provide accurate results.

3.3. Corpus creation

Compiling a corpus means putting together or combining a number of texts into a

corpus, for one specific purpose. In compiling corpora, researchers take into consideration four important issues. Firstly, the language sampled (defined both quantitatively and qualitatively); secondly, the size of the corpus (a big corpus allows the possibility of finding rare words; a small corpus allows the possibility of text manipulation - e. g. tagging); thirdly, the sort of text included (language varieties have a bearing upon results); finally, the proportion of different text types (different use for words emerge in different text types).

The compiler has to deal with practical issues, such as: text collection, permission to use the texts, processing the material, conversion of format and use of the software relevant for the investigation at stake (for detailed discussion of such issues see [http://www.essexhttp://www.essexac.uk/linguistics linguistics)](http://www.essexhttp://www.essexac.uk/linguistics linguistics)

The corpus compiler and the corpus user can be either two separate entities or they can be combined in one single individual. In the current study, the researcher was both the corpus compiler and the corpus user.

3.3.1 The purpose of corpus creation and text collection

The corpus under investigation in the present study was created specially for the purpose of studying specialized language in authentic environments. The core motivation is linked with the assumption that the corpus compiled presents a degree of sublanguage (EST vocabulary used in IT area). Concerning the usefulness of research into sublanguages, McEnery (1992, as cited in McEnery & Wilson, 1997, p. 149) state that - with small finite lexicons and grammar - the task of language understanding may begin to seem manageable in some particular domain associated with the sublanguage.

The notion of sub-language here understood as EST vocabulary in IT texts -

informs the design of activities for Brazilian ESP students, made possible by means of corpus-based methodologies, expected to scaffold a conscious, responsible and collaborative learning. In this study, Nunan's (1989) definition of activities, as a part of work in the class that involves the learners in the comprehension, manipulation, or interaction with the target language, was adopted. Following the discussion, the next subsection explores the issue of permission to use the texts.

3.3.2. Permission to use the texts

The issue of copyright for corpus study is still "uncertain as it is difficult to say for sure how far scholars can go using of other people's intellectual property as a corpus" (Fernandes 2004, p.80). Despite this fact, copyright holders, in the current research, granted permission for the use of this documentation for educational purposes (on this issue, see Appendix 1: The General Public License (GNU(sic)) retrieved in the following site: http://www.linux.org/docs/ldp/howto/HOWTO-INDEX/copyright).

3.3.3. The corpora

This research makes use of two different corpora: the corpus of study - consisting of authentic texts in IT - and a reference corpus – consisting of different text types. The corpus of study was built from machine-readable Information Technology (IT) texts retrieved from Linux Documentation in the following sites: www.tldp.org/ldp/intro-linux.txt and www.tldp.org.documentation. Only one source was used based on two facts: (i) offer of free on line IT texts and (ii) participant's needs. The texts were saved in Rich Text Format (RTF). This procedure was adopted due to the fact that it is ideal

that the files be on TXT format, e.g., that the files contain only keyboard instructions as letters, numbers, and orthographic symbols (Sardinha, 2004 p. 51). Thus, the RTF format suits this pre requisite WordSmith use.

The corpus of study was compiled with written texts, covering the period between 2003 and 2004. As regards size, the corpus has one million words and consists only of manuals and guides of IT related to Linux. The texts are considered authentic, e.g. they are not facilitated for pedagogical purposes. As the objective of corpus creation was to study EST vocabulary, the corpus of study is considered a specialized corpus due to the fact that it contains only IT texts. It is renewable, that is, it is still possible to add texts to it, and it involves just one language, English. Thus, the corpus on which this study is based can then be classified as *monolingual*, *written*, *synchronic*, *balanced* and *specialized* (see section 3.2).

The composition of the specialized corpus is presented in tabular form below:

Table 3:

Extension of the corpus compiled for the study

Texts	text type	Size in terms of words
ACPI	Guide	61290
LDP	Manual	242831
Linux file system	Guide	397114
Linux file system hierarchy	Guide	83640
Scientific Linux	User´s guide	52974
True type fonts	Manual	5515
Videolan project	Manual	51041
VLC	Guide	87098
Intro to Linux	Manual	770
Linux win-nt	Guide	75286
TOTAL		1057559

The corpus of study can be considered a small-medium scale corpus as its size is 1056785 words (this classification is based on Sardinha′s proposal for corpus scale, 2004, p.26). The corpus of study in the present investigation was useful and satisfactory

for the purposes of the research. The activities were based on the corpora of study contrasted with the reference corpus. Data concerning both corpora used in this research are introduced in next table.

Table 4

Extension of specialized corpus and reference corpus

Extension	Corpus of study	Corpus of reference
Size	1056789	3856261

The reference corpus, which has approximately three millions words, was compiled by a group of researchers coordinated by Teacher Dr. Marco Rocha (UFSC). It was compiled following principles of corpus creation, from Internet open-sources and it is based on four specific genres: fiction, science, business, and spoken language. In fact, it is a reference corpus used as examples in courses, called Corpeng.

Scott (information retrieved on February 22, 2005 from http://torvald.aksis.uib.no/corpora_/2003-1/0545.html) states that the procedure to contrast a study and a reference corpus consists of a a method of finding out which words may repay further investigation. The same author explains that in order to process the key-words in WordSmith a reference corpus is necessary. The reference corpus usually needs to be bigger than the specialized corpus. "WordSmith checks to see which is bigger before doing the key-word procedure. But it′s (sic) a bit more complicated than that...″, says Scott (*ibid.*). According to him, such a procedure is designed to study texts, not genres or languages. Therefore, they can be used for collections of texts and they help locate lexical items whose frequency is unusual. The Keyword procedure is related word frequency in both the corpus of study and the reference corpus (see section 3.4). Section 3.3. The use of the software is further divided into three subsections namely: 3.3.1 wordlist; 3.3.2 Concord; 3.3.3 Keywords.

3.4. The use of the software WordSmith Tools

In this subsection, the tools used to deal with corpora compilation and procedures to retrieve the EST vocabulary are WordSmith version 3 (V.3) tools. "WordSmith Tools is an integrated suite of programs for looking at how words behave in texts. It runs under Windows 3.X,95,98, NT″ (see http://www.lexically.net/WordSmith/ for more specific details). This software, developed by Scott (1996) and marketed by Oxford University Press, consists of three sorts of tools which can be used to carry out linguistics analysis, namely Wordlists, Concord and Keywords. They are described in the next subsections.

3.4.1. Wordlist

As its name suggests, the Wordlist elaborates word lists from text files. The program is pre-defined to produce ‒ each time the wordlist is called upon to produce lists ‒ two lists of words; one in alphabetical order (identified by the letter "A″ in brackets); the other classified by the order of frequency of words (identified by the letter "F″ in brackets) with the most frequent words at the top of the list. Each of these lists is presented in a different way ‒ corresponding, respectively, to the alphabetical list (A) and to the frequency list (F). Together with such lists, the program also offers a third window (S), which contains statistics information referring to the data used for the production of the lists. In the statistics window (S), *types* and *tokens* are included, the former referring to the number of "forms″ of words or words and the latter referring to the number of "items″ or occurrences. Information about the number of sentences and

paragraphs are also presented in a third window.

3.4.2. Concord

Concord is the tool that produces concordances, e.g., occurrence lists of a specific item (called "search term" or "node" or "node word" or "target word") together with the co-text around it. The concordance line may come from the beginning, the middle, or the end of the texts. It may be made up of one sentence, part of a sentence or part of two sentences. Each concordance line in a set includes the target word, (e.g. the word being studied).

The target word is always in the middle of the concordance line. This means that when we study a word in a set of concordance lines we can see its context, in other words, the words which are used before it and after it, e.g. collocations. By analyzing a set of concordance lines, we can learn about how often target word investigate is used and in what environment it occurs.

There are other set of this kind of program which can make frequency calculation - files of sentences displayed on the screen. It is the "concordancer". According to Olohan (2004, p.199), concordances provide information about the sort of word which is displayed with the text either side that surround it, usually filling one line of the screen or window or approximately eight characters long. The focused word is called "node" and it is a result of a search presented as a keyword in context KWIC. Such an information of node is a set of concordances, giving evidence to some main collocations, as can be seen in the next table:

Table 5:

Exemplification of the node "patch" as a keyword in context (KWIC)

I was using an old	patch.	
Kernel you will need to	patch	it before
Kernel source files ACPI	patch	that exactly test version of ACPI.
Outlines how to	patch	a kernel for ACPI.
When you	patch	the kernel for ACPI.
ACPI kernel	patch.	
Kernel with an earlier	patch	from....

Collocation analysis can enrich lexical investigation since it introduces the researcher to a different context presenting the uses of a word in many situations. More importantly, if the students themselves discover the meaning of a word by observation it is possible that they will retain the information more easily.

3.4.2. Concord

Concord is the tool that produces concordances, e.g., occurrence lists of a specific item (called "search term" or "node" or "node word" or "target word") together with the co-text around it. The concordance line may come from the beginning, the middle, or the end of the texts. It may be made up of one sentence, part of a sentence or part of two sentences. Each concordance line in a set includes the target word, (e.g. the word being studied).

The target word is always in the middle of the concordance line. This means that when we study a word in a set of concordance lines we can see its context, in other

words, the words which are used before it and after it, e.g. collocations (see Chapter 1, section 4). By analyzing a set of concordance lines, we can learn about how often target word investigated is used and in what environment it occurs.

3.4.2. Keywords

Keywords is a tool that allows for the selection of items from a word list (or more then one list) by means of the comparison of its frequency with a reference list. For instance, a word will be a "keyword" if its frequency is either high or low in comparison to a reference corpus, usually a larger one. The purpose of this program is to locate and identify keywords in a given text. By means of this tool, it is possible to extract keywords from a corpus and then compute collocations of these keywords (for a detailed procedure for the extraction of keywords, see Sardinha, 2004, pp. 97 - 98).

The next section is further divided into two subsections namely: 3.4.1 Procedures to select EST vocabulary; 3.4.2. Procedures to design corpus-based activities.

3.5. Procedures to select EST vocabulary and to design corpus-based activities

The WordSmith tools above – wordlist, concord and keywords – displayed around ninety EST words as a result of procedures applied. I selected twelve as a parameter to design corpus-based activities in the current study, the criteria being the examination of items co-occurring across the specialized corpus compiled for the purpose of this study and the reference corpus.

3.5.1. Procedures to select EST vocabulary

Bell's view (2003) of the term "technical" or "EST vocabulary" as a special kind of

vocabulary composed of lexical units with specific meanings in a particular field – whatever their frequency - is adopted in the present research to select EST items. I examined the co-occurrence of the items selected across the specialized corpus compiled for this study and the reference corpus.

The criteria adopted to define EST vocabulary is based upon the proposal put forward by Trimble (1985), Dudley-Evans and John (1998) Jordan (1997) and Nation (2001), e.g. classifying EST vocabulary as different categories (see Chapter 2, section 2.3.3), as follows: (i) words that were retrieved in the text and at the same time were unfamiliar to the participants; (ii) words found in a specific text.

The procedures for the extraction of EST vocabulary from the corpus can be described as follows:

i) The corpus of study compiled was submitted to WordSmith Tools (Scott 1998) in order to produce a word list; the word list with the most and the least frequent words was the starting point to extract EST vocabulary (see section 3.2).
ii) The meaning of the EST vocabulary items found (around ninety) was investigated in contexts unrelated to information technology (the criterion for the selection of the EST lexical units was based on occurrence - see Chapter 2, section 2.3.3); Nation's (2001) categorization was adopted.

The twelve items are presented in Table 6 below:

Table 6

 WordSmith alphabetical EST units selected

Number of alphabetical order	Word	frequency	%
142	Boot	5	0.06
135	BIOS	9	0.10
1050	Buffer	6	0.15
148	Bug	1	0.01
314	Device	1	0.01
650	Kernel	14	0.02
789	Module	2	0.10
901	Patch	9	0.44
1012	Rate	38	0.08
1080	Root	4	0.16
861	Partition	5	0.09

| 810 | Shell | 14 | 0.07 |

The 4 (four) columns in Table 6 refer to: (i) numerical classification of the items as regards alphabetic order; (ii) selected words; (iii) general frequency of the items in the text; (iv) their percentage of equivalence related to how many times each one appears in the whole corpora.

During the investigation, the reference corpus was found not to have the same words selected; as a result, a third source – *Macmillan Essential Dictionary for Learners of American English* (2004) – with the same words was included, in agreement with Laufer's (1997, as cited in Bell, 2003) suggestion that: "knowing a word implies knowing its form structure, syntactic pattern, meaning, lexical relation and common collocation"(p.37). This dictionary was selected since it is also based on the Bloomsbury Corpus. The intention is to make students aware that these words have different meanings depending on the context in which they are studied, as Laufer (1997, as cited in Bell, 2003) points out.

In fact, around 37% are from the first category (words that rarely appear outside their field) (see Chapter 2, subsection 2.3.1 Table 2). These words were found to have low frequency of occurrence. Thus, it was predicted that they would not offer students any difficulty due to the fact that it has a specific number. Only one of these words was selected (BIOS).

Nation's (2001) categorization was adopted: (i) words that rarely appear outside their field; (ii) words that appear inside and outside their field of study but with different meanings; (iii) EST words that are accessible through their meanings outside the field but that are used the most inside the field; (iv) words that can easily be found outside the field but are better understood by people who know the field.

Nation (2001, as cited in Bell, 2003, p.32) suggested that, in addition to the

computer analysis, specialists in the field could make the selection. Thus, the lexical items selected were explored in the following sites in order to verify their specific meanings. These sites provided satisfactory evidences that the words selected were all EST vocabulary.

Http://searchsecurity.techtarget.com

Http://yourdictionary.com/dictionary.com/dictionary4html#

Http://whatis.techtarget.com/fileFormatA/0289933,sid9,00.html

3.5.2. Procedures to design corpus-based activities

A text was selected from the compiled corpus to design the corpus-based activities. The text was "ACPI: Advanced configuration and Power Interface", by Emma Jane Hogbin. Such a selection was based on the students' interest noticed by means of learner's characteristics questionnaire (see Chapter 3, section 3.5.1 and appendix VII), the number of EST vocabulary found in the text and the number of pages - the text was not too long, since the students' first task was to do the reading (see appendix 6).

In addition, principles upon which to base the design and application of corpus-based exercises were observed (See chapter 2, section 2.5.3.). They were inspired by a variety of sources, most notably Lewis and Hill (2001). Some strategies were considered during the application of corpus-based approaches. The strategies were:

> i) Be aware of all the different aspects of "knowing" words, and of the minimum that learners need to know about a word in order to use it correctly;
> ii) Notice that context is very important: Out-of-context activities are more likely to be useful if the students have originally met the words in authentic contexts as context provides a great deal information about meaning and usage;
> iii) Go from meaning to word when introducing new lexis to students;
> iv) Encourage learners' independence;
> v) Teach learners to use English learners dictionary;
> vi) Encourage students to keep "Lexical Notebooks" in which they record

the chunks;

vii) Encourage students to read widely for pleasure;

viii) Try to test learners on chunks rather than individual words when writing tests: this gives them the message that chunks are basic building blocks of language rather than an "extra", and encourages them to study in this way;

ix) Correct errors through reformulation;

x)   Be aware of Teacher Talking Time (TTT): It is the quality and not the quantity that matters;

The development of corpus-based activities takes into consideration a series of lexical activities and exercises based on Lexical Approaches presented by Lewis (1997). They do not differ from traditional ones (fill in the blanks, words combination, categorization). Nevertheless, the examples and themes to be explored make the difference. According to Lewis (1997) the *"input's"* quality is essential to the learning process. The six activities are briefly presented below:

"Identification of Chunks": This might enhance students learning and acquisition;

"Fill in the blanks": These sort of exercises are similar to combination as the collocation is related to a missing word, which is part of an expression or formulaic expression;

"Categorization": This sort of exercise is based on the principle of word classification or expression established by the students or by teacher, for instance: words with negative connotations, formal or informal expressions;

"Sequence": Exercises in which students have to organize in the most probable sequence;

"Elimination": Exercises with a word list in which the non-correspondent item must be eliminated;

"Combination": Exercises exploring the combination of expressions and formulaic expressions and their possible collocations;

The corpus-based exercises were named "Students' Materials" and were divided into two or sometimes three stages, namely: *vocabulary stud*y, *task one*, *task two and three.* In the Vocabulary Study section, the lexical chunks were introduced. In the first task the objective was to guide students through reflections, e.g., focus on tasks in order

to achieve fluency in reading, focus on form and grammar teaching by learners' discovery. In addition, the exercises were based on collocations or lexical chunks, as Willis (1990,1993) Willis (1993,1997), and Tim Johns (1986) propose.

Willis (1993) reflections inform the questions that I formulated in task one in the corpus-based materials. It is assumed that vocabulary teaching plays an important role in constructing the teaching of grammar, especially when it responds to specific learners' needs.

The second task deals with word's context. This sort of task was designed based on Laufer's (1997) claim that knowing a word implies more than recognizing it in a context. In the same vein, Nation (1982 p.20) suggests that a mixture of approaches should be adopted (see chapter 2, section 2.3.2). The point is that it is easier to remember a seven-word chunk than seven separate individual words. In addition, recent research indicates that learners need to meet a word or phrase about seven times in order to acquire it (Lewis, 2000). The majority of lexical words were found in the text selected for the reading activity.

In the materials designed, the twelve lexical items are presented as "lexical chunks", following Willis's (1993,1997) definition, e.g., any pair or group of words commonly used together. The exercises drew upon this format as can be seen in Appendix 4. They were inspired by similar corpus-based activities found mainly at:

http://eli.ubc.ca/teachers/lessons/reading

http://eli.ubc.ca/teachers/lessons/reading /technology /e.mail.htm#partA,

http://www.eli.ubc.ca/teachers/teach_lexical/advice.html#Awareness

http://www.eli.ubc.ca/teachers/lessons/reading/technology/email.htm#PartB

The reading exercise was based on Islan and Timmis (2004), as can be seen in Appendix 5. It was divided into three stages namely: "*Before you read*", "*As you read*",

and "*After you read*" (see Appendix five for the Reading Activity). The design of the reading activity was inspired by information found in the following site:

http://www.teachingenglish.org.uk/try/readtry/lexicaltry2pdf

The Reading activity and the Students' materials were handled by the researcher - with the teacher's permission - in ten fifty-minute classes for five weeks. Due to time constraints, part of the reading task was done as homework for some of the participants; they were in the Control group. In order to be more specific, the participants were divided in two groups namely: Experimental (EG) and Control Group (CG). In time, the next subsection is devoted to the participants' description and their needs.

