

ENGLISH TEACHING MATERIALS AND CORPUS LINGUISTICS

Thereza Cristina de Souza Lima¹
Rejane Koppe Rolim Ritter²

Abstract: The purpose of the present article is to provide language teachers and students, as well as those interested in linguistic research, with a general overview about corpus-based materials, which many scholars consider the latest tendency in English language teaching. Based on Corpus Linguistics research, such materials intend to meet the needs of a new generation to whom English is essential, as it is part of their lives. Starting with an overview of the learning clientele profile, the text then offers a simplified definition of Corpus Linguistics and its object of study, the language use, presents a short historical background, focuses on the use of its findings by material developers and finally states how learners can benefit from this approach.

Key-words: Corpus Linguistics; corpus-based materials; English language teaching.

Resumo: O objetivo do presente artigo é fornecer aos professores e alunos, bem como aos que se interessam por pesquisa lingüística, uma visão geral sobre materiais baseados em corpus, os quais muitos estudiosos consideram última tendência na área de ensino de língua inglesa. Com base em pesquisas sobre Lingüística de Corpus, tais materiais procuram atender às necessidades da nova geração para a qual o Inglês é essencial, uma vez que faz parte de suas vidas. Começando com uma visão geral do perfil da nova geração de aprendizes, o texto então oferece uma definição simplificada de Lingüística de Corpus e de seu objeto de estudo, o uso lingüístico, apresenta também uma breve explanação histórica, enfoca a utilização de suas descobertas por elaboradores de materiais e, por fim, mostra como os aprendizes podem se beneficiar desta abordagem.

Palavras-chave: Lingüística de corpus; materiais baseados em corpus; ensino da língua inglesa.

¹ Mestre e Doutoranda em Estudos Lingüísticos, Universidade Estadual Paulista (UNESP); Professora de Língua Inglesa na FACINTER e Professora convidada de Lingüística de Corpus no curso de Pós-graduação em Estudos da Tradução da PUC/PR; e-mail: tcslicristina@gmail.com

² Professora do Colégio Militar de Curitiba; Mestre em Língua Inglesa, Universidade Federal do Paraná; e-mail: reja_rolim2@hotmail.com

Introduction

The present article addresses teachers and students of English and those interested in the material design issue. Its main purpose is to provide an overview of the latest tendency in the English teaching area: a corpus based approach. Lately, as long as material design is regarded, there has been a deep concern with the real language use.

English has become a window to the world, an open door, enabling this connected clientele to communicate in real time and access widespread information, written in English on the internet. Such necessity has lead the most well-known material producers look for reliable sources of written and spoken English to be used as raw material for materials production. This way, scholars all over the world have started researching in order to meet the needs of a generation with plenty access to real, updated language through the cable TV, on-line chats and a whole universe on the internet, as learning English has become more than a necessity, it is part of people's lives. This way, language authenticity should be taken into account as an important underlying principle. Commercially speaking, credibility would be a differential to gather the clientele. However, there should be a trustworthy way to deal with such sources. This was the beginning of the link between material design and Corpus Linguistics.

Theoretical base

As the terms “corpus” (and its plural form “corpora”), and “corpus linguistics” are going to be mentioned extensively in this text, some clear term definitions are necessary. According to Kennedy,

“In the language sciences a corpus is a body of written text or transcribed speech which can serve as a basis for linguistic analysis and transcription. Over the last three decades the compilation and analysis of corpora stored in computerized databases has led to a new scholarly enterprise known as Corpus Linguistics.”
(KENNEDY, 1998: P.1)

Corpus linguistics, roughly speaking, is an approach to the study of language which is based on bodies of text (corpora) as the source of evidence for

linguistic description and argumentation. It has also come to embody methodologies for linguistic description in which the quantification of the distribution of linguistic items is part of the research activity. An area which has profited a lot with corpus research is the ESP (English for Specific Purposes), which is an approach to the English teaching/learning as both second and foreign language whose course content is tailored to meet the students' needs, in terms of what language (specific style, register, jargon, etc) is going to be used in specific realms like the reading of scientific reports or for the communication with English speakers. Moving away from specific needs, this concern about the language use has also touched famous publishers such as Oxford, Longman and Cambridge to please a general clientele, who learn English from textbooks, workbooks, listening CDs, CD ROMs, sites, etc, and wish to have an overview of the English varieties used in different contexts.

It is known that before the use of corpus analysis, materials developers used to rely only on their intuitive senses of what students needed to learn. Corpus Linguistics applications have shown a new perspective for decision-making in the difficult task of content selection and grading. Besides that, learners' motivation, an underlying principle for materials development, has also improved since the use of authentic texts about different subjects, illustrating different registers, and contextualized language activities based on real language made materials more attractive, thus abolishing the most common students' complaint that the language of the books was not what they found in the internet chats and cable TV series and other media environments. They are absolutely right – the language was mostly outdated and bookish, considering the primary aim of language learning, which is communication.

One can find in the Wikipedia site that Corpus Linguistics is the study of language as expressed in samples (corpora) or “real world” text, through which it is possible to derive a set of abstract rules by which a language is governed or else relates to another language. Opposite to Noam Chomsky's language study as the result of careful analysis of small speech samples obtained in a highly controlled laboratory setting. Corpus Linguistics scholars believe that reliable language analysis best occurs on field-collected samples, in natural contexts, with minimal experimental experience. As Leech (1992: 107) has noted, the focus of

the study is on performance rather than competence, and on observation of language in use leading to theory rather than vice versa.

