

# Corpus Linguistics in the teaching of ESP and Literary Studies

*Maria José Pereira de Oliveira*

## I. INTRODUCTION

This paper aims at sharing my experience with corpus linguistics and the ways I have been using it as a teacher of both ESP and of graduation courses in literary studies.

As a teacher of English for Specific Purposes (ESP) and a researcher, I have written a few articles on the subject, where I show how much corpus linguistics has been adding to both knowledge and methods of teaching English for science and technology. I have also carried out research on relevant language in the fields of agrarian studies, i.e., the discourse of agrarian science and technology.

Nowadays, saying that most language research and teaching are based on corpus linguistics has become a commonplace. However, those who belong to a younger generation of researchers will probably be unable to evaluate the feelings of those to whom this methodology has turned up as a kind of magic, which opens the wide gates of a new era for language research. Such are my feelings towards corpus linguistics.

## II. DEVELOPMENTS IN THE TEACHING OF ESP AND THE USE OF CORPORA

### 2.1 Some considerations on the teaching of ESP

Since I started teaching ESP – an obligatory discipline in the *curricula* of the agrarian courses in the school where I teach - the Agrarian School in Santarém, Portugal -, the difficulties felt in finding methods and approaches to do my job have been enormous, let alone gaining understanding and collaboration of my colleagues. In fact, most of my colleagues, with degrees in Science and Technology, do not agree with the teaching of English in the kind of courses taught in agrarian schools. Some of them have even stated that they are against it. So, I had to prove that knowing English – especially ESP – is of great advantage, mainly when youngsters start looking for a job. Apart from that, some of the students were not at all in favour of learning ESP.

In contrast, I was aware of how an attractive methodology could be a way of getting students interested in ESP classes. I had to protect my own territory and develop strategies to make my colleagues and students change their minds. To do so, step-by-step, I started to engage teachers

of Science and Technology subjects in the provision of texts for English classes whose content would suit their own disciplines and therefore would be interesting to students. Such a strategy worked and it did not take long for ESP to be finally accepted. From the description above, it can be derived that my first years at the Agrarian School of Santarém, Portugal, were not the easiest in my career.

Another aspect regarding ESP, and as far as my country is concerned, is that quite a few academics experience some difficulty in accepting the idea of ESP out of the context of current English, in spite of it having a fairly long history already (Cunha, M.I, 2001). For those scholars, English is English and that is it. As for myself, following Swales' ideas (1990), I think that there are niches in languages that arise from genres where ESP fits. In my opinion, ESP has to be faced as a set of aspects of the language – English - that have to be considered, analysed and studied apart from the current language, but without neglecting it. In this way, there is a lot to be done by teachers, both in class and as "homework". They have to be sensitive enough to have the ability to extract from the current language the particular aspects of the scientific, technological or academic discourse that mainly characterise it.

From my experience, I must say that several structures are considerably more frequent in certain scientific and technological contexts - medical, agrarian, biological, etc. - than they are in general English. Such structures have been and will be the object of my study, ultimately aiming at students' successful acquisition of specific lexis, which would make their academic and professional lives easier. This has been, in fact, my priority, though students of scientific and technological courses tend to dislike studying languages.

As a result, the task of approaching structures in ESP and identifying specific lexis has been part of my job before and after corpus linguistics has come into my professional life.

In the past, before the regular use of corpus linguistics as a helpful methodology, allowing the observation of concordance lists, the extraction of collocates and further statistic study, among other facilities, it was much more difficult to recognise the recursion of structures and lexis. Corpus linguistics has indeed allowed to confirm that the passive voice, present tense, conditionals and comparisons are the most frequent structures in a specific discourse, where those features are relevant and, as a result, teachers should stress their relevance within the context where they are used. For instance, in a 1.005.517 word corpus of Meat Technology, which I myself have made, it was easy to notice that the verb *to be* has high frequency as an auxiliary of the passive voice, which is exemplified in the following table:

<b>Occurrences of TO BE in the corpus MT</b>				
<b>IS frequencies</b>		<b>ARE frequencies</b>		<b>AM frequencies</b>
<b>total corpus</b>	<b>passives in the 1st</b>	<b>total corpus</b>	<b>passives in the 1st</b>	<b>total corpus</b>
	<b>3000</b> <sup>[1]</sup>		<b>3000</b>	
13.284	685	8.126	694	51

(1,31%)	(22,8%)	(0,80%)	(23,1%)	(0,005%)
---------	---------	---------	---------	----------

In addition to the above structures, there are areas, mainly in specific lexis and terminology [2], which a teacher of English will not be bound to teach in regular classes of current English, because this is the role of teachers of ESP or more widely of LSP. Some linguists, for example, Jennifer Pearson (1998), who have carried out research in these fields, developed the idea of sub-languages, where ESP fits. Fortunately, interest in terminology has been increasing and terminologists are sensitive to the establishing of contexts.

The above overall considerations portray my views on the meaning and teaching of ESP. They present a short description of what has been my experience for the past sixteen years as a teacher of ESP, before I have adopted corpus linguistics.

## 2.2 Corpus Linguistics

As there is a diversity of courses in the school where I teach, to make things worse I am the only teacher of ESP, with a wide range of subjects to deal with. Thus, the methodology of corpus linguistics in language studies has improved the teaching of ESP, in particular by presenting an effective tool in my work with students specialising in different subjects.

The times when all the teaching material was hand-written or presented in transparencies projected from an OHP, which I still use occasionally, are gone. In those past times, I could not, for lack of teaching time (two hours a week), use more than just three or four texts during the whole scholar year. Before I could count on the collaboration of my colleagues, I, with a background in the humanities, had to select specialised texts to study with my students because, being 1<sup>st</sup> year students, they were unable to make a proper selection. When they come to the agrarian school they are not familiar with the scientific and technological subjects they will have to study during the college course.

As a result, when I first came across corpus linguistics and realised its potential, I started to see a light at the end of the tunnel. However, at the beginning most of the students were not familiar with computers as well as they are now. All they used to know was how to play games and chat with friends on the Internet. I am far from thinking that these activities are not important, on the contrary. The knowledge of basic and informal use of computers has proved to be helpful to more formal activities in class, and enables students to adapt themselves more easily to applying it for language learning purposes, and to developing their skills. This is, of course, because when we talk about corpus linguistics today, it is immediately assumed that

[3]  
we are talking about computerised texts, i.e., corpora .

Generally speaking, in many Science and Technology courses, Computer Science is a compulsory subject, but when our students are in their first year in college, they know little about it, if anything at all. Thus, the first step to be accomplished in order to prepare students to work with corpus linguistics is to give them some training on how to use computers for learning purposes, not for fun. They must be familiarised with the software they will be working with to learn ESP, using specialised corpora linguistics. Therefore, my first English classes have become computer classes instead of language classes. This is an imperative even though there still are some students who do not like working with computers as a learning tool, which seems an absurd kind of attitude in present times! Fortunately, those who enjoy working with computers and understand the need for a discipline like ESP in their courses outnumber those

who don't.

In contrast, there is a relevant point that in my opinion deserves to be stressed. I mean what kind of research teachers of ESP have to carry out when willing to work with corpora in class, since they are not experts in other subjects, nor are their students linguists. So, it will be the teacher's task to build up a suitable corpus for each one of the specialisms the language of which he/she will teach. This is a time-consuming task, as most of the time there are no specific corpora available, or corpora that might provide part of the material needed for the ESP class material. But it will not be a waste of time. On the contrary, teachers will be saving time in class.

In cases when teachers are unable to extract specific material from existing corpora, there is the Internet where, fortunately, a lot of articles can be found, and a selection of those can be compiled as a corpus in order to fulfil the needs in the teaching of ESP.