3.6. Describing the participants

Participants in this study were thirty-two students from a State University in the first year of the IT Course. A placement test was applied in order to divide the group into subgroups. In addition, the University regulation offers a sufficiency exam for the students who are fluent in English and can thus be approved without attending English classes. In this exam, two students were successful.

The Experimental Group (EG) was composed of fifteen male students around seventeen years old. I myself was the teacher. The second group – the control group (CG) – consisted of thirteen male and two female students of approximately the same age.

Though the learners were not volunteers, the exercises they did were not graded as course marks. The classes were part of their curriculum. Data drawn from all learners who participated will be reported as Experimental Group (EG) and Control Group (CG). The EG was more fluent than the CG since in the latter group students were beginners,

according to the results of the placement test.

In the CG the corpus based approach was not completely adopted. Another teacher taught their classes with the aid of a course book. The corpus based-activities were applied in nine fifty-minute classes.

The EG group was submitted to nine fifty-minute classes to solve the corpus-based activities. In order to make students from EG familiar with corpus-based approaches, they were given exercises exploring methodologies other than those developed for the purpose of this research. Those exercises can be found in the following sites:

http://www.uefap.co.uk/uk/vocab/exercises/exercise

 http://www.er.uquam.ca/nobel/r21270/textools/webvphtml

http://www.1.2.3.208.224.131http://www.1.2.3.208.224.131

The next section describes the Needs Analysis questionnaires.

3.6.1. Needs analysis questionnaires

In the study, two questionnaires were designed to: (i) determine the characteristics of the learners participating in this study; (ii) verify learners' acting in response and receptivity to the corpus based approach informing the study (see Appendices 7 and 8). The first questionnaire contained questions related to learners' expectations as regards EST vocabulary. The second questionnaire aimed at capturing students' receptivity as regards corpus-based approaches; it is composed of three parts with closed questions, semi-open questions and open questions. All questions are related to the theories of language, theory of learning, activities, teachers and learners roles (see chapter 2, section 2.6). The latter questionnaire was designed by Jacobi (2001) who carried a corpus-based study in Spanish dealing with collocations. Jacobi's questionnaires were

adopted because there is some lack of research on how to evaluate the application of corpus-based approach for Brazilian ESP students in IT.

The Pre-test and the posttest applied are described in the next subsection.

3.7.   The pre- and the post test

The participants answered a worksheet related to EST lexical items found into the corpus of study. In order to know the average percentage of unfamiliar words for the participants, a worksheet was answered (Appendix 2). I have explored in pilot studies those words which were unfamiliar to the participants. EST lexical items that were not explained in the IT context or had a vague explanation of meaning were considered a mistake.

Students were found to make mistakes due to the lack of translational competence in finding equivalence between the first language (L1) and second language (L2) as Lewis (1993) pointed out. For instance, they had a vague idea about the meaning only related to IT context. According to Kavaliauskienë and Janulevièienë (2001, see section 2.3) there is another dimension to vocabulary teaching that should be considered, e.g. the combinatory possibilities of words. Thus, I decided to teach those twelve words despite the fact that some participants were, to some level, familiar with them. Thus, the pre test determined the target EST lexical items selected to design the corpus-based studies.

I have considered issues of sublanguage subdivided into categories (section 4.1.1). Thus, I have considered that it was relevant for the participants to recognize that the same word acquired new meanings in other contexts.

Collocates, or *"lexical phrases"*, *"lexical items"*, *"multi-word"* or *"polywords"*, concentrate an important feature in language use and language acquisition and offer

advantages for language teaching, specially for ESP learners (Kavaliauskienë and Janulevièienë, 2001). Thus, the selection from a different source of lexical items can provide substantial evidence for recognition of context. In fact, eight words from the twelve are introduced in other contexts. The lexical chunk chosen was taken from Macmillan dictionary (2004). The activity layout was based on Sandra Haywood (2003), in the following web page: http://titania.collins.co.uk

In order to verify whether learners have acquired the target EST lexical items after solving corpus-based activities some sort of material was needed to check the average percentages of their familiarity with the lexical items. In this specific task, I was inspired by corpus-based activities found on the Whatis website:

http://whatis.techtarget.com/fileFormatA/0,289933,sid9,00html.

I took the lexical units definition and prepared a "fill in the blanks" post-test (see Appendix 9). The post-test was applied without prior notice in two fifty-minute classes, after twenty one days – for the EG – and after twenty eight days – for the CG – after administering corpus-based exercises. I took the lexical units definition and prepared a "fill in the blanks" post-test (see Appendix 9). The post-test was applied without prior notice in two fifty-minute classes, after twenty one days – for the EG – and after twenty eight days – for the CG – after administering corpus-based exercises. The post-test was a " fill-in-the-blanks" exercise in which the EST vocabulary selected were the missing words in the blanks. Such words were specific definitions of EST lexical items.

3.8. Final comments

This chapter outlined the method applied in Corpus design and construction procedures for the analysis of the corpus retrieved data. The method complied with the ideas

underlying the epigraph beginning the chapter: it is time teachers had the opportunity (and qualification) to design their own courses and produce their own classroom syllabuses, which can be supported by corpus-based approaches to language learning and acquisition. The next chapter will discuss the findings of this research.

Chapter Four

Results and Discussion

Tell me and I forget, teach me and I remember, involve me and I learn.
(Author Unknown)

4.1. Initial Remarks

This chapter presents (i) a comparative analysis between the specialized corpus compiled for this study and the reference corpus selected; (ii) a study of lexical frequency in the specialized corpus, carried out with a view to verifying the behavior of words in specialized fields; (iii) a discussion of the pedagogical activities developed for word acquisition in IT; (iv) a description of relevant details in the application of activities; (v) some observations as regards students' receptivity to corpus-based methodologies; (vi) a discussion of the findings; and (vii) some final remarks.

4.2. Comparative analysis between the specialized corpus and the reference corpus

This subsection presents the results of the comparative analysis between the specialized corpus compiled for this study and the reference corpus selected for comparison purposes. Around 90 (ninety) EST lexical items were found when the two corpora were contrasted. The expectation was that some lexical items could co-occur in both corpora. However, this expectation was not confirmed: EST lexical items were present only in IT texts.

However, this "sampling frame", the ninety lexical items, was sufficient for the purposes of the current research. Sardinha (2004, p. 146, my translation) explains a

"sampling frame" as an operational definition related to the population. In the case of the present study, the population represents a sub-language in the Guides from Linux, thus constituting "the sample" of a specific language.

On the other hand, the reference corpus also represents a sampling frame in which unfortunately the same lexical items co-occurring were not found. Sardinha (2001) confirms this lack of co-occurrence as a complex activity resulting from two important issues. First, there is no minimum criteria universally acceptable for corpus compilation as regards selection parameters for the texts. Secondly, corpus representation is an issue under investigation, e.g. there is not an objective criteria to determine corpus representativeness strictly. Thus, the specialized corpus contrasted with other kinds of corpus other than the Corpeng (see chapter 3, section 3.2.3.) may produce a wordlist with items co-occurring in both corpora. Nevertheless, due to time constraints this procedure was not adopted.

The next section discusses lexical frequency in the specialized corpus so as to verify the behavior of words in specialized fields.

4.3. Lexical frequency study and the behavior of EST lexical items in specialized fields

Frequency is the relevant point here: Frequency analysis allows to determine which words are common and which words are rare. From the specialized corpus - with 1056789 words, and the reference corpus, with 3856261 words - ninety lexical items taken to be EST vocabulary is a figure which can be considered statistically low, in terms of occurrences. The researcher's expectation was to find more than ninety words. The next sections address a study of lexical frequency and EST lexical items in

specialized fields in an attempt to clarify the main reason related to the expectation that was not confirmed.

4.3.1. Lexical frequency study

The reason for the compiled specialized corpus having a low degree of EST vocabulary might be explained in terms of the sort of sub-language it contains. This sub-language shows a high degree of closure at various levels of description (closure meaning that a sub-language tends towards being finite). In this case, it is possible to study the most common collocations (a simple list of sentences type rules) for the sub-language (McEenery & Wilson, p.148, 1997). I agree with this statement, as some lexical items in this research have specific collocations. For instance, possible collocations for the lexical item " *root"* are " *file"* or " *directory"*, while the lexical item " *partition"* collocates best with " *Linux"* and " *Windows."* Thus, these lexical items have a degree of closure.

On the other hand, lexical items that accept prefixes, suffixes changing an initial structure also accept more collocations. The point is that for the lack of specific terms or representation, simple nouns are used (boot, bug, patch, shell, buffer). Thus, in the context of IT such lexical items assumed new meanings and prefixes and suffixes are added to them (boot, reboot and rebooting). At this point, they behave as nouns and verbs. Thus, from an initial form, such words suffered a changing in meaning, morphology and syntax.

Closure of words can also be verified in the occurrence of acronyms. In fact acronyms represent 30% of the lexical items. They comprise a complex meaning that requires co-referencing of one expression with its antecedents. The lexical item *"BIOS"*

- an acronym selected for the study - scored a good percentage in learning, (100%) which might be related to its specific collocations. Thus, at this point the good score in learning suggests that participants were found to notice the level of closure of this lexical item, e.g. its specific collocation. In other words, learners noticed the co-references and its antecedents.

EST vocabulary, which represents a sub-language, shows a low frequency and behave by changing its meaning, morphology, syntax and have specific collocations. It is relevant to teach ESP learners how to deal with this sort of vocabulary. Students need to understand how EST vocabulary behaves so as to learn them. Lexical combination is context-dependent, specially those with (i) low frequency and (ii) complex meanings that requires co references.

4.3.2. EST lexical items in specialized fields

Text genre is important to determine the sort of language used in the texts. Thus in order to search for EST vocabulary text genre was also considered a concern in the process of corpus compilation, as regards EST vocabulary frequency. The main concerns were (i)the sort of text genre that could concentrate high EST vocabulary frequency; (ii) to verify if this text genre is available to be compiled. According to Lee (2001), teachers and researchers need to know exactly what kind of language they are examining or describing.

There seems to exist lack of research on the kind of sub-language used in IT texts. Studies addressing text genre in the IT context were found in the area of Computational Linguistics, but they are related to areas of Natural Language Processing (NLP). Text genre is, somehow, considered an issue of sub-languages for CL methodology, though

not completely covered by CL. Thus, any other sort of text genre related to IT compiled for the specialized corpus in the current study may have unpredictable results.

On the other hand, there seems to exist an indication that Linux guides, the specialized corpus compiled for the study, has low frequency of EST vocabulary, which was not an impediment to work with EST vocabulary. In the sense that it was possible to notice collocations and design pedagogical material based on corpus-based approaches. The next subsection discusses the pedagogical activities developed for EST vocabulary acquisition.

4.4. Pedagogical activities developed for EST vocabulary acquisition in IT

After the study of the EST vocabulary in the compiled corpus and selecting target EST lexical items, the next objective of this research was the design of pedagogical activities with the selected EST lexical items. To this end, first, a reading class was planned and then corpus-based activities were designed. The design of both activities is based on Dave Willis (1990, 1993), Michael Willis (1993, 1997) and Tim Johns (1986) and on Lexical Approaches (see chapter 2, section 2.5.3). Both sorts of activities can be seen in Appendixes 4 and 5.

The pedagogical materials developed for the purpose of this study are divided into stages namely: Reading and Students Materials. The former was applied before the latter. Thus, the discussion about them take place in the two following subsections.

4.4.1. Reading tasks

The text selected for the reading tasks was a guide describing procedures to deal with Advanced Configuration and Power Interface (ACPI), written by Emma Jane Hogbin. It

is divided into three tasks, namely: *"Before you read", "While you read"* and *"After you read."* as can be seen in appendix six. The discussion about these tasks will be reported addressing the three headings above. Only comments will be made, no average percentage about the results is presented. This view was adopted due to the objective of this study, which does not focus specifically on reading issues. This reading activity was designed because   learners find EST vocabulary by reading materials and such vocabulary needs to be contextualized, as corpus-based approaches suggest.

While solving the first stage from the reading worksheet*, "Before you read,"* the learners discussed the issue and reached some conclusions about the exercises as related to the theme selected, i. e. Advanced Configuration and Power Interface (ACPI) (see chapter 3, section 3.4.2).They interacted in both groups, e.g. they shared information about the topic based on their empirical knowledge.

In the second stage, *As you read,* learners concentrated on the reading and some of them forgot to highlight the author's comment, especially in the CG. Finally in the last stage, *After you read,* as regards question four, which asked students if they had the same experience described in that text, some students reported that they had never had that kind of experience. Regarding question five, learners did not reach the same conclusions, some found it easy, others found it difficult in both groups. The last question asked for a summary of the process and they understood how to do the process.

Both groups complained about the text extension and about the language used in the text selected for the reading class. As the language used was not only academic, metaphors or comments proved to be difficult for the participants. Therefore, the amount of language that can be covered in the classroom will almost always be less than what the student meets or needs, as Lewis points out (2000).

As regards the issue of teaching of foreign language reading, I share Braga and

Busnardos's view (1993, as cited in Xatara, 2001): the teaching of reading foreign language needs to include strategies of comprehension and acquisition. In addition, it is essential that the teacher equips the students with skills which enable them to discover significant collocations by themselves, in both the language they meet in or outside the classroom (Lewis, 2000). In the present experiment, the most successful readers were found to be learners who used strategies in order to deal with their difficulties in vocabulary. This means that they were active during the reading process: they used reading strategies, e.g. they read the text as a whole, they established associations with previous knowledge, they highlighted unfamiliar words, they used dictionaries, they took notes, they hypothesized about the meaning or behavior of words through association.

As regards the difference in the level of proficiency in English, participants in both groups as effective readers. Echoing Eskey's (1988) statement (see Chapter 2 section 2.4), I view effective reading as the ability to notice the behavior of words. In this sense, it can be said that learners made inferences while doing the reading tasks. They were able to understand the text due to the fact that they used their linguistic knowledge regarding how to recognize "markers" in the text. The prediction is that their lack of knowledge concerning vocabulary meanings will be solved by practice in reading. The next subsection deals with the corpus-based activities designed for the purpose of this study.

4.4.2. Students Materials

As discussed in Chapter 3 (section 3.4.1), twelve EST lexical items were chosen as target EST lexical items to design corpus-based activities. The process of selecting

those words was based on students average percentage of unfamiliarity with the target words, e.g. words whose meanings they do not know. A pre-test determined the target EST lexical items selected to design the corpus-based activities (see Appendix 2).

While applying the pilot test I noticed that, as a group, the participants knew some of the EST vocabulary presented in the test with the twelve words. On the other hand, as individuals, they had a vague idea about the meaning of the lexical items. This study aims to explore the application of corpus-based activities to expand the learning of EST vocabulary. Thus, I decided to teach them as it seems to be difficult to find unfamiliar EST vocabulary that may be unfamiliar to all learners at the same time. Figure 7 presents the percentage of familiar lexical units from the perspective of the learners; then a discussion about the percentages will be carried out.

Figure 1

Percentage of familiar lexical items from learners perspectives – **EG** –**CG**
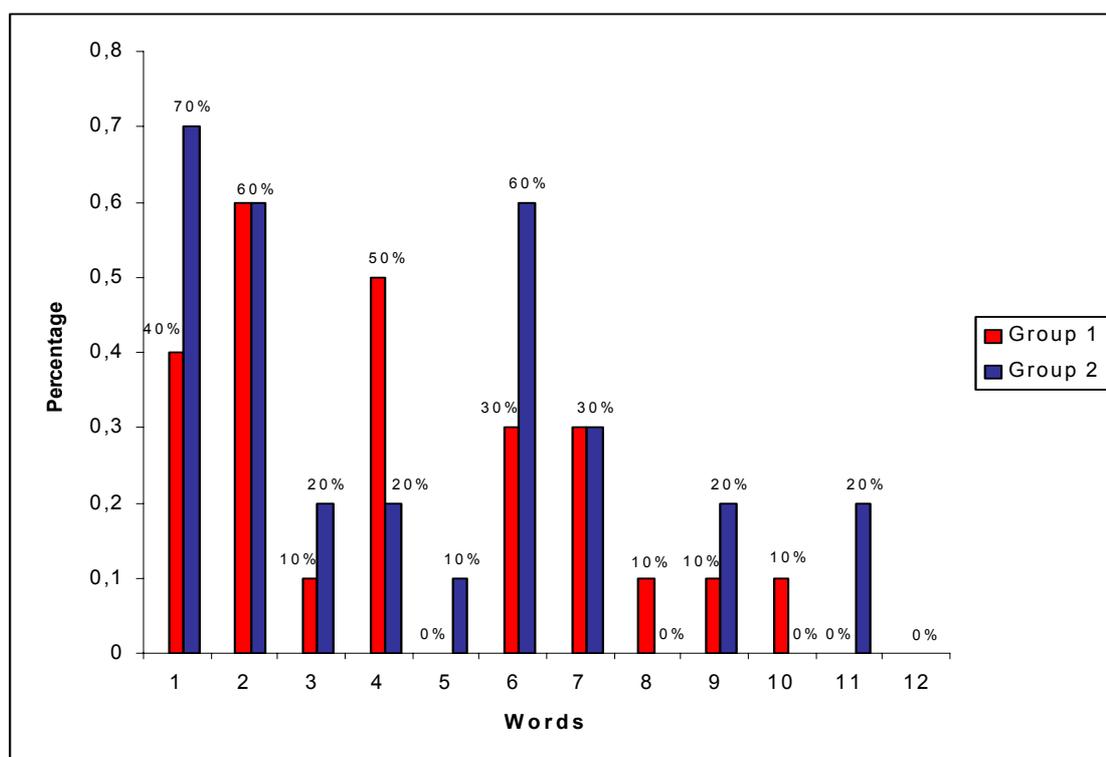
1	2	3	4	5	6	7	8	9	10	11	12	13
Words	boot	bug	module	Bios	Patch	Kernel	Partition	device	root	shell	buffer	rate

The percentages revealed that only one word was unfamiliar to all learners, since the word "*rate*" had a percentage of familiarity of 0%. On the other hand, it can be seen that the levels of unfamiliarity are in close proximity when the words "*module*", "*patch*", "*device*", "*root*", "*shell*", "*buffer*" are taken into consideration for both groups. The words "*BIOS*" (50%) and "*bug*" (60%) were the most unfamiliar items for the EG, while the words "*device*" (0%) and "*shell*" (0%) were the most unfamiliar for the CG. The CG took advantage in the words "*boot*" and "*kernel*" while the EG took advantage in the word BIOS.

