

PERSONAL ADVISORS FOR BUSINESS PROFESSIONALS: IMPROVING COMMUNICATION SKILLS WITH ONLINE CORPORA

CONSULTORES PESSOAIS PARA PROFISSIONAIS DE NEGÓCIOS: DESENVOLVIMENTO DAS APTIDÕES COMUNICATIVAS DE CORPORA ONLINE

NATALIYA GODINHO SOARES VIEIRA, PhD Researcher, Portuguese Centre for Global History/CHC, Faculty of Social Sciences and Humanities, NOVA University of Lisbon, Portugal

ABSTRACT:

The business sphere is a multilingual world where foreign language communication skills are crucial in international relations. It makes employers look for business professionals who have a high level of linguistic competences. Language proficiency increases the chances of negotiation among partners. There are mainly two obstacles that make barriers in formal communication in a foreign language: lack of knowledge of specific linguistic structures or terminology and frequent transitions from one language to another.

This paper contributes to the quest for quick access to a wide range of English, Spanish and Russian online databases that provide authentic language samples. Their application may improve communication skills and facilitate preparation for business discourse.

KEYWORDS: Business communication; Foreign languages; Online corpora; Communication skills; Translation equivalents.

RESUMO

A esfera dos negócios é um mundo multilingue, onde a capacidade de comunicação com recurso a línguas estrangeiras é crucial nas relações internacionais. Por esta razão os empregadores procuram profissionais com um elevado nível de competências linguísticas. A proficiência nas línguas aumenta as possibilidades de negociação entre parceiros. Existem dois obstáculos principais criadores de barreiras numa comunicação formal em língua estrangeira: a falta de conhecimento das estruturas linguísticas específicas ou de terminologia e transições frequentes de uma língua para outra.

Este artigo contribui para a busca de um acesso rápido a uma vasta gama de bases de dados *online* em inglês, espanhol e russo, que contenham exemplos de língua autêntica. A sua aplicação pode melhorar a proficiência comunicativa e facilitar a preparação do discurso empresarial.

PALAVRAS-CHAVE: Comunicação empresarial; Línguas estrangeiras; Online corpora; Proficiências comunicativas; Equivalentes de tradução.

1. INTRODUCTION

International relationships lead business professionals to confront situations where they need to consult foreign language dictionaries, reference books or spend some time on the Internet, searching for the proper key terms, lexical compatibility of word forms, well-formed grammatical structures, etc. In most cases, in cooperation with foreign markets, business communication occurs via e-mail. A variety of business correspondences such as commercial documents and newsletters require translation into foreign languages. Some such problems may be resolved with the help of online dictionaries or specific programs, for instance, *Google Translator* (<<http://translate.google.com/>>), *Thesaurus* (<<http://thesaurus.com/>>), *Linguee* (<www.linguee.com>), *Collins* (<<http://www.collinsdictionary.com/>>), *SDL Trados Studio* (video tutorial: <http://www.youtube.com/watch?v=w0rAA8baU_Y>), etc.

In the last decade a large number of studies in the field of Corpus linguistics have confirmed the advantages of the application of online corpora in variety of works related to foreign language perception and production (Bowker, 2000; Aston, 2001; Bowker & Pearson, 2002; Laviosa, 2002; Anderson & Corbett, 2009).

Corpus (plural: corpora) is “a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria” (Bowker & Pearson, 2002: 9). There are mainly four types of online corpora: monolingual, parallel, multilingual and multimedia.

Monolingual corpora provide users with access to a single-language electronic database. These corpora may be called *national corpora* (for example, *British National Corpus*, *National Corpus of Polish*, *Academia Sinica Balanced Corpus of Modern Chinese*, *Russian National Corpus*, *Corpus of Modern and Diachronic Spanish of the Royal Academy*, *Corpus of Contemporary American English*, *Corpus of Spanish*).

Bidirectional parallel corpora contain aligned concordances in two languages (for instance, *COMPARA – bidirectional parallel corpus of English and Portuguese*). Multidirectional parallel corpora provide aligned concordances in more than two languages (for example, *CLUVI – Linguistic Corpus of the University of Vigo*, *OPUS – collection of translated texts from the web*).

Multilingual corpora are collections of individual monolingual corpora in several languages (for example, *The PolyU Language Bank*).

Multimedia corpora may consist of audio materials as well as video recorded files with their transcriptions (for example, *ELISA – English Language Interview Corpus as a Second-Language Application*, *SCOTS – Scottish Corpus of Texts & Speech*, *BACKBONE – pedagogic corpora of video-recorded interviews*).