## 2.3 Corpora to be used in the classroom

At this point a question can be put forward: what kind of corpora should be built up – either from the existing or new material – to be relevant and beneficial for the students in an ESP class? This question, in turn, raises another one: how about the parameters the corpus should observe, if one wants to follow the “rules” various linguists have suggested for developing a corpus? Is it reasonable to expect that a teacher collects a certain amount of texts and calls them a corpus in order to meet his/her classroom needs?

My answer is ‘yes’, and, in my opinion, a corpus to be used in the classroom should take account of two fundamental aspects:

- The area of the course where ESP is taught;
- Variation.

Variation is one of the parameters given much attention to by linguists, Biber (1988)

<sup>[4]</sup> among others, who point out dimension, balance and representativeness as parameters that have to be taken into consideration for a corpus to be considered as such. In fact, variation may be one of the easiest parameters to observe, when talking about small corpora. The great difficulty with maintaining variation is faced when dealing with large corpora, which should follow all the parameters referred above.

The use of the Internet, as I said before, may help to solve the problem. A collection of articles on a number of topics found therein can produce an acceptable sample of the target discourse. These materials are published by experts on their own web pages or by entities in the fields of agriculture (I am particularly considering the case of an agrarian school, but it can be applied to other areas), such as the Ministry of Agriculture from any English speaking country; agricultural associations, radio or TV communications, and transcripts of interviews. Such articles can be a reliable source for the compilation of a corpus to be used in the classroom. Doing so, and disregarding representativeness – considered to be important when a corpus in

<sup>[5]</sup> being built by several linguists, teachers will be able to find enough material, with a minimal effort, for their ESP classes, applying corpus methodology from the beginning of the academic year.

In the case of corpora for teaching purposes, representativeness may be solely implied by the subject of the course to be taught. However, caution must be used regarding the concept of representativeness, and teachers should be aware of cases when insufficient material is treated as a corpus that represents any language or part

of it. But this should not prevent us from using a small amount of text, which may still suit our purposes. This is possible because all that is required of a corpus to be used in the classroom is that it demonstrate characteristics of both recursion of scientific and technological discourse structures, as well as the specific lexis and terminology in the subject area. It should allow for the building up of exercises that are given to students as paradigms of the scientific or technological discipline they are studying. The role of the teacher will be to draw the student's attention to those aspects that may be taken as paradigmatic of a discourse that is only a niche of the language (Swales, 1990).

Students, in turn, should be able to know how to use suitable software to extract further information on the subjects of their field of studies by the end of the academic year, and apply their knowledge whenever necessary. Special attention should be given to ways of identifying terminology, when structures are detected as scientific and technological discourse in general. Having practised such aspects, students themselves should become researchers at a certain point, not too far from the beginning of the academic year, and the teacher should then become a supervisor, directing the class and providing guidance to students in their research tasks. This means that they should be prepared to go on using corpus linguistics as a tool in their future professional life, including the ability to select suitable material. Besides, they will find out how easy and fascinating it is to work with corpus linguistics, a methodology that can provide the reward of new findings in the language they are learning.

As a researcher and a teacher I have been working with corpora for about eight years, and all I have said so far results from my experience. From my work with students of different academic specialisms at Santarém, Portugal, I can say that I have worked with both large and small corpora, large corpora - for research purposes and small corpora mainly for classroom use. As a researcher in the fields of lexis in meat-processing technology, I have built up a corpus, the "meat technology corpus", with 1.005.517 words. It would make no sense if I did not use this corpus to the benefit of my students. I have been using it in my classes since 2001, mainly as a complementary tool to dictionaries where technical terminology in the domain of meat technology would be hardly found.

As an illustration, I am presenting an exercise prepared with material from the corpus of meat technology, and which I gave my students in a test.

#### EXERCISE:

1. Fill in the gaps with a word from the list: *T-bone; tenderness; marbling; stunning; tumbling.*
2. Write down the numbers of the lines that belong to the context of meat technology; justify your answers.