The students in the EG were expected to have some advantages in comparison with the CG as the formers were not beginners in English. In average, the rate of unfamiliarity with words was around 92%.

After this study about the students' level of familiarity with the lexical items at issue, the application of corpus-based activities aimed to explore EST vocabulary and also related collocations. Participants formulated their own rules, while solving task one. For instance, one participant from CG formulated the rule as follows: "*Onde tem loop usa device.*" Equally important was the fact that they worked in groups, in a kind of focused interaction in which classmates or teacher answered questions. The introduction of the EST word in a different meaning, which was task two, presented two important aspects. Firstly, participants from both groups scored 88% while solving this task, which can be considered a high percentage of learning. In both groups, participants were astonished to notice that words had different meanings while attending a different context. They used dictionaries, consulted with classmates or the teacher in

order to learn the new meaning. Secondly, they were found to be able to recognize a different context, since they did it quickly.

A factor influencing these results in the process of learning or familiarization can be the exercise complexities. The rate of difficulty in learning is observed in the most complex exercises. On the other hand, less complex exercises showed a high average of learning. Nevertheless, as this kind of exercise is based on frequency of lexical EST words, the most frequent words gave the researcher more options to design activities while the less frequent did not. Thus, the word "*rate*", which is less frequent, had a learning percentage of 100% for the CG, while the word "*shell*" had a percentage of learning of 90% as can be seen in figure 2:

Figure 2

Percentage of familiarity after the application of corpus-based activities. **EG** – **CG**
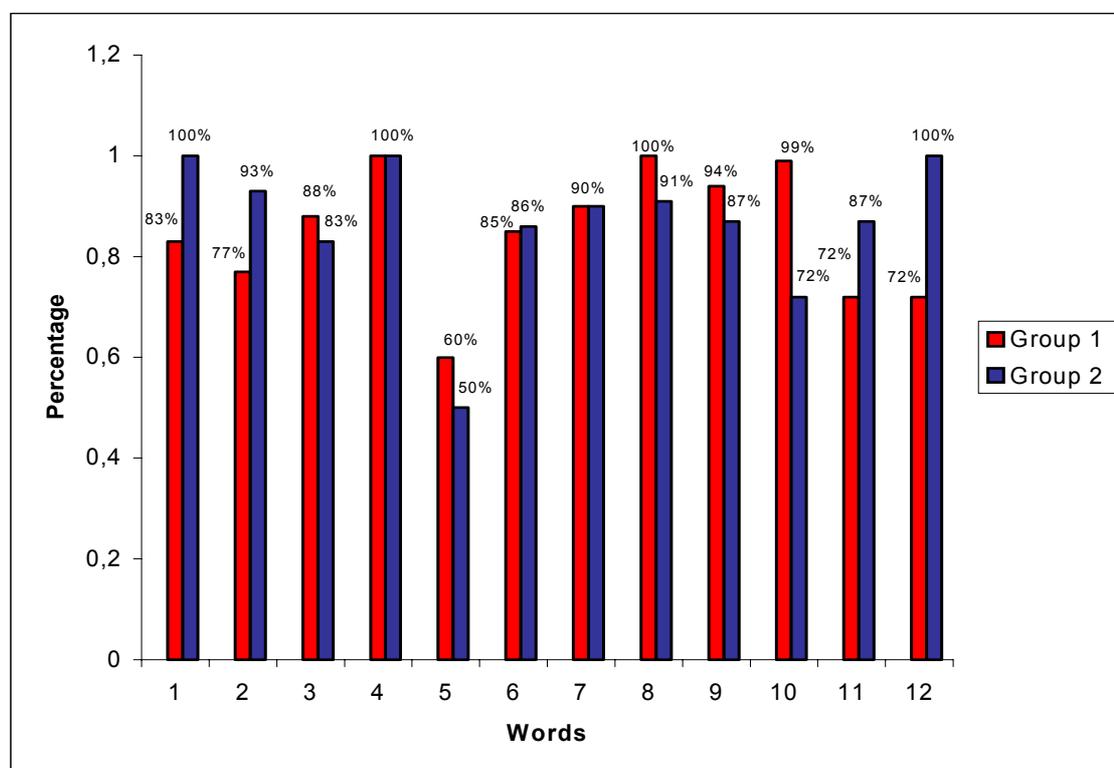
1	2	3	4	5	6	7	8	9	10	11	12	13
Words	boot	bug	module	Bios	Patch	Kernel	Partition	device	root	shell	buffer	rate

Participants were confused with one direction of the exercise (vocabulary study buffer) as the word "opposite" was in the title and also in the middle of lexical chunks. Nevertheless, the words in this specific instance can be considered to have had a positive score in learning (72%- EG and 87% - CG).

4.5. A description of relevant details in the application of activities

From both initial groups of thirty participants, only twelve from the EG and thirteen from CG attended the corpus-based exercises classes. This rate decreased, as eventual participants who did not take part in the whole study were not considered – as regards needs analysis questionnaire, reading class, students materials, the pre- and post written tests. Thus, data drawn from only twenty learners who participated were reported.

Participants in the EG were familiar with corpus-based activities. Thus, for the EG group the task was somewhat easy to solve, while for the CG group the same activity was found to be difficult; students, however, were able to solve the exercise after a brief explanation.

 A major concern after the application of students materials was to observe whether corpus based approaches promoted learning or acquisition of EST lexical units. To this end, a sort of post-test text was applied so as to explore such an issue (see Chapter 3, section 3.6.). The purpose of this post-test was to observe whether participants were able to recognize EST vocabulary in other contexts.

There were learners who could recognize the target EST lexical units in the post-test but were not able to match them to their correspondent picture appropriately. Nevertheless this issue takes into account that learners were expected only to recognize

the target lexical units in reading events and also to know their collocations in the particular context of IT.

When post-tests were taken as one single test,corpus based-activities promotes a high acquisition in comparison to EG (95.7%), while in the CG the average of learning decreased (63.9%). As the figure shows, more lexical units were acquires in applying corpus-based activities for the EG than for the CG. Despite the fact that in the CG the rate decreased, the result can be considered a positive outcome, for at least two reasons. Firstly, participants were not so fluent in reading because of their level of English and they did not use dictionaries while solving the post-test. Secondly, they were not taught by corpus-based approaches for some time as the control group was. In fact, participant three in the CG contributed for such a decrease as he scored only 16% of learning. Thus, taking this into consideration, the percentage for the control group is also positive.

The calculation of the results for the participants' outcomes considered only EST lexical units correctly placed into the related corresponding definition. Thus results were awarded by percentages and participants were reported as individuals. They solved the post-test by themselves, not in groups, and did not use dictionaries or annotations

After reading authentic texts, participants in both groups found reading easier in the post-test as it makes use of a facilitated language. In sum, EST vocabulary acquisition can thus be said to vary between the EG and the CG, the difference being 31.8%. Therefore, the corpus-based approaches could be said to be effective in terms of EST vocabulary instruction. The EG presented an advantage over the CG considering participants' different level of fluency in English. . Figure 3 presents their percentage of acquisition of the EST lexical items.

Figure 3

Percentages from **EG** and **CG**- the post test
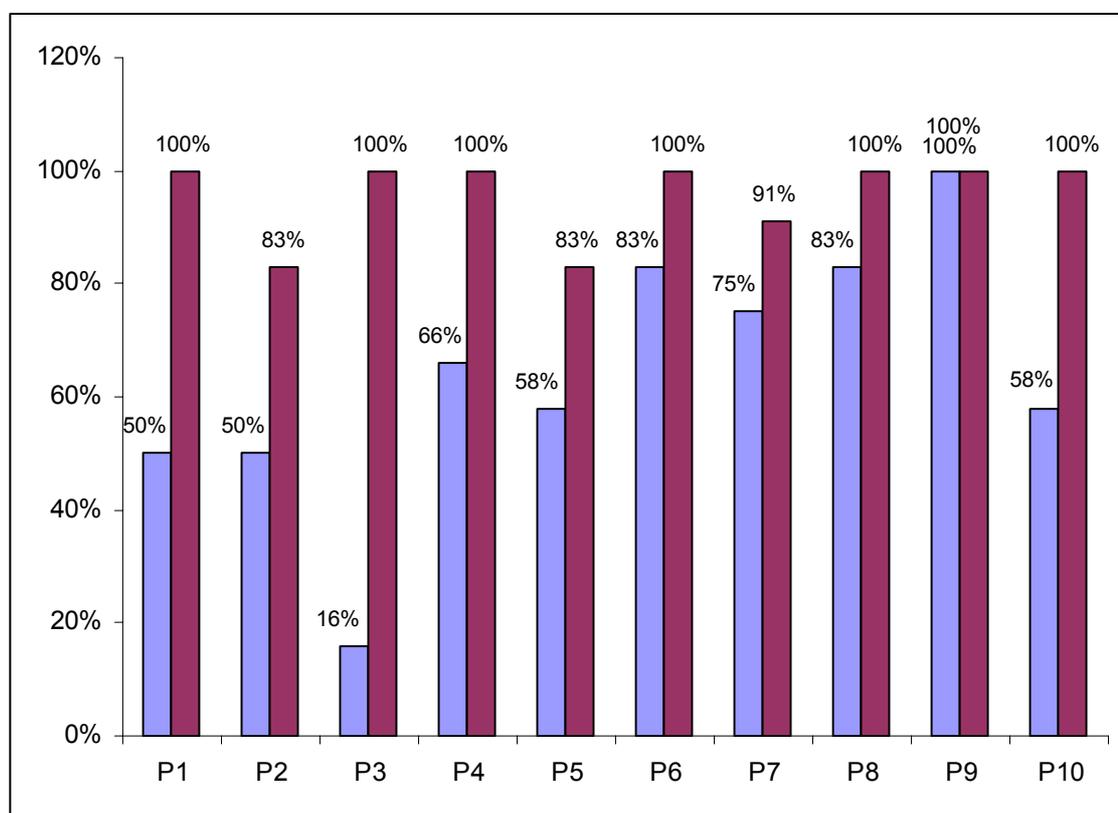
The next section presents a view of the students' receptivity to corpus-based approaches. Initially, it sums up the main purpose, objectives and findings of the study. Next, it discusses the results obtained speculating on some factors that might have influenced them.

4.6. Students' receptivity to corpus-based methodologies

The purpose of collecting data about students receptivity to corpus-based approaches was to verify in what senses CL methodology is an adequate methodology to produce ESP materials and what the receptivity to them is from the learners' perspectives. Data were analyzed quantitatively, that is they were analyzed in terms of learning outcomes.

The purpose of needs analysis questionnaires is to take into consideration learners' perspectives as regards the learning process. The objectives are to verify whether the participants were aware that they were submitted to a different sort of approach and to verify positive and negative factors in corpus-based approaches.

Despite the fact that students materials were based on EST vocabulary, grammar and collocation was taught. Thus, regarding grammar in the first part of the questionnaire, which had the open questions, the majority of learners (95%) reported not to like grammar. In fact, only one participant (5%) out of twenty was found to enjoy grammar. Nevertheless, when answering the second question, only one (5%) reported that grammar is unnecessary. Thus, they did not like grammar but they seem to know it is important.

In answering question four - about the nature and quality of the exercises - fifteen (75%) participants found them interesting, while two (10%) found them easy, against one (5%) that found them difficult. One participant (5%) reported that he did not understand anything, while one (5%) preferred grammar activities. Six participants (30%) did not like the exercises, as they spent a long time solving the tasks. In fact, they reported to prefer textbooks. Six students (30%) complained about incomplete sentences and four (20%) complained about the examples. Two participants (20%) did not understand the objective and two (20%) did not find the task interesting.

Eleven participants (55%) worked in groups and eight (40%) worked partially in groups or alone and only two (10%) worked alone. Despite this fact, ten (50%) did not know whether they liked to solve corpus-based exercises or not, eight (40%) participants answered positively, while two (10%) did not like the activity at all.

The application of the second page of the questionnaire - with the semi open questions - revealed interesting results. A great amount of participants indicated that they learned: (i) EST vocabulary and grammar, (ii) how to understand words, (iii) word meanings; (iv) how to use language; (v) interpretation; (vi) English and IT; and (vii) new contexts. Only two (10%) indicated that they learned unimportant things.

As regards skills, results revealed that students learned skills related to putting questions, reading by using dictionaries, practicing by discovering, making relations among words, working with classmates. Nevertheless, they reported that they learned more in groups. Participants (70%) felt that the teacher's explanation is important. However they also reported that it is necessary to discover by themselves. Thus, this statement suggests that they like to act as independent learners.

Regarding the teacher's role, the participants (60%) reasoned that a mixture of strategies are necessary. They felt that teacher can steer them towards the language research, which shows that they liked to learn by discovering for themselves. Despite this fact, few (20%) participants agreed that the teacher has to explain in order to save their time and efforts.

Learners also reported positive and negative factors in the activities. As negative factors half of them (50%) did not point out any. However, others (20%) illustrated as negative factors: (i) the fact that the activities were tiresome; (ii) the great amount of reading; (iii) difficulty in understanding the directions of the tasks; (iv) the level of difficulty as regards the words themselves; (v) the little amount of explanation from the

teacher; (vi) the reduced focus of the task (only twelve terms were focused); (vii) the small number of classes for corpus-based approaches. Nevertheless, I did not consider the last two statements as completely negative. In fact, they answered that it could be positive to learn more lexical units and have more Lexical-Approach classes.

Among positive factors, they emphasized: (i) learning; (ii) vocabulary expansion; (iii) updated and interesting texts that were compatible with IT;   (iv) reflective activities; (v) learning of English and IT; (vi) classmates' explanations (peer interaction); (vii) expanded knowledge in vocabulary and grammar. The next section discusses the finding.

4.7.   Research Questions

The more specific objectives of the current study were translated into three Research Questions (RQ), which I want now to revisit and answer. These three research questions dealt with EST vocabulary, corpus-based materials and students' response to corpus-based approaches.

4.7.1 RQ 1 - EST vocabulary behavior in IT texts
RQ 1 - How does EST vocabulary behave in some IT texts of Linux guides?

Students need to understand how EST vocabulary behaves. Sublanguages - here used interchangeably with EST vocabulary - will show a high degree of closure at various levels of description. The subdivision of EST vocabulary in categories allows noticing that closure is context dependent. Frequency is the key point. It was possible to notice a simple list of sentence type rules for the sublanguage as some lexical items in this

research had specific combinations. Those lexical items that accept changes in their form or structure - that is, those accepting prefixes and/or suffixes - were found to accept more collocations. Lexical combinations are context-dependent, especially those with few occurrences. In fact, collocations determined the degree of closure of related EST vocabulary. Acronyms represented a great percentage of EST vocabulary.

4.7.2. RQ 2 : production of ESP materials in IT

RQ 2 - What are the advantages of using CL methodology to produce ESP materials in IT?

Three important aspects as regards the production of corpus-based materials became evident in the study. Firstly, the possibility of the use of authentic material in class and of dealing with real language in a way that meets teachers' and students' expectations. Secondly, the usefulness of material designed for the participants/students - "useful" referring to participants' (i) use of the data to observe the behavior of EST vocabulary; (ii) involvement in the learning situation; (iii) learning how to deal with words; and (iv) learning grammar and vocabulary. Finally, as regards the teacher's opportunity to design her own course materials, the study demonstrated that this can result in satisfaction, both for teachers and learners alike.

4.7.3. RQ 3 : participants' response to corpus-based approaches

RQ 3 - How do learners act in response to corpus-based activities designed focusing on vocabulary instruction?

Participants were found to respond positively as regards corpus-based approaches. This

positive response is important in the sense that it revealed an attitude contrary to that observed in the participants in the beginning of the research. Corpus-based exercises assumed a different position especially for the reason that their starting point was the word-grammar. Learners' styles was surfaced in response to corpus-based approaches. Some participants were found to prefer to act as dependent learners, others as independent learners. Those participants acting in response to corpus-based exercises as independent learners asked for more corpus-based classes and pointed that they need to learn more EST lexical items than those twelve focused on the corpus-based exercises.

4.8.   Discussion of the findings

In this subsection I presented the data analysis. First I presented the analysis of a comparative study between the specialized corpus compiled for this study and the reference corpus selected for comparison purposes (section 4.2); secondly, I studied lexical frequency in the specialized corpus so as to verify the behavior of words in specialized fields (section 4.3); thirdly, I discussed the pedagogical activities developed for word acquisition in IT(section 4.4); then, I presented a description of relevant details in the application of such activities (section 4.5.); after that, I made some observations as regards students' receptivity to corpus-based methodologies (section 4.6); I answered the research questions (section 4.7) and, finally, the findings will be discussed. In order to be objective, the discussion of the findings will be addressed under the headings mentioned above.

The analysis of a comparative study between the specialized corpus compiled for this study and the reference corpus selected for comparison purposes highlighted the fact that EST lexical items were present only in IT texts. Linux Guides show low

frequency of EST vocabulary. A sampling frame demonstrates that EST vocabulary seems to have few occurrences.