Every corpus has its own size, structure and design. For instance, the *Corpus of Contemporary American English (COCA)* consists of more than 425 million words; the *Business Letter Corpus (BNC)* provides 1 million specified words or phrases. Some large corpora may consist of various sub-corpora of the above-mentioned types. For example, the *Russian National Corpus* includes a historical corpus, dialectal corpus, multimedia corpus, spoken corpus, poetic corpus, as well as multidirectional parallel corpora, etc.

Corpora are annotated with morphological, syntactic and semantic information. The specific tagging systems help to describe lexical items (word boundary tagging, part of speech tagging, sense tagging, syntactic relation tagging and semantic relation tagging). As corpora are deliberately selected collections of texts, they may provide multifunctional assistance related to foreign language production, specifically, to identify correct grammatical forms or usage of articles, to select a set of descriptive adjectives, to check translation equivalents, and so on.

Corpora may serve as personal foreign language advisors or tutors. Corpora help to identify a set of concordances retrieved from different kinds of authentic discourses. This information may be applied in error corrections or in preparations of some papers.

Taking into account the fact that electronic corpora are multifunctional tools (they can include statistical, bibliographic, sociolinguistic data and other extra materials), users have to adapt to the ways they work at the initial stages. In most cases, the process of adaptation does not take much time, because the front page of electronic corpora provides visual and practical instructions.

2. GENERAL STRATEGIES FOR NAVIGATION OF ONLINE CORPORA

The nature of any one electronic corpus is different from the nature of any other one. This has to do with the purposes that each corpus serves. However, electronic corpora are applied in many different practical ways – to search for lexical items, word-forms, other language structures, to listen to audio or video records and to access their transcriptions, etc.

Technically, corpora provide a quick search of information that is systematized, classified and visible. The most electronic corpora have free access; some of them require a simple registration. There is not a definite methodology for work with corpora. It is essential to be familiar with their functions and possibilities as well as the terminologies that are frequently used in their databases. There are some important definitions:

Concordancer “is able to recover from text all the contexts for a particular item (morpheme, word or phrase) and to print them out in a way which facilitates rapid scanning and comparison. The most usual format is the keyword-in-context (*KWIC*) concordance in which the keywords

are arranged one below the other down the centre of the page, with a fixed number of characters of context to the left and to the right" (Johns, 1991: 2). *Word list*, is "a list of all of the lemmas or word-forms in a corpus, ordered alphabetically or by their frequency of occurrence" (Anderson and Corbett, 2009: 200).

KWIC display is an essential option of electronic corpora. It helps to visualise all concordances that contain a necessary item that is usually marked by another colour or specifically arranged. For instance, if you are making requests in e-mails or letters in English, you can ask advice from the *Business Letter Corpus* that is also called *Online BLC KWIC Concordancer*, following some simple steps:

- type *request* in a search string;
- click Submit;
- select the concordances you are interested in;
- apply ready models to your work.

Table 1. Some selected concordances that contain *request* from the *Business Letter Corpus* (<<http://www.someya-net.com/concordancer>>).

<div> Online BLC KWIC Concordancer Search Result Search String: contain "request" Search Corpus: 01. Business Letter Corpus (BLC2000) No. of Hit(s): 987 </div>		
18 For the afternoon of June 12.I'd like to	request	a private meeting room for ten people
19 tributations and newly acquired skills. I'd like to	request	a promotion to a newly created position of
20 As a result. I have instructed Jim Hart to	request	a refund of \$26.005.23 in sales-tax
21 a branch in (Name of community/neighbourhood). I	request	a transfer to the new branch.
22 For obvious reasons. I must	request	absolute confidentiality
23 We	request	an amendment to increase the insured amount of
24 I wish to	request	an appointment with (name or title of person)
25 We consider total compliance with this	request	an essential element of our business relationship
26 in the above Madra Isle property phoned to	request	an explanation of why he is still being billed
27 and this leaves us in the position of having to	request	an extension of credit from you

There are some examples that demonstrate how to explore languages with the help of electronic corpora.