1	90% of the plants were able to do this.		was scored in 41 Federally inspected
2	Historical repeatability of sensory		ratings also was provided for comparison
3	old whether it be		over rocks and splashing
4	Same technology may be used to predict		in the longissimus muscle on a
5	That she could use her		as skill. To convince
6	an excellent performance of a		concerto. ROBIN NEWTON
7	The lumbar vertebra has the typical		configuration formed by the late
8	8 of massaging and tumbling. In general,		is recommended for firm-textured meats
9	and I mean they had two		looking girls right, but

10	Nd products from different species Since		has generally not been a concern
11	the nineteen nineties with the		costs of hardware we
12	Efficacy, insensibility of animals hanging		variables evaluated at each plant were
13	among the various muscle groups.		provides mechanical energy to the food sy
14	s , exchanging for their long-held		their new-found passion.
15	percent less cholesterol than broiled		Steak, and 35 percent less fat
16	Guidelines for visual assessment of pork		have been clearly outlined Jones et

### III. CORPUS LINGUISTICS IN LITERARY STUDIES

Another area where I have obtained interesting and successful results, and which I would like to give a brief account of, is language analysis for literary studies.

Recently, I have been teaching graduate courses (MPhil courses) to students who are mainly interested in literary analysis. Some of these students have to analyse literary works in order to characterise them, decode the authors' message and the motivation behind the writing of the work, relevance of the characters for the development of the plot, and even the description of attitudes by analysing the language used.

Others are analysing cultural aspects of literary works, focusing on the way their authors deal with conventions and customs at a given time; while other learners concentrate on such themes as the usage of language by politicians, trying to analyse how they make their speeches in order to show power, or fear or even dominate or engage the addressees in the subject they develop. Firstly, I had to introduce those students to corpus linguistics, even though they already had a degree in modern languages and were also teachers. I taught a 60-hour course, at the end of which I perceived that the students were keen on corpus use. In fact, they enjoyed all the activities that corpus linguistics could offer them: extracting and using frequency lists, keywords, clusters, concordance lists and collocations. Using computers and suitable

[\[6\]](#)

software enabled them to achieve conclusions in a way they had never thought of. Moreover, they were absolutely overwhelmed by unexpected findings, which were of great relevance to the work they were performing. The following examples of the themes analysed through corpus linguistics will illustrate my words.

The essays graduate students produced were done as a basis for future development in their dissertations. In the essay *Personal pronouns in the political speeches of President Bush*, a student intends to prove that linguistic choices in George W. Bush's speeches (regarding the use of personal pronouns) have a specific purpose and are socially and ideologically loaded. After analysing frequency lists of Bush's speeches compiled from the Internet, and organised into a small corpus, the graduate student came to the conclusion that it would be relevant to choose the pronoun *we* as a node word – *we* includes the person who is addressing the nation (or the world) and the entire nation (or the world). As Melrose (1995) puts it, «as the name suggests, the use of inclusive “we”, presupposes common ground between text producers and assumed readers/listeners.»

By analysing concordance lists and clusters, and all the material she was able to obtain, my student could show people's anxiety about the way the world is moving towards violence and insecurity, as well as the naturalization of hegemonic ideologies that contribute to the growth of some countries' power. Another example in the use of corpus linguistics for literary studies is an essay on *The loss of identity in Brave New*