Regarding the study of lexical frequency in the specialized corpus, to verify the behavior of words in specialized fields, the present study corroborates the assertions that: (i) a sub-language shows a high degree of closure at various levels of description also in IT texts; (ii) closure in IT texts means that this sub-language is tending towards being finite due to the fact that Lexical items have specific collocations. Nevertheless, this closure tends to be broken when some items changed their initial form or structure and as a result accepted more collocations. Lexical combination is context-dependent, specially those with few occurrences. Thus, it is relevant to teach how to deal with Lexical combination.

As regards the discussion of the pedagogical activities developed for word acquisition in IT, the reading tasks presented complexities for learners due to the fact that the authentic texts were long and language use was not easy for the participants. Nevertheless, after dealing with these complexities learners found IT easier to read other kind of texts.

In the description of relevant details in the application of corpus-based activities, findings refer to (i) the difficulty to search for unfamiliar words to all learners at the same time; (ii) the fact that participants are able to formulate their own rules and that they prefer to interact with classmates and teachers as the approach suggests; (iii) the fact that participants understood that words had different meanings while attending a different context and that this fact could be related to the comprehension of different collocations. For the same reason, they were able to recognize a different context as the results of the post-test showed. Results from the application of the corpus-based approaches proved to be an effective method to teach EST vocabulary; (iv) the fact

that Grammar of collocation was taught, despite the fact that participants did not like grammar they gave positive answers about the corpus-based activities.

Regarding students' receptivity to corpus-based methods the findings show that corpus-based methods were found to be effective to the teaching of EST vocabulary, as the majority of students answered positively to the approach.

4.9.   Final Remarks

Irrespective of the drawbacks and of benefits derived from the application of corpus-based methods to vocabulary teaching discussed above, this study had a very important result: it gave me the opportunity to design my own exercises supported by an effective methodology, in response to the assumption that different situations of teaching - as the ones faced by ESP teachers, ask for different materials, methods, activities and strategies and CL can be an efficient methodology.

In Chapter 5, I present a summary of the study, carry out a discussion and a comparison of the findings. The limitations and implications of the study, and suggestions for further research are also pointed out.

Chapter Five

Final Considerations

> The woods are lovely dark and deep.
> But I have promises to keep,
> And miles to go before I sleep,
> And miles to go before I sleep.
> Robert Frost

5.1. Final Remarks

The epigraph used to introduce this last chapter reflects what I feel in relation to having come to the end of this MA thesis: a sense of incompleteness. In fact, I realized that there is a great amount of research to carry out at the interface among ESP, EST, and CL. This feeling of incompleteness can lead me to the further investigating on the topic so as to shed light on the many aspects left unexplored. However, a seemingly paradoxical feeling coexist with incompleteness: that of achievement. I realize that at individual level I have come a long way, improving both as a researcher and as a teacher; at the academic level, I expect this study has contributed some insights to foster the field of vocabulary acquisition using corpus-based methods.

This final chapter is organized as follows. First, it sums up the main purposes of the study, its objectives, the research methods used and the findings obtained. Secondly, it discusses and the results obtained exploring issues of students' perceptions as concerns the application of corpus-based approaches in the teaching of EST vocabulary. Next, it looks at the limitations of the research in terms of the issues left under- or unexplored in the investigation. Finally, some suggestions for further research are presented, on the basis of the limitations of the study.

5.2. Summary of the study

The purpose of the study was to investigate the behavior EST vocabulary, the design of corpus-based materials and students' attitude as regards corpus-based approaches. CL informed this investigation.

Chapter one dealt with the composite framework, TEFL, ESP and CL, interfacing in the application of corpus based approaches to the teaching of EST vocabulary. After that, some considerations were made on the problems in the teaching of vocabulary instruction, especially from ESP learners' perspective. The issues of sublanguages were also discussed. Finally, CL was pointed out as an effective approach to investigate the EST vocabulary behavior in the field of IT.

In Chapter Two, the composite theoretical and methodological framework informing the study was presented and a brief overview was given of the fields explored, namely: TEFL, EST and the related EST field and CL.

In Chapter Three, the issues considered in corpus construction were discussed. These included procedures to collect corpora and for retrieving standards of language used in authentic texts; the preparation and application of activities from corpus-based methods; and, finally, procedures for data analysis.

In Chapter Four, the results obtained were presented and discussed. The application of corpus-based activities was considered and the research question were answered. Finally, by way of conclusion, this chapter highlights some of the more salient issues raised in the research, discusses the limitations of the study and touches upon potential future developments in the use of corpora in vocabulary acquisition in IT. In addition, this chapter presents a critical review of whole study.

5.4. Pedagogical implications

The current interest in vocabulary instruction is a fertile area for language researchers, materials writers, and teaching methodologies in this and in the next decades. In this area, descriptive interests inform the investigation into CL related to vocabulary. Such a descriptive interest is based on the desire to clarify obscure points found in monolingual and bilingual dictionaries and grammars, particularly those related to issues of authenticity. This sort of descriptive investigation had its origin in the teaching field.

In the sense that CL methodology provides corpora (empirical data) to support research dealing with descriptive interests, the term corpora "exploration" is applied. Nevertheless, corpora are also being used to provide data, e.g. samples of language use, by means of collocations. This is the turning point from corpora "exploration" to corpora "exploitation" (Sardinha, 2004, p. 262).

CL has potentiality to change the methods of teaching languages, due to the following points: (i) It informs materials production; (ii) its resources (e.g. Concordancers) are assuming relevant positions in the production of materials; (iii) it focuses on grammar and lexis placing both as inseparable issues named lexico-grammar.

The impact of CL upon language teaching has moved language study from the traditional introspective perspective to bases that are more empirical. An illustration of the impact of CL upon language teaching is the Collins Cobuild English Course, a series of pedagogical books published in the early 1990´s. The series was developed by Dave Willis (1990) and it makes use of concepts and methods of CL. It is known as Lexical Syllabus (Sardinha 2004, p.282, my translation).

A distinction which can be made along these lines is that proposed by Sinclair (1991, as cited in Schmitt 2000), between the open-choice principle and the idiom principle. The open choice principle underlies the idea that language is creative and in most instances, there is a variety of possible words that could be put in any slot. Traditional grammars tend to be constructed under this principle. On the other hand, the idiom principle has it that language also has a systematization that constrains vocabulary choices in discourse, constraints that the open-choice principle does not capture.

It is the idiom principle which can explain the fact that there are regularities in the way words co-occur. Such a co-occurrence is defined as "collocation" that can be explained as "words which are statistically much more likely to appear together than random chance suggests" (Lewis, 2000 p. 29).

Traditional pedagogical material designed to Teaching English as a Foreign Language (TEFL) are still based on the former principle. Thus teaching materials and specially textbooks for TEFL seem to be unaffected by the findings of corpus-based approaches, as Mindt (1996) points out. However, teachers do not need to wait for corpus-based materials to arrive into bookstores. The possibility to analyze language performance by means of a corpus was solved with personal computers, which allows storing and processing of a large amount of data without human error. By this process, it is possible to observe some special language occurrences and quantify the language thus predicting what the probabilities are by the observation of the existence of frequencies.

The point is that learners become aware of the uses of such words by seeing collocates, i.e. words that go together. The development of students' awareness is

supposed to make the learning process more experimental (Johns, 1991), in the sense that they turn into more independent learners.

This study investigated the behavior of EST lexical items in the context of ESP teaching in the IT field through a corpus-based approach. The research emphasized the need for supervision, coordination and application of activities as an essential stage in the development of corpus-based activities.

To teach ESP learners is not the same as to teach general English learners, as realia for ESP learners means authentic texts. Thus, I suggested that corpus-based methodologies could promote the use of real language for ESP learners.

An interesting fact resulting from the use of corpus-based approaches to the design of activities was that students learned the lexical behavior of items which were not in the texts explored during the classes. For example, the words "buffer", "partition" and "shell" were not in the text chosen for the reading class in the study; however participants learned their lexical behavior by studying their collocations. This fact provides support to two considerations. Firstly, they learned the words by being exposed to their lexical complexities. Secondly, students were not aware of what kind of words were difficult for them while reading the text. Thus, they need to be guided by strategies which focus on vocabulary instruction objectively as opposed to learn vocabulary accidentally. They need to learn what are the target words in specific contexts. Thus, the teaching of ESP vocabulary can be supported by corpus based approaches.

I subscribe to Nation's (1982) statement as regards the fact that learning occurs along a line or continuum with no clearly marked transition stages and that a mixture of approaches should be adopted. Thus drawing upon Bell´s (2003) statements that approaches to EST have derived from different perspectives, I suggest the corpus-based approach as an effective method to teach EST vocabulary for ESP learners. As the

findings show, activities exploring EST lexical items and concordances from this perspective seem to stimulate conscious and unconscious knowledge construction, turning learners into more independent language users.

This research contributes to demonstrate the advantages of the application of authentic material to EST vocabulary instruction in helping learners do systematic language investigation and thus improve their reading skills. The findings suggest that ESP teachers should focus on the issue of EST vocabulary as participants in this study manifested interest in learning more about lexical items together with their specific collocations in context. Thus, the activities carried out in this study were also developed with the purpose of demonstrating to ESP teachers that learning EST vocabulary can also contribute to improve reading skills.

A positive point deriving from the research is that corpus-based approaches allow active students' participation in the process of acquiring and discovering new knowledge. This fact is related to changing paradigms as regards teachers' role - now that of advisors, facilitators or coordinators as opposed to bearers of the information to be "transferred" to students. In this sense, students' learning becomes constant, dynamic and independent as they can learn how to learn, are actively involved in their learning process and come to enjoy being able to learn by themselves.

5.5. Limitations of the study and suggestions for further research

The study carried out suffers from some limitations, since - as common in research - some issues are not fully explored. Being aware of these limitations I discuss briefly the issues which were either unexplored of left untouched upon:

    i)   It is a well-known fact that it is difficult to determine strictly and objectively

the proficiency level of knowledge. Thus, the EST lexical items that learners were supposed to acquire could be learned during other classes focusing on those lexical words. It was possible that learners were aware of this fact. On the other hand, ESP learners asked for more lexical words and reported having their vocabulary repertoire expanded.

ii) As regards corpus extension, the one in this study can be considered a small-scale corpus, according Sardinha (2004). A larger and more varied corpus could alter the results: the field would benefit from a more systematic investigation on a large-scale corpus basis so as to allow a more complete investigation of EST vocabulary behavior.

iii) Another limitation is related to the texts chosen for the compiled corpus of study. No on line source other than Linux was found for the collection of a free corpus. A second or third source to compile a specialized corpus may contribute to the expansion of EST vocabulary study in IT. For instance, some fields of ESP expanded a great amount due to the fact that there are different sources to compile authentic material to explore EST vocabulary. While researching into the specific sites dealing with ESP learners, I observed that only a few of them explore the IT field. In fact, most privilege interfaces with business, medicine, engineering ,etc. and ignore IT.

iv) Some caution must be taken in extrapolating the results of this study since, as explained earlier, the number of learners participating in the study was limited, when compared to other research in the field of CL (see McEnery and Wilson, 1997). I am conscious that I have used a small number of learners and that they had different levels of language proficiency. Both factors may have a bearing in the interpretation of the results:

generalizations cannot be made. This is not the ideal situation to compare them. One solution would be to consider the time of exposure to corpus-based approaches: the CG had less time to be familiarized with the method. I did not expose them to the same number of lexical approach classes because I could not use more than ten classes for the research. Therefore, I would not be able to compare them negatively, since the EG was more exposed to the approach.

v)   Learners are familiar with ready made didactic materials focusing on grammar teaching, so to change the paradigm in teaching vocabulary was not an easy task. Some of them complained about time, pointing out that the exercise was not as practical as the one presented in other pedagogical materials, such as text books. Despite the fact that corpus-based strategies were used, not all of the students behaved as researches as the method suggested. For instance, not all of them brought dictionaries to the classes or took notes in notebooks. Thus, I suggest that in future research, some kind of classification can be considered as regards participant's learning styles as kynestesic, visual, mathematical as Psychoneuro Linguistics suggests (for example, Gardner, 1999).

In spite of the limitations pointed out, this study may serve as a starting point for further research integrating TEFL, ESP and CL. In this regard, I would like to suggest first that and investigation about EST vocabulary analyze whether other lexical items show a degree of closure and how collocations occur for them.

Second, as regards Corpus extension, a larger and more varied corpus could be compiled in order to carry out research allowing generalizations concerning the extent and range of vocabulary collocates; in addition, EST vocabulary collocations could be

also explored in other fields.

Third, it would be interesting to observe a greater amount of ESP learners in different fields (Chemistry, Business, Physics) from a corpus-based perspective in order to verify participants' response and the value of corpus-based approaches as an effective toll for instruction in EST vocabulary.

As a final word, I would like to point out that the study carried out tried to contribute the filed of EST as it proposed the design of pedagogical material through EST collocations from a corpus-based perspective. I hope the findings of the present research may lead us, ESP teachers and researchers, towards more reflection upon the sort of pedagogical materials employed to EST teaching in Brazil, especially in the IT field and upon the role of teachers in the learning process.

REFERENCES

Ball,C. N. (1995). Haet! A corpus-based hypermedia resource for Old English *Lexical Acquisition Language Learning & Technology.* Retrieved February 10, 2004 from: http:// web.bham.ac.uk/johnstf/biblio

Bell C.D.S. (2003)Investigating the effectiveness of two approaches to EST vocabulary: an exploratory study. Unpublished doctoral dissertation. PGI, UFSC.

Braga, D.B. (1997) Ensino de língua via Leitura: uma reflexão sobre a elaboração de um material didático para auto instrução. *Trab.Ling.Apl.,* Campinas, (30):5-16, Jul./dez.

Braga , D.B. & Busnardo, J. (1993) Metacognition and foreign language reading: fostering awareness of linguistic form and cognitive process in the teaching of language though text Lenguas Modernas 20, 129-149.

Borg S. (1999) *Studying teacher cognition in second Language grammar teaching.* London: Pergamon.

Carrazzai, M.P.(2002). Grammar and Grammar teaching: A qualitative study of EFL teachers' beliefs and behaviors. Unpublished master thesis PGI, UFSC.

Carrell, P. (1984) *Evidence of a formal schema in second language comprehension.* London:Language Learning (2): 183- 201.

Chomsky, N. (1957) *Syntactic Structures.* DenHaag:Mouton.

Coob,T.(2002) *Learning vocabulary in another language review of Paul Nation, learning vocabulary in another language.* Cambridge: Cambridge University Press.

Cyrino,S.M.L. (1999) Issues in second Language Acquisition. *Cad.Est.Ling.,* Campinas, (36):177-191, Jan/Jun

Dudley-Evans, T. & St John, M. (1998) *Developments in English for Specific Purposes.* A multidisciplinary approach. New York: Cambridge University Press.

Eskey, D.E. (1998). Holding in the bottom: An interactive approach to the language problems of second language readers. In P. Carrel, J. Devine, e D.Eskey(eds), *Interactive approaches to second language reading.* London: Routledge

Fernandes, L. (2005) Translating children's fantasy literature in Brazil: a corpus-based study. Unpublished doctoral dissertation. PGI, UFSC.

Frost, R. (1963) Stopping by woods on a snowy Evening.In: More poetry Please (1997): London:Orion

Ghadhessy, M., Henry, A.& Resebery R.L. (2001) *Small Corpus Studies and EFL-theories and practice.* Amsterdam/Philadelphia:John Benjamin.

Granger, S. & Tribble C. (1998) "Learner corpus data in the foreign language classroom:Form focused instruction and data-driven learning". In: S.Granger (ed.) *Learners English on Computer* (199-2009). New York: Longman.

Graves, K. (1996). *Teachers as course developers.* Cambridge: Cambridge University Press.

Greenbaun, S. (1987) *Reference grammars and pedagogical grammars.* World Englishes, Vol.6 no 3, 191-197.

Grishman,R.& Kittredge,R.(1986)*Analysing language in restricted domains: Sublanguages description and processing.* Hillsdale, NJ: Lawrence Erlbaum.

Griva (2002).*A needs of analysis project for the E.A.P.* Students of Florina Pedagogical School. Greece TESOL.GREECE/EAP SIG 2002

Goodman , K. S. (1967) Reading: A psycholinguistic guessing game. *Journal of the Reading* Specialist 4: -135

Guariento W. & Morley, J.(2001) Text and task authenticity in the EFL classroom. *ELT journal* Vol 55 no4, pp 1- 4

Halliday, M.A.K. (1994). *An introduction to functional grammar.* London: Edward Arnold.

Hutchinson,T.&Waters, A. (1987) *English for Specific purposes.* Cambridge:Cambridge

Jacobi, C.C.B.(2001) *Lingüística de Corpus e ensino de espanhol a brasileiros: descrição de Padrões e preparação de atividades didáticas.*Unpublished Masters thesis. PUC.

Kavaliauskienë G.& Janulevièienë V. (2001) Using the lexical approach for the acquisition of ESP vocabulary (I-TESL-J) current. *The Internet TESL Journal,* vol. VII,no3.Retrieved, February 20, 2004, from: http://iteljs.org/articles/kavaliauskkiene-LA html

Kennedy, C.& Bolitho R. (1984) *English for Specific Purposes.* Hong Kong: Modern English Publications

Laufer, B. (1997) What´s in a word that makes it hard or easy: some intralexical factors that affect the learning of words. In: Schmitt N. & McCarthy M.(1997). *Vocabulary description, Acquisition and Pedagogy.* New York:Cambridge.

Leech, G. (1997) *Teaching and Language Corpora: a convergence. Applied Linguistics and Language study.* Longman: New York.

Lewis M. (1993) *The Lexical Approach: The state of ELT and a way forward.* Hove, England: Language Teaching Publication.

Lewis, M. & Hill J. (1999) *Practical Techniques for Language Teaching.* London: Language.

Lewis,M. (2000) *Teaching Collocations: Further developments in the Lexical Approach.* London: Language Teaching Publications.

Ljung M. (1996) Corpus-based Studies in English- papers from the seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17 Stockolm, may 15-19, Rodopi Amesterdan – Atlanta , GA

Macmillan Essential Dictionary (1st ed).(2003). Oxford: Macmillan Education.