In modern political, economic and commercial discourse in different languages we can find a lot of 'anglicisms' ("words or phrases borrowed from English into a foreign language", Oxford dictionaries: <<http://www.oxforddictionaries.com>>). The use of anglicisms may cause problems about their functioning in the texts that are not written in English. If we need to obtain statistical data that provide text samples on application of the term *leasing* in the Spanish-speaking world of policy, economics, commerce and finance, the information would be acquired with the help of the *Corpus of Modern Spanish of the Royal Academy* (*Real Academia Española. Corpus de Referencia del Español Actual*). This corpus is very suitable for verification of lexical items in the texts related to a large number of fields that are selected from different sources belonging to Spanish-speaking countries.

Table 2- A front page of the *Corpus of Modern Spanish of the Royal Academy* (<<http://corpus.rae.es/creanet.html>>): typing *leasing* in a search line.

Real Academia Española - Corpus de Referencia del Español Actual (CREA)

Consulta:

Criterios de selección:

Autor:

Obra:

Cronológico:

Medio:

Geográfico:

Tema:

To find concordances that contain the specific term *leasing*, it is necessary to type this item in a search line (Table 2), select *Política, economía, comercio y finanzas* in the option *Tema* and click on *Buscar*.

Table 3- Search results: a term *leasing* in the spheres of policy, economics, commerce and finances in the Spanish-speaking countries.

Resultado de la consulta al banco de datos

Cómo citar el CORPUS

Consulta:	<i>leasing</i> , en todos los medios, en CREA , en Comercio y finanzas
Resultado:	81 casos en 41 documentos.

Ver estadística

As a result we have obtained 81 cases of the use of the term *leasing* in 41 documents of the specific field (Table 3). After that we can click on *Ver estadística* and find out more detailed information about the functioning of this term in Spanish-speaking countries (Table 4).

Table 4- Statistical data: the functioning of a term *leasing* in the spheres of policy, economics, commerce and finance in Spanish-speaking countries.

OBTENCIÓN DE EJEMPLOS

Recuperar

Concordancias
Normal.

Estadísticas								
Año	%	Casos	País	%	Casos	Tema	%	Casos
2000	27.84	22	ESPAÑA	50.63	40	3.- Política, economía, comercio y finanzas.	100.00	81
1997	20.25	16	EL SALVADOR	25.31	20			
1995	15.18	12	ARGENTINA	5.06	4			
1996	10.12	8	COLOMBIA	3.79	3			
1992	6.32	5	PERÚ	3.79	3			
1994	6.32	5	COSTA RICA	2.53	2			
1989	3.79	3	PARAGUAY	2.53	2			
1999	3.79	3	PUERTO RICO	2.53	2			
1986	2.53	2	CHILE	1.26	1			
Otros	3.79	3	Otros	2.53	2			

Clicking on *Recuperar*, it is possible to obtain concordances, their contexts and data sources (Table 5). The corpus allows retrieval of the whole abstracts that contain lexical items which are being searched and provides full references to them.

Table 5- Some selected concordances: *leasing* in the *Corpus of Modern Spanish of the Royal Academy* (<<http://corpus.rae.es/creanet.html>>).

Nº CONCORDANCIA			AÑO
1 sobre dos vehículos suscritos bajo la fórmula de	leasing ,	un sistema de alquiler a plazo a fijo, que d**	1999
2 entregaron los vehículos a las propias empresas de	leasing	para que los vendieran y de esta forma saldar**	1999
3 tituirán a las actuales financieras, compañías de	leasing .	factoring y sociedades de crédito hipotecari **	1995
4 del impuesto de sociedades prevé la supresión del	leasing	para las pymes. MADRID.- El endurecimiento de **	1995
23 argo del temario. El Diario de Hoy El sistema de	leasing	(arrendamiento financiero) como alternativa **	2000
24 en esta capital, un seminario internacional sobre	leasing	el cual se enmarca dentro del proyecto de p **	2000
25 El objetivo del evento es impulsar la figura del	leasing	como una alternativa innovadora de financiam **	2000

As we see, electronic corpora may be successfully integrated into the work required for continuous improvement of communication skills, helping to update and consolidate language knowledge.

There are many ways to apply electronic corpora for different purposes. We will present some of them, paying more attention to the implementation of corpora as effective foreign language informational advisors and personal language trainers for business professionals.


3. PREPARATION FOR ORAL COMMUNICATION

Participating in negotiations and keeping in touch with foreign colleagues cause situations when business professionals need to adapt to non-native accents or different forms of speaking in the same language. Some types of audiovisual electronic corpora may help in speech-perception trainings.

Audiovisual corpora are original softwares that allow exploration of authentic oral discourse in visual or verbal forms. Audiovisual corpora contain both audio and video files. Audio corpora consist of audio files and text transcriptions without their demonstration in visual form.