*World by Aldous Huxley*. The aim of this essay is to prove that the loss of identity is one of the main themes Aldous Huxley approaches in *Brave New World*. In order to prove lack of identity in this novel, four words in a 64.971 token corpus were analysed: "God", " Ford", " John" and "Savage". The corpus is the work itself, and each chapter was saved in a different file in order to facilitate the analysis. Then, the study of the pairs *God/ John* and *Ford/ Savage* is concerned with the two distinct worlds that are described in the novel, where the brave new world would never exist if it was not the creation of mass production by Ford, who, on the other hand, nullified the need of God. Such findings were obtained through the observation of concordance lines and further analysis of collocates of the above pairs. Of the many essays produced by my students in literary studies, I shall give one last example. It is an essay on *The Portrait Of Women In Shakespeare's Comedy Plays* where the student tries to prove that the female characters in Shakespeare's plays are confined to the domestic sphere, that is, regarding their approved roles in the structure of society. Women were portrayed as helpless beings and were cast in the shadow of a patriarchal society. Language and literature were seen as symbolic representations of this social and historical reality. The student's aim is present a semantic overview of how women are defined in Shakespeare's comedies. He compares samples of the language used to portray women in relation to men and collects words that co-occur with his chosen node words, vital for justifying the analysis. Therefore, through certain expressions used in Shakespeare's writing, he tries to prove that some terms and expressions are more often used to define women. He compiled a corpus for his analysis from three Shakespeare's comedies, *The Twelfth Night*, *The Taming of the Shrew* and *Much Ado about Nothing*. The collected collocations and clusters enabled him to show how women were designated according to the relationship established with the opposite sex. Through the node words, such as, *woman, lady, daughter, wife* and *mother* and of course, *man, son, husband* and *father*, the female characters in Shakespeare's plays are confined to their approved roles in the society of the time.

#### **IV. Conclusion**

I hope I have been able to convey part of my experience as a teacher who has enthusiastically adhered to the methodology of corpus linguistics. It has been of great help in my classes of ESP and graduation courses. But most relevant of all is that I have managed to pass my enthusiasm on to my students, mainly those who are already teachers and share my feelings about the adoption of corpus linguistics for both purposes - teaching and research. Their work has demonstrated how useful corpus linguistics can be in areas other than Linguistics.

#### **BIBLIOGRAPHY**

BEX, Tony (1996). *Variety in Written English, Texts in Society: Societies in Text*. London: Routledge

BIBER, Douglas (1993a). "Using Register-Diversified Corpora for General Language Studies". In *Computational Linguistics*, volume 19, n° 2, pp. 219-240

---- (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press

CHAFE, W. (1992). "The importance of Corpus Linguistics to Understanding the Nature of Language". in : Svartvik J. (ed.), *Directions to Corpus Linguistics. Proceedings of the Nobel Symposium 82, Stockholm*. Berlin/New York: Mouton de Gruyter.

CUNHA, M.I. (2000). *Estudo diacrónico do IFE em Portugal: a problemática dos erros lexicais dos alunos dos ISCAS* (PhD theses. Unpublished)

KENNEDY, Graeme (1998). *An Introduction to Corpus Linguistics*. London: Longman

McENERY, A. M. and Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press

McKENNY, John. (2000). "Wordsmith Tools: The Vademecum of the Intrepid Traveller through Academic Domains". In *Actas do VIII Congresso Luso Espanhol de Línguas Aplicadas às Ciências e às Tecnologias*, 8-11 de Julho de 1998, Santarém: ESAS, pp.211-212

MELROSE, Robin, (1995). *The Communicative Syllabus: A Systemic-Functional Approach to Language Teaching*. London: Pinter

PEARSON, J. (1998). *Terms in Context*. Amsterdam: John Benjamins Publishing Company

PEREIRA DE OLIVEIRA, M.J. (2000) *O Recurso a corpora linguísticos para o estudo de léxico específico: o léxico da tecnologia da carne em Inglês*. (Universidade de Aveiro - PhD theses) unpublished

SINCLAIR, J.M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press

---- (1999). *New Developments in Corpus Linguistics*. Birmingham and Tuscany: University of Birmingham and TWC

SWALES, John M.( 1990). *Genre Analysis*. Cambridge: Cambridge University Press

## NOTES

---

[1] I have used occurrences within the first 3000 words, an arbitrary choice.

[2] I distinguish both specific lexis and terminology because I have come across words which are greatly used in scientific and technological texts that have a different meaning from current language use and which are not terms.

[3] For further information on developments on corpus linguistics see Kennedy 1998

[4] Sinclair (1991); Kennedy (1998); McEnery and Wilson (1994).

[5] MacEnery and Wilson (1994); Bex (1996)

[6] The software used in my classes has been Mike Scott's *WordSmith Tools*





jñ Web jñ esp-world.info