McCarthy M.(1997). *Vocabulary description, Acquisition and Pedagogy.* New York:. Cambridge.

McEnery,T. & Wilson, A. (1996). *Corpus Linguistics.* Edinburgh: C.T. University Press.

Meara, P. (1984) Towards a new approach to modeling vocabulary acquisition. In: Schmitt N. & McCarthy M.(1997). *Vocabulary description, Acquisition and Pedagogy.* New York: Cambridge.

Mindt, D. (1996) English Corpus Linguistics and the foreign language teaching syllabus. In: J. Thomas& M. Short (eds.), *Using corpora for Language Research* (232-247) Harlow, England: Longman.

Nation, P. (2003). *Como estruturar o aprendizado de vocabulário.* São Paulo: SBS.

Neves, M.H.M.(1997) *A gramática funciona*l. São Paulo: Martins Fontes.

Nunan, D.(1989). *Second Language Teaching and Learning.* London: Heinle & Heinle.

Nunan, D.(1999).*Second Language Teaching and Learning.* Boston MA: Heinle and Heinle. In: The Cambridge Guide to Teaching English to Speakers of Other Languages.

Oliveira,A.M.P.P.& Isquerdo,A.N.(2001)*As ciências do léxico: Lexicologia, lexicografia, terminologia.* (2 ed.) Campo Grande: MS Ed UFMS.

Olohan, Maeve (2004). *Introducing Corpora in Translation Studies.* London and New York: Routledge.

Petitclerck, A. (1998) Corpora for concordances. ESL teacher center Retrieved, January 10, 2004, from:http://station05.qc.ca/css/cybersite/teachref/theory

Robinson, P. (1980) *ESP. English for Specific Purposes.* Pergamon:UK

Rocha, M., Moura, H. M.M. (2001) O uso de corpora no ensino de língua portuguesa:

metodologia e avaliação. *Lingüística e Ensino: Novas Tecnologias*,vol 2 137-155.

Sardinha,T.B.(1999) Usando o WordSmith tools na investigação da linguagem. Retrieved July 20, 2004 from: http//sites uol.com.br/tony/homepage.html direct papers.

Sardinha, T. B. (2000) *Lingüística de Corpus: Histórico e Problemática.* LAEL, PUC-SP.DELTA. 16/2, 323-367.

Sardinha, T.B. (2004) *Linguistica de Corpus.* Barueri, SP: Manole

Schmitt N. & McCarthy M.(1997). *Vocabulary description, Acquisition and Pedagogy.* New York. Cambridge. Scott, M.R. (1996)*WordSmith Tools. Software for text Analysis.* Oxford:CUP.

Schimitt N. (2000) *Vocabulary in Language Teaching.* Cambridge :CUP

Scott, M.R. (1996)*WordSmith Tools. Software for text Analysis.* Oxford:CUP.

Sinclair J. (1992) The automatic analysis of Corpora .In I. Svatvik (Ed.) *Directions in Corpus Linguistics* Proceedings of Nobel Symposium 82 Stockholm, 4-8 .Berlin.

Sinclair (1996) The empty Lexicon. *International Journal of Corpus Linguistics*,Vol 1, No.1:101-119.

Sinclair, J. M. (1997) *Corpus evidence in Language D. Teaching and Language Corpora. Library of Congress.*

Söken, A. (1997) A current trends in teaching second language vocabulary. In Schmitt, N. & McCarthy, M.(1997) *Vocabulary description, Acquisition and Pedagogy.* New York:Cambridge.

Stubbs M.(1996) *Text and Corpus Analysis*. Oxford:Blackwell.

Svartivik J. (1996) *Corpora are becoming a mainstream. Using corpora for Language research.* London: Longman

Swales, J. M. (1990) *Genre Analysis.* Cambridge: Cambridge.

Swan (1985), A critical Look at the communicative approach . *ELT Journal*. Vol 39.pp 35-66.

Thompson, P.&Trimble, C. (2001) Looking at citations: Using corpora in English for Academic purposes *LLT* Vol 5 No3 September 2001pp91-105

Tomlinson, B.(1998) *Materials Development in Language Teaching.* Cambridge: Cambridge.

Trimble,L.(1985) *English for Science and Technology: a discourse approach.* Cambridge:Cambridge.

Trimble, C. (1999) Practical uses for Language Corpora in ELT. In P. Brett & G Motteram (eds). *A special interest in Computers learning and teaching with information and Communications technology*. IATEFL 2000.

Vian, O. JR. (1999, p.) Inglês Instrumental, Inglês para negócios e Inglês Instrumental para negócios. *LAEL,*vol 15.

Widdowson, H. (1981) Research and Practice in teaching English as a second language. *TESOL* and English for Special Purposes v. 15: 145-166.

Willis, D. (1993) Grammar and Lexis: some pedagogical implication. In Sinclair,(1991) J.M. Hoeye M. and Fox, G. *Techniques of description – spoken and written discourse.* London: Rutledge.

Xatara C.M. (2001) O ensino do léxico e as epressões idiomáticas.*Trab.Ling.Apl.* Campinas, Vol.37: 49-59.

Xavier,A.C.S.(2001) Processos de Referenciação de Hipertexto. *Cad.Est.Ling.*, Campinas, (41):165-176, Jul/Dez 2001.

Yule, G. (1996). *The Study of Language.* 2nd edition. Cambridge: Cambridge University Press.

Web sites

Corpus compilation. Retrieved on January 16 2005,from:
 http//www.essexac.uk/linguistics linguistics
Corpus of study:Retrieved on February 22, 2005 from:
www.tldp.org/ldp/intro-linux.txt .
WordSmith Tools:Retrieved on March, 2 2005from:
 http://www.lexically.net/WordSmith/
Specialists:Retrieved on August 2004 from:
Http://searchsecurity.techtarget.com
Http://yourdictionary.com/dictionary.com/dictionary4html
Http://whatis.techtarget.com/fileFormatA/0289933,sid9,00.html

Lexical Exercises:Retrieved on June 2004 from:
http://eli.ubc.ca/teachers/lessons/reading /technology /e.mail.htm#partA,
http://www.eli.ubc.ca/teachers/teach_lexical/advice.html#Awareness
http://www.eli.ubc.ca/teachers/lessons/reading/technology/email.htm#Par

Reading Activities -Islan and Timmis (2004).Retrieved on June 2004 from:
 http://www.teachingenglish.org.uk/try/readtry/lexicaltry2pdf

EAP vocabulary:Retrieved on April 2004:
http://www.uefap.co.uk/uk/vocab/exercises/exercise
http://www.er.uquam.ca/nobel/r21270/textools/webvphtmlhttp://www.er.uquam.ca/nobel/r21270/textools/webvphtml
http://www.1.2.3.208.224.131http://www.1.2.3.208.224.131

Activities based on Sandra Haywood (2003):Retrieved on April 2004 from:
 http://titania.collins.co.uk

APPENDIX I

GNU Free Documentation License
Version 1.2, November 2002

0. PREAMBLES

The purpose of this License is to make a manual, textbook, or other functional and
useful document "free" in the sense of freedom: to assure everyone the effective
freedom to copy and redistribute it,with or without modifying it, either commercially or
non commercially.
Secondarily, this License preserves for the author and publisher a way to get credit for
their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document
must themselves be free in the same sense. It complements the GNU General Public
License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because
free software needs free documentation: a free program should come with manuals
providing the same freedoms that the software does. But this License is not limited to
software manuals; it can be used for any textual work, regardless of subject matter or
whether it is published as a printed book. We recommend this License
principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that
contains a notice placed by the copyright holder saying it can be
distributed under the terms of this License. Such a notice grants a
world-wide, royalty-free license, unlimited in duration, to use that
work under the conditions stated herein. The "Document", below,
refers to any such manual or work. Any member of the public is a
licensee, and is addressed as "you". You accept the license if you
copy, modify or distribute the work in a way requiring permission
under copyright law.

A "Modified Version" of the Document means any work containing the
Document or a portion of it, either copied verbatim, or with
modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material

this License requires to appear in the title page. For works in
formats which do not have any title page as such, "Title Page" means
the text near the most prominent appearance of the work's title,
preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose
title either is precisely XYZ or contains XYZ in parentheses following
text that translates XYZ in another language. (Here XYZ stands for a
specific section name mentioned below, such as "Acknowledgements",
"Dedications", "Endorsements", or "History".) To "Preserve the Title"
of such a section when you modify the Document means that it remains a
section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which
states that this License applies to the Document. These Warranty
Disclaimers are considered to be included by reference in this
License, but only as regards disclaiming warranties: any other
implication that these Warranty Disclaimers may have is void and has
no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either
commercially or noncommercially, provided that this License, the
copyright notices, and the license notice saying this License applies
to the Document are reproduced in all copies, and that you add no other
conditions whatsoever to those of this License. You may not use
technical measures to obstruct or control the reading or further
copying of the copies you make or distribute. However, you may accept
compensation in exchange for copies. If you distribute a large enough
number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and
you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have
printed covers) of the Document, numbering more than 100, and the
Document's license notice requires Cover Texts, you must enclose the
copies in covers that carry, clearly and legibly, all these Cover
Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on
the back cover. Both covers must also clearly and legibly identify
you as the publisher of these copies. The front cover must present
the full title with all words of the title equally prominent and
visible. You may add other material on the covers in addition.
Copying with changes limited to the covers, as long as they preserve
the title of the Document and satisfy these conditions, can be treated

as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit
legibly, you should put the first ones listed (as many as fit
reasonably) on the actual cover, and continue the rest onto adjacent
pages.

If you publish or distribute Opaque copies of the Document numbering
more than 100, you must either include a machine-readable Transparent
copy along with each Opaque copy, or state in or with each Opaque copy
a computer-network location from which the general network-using
public has access to download using public-standard network protocols
a complete Transparent copy of the Document, free of added material.
If you use the latter option, you must take reasonably prudent steps,
when you begin distribution of Opaque copies in quantity, to ensure
that this Transparent copy will remain thus accessible at the stated
location until at least one year after the last time you distribute an
Opaque copy (directly or through your agents or retailers) of that
edition to the public.

It is requested, but not required, that you contact the authors of the
Document well before redistributing any large number of copies, to give
them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under
the conditions of sections 2 and 3 above, provided that you release
the Modified Version under precisely this License, with the Modified
Version filling the role of the Document, thus licensing distribution
and modification of the Modified Version to whoever possesses a copy
of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct
from that of the Document, and from those of previous versions
(which should, if there were any, be listed in the History section
of the Document). You may use the same title as a previous version
if the original publisher of that version gives permission.
B. List on the Title Page, as authors, one or more persons or entities
responsible for authorship of the modifications in the Modified
Version, together with at least five of the principal authors of the
Document (all of its principal authors, if it has fewer than five),
unless they release you from this requirement.
C. State on the Title page the name of the publisher of the
Modified Version, as the publisher.
D. Preserve all the copyright notices of the Document.
E. Add an appropriate copyright notice for your modifications
adjacent to the other copyright notices.
F. Include, immediately after the copyright notices, a license notice

giving the public permission to use the Modified Version under the
terms of this License, in the form shown in the Addendum below.
G. Preserve in that license notice the full lists of Invariant Sections
and required Cover Texts given in the Document's license notice.
H. Include an unaltered copy of this License.
I. Preserve the section Entitled "History", Preserve its Title, and add
to it an item stating at least the title, year, new authors, and
publisher of the Modified Version as given on the Title Page. If
there is no section Entitled "History" in the Document, create one
stating the title, year, authors, and publisher of the Document as
given on its Title Page, then add an item describing the Modified
Version as stated in the previous sentence.
J. Preserve the network location, if any, given in the Document for
public access to a Transparent copy of the Document, and likewise
the network locations given in the Document for previous versions
it was based on. These may be placed in the "History" section.
You may omit a network location for a work that was published at
least four years before the Document itself, or if the original
publisher of the version it refers to gives permission.
K. For any section Entitled "Acknowledgements" or "Dedications",
Preserve the Title of the section, and preserve in the section all
the substance and tone of each of the contributor acknowledgements
and/or dedications given therein.
L. Preserve all the Invariant Sections of the Document,
unaltered in their text and in their titles. Section numbers
or the equivalent are not considered part of the section titles.
M. Delete any section Entitled "Endorsements". Such a section
may not be included in the Modified Version.
N. Do not retitle any existing section to be Entitled "Endorsements"
or to conflict in title with any Invariant Section.
O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or
appendices that qualify as Secondary Sections and contain no material
copied from the Document, you may at your option designate some or all
of these sections as invariant. To do this, add their titles to the
list of Invariant Sections in the Modified Version's license notice.
These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains
nothing but endorsements of your Modified Version by various
parties--for example, statements of peer review or that the text has
been approved by an organization as the authoritative definition of a
standard.

You may add a passage of up to five words as a Front-Cover Text, and a
passage of up to 25 words as a Back-Cover Text, to the end of the list
of Cover Texts in the Modified Version. Only one passage of
Front-Cover Text and one of Back-Cover Text may be added by (or
through arrangements made by) any one entity. If the Document already

includes a cover text for the same cover, previously added by you or
by arrangement made by the same entity you are acting on behalf of,
you may not add another; but you may replace the old one, on explicit
permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License
give permission to use their names for publicity for or to assert or
imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this
License, under the terms defined in section 4 above for modified
versions, provided that you include in the combination all of the
Invariant Sections of all of the original documents, unmodified, and
list them all as Invariant Sections of your combined work in its
license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and
multiple identical Invariant Sections may be replaced with a single
copy. If there are multiple Invariant Sections with the same name but
different contents, make the title of each such section unique by
adding at the end of it, in parentheses, the name of the original
author or publisher of that section if known, or else a unique number.
Make the same adjustment to the section titles in the list of
Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History"
in the various original documents, forming one section Entitled
"History"; likewise combine any sections Entitled "Acknowledgements",
and any sections Entitled "Dedications". You must delete all sections
Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents
released under this License, and replace the individual copies of this
License in the various documents with a single copy that is included in
the collection, provided that you follow the rules of this License for
verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute
it individually under this License, provided you insert a copy of this
License into the extracted document, and follow this License in all
other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate
and independent documents or works, in or on a volume of a storage or
distribution medium, is called an "aggregate" if the copyright
resulting from the compilation is not used to limit the legal rights
of the compilation's users beyond what the individual works permit.
When the Document is included in an aggregate, this License does not
apply to the other works in the aggregate which are not themselves
derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these
copies of the Document, then if the Document is less than one half of
the entire aggregate, the Document's Cover Texts may be placed on
covers that bracket the Document within the aggregate, or the
electronic equivalent of covers if the Document is in electronic form.
Otherwise they must appear on printed covers that bracket the whole
aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may
distribute translations of the Document under the terms of section 4.
Replacing Invariant Sections with translations requires special
permission from their copyright holders, but you may include
translations of some or all Invariant Sections in addition to the
original versions of these Invariant Sections. You may include a
translation of this License, and all the license notices in the
Document, and any Warranty Disclaimers, provided that you also include
the original English version of this License and the original versions
of those notices and disclaimers. In case of a disagreement between
the translation and the original version of this License or a notice
or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements",
"Dedications", or "History", the requirement (section 4) to Preserve
its Title (section 1) will typically require changing the actual
title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except
as expressly provided for under this License. Any other attempt to
copy, modify, sublicense or distribute the Document is void, and will
automatically terminate your rights under this License. However,
parties who have received copies, or rights, from you under this
License will not have their licenses terminated so long as such
parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions
of the GNU Free Documentation License from time to time. Such new
versions will be similar in spirit to the present version, but may
differ in detail to address new problems or concerns. See
http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number.
If the Document specifies that a particular numbered version of this
License "or any later version" applies to it, you have the option of
following the terms and conditions either of that specified version or
of any later version that has been published (not as a draft) by the
Free Software Foundation. If the Document does not specify a version
number of this License, you may choose any version ever published (not
as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of
the License in the document and put the following copyright and
license notices just after the title page:

Copyright (c) YEAR YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.2
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts,
replace the "with...Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other
combination of the three, merge those two alternatives to suit the
situation.

If your document contains nontrivial examples of program code, we
recommend releasing these examples in parallel under your choice of
free software license, such as the GNU General Public License,
to permit their use in free software.

APPENDIX II

Unfamiliar words

Put an (X) if you know the meaning of the words

( ) Boot
( ) Device
( ) Module
( ) BIOS
( ) Kernel
( ) Partition
( ) Bug
( ) Root
( ) Default
( ) Buffer
( ) Rate
( ) shell

Explain the meaning (related to IT) of the words that you make a (X)

1_____

2_____

3_____

4_____

5_____

6_____

7_____

8_____

9_____

10_____

11_____

12_____

APPENDIX IV

Students' Materials
<u>**Look at the language:**</u>

VOCABULARY STUDY

Task one
**2)Read the lexical chunks below and underline a kind of word that is always
   repeated in different ways:**

After your machine boots. You can do this.
The first time I/ rebooted/loaded them......
When your system rebooted, do this: dms
Reboot and test, at this....
Modules each time you reboot you can do one...
Padding, disk name boot block, etc...
Lib modules directory, I boot/ vmlinuz,/usr/src/lin
At this point you should reboot your machine
module each time you reboot
a pair of black boots
13 reboot and test
after your machine reboots.
Figuration often involves rebooting

A) Which is the repeated word ?

_____

B) What happened with this word ?

_____

C) Try to explain the sort of changes that happened with this word and how these
process infer in the meaning.

_____

Task two: Underline the lexical chunk in which the word *boot* is not used in the context
of Information Technology (IT)

_____

VOCABULARY STUDY: bug

Look at the lexical chunks and pay attention to the word bug.

It really ***bugs*** me when people smoke in ....
I have been released to fix ***bugs*** . You can find this....
Bug fixes and enhance....
I sometimes have minor ***bugs*** fixes as a patch
he ***bug*** got fixed on this data.
This is because of a ***bug*** in the large file.
x series until more ***bugs*** are ironed out...

Task one: answer the questions below:

A) Is the word bug used in the same way in all the sentences?