According to John Sinclair (1961), when the first research about Corpus Linguistics started, it was considered impossible to process texts with millions of words; later people used to think it was possible, but still lunatic. Today we may say it is quite popular. A landmark in modern corpus linguistics was the publication by Henry Kucera and Nelson Francis of the *Computational Analysis of Present-Day American English* in 1967, a work based on the analysis of Brown Corpus, a carefully compiled selection of current American English, totaling about a million words drawn from a wide variety of sources. A further key publication was Randolph Quirk's "Toward a description of English Usage" (1960) in which he introduced *The Survey of the English Usage*".

The first dictionary compiled using corpus linguistics was the *American Heritage Dictionary*, an innovative step of combining prescriptive elements (how language should be used) and descriptive information (how it actually is used). Following this path, the British publisher Collins' COBUILD monolingual learner's dictionary, designed for users learning English as a foreign language, was compiled using the Bank of English. The *Survey of English Usage Corpus* was used in the development of one of the most important corpus-based grammars, the *Comprehensive Grammar of English* (Quirk et al. 1985).

Recently, well known publishers like Oxford, Cambridge and Longman have updated their English materials, producing corpus-based materials, such as dictionaries, grammars and textbooks containing authentic texts, contextualized exercises and real language situation dialogues, updated recorded conversations in listening CDs, attractive activities in CD ROMs and internet sites.

Being corpus based, that is, built on the basis of real language transcription and analysis, English materials provide authentic language taken from huge and trustworthy data banks. This, in turn, enables the present learner generation with specific English language needs, to communicate, access information (and exchange) and enjoy entertainment. Besides, there are readings and exercises based on real examples and authentic dialogues and texts to offer students an opportunity to discover features of language use. In other words, Corpus Lin-

guistics applications provide English students with access to the facts of authentic language, which come from real contexts rather than being constructed for pedagogical purposes.

A great amount of time could be spent on examples to illustrate the importance of Corpus Linguistics in language teaching. However, due to several factors, including time, two examples will be presented here.

The first one is related to the study of collocations. According to Berber Sardinha (2004), “collocation is an association of lexical items which does not happen at random, or aleatorily, that is to say, this association occurs more often than expected³.” For example, when we talk about dairy products, such as butter, which have become too old to be eaten, the collocation “rancid butter” is generally used, so it is very useful for students to learn both words together, once the possibility of listening, reading and using them together has been proved by Corpus Linguistics to be very strong.

The second example is related to positive/negative connotation of words, whose research concerning Corpus Linguistics is named Semantic Prosody, which John Sinclair defines as

... attitudinal, and on the pragmatic side of the semantics/pragmatics continuum. It is thus capable of a wide range of realization, because in pragmatic expressions the normal semantic values of the words are not necessarily relevant. But once noticed among the variety of expression, it is immediately clear that the semantic prosody has a leading role to play in the integration of an item with its surroundings. (Sinclair 1996: 87)

For instance, when foreign students use the verb “happen”, mainly in the question, “What happened?”, they may be talking about a positive or a negative situation. However, according to Corpus Linguistics, the verb “happen” is much more frequently used in a negative situation than in a positive one. But how do foreigners, that is, our language students, realize this? In fact, they do not, as it was found out by Corpus Linguistics researchers.

3 Tradução nossa.

Final considerations

The primary aim of this study was to present a quick outline on Corpus Linguistics and its contribution to the language teaching/learning area as far as English materials were concerned. Its focus was the application of Corpus Linguistics findings to enrich textbooks, grammars, dictionaries, listening CDs, CD ROMs, etc. with real language, enabling those learners whose main access to English was through didactic materials to acquire the desired knowledge in order to enjoy their learning, and interact with a whole universe available in English.

According to the Brazilian culture, one may know who someone else is by looking into his/her friends. Based on this old Brazilian proverb, it is possible to have an analogy with Firth's words "You shall know a word by the company it keeps" (FIRTH, 1968), which is the key to the studies of collocation. In addition to investigating collocation, we have seen that Corpus Linguistics may be really considered an open window or even an open door to the future of language teaching.

BIBLIOGRAPHICAL REFERENCES

BERBER SARDINHA, Linguística de Corpus. São Paulo: Editora Manole Ltda., 2004.

FIRTH, J. R. Linguistics and Translation. London: Palmer Press, 1968.

KENNEDY, G. An Introduction to Corpus Linguistics. New York: Addison Wesley Longman Limited, 1998.

LEECH, G. An Introduction to Corpus Linguistics, Cambridge: Cambridge University Press, 1992.

PARTINGTON, A. Using Corpora for English Language Research and Teaching. Amsterdam: John Benjamins, 1998.

SINCLAIR, J. (1966) Beginning the Study of Lexis, In Memory of F. R. Firth, London: Longman, 1966.

SINCLAIR, J. Corpus Concordance Collocation, Oxford: Oxford University Press, 1991.

www.wikipedia.com.br (Last access on 28th July, 2009)