_____

B) Why are the words **" the "** and **" a "** placed before the word bug?

_____

C)What happens when do the words **" a "** is not used with the word **bug** in most of sentences?

_____

Task two: Underline the lexical chunk in which the word *bug* is not used in the context of Information Technology (IT)

VOCABULARY STUDY: Module

Study the lexical chunks

O extension on a ***module*** filename
Even want it as a ***module*** unless you really
nd processor is the ***module*** name is the bit
The ***module*** name is the bit
The last ***module*** is my graphics
A loadable kernel ***module*** use: insmod.
Additional ***modules*** .*CD ____ /usr)sr
you have additional ***modules*** that are not ...
add ***modules_*** images when ...
You need to install your ***modules*** and then...
deb files for your ***modules*** you should.
extension on a ***module*** filename

Even want it as a *module* unless you really

And processor is the *module* name is the bit

The *module* name is the bit

The last *module* is my graphics

A loadable kernel *module* use: insmod.

Additional *modules*. *CD _____/usr)sr

you have additional *modules* that are not ...

add *modules_* images when ...

You need to install your *modules* and then...

deb files for your *modules* you should.

Taks one

Underline the weakest verb and noun collocation for the word module

Example:

Exam	**take/ pass/ fail/ study for / sit/ <u>revise</u>**
Module(s)	add / is /use / should
Module (s)	images / kernel / name /

VOCABULARY STUDY: BIOS

Study the lexical chunks

	BIOS settings for the A
s.com/au/pri/item_instr_	*bios* 7521N.asp
Previously been in the	*BIOS* configuration
Karl also notes that a	*BIOS* update was also....
The DSTDT table from	*BIOS* the kernel uses ....
The Acer TM (Phoenix	*BIOS*) http:// help. Nic
The kernel to ignore the	*BIOS*-supplied ....

Task one: **Circle the word that does not collocate with BIOS**

SETTINGS - THAT - CONFIGURATION - UPDATE - SUPPLIED - BEATIFUL - FROM - KERNEL

VOCABULARY STUDY: PATCH

1. I was using an old	*patch*.	
2. Kernel you will need to	*patch*	it before
3. Kernel source files ACPI	*patch*	that exctly test version of ACPI patch.
4. Outlines how to	*patch*	a kernel for ACPI.
5. When you	*patch*	the kernel for ACPI.
6. ACPI kernel	*patch.*	
7. kernel with an earlier	*patch*	from....
8. You need to get the	*patch*	that exactly...
9. make sure the correct	*patch*	was installed.
10. Of packages required to	*patch*	a kernel for ACPI.
11 A kernel that is already	*pre-patched* kernel.	
12. directory for the newly	*patched*	kernel.
13. kernel will need to be	*patched*.	
14. your kernel needs to be	*patched,* do so now.	
15. After you've	*patched*	the kernel add...
16. If a kernel has had other	*patches*	applied to it.
17. Depending on the	*patches*	that have been...
18. If there are any	*patches*	available.
19. This is the actual	*patching*	part.
20.	*Patching*	old kernels.
21. Once you've finished	*patching*	your kernels,
22 an eye	*patch*	
22. Instructions on	*patching*	your kernel are....

Task one: Study the concordance lines.

Which is the preposition that often follow the word **PATCH ?**

_____

List the adjectives that are used with **PATCH.**

_____

Task two: Practice

Answer the questions below:

Is the word "**patch**" used as a verb? In which lexical chunks?

_____

Is the word "**patch**" used as a noun? In which lexical chunks?

_____

How does the word "**patch**" best collocates after *needs to be* ?

_____

Write a sentence using the word "**patch.**"

_____

Task three:

Underline the lexical chunk in which the word patch is not used in the context of Information Technology (IT)

VOCABULARY STUDY: partition - kernel

Study the vocabulary

My current	***kernel*** ,2.6.6 is
to the new	***kernel*** and then return to lilo
patch the	***kernel***
you are using a new	***kernel.***
Configure the new	***kernel.***
You are using the new	***kernel*** you will need.
To pick the right	***kernel*** you will need.
You are using a different	***kernel*** make sure
starting with the	***kernel*** build "alien"
load the new	***kernel***
distinguish	***kernel*** builds
the	***kernel*** of the argument...
note	***kernel*** compile
series	***kernel,*** kernel source
in the 2.6.x series	***kernels*** you should check
(and before) series	***kernels*** this will quite
update the 2.6. series	***kernel*** with Debian´s kernel
Creating custom	***kernels*** with Debian´s kernel.

Adjectives plus a noun collocation.

Task one:

 Look at the lexical chunks above:

a) What is the most frequent adjective collocation for the noun **"kernel"** ?

( ) new   (   ) right (   ) different

b) Complete the gaps with the most common collocation before an adjective:

_____new kernel ;   _____ right kernel;   _____ different kernel;

c) Observe the verbs that are using **ing** and rewrite them:

_____

d) Observe the verbs "**use** " and **"run"** and colocate the most frequent verbs that are used before verbs ending with **ing.**

you _____ using a new **kernel.**

you _____ already running a **kernel** that...

e) Why do **the words "creating"** and **"starting"** do not have the same collocation as the exercise d ?

_____

f) The words **"Linux"** and **"Windows"** are common collocations for the word " **partition**." Based on this information complete the lexical chunks with " **partition"** or " **kernel".**

   1) Configure the new _____ **.**

   2) directory of your Linux _____ is considered ....

   3) to get your Windows _____ detected and...

   4) You are using the new _____ you will need.

   5) To pick the right _____ you will need.

   6) point of your windows _____ .

   7) You are using a different _____ make sure

   8) starting with the _____ build "alien"

   9) load the new _____.

Task two:
 Underline the lexical chunk in which the word *kernel* is not used in the context of Information Technology (IT)

VOCABULARY STUDY: ROOT - DEVICE

Linux xrc mounts the "real " *root* file system
The previous *root* (from initrd) is then moved to a.
This RAM disk can be then mounted as the *root* file system
initrd is mounted read-write as *root.*
Linuxrc creates and populates the *root* file system (r.01)
Note that changing the *root* directory does not involve
unmounting it.
Linuxrc places the *root* files system at the root directory
using...
the pivot _ *root* system call.
Olive trees have deep *roots...*
It is performed on the *root* file system.
The "real" *root* file system.

An unmounted loop *device*. All you people*....*
A loop *device* you may need to

Task one: study the lexical chunks with the words **" root"** and **"device"**

A) Can you use other nouns after the word   **" root"** ? Which ones?

_____

B) Complete the gaps above with the words "**root"** or "**device"**

1) Linux rc creates and populates the _____ file system (r.01)
2) A loop _____ you may need to...
3) An unmount the loop _____.
4) Note that changing the _____ directory does not involve unmouting it.
5) Linux rc places the _____files system at the root directory using...
6) The ACPI also has _____ drivers and using rsyn.

Task two:
 Underline the lexical chunk in which the word *root* is not used in the context of
Information Technology (IT)

VOCABULARY STUDY: SHELL

Study the lexical chunks

earning the bash   *shell* by Camerom
Open a standard   *shell,* see section...
The complete Linux   *Shell* programming
Rebuild the bash   *shell*
using   *shell* features
The   *shell*   can substitute rang...
and rebuild it into a *shell*
while running *shell*   commands
The   *shell*   contains a build-in...

```
          collecting   shells  on the beach
       While running   shell   commands
 difficulties to assimilate  shell   commands
They are not already set as   shell   variables
            or any   shell   command
    as a very thin hard   shell ,  a perfect...
     As we saw in the   shell    configuration
         Including   shell   scripts
```

Task one: Circle the weakest collocation for the word **"shell"**

COMMANDS	THE	BASH	INCLUDING	RUNNING	USING	
	DOING	IN	ASSIMILATE			

 Task two: Underline the lexical chunk in which the word "**shell**" is not used in the context of Information Technology (IT)

VOCABULARY STUDY: Buffer

*Task One: R*ead the lexical chunks below and highlight two opposite adjectives which collocate before the word " **buffer":**

 1. Finally the age   ***buffer*** and supper parameters govern the maximum time.
 2. Opposed to a clean   ***buffer*** which can just...
 3. As opposed to a clean   ***buffer*** it can...
 4. Before writing out a dirty   ***buffer*** to disk.
 5. Of the socket send   ***buffer.***
 6. Would be used for   ***buffer*** memory.
 7. MM subsystem will prune the ***buffer*** cache more heavily than memory to compensate...
 8. Age buffer time for normal   ***buffer*** to age before flush.
 9. The air bag act as a   ***buffer*** between the driver and wheal tearing wheel.

Task two:
Underline the lexical chunk in which the word **" buffer"** is not used in the context of Information Technology (IT)

VOCABULARY STUDY: Rate

*Task One: R*ead the lexical chunks below and highlight two lexical chunks in which the word **"rate"** is used as a noun.
 1. Controls the   ***rate*** at which expires are checked
 2. Out   ***rate***   sets the audio output frequency
 3. Include speed, baud   ***rate***, port, irq, and type.
 4. They enforce a   ***rate***   limit to make a denial-of-service attack
 5. The   ***rate***   of unemployment is high.

Task Two: The word **" rate"** has more than one meaning. Highlight the lexical chunk in which the word rate is not used in the content of Information Technology (IT)

APPENDIX V

Reading

ACPI: Advanced Configuration and Power Interface
Emma Jane Hogbin   emmajane@xtrinsic.com>

Before you read

1. You are going to read an article about Advanced Configuration and Power Interface (ACPI).Read these sentences from the article and discuss what kind of problem the author could have.

When I first started the switch from APM to ACPI I didn't realize the kernel needed to be patched. My problem (insanely loud fan) was fixed just by upgrading to 2.4.20 (Debian packaged kernel with an earlier patch from %5Bhttp: //acpi.sourceforge.net%5D acpi.sourceforge.net).

2. Read these sentences from the article and decide whether do you think they come at the beginning, in the middle or at the end of the article.

ACPI allows control of power management from within the operating system. The previous industry standard for power management, Advanced Power Management (APM), is controlled at the BIOS level. APM is activated when the system becomes idle--the longer the system idles, the less power it consumes (e.g. screen saver vs. sleep vs. suspend). In APM, the operating system has no knowledge of when the system will change power states.

As you read :

3. Highlight the author´s comments and observe whether they have a positive or negative position about the topic that is discussed.

After you read

4. Tell your classmate if you have had the same experience.
5. Are the author´s explanation clear and easy to understand?
6. Make a brief summary of the process with your classmate.

APPENDIX VI

ACPI: Advanced Configuration and Power Interface
Emma Jane Hogbin
emmajane@xtrinsic.com>

Minor updates for the 2.6.6 kernel and corrections regarding which kernels
need patching.
Revision v1.4 2004-05-12 Revised by: ejh
Initial thoughts on the 2.6.5 kernel; includes information on battery
monitoring applications causing touchpad lockup problems.
Outlines how to patch a kernel for ACPI support.

Table of Contents

1. About this document

When I first started the switch from APM to ACPI I didn't realize the kernel needed to be patched. My problem (insanely loud fan) was fixed just by upgrading to 2.4.20 (Debian packaged kernel with an earlier patch from %5Bhttp: //acpi.sourceforge.net%5D acpi.sourceforge.net). Unfortunately after the first upgrade I wasn't able to halt my computer without using the power switch to power-down my computer. It wasn't until later that I realized I had an old, ineffectual ACPI patch. This HOWTO was written to summarize the install process for myself, and hopefully help others who are also having a hard time finding information about ACPI. Please note: the main article outlines %5Bhttp: //www.debian.org%5D The Debian Way of doing things. There is also generic information in the Appendix B for those of you who prefer ... the generic way.

----

2. Copyright and License

Copyright (c) 2003, 2004 Emma Jane Hogbin.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.1 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and with no Back-Cover Texts. A copy of the license is included in Appendix C.

----

3. Translations

This document is also available in the following languages:
* English version 1.2 translated to: %5Bhttp://www.traduc.org/docs/HOWTO/ lecture/ACPI-HOWTO.html%5D Francais. Merci a Guillaume Lelarge et Vanessa Conchodon pour le traduction%21

----

4. About ACPI

In the world of power management ACPI is relatively new to the game. It was first released in 1996 by Compaq/Hewlett-Packard, Intel, Microsoft, Phoenix and Toshiba. These developers aimed to replace the previous industry standard for power management. Their %5Bhttp://www.acpi.info%5D ACPI.info site contains the official specifications, a list of companies that support ACPI and a number of other goodies. This is definitely not required reading, but may be of some interest to the insanely curious.

ACPI allows control of power management from within the operating system. The previous industry standard for power management, Advanced Power Management (APM), is controlled at the BIOS level. APM is activated when the system becomes idle--the longer the system idles, the less power it consumes (e.g. screen saver vs. sleep vs. suspend). In APM, the operating system has no knowledge of when the system will change power states.

ACPI can typically be configured from within the operating system. This is unlike APM where configuration often involves rebooting and entering the BIOS

configuration screens to set parameters.

ACPI has several different software components:

* a subsystem which controls hardware states and functions that may have previously been in the BIOS configuration

These states include:

%2B thermal control

%2B motherboard configuration

%2B power states (sleep, suspend)

* a policy manager, which is software that sits on top of the operating system and allows user input on the system policies

* the ACPI also has device drivers that control/monitor devices such as a laptop battery, SMBus (communication/transmission path) and EC (embedded controller).

If you would like more information on power management in laptops, check out the resources on %5Bhttp://www.tuxmobil.org%5D tuxmobil.org. Specifically: Power

Management with Linux - APM, ACPI, PMU and the %5Bhttp://tuxmobil.org/Mobile-Guide.db/mobile-guide-p2c1-hardware-in-detail.html%5D Hardware in Detail

section of the Linux Mobile Guide.

----

5. Why switch?

Not all systems support both APM and ACPI. I switched because my system only supported ACPI. Pretty easy decision really. If you're switching to get %5Bhttp://acpi.sourceforge.net/documentation/sleep.html%5D S3 (suspend to RAM)

support and you're using a 2.4.x kernel, don't bother. It is %5Bhttp://lists.debian.org/debian-laptop/2003/debian-laptop-200304/msg00418.html%5D not supported. Period.

Not sure if your system is supported? ACPI4Linux has a list of supported machines/BIOSes started on their Wiki. Please contribute to the list if you've installed ACPI%21 They also have a list of machines that are %5Bhttp://acpi.sourceforge.net/documentation/blacklist.html%5D not supported.

For more information about the power management in laptops you may (also) find the %5Bhttp://www.tldp.org/HOWTO/mini/Battery-Powered/index.html%5D Battery

Powered Linux Mini-HOWTO useful.

----

6. DSDT: Differentiated System Description Table

Thanks to %5Bhttp://www.vitavonni.de/%5D Erich writing this section.

You might need to override the DSDT when certain features like battery status are incorrectly reported (usually causing error messages to syslog). DELL laptops usually need this kind of override. Fixed DSDT for many systems are available on the %5Bhttp://acpi.sourceforge.net/dsdt/index.php%5D DSDT page, along with a patch that tells the kernel to ignore the BIOS-supplied table but use the compiled-in fixed DSDT.

Basically you need to copy the fixed table into your kernel source with a special filename (or modifying the filename in the patch supplied at the

%5Bhttp://acpi.sourceforge.net/dsdt/index.php%5D DSDT page) This override is quite easy: instead of loading the DSDT table from bios, the kernel uses the compiled-in DSDT table. That's all.
----

7. Installing from scratch

ACPI is constantly being revised. It is available in later versions of the 2.4.x series kernel (2.4.22 and higher), and all 2.6.x series kernels. If you would like to use a kernel before 2.4.22, you will need to patch your kernel source to add ACPI functionality. If at all possible you should use the latest stable version of the kernel. Patches are available from %5Bhttp://acpi.sourceforge.net%5D acpi.sourceforge.net.

Red Hat Fedora Core 2 now ships with ACPI enabled by default%21 This is big progress for the ACPI development team. Congratulations to everyone.

Note Even the latest kernel will sometimes have minor bug fixes available as a patch. You should check the ACPI4Linux web site to see if there are any patches available.

You need to get the patch that exactly matches the version of the kernel that you are running. Since this is the "install from scratch" section I will assume you know exactly which kernel you will be installing.

7.1. Choosing a kernel

This document was originally written for the 2.4.20 kernel and has been updated since to include information about the 2.6.x series kernels. At the time of this update the 2.6.x series kernels are proving easy for some and harder for others. (I personally cannot properly power down my computer with the 2.6.5 kernel.)

If you can, I would recommend waiting to upgrade your kernel to the 2.6.x series until more bugs are ironed out. There are a lot of changes in the 2.6.x series kernel. When I upgraded to 2.6.5 to update this document I ran into problems with my wireless connection, my nvidia graphics card, and with ACPI. Your mileage may vary. I personally had good success with the 2.4.20 with the latest patch and the 2.4.22 kernel with no patch. A Google through your distribution's mailing list, and the acpi-devel mailing list should help you to pick the right kernel.

Note This document uses the 2.4.20 kernel as an example for 2.4.x series kernels. Substitute your own kernel version as appropriate.

Regardless of which kernel you choose, if it is a kernel that requires patching, it is important to use the latest version of the ACPI patch. Some distributions have already patched their kernels. This is the case for Debian, and may be the case for others. For more information on the patches that have been applied to the Debian kernel source package scan through: /usr/src/kernel-source-<version>/README.Debian. If you are not using Debian you will probably still be able to find an equivalent file for your distribution.

A user on acpi-support confirmed that I shouldn't need any of the additional patches that have been applied to the kernel to run my laptop. If you are running a production-level server and/or are serving web pages to the internet, you should really apply any additional security patches.

Warning If a kernel has had other patches applied to it, you may have problems applying the ACPI patch. Of course, an ACPI patch should not be applied to a kernel that is already patched for ACPI. As long as

there has not been an ACPI patch applied to the kernel it should be
possible to apply one now. Depending on the patches applied, you may
need to modify some of the Makefiles for your patch to be successful.
This is beyond my current grasp of reality so it is not covered in
this document.

7.1.1. Debian-ized pre-patched kernel
If you would prefer to use a Debian-ized kernel instead of a fresh one,
%5Bhttp://people.debian.org/~maxx%5D maxx has provided a pre-patched
kernel-source package with the latest patch for the 2.4.20 kernel. This would
be instead of downloading a fresh (non-patched) kernel from %5Bhttp://
www.kernel.org%5D www.kernel.org. He sent me an email with the following
details:
I took the kernel-source 2.4.20-8 from unstable, removed the
ACPI changes %5Be.g. the old patch%5D and applied
acpi-20021212-2.4.20.diff.gz from acpi.sf.net since the
vanilla 2.4.20 HAS several security leaks (ptrace, hash
table, ...).
You can find the package at %5Bhttp://people.debian.org/~maxx/
kernel-source-2.4.20/%5D http://people.debian.org/~maxx/
kernel-source-2.4.20/ (I didn't upload the .orig.tar.gz
since you can get it from any debian mirror and the .deb is
already big enough)
--%5Bhttp://people.debian.org/~maxx%5D maxx
Warning I have not tested these packages. You may or may not have any luck
with them. Please don't email me asking about them, ask maxx instead.
----
8. Backups
If you are already running a kernel that is the same version of the one you
are about to patch I recommend creating a fresh directory for the newly
patched kernel. Remember that backups are never a bad thing. These are the
files that I back up:
* /etc/lilo.conf
* /usr/src/*.deb (Debian-specific)
* /etc/modules
* /etc/modutils/aliases
* /usr/src/linux/.config
* If you are not doing things The Debian Way you should also back up the /
lib/modules directory, /boot/vmlinuz, /usr/src/linux/arch/i386/boot/
bzImage and /usr/src/System.map. It's possible my notes on the location
of these files differs. Do a locate <file> if they're not where I've
stated they should be.
----
9. Download and Unpack the New Kernel
9.1. Required packages
The following is a list of packages required to patch a 2.4.x series kernel.
I am still working on the notes for a 2.6.x series install.
2.4.x series kernels
* kernel source files
* ACPI patch that exactly matches the kernel version

* debian packages: make, bzip2, gcc, libc6-dev, tk8.3, libncurses5-dev,
kernel-package
* after you've patched the kernel add the debian packages: acpid, acpi
(this last package is available in testing and unstable versions of
Debian, but not stable)

9.2. Unpack
We need to unpack the bz2 file (bzip2) and shuffle the directories around a
bit. /usr/src/linux probably points to your current kernel. We need it to
point to the new kernel, so we'll do that as well.
* cd /usr/src
* mkdir kernel-source-<version> (use an alternate name if you already
have a version of this kernel installed)
* cp linux.<version>.tar.bz2 /usr/src/kernel-source-<version>
* cd /usr/src/kernel-source-<version>
* tar xjfv linux.<version>.tar.bz2
* mv linux.<version> /usr/src/linux-<version>
* rm linux (assuming that's a link to your old kernel)
* ln -s /usr/src/linux-<version> linux
If your kernel needs to be patched, do so now. Instructions are available
from Appendix A.
----
10. Configure the new kernel
Note Patch Your Kernel First
If you are using an old kernel you will need to patch it before you can
proceed. Instructions on patching your kernel are available from
Appendix A. The 2.6.x series kernels do not need to be patched.
Now instead of using make menuconfig, I have an excellent alternative. Check
this out: copy your current .config file into /usr/src/linux. Now use "make
oldconfig". It will run through your old config file and see what's been
updated so that you don't have to find all the new options. For everything to
do with ACPI and your specific hardware (Toshibas choose the Toshiba options,
Asus choose the Asus options) choose M for module. There are about ten
different ACPI related options that you will need to select.
In point form, this is how the kernel should be configured:
* cd /usr/src/linux
* cp /usr/src/<oldkernel-source-directory>/.config .config
* make oldconfig (say M to all new options for ACPI--you can also say "Y"
if you prefer to compile it directly into your kernel)
Now go in to the config file with make menuconfig. I want you do check and
make sure you have your APM (the old stuff) turned off. Under "General Setup"
, make sure that:
* Power Management Support is ON
* APM (Advanced Power Management) is OFF (this is the old one--you don't
even want it as a module unless you really know what you're doing. And if
you really know what you're doing you're probably not reading this.)
* everything to do with ACPI should be M (modules) or * (compiled directly
into the kernel). Read the list carefully. Some options will not apply to
your hardware.
exit and save the new configuration

----
11. Compile the new kernel
If you have additional modules that are not part of the main source tree, you
will need to add modules_image when you make your Debian packages. This is
almost inevitable if you're using a laptop and an older kernel. Only my
nvidia graphics card now requires additional modules.
* cd /usr/src/linux
* make-kpkg clean
* make-kpkg --append-to-version%3D.<date> kernel_image modules_image
Note Naming kernel builds
I no longer use .date to distinguish kernel builds. It was too
frustrating to have 030627a, 032627b (etc) as I tried to figure things
out. I now use names, in alphabetical order, starting with the kernel
build "alien". I'm going to leave the date option in though as I still
think it's a good way to do things.
My current kernel, 2.6.6, is "Elrond." The machine itself is "Smeagol."
Note Kernel compile help
For non-Debian instructions see the Appendix "Appendix B".
For more information on how to compile the kernel The Debian Way please
read Creating custom kernels with Debian's kernel-package system
----
12. Install the new kernel
I like to configure lilo on my own, but do whatever tickles your fancy.
* cd /usr/src
* dpkg -i kernel-image-<version>.<date>_10.00.Custom_i386.deb At this
point I decline all the lilo updates and configure it myself by hand.
* configure lilo by hand: vi /etc/lilo.conf
* load the new kernel into lilo: lilo
* If you have any other deb files for your modules you should install
them now as well. If you're not sure check /usr/src for additional .deb
files.
Note Kernel compile help
For non-Debian instructions see the Appendix "Appendix B".
For more information on how to compile the kernel The Debian Way please
read Creating custom kernels with Debian's kernel-package system
----
13. Reboot and test
At this point you should reboot your machine. When your system comes back up
(assuming of course that everything went well and you still have a system),
check to see what kernel you're running with uname -a. It should show you the
one you just built. You also need to make sure the correct patch was
installed. You can do that with dmesg %7C grep ACPI.*Subsystem%5C revision .
It
should give the output: ACPI: Subsystem revision 20021212. The revision is
the date the patch was released. This number will be different than mine if
you are not using the 2.4.20 kernel. To look at all ACPI-related bits that
were loaded/started when your system rebooted, do this: dmesg %7C grep ACPI .
dmesg prints your boot messages and grep ACPI makes sure that only
ACPI-related messages are printed.
You can also check to see what version you're using with cat /proc/acpi/info.

Don't believe everything you read though. My output says that S3 is a supported state, but we already know it's not. It does give the correct version though, which is useful.

--

14. Load related modules
Check to see that each of the ACPI modules have been loaded after your machine boots. You can do this with the command lsmod. You are looking for the following options: button, battery, fan, ac, thermal and processor. If you chose "Y" instead of modules when you compiled your kernel, you will not see this list. The output on my computer looks like this:
Module Size Used by Tainted: P
button 2420 0 (unused)
battery 5960 0 (unused)
ac 1832 0 (unused)
fan 1608 0 (unused)
thermal 6664 0 (unused)
processor 8664 0 %5Bthermal%5D
NVdriver 945408 11
The last module is my graphics card, which uses proprietary drivers. This is why I have a "P" next to Tainted on the top line.
If you compiled ACPI support in as "M"odules and you don't see the ACPI modules listed you will need to load the modules by hand. The modules should be in /lib/modules/<version>. <date>/kernel/drivers/acpi/, and are as follows:
-rw-r--r-- 1 root root 4.1k Jun 3 23:57 ac.o
-rw-r--r-- 1 root root 9.5k Jun 3 23:57 battery.o
-rw-r--r-- 1 root root 5.2k Jun 3 23:57 button.o
-rw-r--r-- 1 root root 3.7k Jun 3 23:57 fan.o
-rw-r--r-- 1 root root 14k Jun 3 23:57 processor.o
-rw-r--r-- 1 root root 11k Jun 3 23:57 thermal.o
-rw-r--r-- 1 root root 6.2k Jun 3 23:57 toshiba_acpi.o
Note Extensions on Modules
The module name is the bit before .o extension on a module filename. processor.o is the file, and processor is the module name. To install a loadable kernel module use: insmod processor.
The 2.4.x series kernels use the extension .o; however, the 2.6.x series kernel use the extension .ko.
The first time I rebooted I loaded them all by hand, typing insmod < modulename>. I personally load processor first, although there are mixed feelings on whether or not the order matters.
Note Operating System Power Management (OSPM)
The first time I tried this the modules were all in separate directories and were ospm_<name>. This was probably because I was using an old patch, but it is something to be aware of. The OSPM modules are now deprecated so hopefully you won't see them.
To prevent having to load the modules each time you reboot you can do one of two things: compile them directly into the kernel (bit late for that though, eh?), or add them to your /etc/modules file. If you don't already have a copy of the file just create a new one and add each module name (remember, no dot-o) on a separate line. You can also try running update-modules which

should automatically update your /etc/modules.conf configuration file.

----

15. Switching from APM to ACPI

Do not let apmd and acpid run at the same time unless you REALLY know what you're doing. Debian will not make sure only one is running at a time. You will have to check. APM will try to put your system into S3. On the 2.4.x (and before) series kernels this will quite probably hang your machine. S3 is not supported until at least 2.5.x. Even the patch won't provide support for S3 in the 2.4.x series kernels. I personally did an apt-get remove apmd to solve the hanging problem.

You should also be aware of another little glitch I discovered. The XFree86 server has an option for DPMS (Energy Star) features. The DPMS can states can be one of standby, suspend, off or on. Since the 2.4.x kernels cannot suspend to disk, this can cause problems. I fixed my system by doing two things:

* xset -dpms (disables DPMS features)
* In /etc/X11/XF86Config-4 I commented out the line Option "DPMS" under Section "Monitor".

Warning Lost Touchpad and Keyboard Interrupts

Karl Hegbloom emailed me to say that "keyboard keys sticking, touchpad pointer jumping suddenly across the screen, lockups under heavy network I/O %5Bwas%5D caused by polling the battery state via ACPI and the proc file system." This problem caused, in part, by a delay between the initial request for information about the battery's status and the response. In that delay interrupts may be locked out and synchronization with the keyboard and touch pad may be lost. "The overnight solution is to either turn off the battery applet, or reduce its polling frequency." Karl also notes that a BIOS update was also able to solve the problem. Thanks Karl%21

I did not find this to be a problem for me in the 2.4.x kernels, however, it was a problem when I upgraded to 2.6.5. I use wmacpi to monitor my battery status. By setting the polling frequency to 1 (the lowest possible number), I seem to have eliminated the touchpad lockups. In my .xinitrc file I use:

/usr/bin/wmacpi -s 1

The -s 1 represents the polling frequency (sample rate) of "once per minute." The default is 20.

----

16. Using ACPI

There are a few different applications/daemons you will want to install on your system: acpid (the daemon that will control your hardware states), and acpi (the interface to monitor events and states) are the base install. The acpi Debian package is only available in testing and is unstable. If you're running stable you won't be able to install it without playing around with apt and your list.sources file. You can probably also compile from source. If you do get acpi installed you can use it to monitor your system like this: acpi -V. The output will tell you about your system. Mine looks like this:

Thermal 1: ok, 47.1 degrees C
Thermal 2: ok, 45.1 degrees C
AC Adapter 1: off-line <-- running off battery
AC Adapter 1: on-line <-- running off AC power

Unfortunately, the -V "full version" doesn't work for me. Fortunately I can still look in each of the acpi files individually for information about my system. Check in the /proc/acpi directory for various things of importance. If I want to check my battery I read the following file like this: cat /proc/acpi/battery/BAT0/state. The output is as follows:
present: yes
capacity state: ok
charging state: discharging <-- running off battery
present rate: unknown
remaining capacity: 3920 mAh <-- watch this number
present voltage: 14800 mV
present: yes
capacity state: ok
charging state: discharging
present rate: unknown
remaining capacity: 3840 mAh <-- capacity getting smaller
present voltage: 14800 mV
present: yes
capacity state: ok
charging state: charging <-- AC adapter plugged in
present rate: unknown
remaining capacity: 3840 mAh
present voltage: 14800 mV
If I want information about my battery in general I check it out like this:
cat /proc/acpi/battery/BAT0/info
present: yes
design capacity: 3920 mAh
last full capacity: 3920 mAh
battery technology: rechargeable
design voltage: 14800 mV
design capacity warning: 30 mAh
design capacity low: 20 mAh
capacity granularity 1: 10 mAh
capacity granularity 2: 3470 mAh
model number: Bat0
serial number:
battery type: Lion
OEM info: Acer
You're smart people. You can probably figure it out from here. :)
----
17. References and Resources
The following URLs were incredibly useful in writing this HOWTO and generally getting ACPI up and running.
HOWTOs
HOWTO install ACPI under Linux
http://sylvestre.ledru.info/howto/howto_acpi.php
Linux ACPI-HOWTO
http://www.columbia.edu/~ariel/acpi/acpi_howto.txt
Linux on the road, formerly: Linux Laptop HOWTO
http://tuxmobil.org/howtos.html You'll need to scroll a bit, or use

the HTML version: http://tuxmobil.org/Mobile-Guide.db/Mobile-Guide.html
Hardware in Detail (part of Linux on the road)
http://tuxmobil.org/Mobile-Guide.db/
mobile-guide-p2c1-hardware-in-detail.html
Power Management with Linux - APM, ACPI, PMU
http://tuxmobil.org/apm_linux.html
Battery Powered Linux Mini-HOWTO
http://www.tldp.org/HOWTO/mini/Battery-Powered/
Creating custom kernels with Debian's Kernel-Package system
http://newbiedoc.sourceforge.net/system/kernel-pkg.html
Hardware-specific Install Reports and Info
Installation Reports
http://acpi.sourceforge.net/wiki/index.php/InstallationReports
Blacklist
http://acpi.sourceforge.net/documentation/blacklist.html
DSDT: Overview
http://acpi.sourceforge.net/dsdt/index.php Includes links to patched
DSDTs and HOWTOs about patching your own DSDT.
BIOS Settings for the AcerTM (Phoenix BIOS)
http://help.nec-computers.com/au/pri/item_instr_bios_7521N.asp
Software Development Groups
ACPI4Linux
http://acpi.sf.net
ACPI Special Interest Group
http://www.acpi.info/
Intel
http://developer.intel.com/technology/iapc/acpi/
ACPI articles
Fan Speed Control Techniques in PCs
http://www.analog.com/library/analogDialogue/archives/34-04/fan/
Mailing List Threads
debian-laptop thread: can't restore from suspend
http://lists.debian.org/debian-laptop/2003/debian-laptop-200304/
msg00367.html
acpi-support thread: newbie HOWTO and debian patching
http://sourceforge.net/mailarchive/forum.php?forum_id%3D7803&max_rows%3
D
25&style%3Dflat&viewmonth%3D200304&viewday%3D17
debian-laptop thread: acer 634 acpi & apm
http://lists.debian.org/debian-laptop/2002/debian-laptop-200212/
msg00242.html
ACPI packages and related software
The Kernel
Remember to choose "F" for full when you download your kernel source.
http://www.kernel.org
Debian-ized kernel
maxx's pre-patched 2.4.20-8 kernel source package. For more information
see maxx's notes. http://people.debian.org/~maxx/kernel-source-2.4.20/
ACPI kernel patch
You'll need to pick the version that exactly matches the kernel you're

using. http://sourceforge.net/project/showfiles.php?group_id%3D36832
acpid
the daemon http://sourceforge.net/projects/acpid
acpi
text interface http://grahame.angrygoats.net/acpi.shtml
Kacpi
graphical interface for KDE http://www.elektronikschule.de/~genannt/
kacpi/download.html
aKpi
another KDE interface http://akpi.scmd.at/
wmacpi
WindowMaker DockApp (another GUI) http://www.ne.jp/asahi/linux/timecop/
wmacpi%2Bclecourt
WindowMaker DockApp (another graphical interface). Handles two battery
slots. http://open.iliad.fr/~clecourt/wmacpi/index.html
----

18. Thanks
Much thanks goes out to the following:
* %5Bhttp://acpi.sourceforge.net/mailinglists.html%5D acpi-support (note: the
discussion list for ACPI4Linux is now at acpi-devel)
* %5Bhttp://lists.debian.org/debian-laptop/%5D debian-laptop
* %5Bhttp://lists.debian.org/debian-user/%5D debian-user
* %5Bhttp://linuxchix.org/%5D techtalk
* TLDP mailing lists (discuss and docbook)
* Sebastian Henschel for reminding me I'd promised to write it all down
* Erich Schubert for writing the section on DSDTs
* Werner Heuser for suggesting I submit the document to The LDP
* Tabatha Marshall for editing and generally being very enthusiastic about
learning DocBook

A. Patching Old Kernels
If you are using a 2.4.x series kernel, the kernel will need to be patched
before you can add ACPI support. Although ACPI is included in the 2.6.x
series kernels you should check to see if any patches have been released to
fix bugs. You can find this information on the ACPI4Linux site.

A.1. Getting the Source Files
Download a fresh kernel from %5Bhttp://www.kernel.org%5D www.kernel.org.
You
need
to make sure you get a full kernel. Find the "latest stable version of the
Linux kernel" and click on F for FULL. Wait patiently. A bzipped kernel is
about 26M. If you're feeling particularly geeky you could also wget http://
kernel.org/pub/linux/kernel/v2.4/linux-<version>.tar.bz2.
Tip You may or may not want the latest stable version. For more information
read the Section 7.1 section of this document. If you decide to use a
version of the kernel that is not published on the front page, use the
%5Bhttp://www.kernel.org/pub/linux/kernel/%5D /pub/linux/kernel directory on
the %5Bhttp://www.kernel.org%5D kernel.org site to find the kernel you'd
like.

While you're waiting, grab a copy of the patch as well. For the 2.4.20 kernel use the 2.4.20 patch. It's dated 2002.12.12. You'll need to know that number later when we check to make sure the patch worked. If you are using a different kernel version make sure you take note of the date of your patch. Your numbers will differ slightly from the one I use later on.

Once you've got those two files (the kernel and the patch) unpack them and patch the kernel.

A.1.1. Patch

Now we're going to actually patch the kernel. I take one extra step from %5Bhttp://acpi.sourceforge.net/download.html%5D the instructions at ACPI4Linux.

Instead of gunzipping and patching in the same line, I use two lines. This is purely a matter of preference. When you patch the kernel you want to make sure there are no error messages. (There is no "yay" line, instead look for the absence of errors.)

* cd /usr/src/linux
* cp acpi-20021212-2.4.20.diff.gz /usr/src/linux/. (Your patch filename will be different if you're not using the 2.4.20 kernel.)
* gunzip acpi-20021212-2.4.20.diff.gz
* patch -p1 < acpi-20021212-2.4.20.diff (this is the actual patching part)

Once you've finished patching your kernel, continue reading at Section 10.

B. ACPI the Non-Debian Way

There is very little difference between The Debian Way and the generic way. In fact it's probably only 10 or so lines of difference.

B.1. Compile the kernel

The "normal" way of compiling a kernel does not use make-kpkg. Instead, it uses the following steps:

* cd /usr/src/linux which should point to the 2.4.20 kernel (unzipped) files
* make dep
* make clean
* make bzImage
* make modules (remember to unpack your modules first)

B.2. Install the new kernel

In The Debian Way, you create a deb file which contains information about where the kernel is (and makes the kernel and yada-yada). In the "normal" way, you put things where they need to be right away. You need to install your modules and then configure lilo to point to the new kernel and then run lilo. If you are not doing things The Debian Way your "install" will look like this:

* cd /usr/src/linux
* make modules_install
* cp arch/i386/boot/bzImage /boot/vmlinuz.<date>
* vi /etc/lilo.conf and copy the structure of your existing kernel. Do NOT delete the reference to your existing kernel%21 You need to point lilo

to the "vmlinuz" file that was created when you compiled the kernel above
* lilo (yup, just exactly like that.) Lilo will let you know if it's going
to have major problems loading the new kernel.
Warning Do NOT forget to run lilo before rebooting. Type lilo. It's that easy
(and that easy to forget).

B.3. Software packages
You can still use all of the software mentioned in this HOWTO even if you're
not using Debian. Unfortunately it will take a little more effort on your
part to download and install everything. Fortunately it's really not that
difficult. Most software packages include a README file when you gunzip them
which will explain what you need to do to get things working on your system.
Tip Software downloads
For more information about software for ACPI, please use the ACPI
packages and related software.

C. GNU Free Documentation License
Version 1.1, March 2000
Copyright (C) 2000 Free Software Foundation, Inc. 59 Temple Place, Suite
330, Boston, MA 02111-1307 USA Everyone is permitted to copy and
distribute verbatim copies of this license document, but changing it is
not allowed.

C.1. PREAMBLE
The purpose of this License is to make a manual, textbook, or other written
document "free" in the sense of freedom: to assure everyone the effective
freedom to copy and redistribute it, with or without modifying it, either
commercially or noncommercially. Secondarily, this License preserves for the
author and publisher a way to get credit for their work, while not being
considered responsible for modifications made by others.
This License is a kind of "copyleft", which means that derivative works of
the document must themselves be free in the same sense. It complements the
GNU General Public License, which is a copyleft license designed for free
software.
We have designed this License in order to use it for manuals for free
software, because free software needs free documentation: a free program
should come with manuals providing the same freedoms that the software does.
But this License is not limited to software manuals; it can be used for any
textual work, regardless of subject matter or whether it is published as a
printed book. We recommend this License principally for works whose purpose
is instruction or reference.

C.2. APPLICABILITY AND DEFINITIONS
This License applies to any manual or other work that contains a notice
placed by the copyright holder saying it can be distributed under the terms
of this License. The "Document", below, refers to any such manual or work.
Any member of the public is a licensee, and is addressed as "you".
A "Modified Version" of the Document means any work containing the Document
or a portion of it, either copied verbatim, or with modifications and/or
translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (For example, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, whose contents can be viewed and edited directly and straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup has been designed to thwart or discourage subsequent modification by readers is not Transparent. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML designed for human modification. Opaque formats include PostScript, PDF, proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

C.3. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

C.4. COPYING IN QUANTITY

If you publish printed copies of the Document numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a publicly-accessible computer-network location containing a complete Transparent copy of the Document, free of added material, which the general network-using public has access to download anonymously at no charge using public-standard network protocols. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

C.5. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has less than five).

C. State on the Title page the name of the publisher of the Modified

Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section entitled "History", and its title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. In any section entitled "Acknowledgements" or "Dedications", preserve the section's title, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section as "Endorsements" or to conflict in title with any Invariant Section.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover,

previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

C.6. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name the patch ent contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections entitled "History" in the various original documents, forming one section entitled "History"; likewise combine any sections entitled "Acknowledgements", and any sections entitled "Dedications". You must delete all sections entitled "Endorsements."

C.7. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

C.8. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, does not as a whole count as a Modified Version of the Document, provided no compilation copyright is claimed for the compilation. Such a compilation is called an "aggregate", and this License does not apply to the other self-contained works thus compiled with the Document, on account of their being thus compiled, if they are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one quarter of the entire aggregate, the Document's Cover Texts may be placed on covers that surround

only the Document within the aggregate. Otherwise they must appear on covers around the whole aggregate.

C.9. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License provided that you also include the original English version of this License. In case of a disagreement between the translation and the original English version of this License, the original English version will prevail.

C.10. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

C.11. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See %5Bhttp://www.gnu.org/copyleft/%5D http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

C.12. How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (c) YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.1 or any later version published by the Free Software Foundation; with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have no Invariant Sections, write "with no Invariant Sections" instead of saying which ones are invariant. If you have no Front-Cover Texts, write "no Front-Cover Texts" instead of "Front-Cover Texts being LIST"; likewise

for Back-Cover Texts.
If your document contains nontrivial examples of program code, we recommend
releasing these examples in parallel under your choice of free software
license, such as the GNU General Public License, to permit their use in free
software.

APPENDIX VII

LEARNER'S CHARACTERISTICS

Universidade Federal de Santa Catarina
Centro de comunicação e expressão

Curso de Pós-Graduação em Letras/Inglês e Literatura correspondente

Este questionário faz parte de um projeto de pesquisa do curso de Mestrado da Universidade Federal de Santa Catarina – UFSC. Tal projeto, cujo título é "A corpus-based-approach to ESP: EST vocabulary in Information Technology."("Uma abordagem baseada em corpus para alunos que estudam inglês com propósitos específicos: vocabulário de inglês para ciência e tecnologia em Análise de Sistemas .")tem por objetivo investigar os resultados da aplicação de uma metodologia baseada em estudo de Corpus relacionada à aquisição de vocabulário técnico. Por isso, o preenchimento correto e preciso deste questionário é de vital importância para a confiabilidade dos resultados da pesquisa. Agradeço, antecipadamente, a sua participação, o que em muito contribuirá para a investigação da aquisição de vocabulário técnico em língua inglesa como língua estrangeira. As informações pessoais aqui contidas, bem como nome desta instituição de ensino permanecerão em absoluto sigilo e não serão revelados na apresentação dos resultados.

Nome: _____ Data_____/_____/____

Idade:_____ Natural de:_____

Onde concluiu o Ensino Fundamental e Médio?

Estudou Inglês em cursos particulares? Sim ( ) Não( )

Morou fora do Brasil onde o idioma oficial era o inglês? Sim( ) Não ( )

Teve/tem contato com falantes nativos de inglês? Sim ( ) Não ( )

Domina outra língua além do português? Sim ( ) Não ( ) Qual?_____

Como você classificaria o seu atual nível de domínio do idioma inglês?

Nenhum( ) Básico( ) Pré intermediário( ) Intermediário( ) Pré avançado( )

Avançado( )

Como você classificaria o seu atual nível de domínio de termos técnicos relacionados à Análise de Sistemas?

Nenhum ( ) Básico ( ) Intermediário( ) Pré avançado( ) Avançado ( )

Faz outro curso, tem outra graduação?

 Sim ( ) Não ( ) Qual? Por favor, especifique _____

Tinha algum conhecimento na área de Análise de Sistemas antes de iniciar a graduaçãₒ?

Sim ( ) Não ( ) Por favor, especifique em caso afirmativo._____

Por que meio você tem contato com vocabulário técnico em inglês?

Internet( ) Revistas( ) Manuais( ) Livros( ) Outros( )especifique _____

Você reconhece a importância do conhecimento de termos técnicos em língua inglesa?

Sim ( ) Não ( ) Por favor, justifique _____

Como o conhecimento de termos técnicos poderia ajudá-lo? Por favor, justifique.

Que tipos de livros, revistas ou material você lê na área de Análise de Sistemas. Qual material/materiais é mais difícil de ser lido/entendido por você? Por quê?

Que tipos de textos você teria interesse/necessidade de serem trabalhados na disciplina de Inglês Instrumental?

Você utiliza livros e revistas em inglês disponíveis na biblioteca? Se não por quê?

APPENDIX VIII

LEARNERS ACTING IN RESPONSE

Questionário

1) Estudar gramática:
   ( ) gosto.   ( ) mais ou menos   ( )não gosto.

2) O estudo da gramática parece ser:
   ( ) necessário   ( )as vezes é necessário   ( ) não é necessário.

3) Qual é sua opinião dos exercícios que resolveu?
( ) interessante
( ) não entendi nada
( ) difícil
( ) prefiro que o professor repasse conteúdos de gramática.
( ) fácil
( ) outros.

4) Marque uma opção de cada coluna:

a) gostei desta atividade porque: b) Não gostei das atividades porque:
  ( ) os textos são autênticos   ( ) há muitos exemplos
  ( ) me faz pensar   ( ) é difícil devido as orações incompletas
  ( ) entendi   ( ) perde-se muito tempo resolvendo
  ( ) é mais difícil de entender   ( ) não são interessantes
  ( ) descobri as regras sozinho(a)   ( ) não entendi o objetivo

5) Como trabalhou:
   ( ) o tempo todo sozinho(a)
   ( ) sempre em grupo
   ( ) às vezes sozinho, às vezes em grupo

6) Gostaria de resolver mais exercícios deste tipo:
   ( ) sim   ( ) não   ( ) não sei

Questionário n 2

O que aprendeu com estes exercícios?

Como aprendeu? Que estratégias usou?

De que maneira aprendeu melhor?

O que prefere a explicação do professor ou descobrir sozinho? Explique por quê?

Questionário 3

Escreva 3 coisas que considere positivas e três que considere negativas com relação as atividades que resolveu:

APPENDIX IX

POSTTEST

Student_____

Fill in the blanks with the following words:

BIT RATE /PATCH/ BUFFER/ BUG / SHELL/ KERNEL/ DEVICE/ BIOS/MODULE /BOOT/PARTITION/ROOT

1)

A _____ (sometimes called a "fix") is a quick-repair job for a piece of programming. During a software product's beta test distribution or try-out period and later after the product is formally released, problems (called ____)will almost invariably be found. A_____ is the immediate solution that is provided to users; it can sometimes be downloaded from the software maker's Web site. The _____ is not necessarily the best solution for the problem and the product developers often find a better solution to provide when they package the product for its next release. A _____ is usually developed and distributed as a replacement for or an insertion in compiled code (that is, in a binary file or object module). In larger operating systems, a special program is provided to manage and keep track of the installation of _____.

2

In digital telecommunication, the _____ is the number of bits that pass a given point in a telecommunication network in a given amount of time, usually a second. Thus, a _____ is usually measured in some multiple of bits per second - for example, kilobits, or thousands of bits per second (Kbps). The term _____ is a synonym for data transfer _____ (or simply data _____)._____ seems to be used more often when discussing transmission technology details and data transfer rate (or data rate) when comparing transmission technologies for the end user.

3)

A _____ is a data area shared by hardware devices or program processes that operate at different speeds or with different sets of priorities. The _____ allows each device or process to operate without being held up by the other.Like a cache, a _____ is a "midpoint holding place" but exists not so much to accelerate the speed of an activity as to support the coordination of separate activities.This term is used both in programming and in hardware.In programming,_____ sometimes implies the need to screen data from its final intended place so that it can be edited or otherwise processed before being moved to a regular file or database.

4)

_____ is a Unix term for the interactive user nterface with an operating system. The_____ is the layer of programming that understands and executes the commands a user enters. In some systems, the _____ is called a command interpreter. A _____ usually implies an interface with a command syntax (think of the DOS operating system and its "C:>" prompts and user commands such as "dir" and "edit"). As the outer layer of an operating system, a _____ can be contrasted with the kernel, the operating system's inmost layer or core of services.

5)

_____ (basic input/output system) is the program a personal computer's microprocessor uses to get the computer system started after you turn it on. It also manages data flow between the computer's operating system and attached devices such as the hard disk, video adapter, keyboard, mouse, and printer. _____ is an integral part of your computer and comes with it when you bring it home. (In contrast, the operating system can either be preinstalled by the manufacturer or vendor or installed by the user.) _____is a program that is made accessible to the microprocessor on an eraseable _____ programmable read-only memory (EPROM)chip. When you turn on your computer, the microprocessor passes control to the _____ program, which is always located at the same place on EPROM. When _____ boots up (starts up)your computer, it first determines whether all of the attachments are in place and operational and then it loads the operating system (or key parts of it) into your computer's random access memory (RAM) from your hard disk or diskette drive. With_____ your operating system and its applications are freed from having to understand exact details (such as hardware addresses)about the attached input/output devices. When device details change, only the _____program needs to be changed. Sometimes this change can _____ be made during your system setup. In any case, neither _____ your operating system or any applications you use need to be changed.

6)

To _____ (as a verb; also "to _____ up") a computer is to load an operating system into the computer's main memory or random access memory (RAM). Once the operating system is loaded (and, for example, on a PC, you see the initial Windows or Mac desktop screen), it's ready for users to run applications. Sometimes you'll see an instruction to "_____" the operating system. This simply means to reload the operating system (the most familiar way to do this on PCs is pressing the Ctrl, Alt, and Delete keys at the same time). On larger computers (including mainframes), the equivalent term for "_____" is "initial program _____" (IPL) and for "re_____" is "re-IPL." _____ is also used as a noun for the act of _____, as in "a system _____."

7)

In computer technology, a _____ is a coding error in a computer program. (Here we consider a program to also include the microcode that is manufactured into a microprocessor.) The process of finding _____ before program users do is called _____. _____ starts after the code is first written and continues in successive stages as code is combined with other units of programming to form a software product, such as an operating system or an application. After a product is released or during public beta testing, _____ are still apt to be discovered. When this occurs, users have to either find a way to avoid using the "_____" code or get a patch from the originators of the code.A _____is not the only kind of problem a program can have. It can run _____-free and still be difficult to use or fail in some major objective. This kind of flaw is more difficult to test for (and often simply isn't).It is generally agreed that a well-designed program developed using a well-controlled process will result in fewer _____ per thousands of lines of code. The term's origin has been wrongly attributed to the pioneer programmer, Grace Hopper.

8)

The _____ is the essential center of a computer operating system, the core that provides basic services for all other parts of the operating system. A synonym is nucleus. A _____ can be contrasted with a shell, the outermost part of an operating system that interacts with user commands. _____ and shell are terms used more frequently in Unix operating systems than in IBM mainframe or Microsoft Windows systems. Typically, a _____ (or any comparable center of an operating system) includes an interrupt handler that handles all requests or completed I/O operations that compete for the _____'s services, a scheduler that determines which programs share the _____'s processing time in what order, and a supervisor that actually gives use of the computer to each process when it is schedule._____ may also include a manager of the operating system's address spaces in memory or storage, sharing these among all components and other users of the _____'s services. A _____'s services are requested by other parts of the operating system or by application programs through a specified set of program interfaces sometimes known as system calls. Because the code that makes up the _____ is needed continuously, it is usually loaded into computer storage in an area that is protected so that it will not be overlaid with other less frequently used parts of the operating system.The _____ is not to be confused with the Basic Input/Output System (BIOS).Some _____ have been developed independently for use in any operating system that wants to use it.A well-known example is the Mach _____, developed at Carnegie-Mellon University, and currently used in a version of the Linux operating system for Apple's PowerMac computers.

9)

In personal computers, a _____ is a logical division of a hard disk created so that you can have different operating systems on the same hard disk or to create the appearance of having separate hard drives for file management, multiple users, or other purposes. A _____ is created when you format the hard disk. Typically, a one-_____ hard disk is labelled the "C:" drive ("A:" and "B:" are typically reserved for diskette drives). A two-_____ hard drive would typically contain "C:" and "D:" drives. (CD-ROM drives typically are assigned the last letter in whateversequence of letters have been used as a result of hard disk formatting, or typically with a two-_____, the "E:" drive.)

10)

In a computer file system that is organized as a hierarchy or tree, the _____ directory is the directory that includes all other directories. (Unlike a real tree, a tree file system has only one _____!) In Unix-based as well as in other operating systems, the _____ directory has no name. It is simply represented by the special character that separates directories in a file system .

11)

In general, a _____ is a machine designed for a purpose. In a general context, a computer can be considered a _____. 2) In the context of computer technology, a _____ is a unit of hardware, outside or inside the case or housing for the essential computer (processor, memory, and data paths) that is capable of providing input to the essential computer or of receiving output or of both. When the term is used generally (as in computer _____) it can include keyboards, mouses, display monitors, hard disk drives, CD-ROM players,] printers, audio speakers and microphones, and other hardware units.

12)

In computers, a _____(pronounced MAHD-yewl, from Latin modulus, diminutive of modus, the measure or way of being) is, in general, a separate unit of software or hardware. The term was first used in architecture.1) In computer programming, especially in older languages such as PL/1, the output of the language compiler was known as an object _____ to distinguish it from the set of source language statements, sometimes known as the source_____.