In preparation for contacts with English-speaking partners, customers, buyers, etc., we can ask for help at the *English Language Interview Corpus as a Second-Language Application – ELISA* (<http://www.uni-tuebingen.de/elisa/html/elisa_index.html#topic_keys>). This corpus includes video interviews with native English speakers from various regions of the English-speaking world. There are the interviews with a city councillor of the community of New Mexico/US, a travelling businesswoman who works in publishing and advertising in Bermuda, the Caribbean, South and Central America and the Pacific, a tour guide from Australia, a language teacher from Oxford University, a photographer from Birmingham, a project manager from the West Midlands/UK, etc. (Table 6). The rich database of *ELISA* helps gain familiarity with varieties of English and to develop language awareness. This corpus lets the user observe differences in articulation of the same sounds or in the use of weak forms, as well allowing us to analyse how speakers divide their utterances into units, characterised by falling, rising and level tones and to listen to a variety of intonation types in which speakers make statements, exclamations, etc. *ELISA* provides transcribed speech of all verbal interviews which is an important aspect for self-test listening comprehension.

Table 6- *ELISA* – Browse interviews: a video interview with the owner of the horse-back riding company.

	<p>→ The 'Broken Saddle' Riding Company</p> <p>Harold is the owner of the 'Broken Saddle' horse-back riding company in New Mexico (US). He tells the story of his company, talks about his clients and ... his horses. Then he lets us join one of his riding lessons.</p>	<p>[warm-up]</p> <p>[full video] [word lists] [txt] [xml]</p>
---	---	--

Dealing with Russian-speaking partners, non-native Russian-speakers may have difficulties in pronunciation of some sounds that are not typical for the languages they are familiar with. The *Multimedia sub-corpus – MURCO* of the *Russian National Corpus* (<<http://www.ruscorpora.ru/search-murco.html>>) provides concordances from the video fragments of Russian feature films and documentaries.

For instance, *MURCO* may retrieve relevant video files with necessary lexical items or word forms in a context of full dialogs that are useful to listen to the pronunciation of the Russian initial, middle and final clusters consisting of more than three consonants (а~~ре~~нтство (agency), произво~~д~~ство (manufacture), партне~~р~~ство (partnership), филосо~~ф~~ствовать (philosophize), приве~~т~~ствовать (greet), etc).

Table 7- *MURCO* – Browse concordance of *приветствовать* (greet).

<p>11. Георгий Шенгелия и др. Классик, к/ф (1998) [омонимия не снята]</p> 	<p>[На площади] [Савицкий, Юозас Будрайтис, муж, 58, 1940] А... Дóбрый де нь... господи нписа тель. К сожалёнию/ мы... с ва ми вчера не успе ли познако митьс я/ поэ тому... не име ю возмо жности приве тствовать ва с по и мени-о тчеству [Горский, Сергей Никоненко, муж, 57, 1941] Никола й Васи лич [Савицкий, Юозас Будрайтис, муж, 58, 1940] О! Пря м как Го голь! [Горский, Сергей Никоненко, муж, 57, 1941] О! [Георгий Шенгелия и др. Классик, к/ф (1998)] [омонимия не снята] ...</p>
---	---

The abovementioned examples confirm that audiovisual corpora are very practical in their application. They response to the problems that individuals have, trying to speak foreign languages.

4. SEARCHING FOR TRANSLATION EQUIVALENTS

Corpus-driven methodology in translation is a current subject of academic researches (Baker, 1993; Laviosa, 2002; Zanettin, Bernardini & Stewart, 2003). More than two decades ago Sinclair argued that corpora were potential tools to increase the quality of translations: “The new corpus resources are expected to have a profound effect on the translations of the future. Attempts at

machine translation have consistently demonstrated to linguists that they do not know enough about the languages concerned to effect an acceptable translation. In principle, the corpora can provide the information” (Sinclair, 1992: 395).

Parallel electronic corpora have been more and more widely used by linguists, translators and specialists of other fields as good sources of translation equivalents.

There are some examples illustrating the effectiveness of application of parallel corpora in professional spheres for the purpose of translation.

CLUVI – Linguistic Corpus of the University of Vigo (<<http://sli.uvigo.es/CLUVI/index.html>>) consists of 23 million words. It has various thematic sub-corpora. For instance, the sub-corpus *CONSUMER* provides aligned concordances in Spanish, Galician, Basque and Catalan in the field of consumption (economics).

To identify the translation equivalents for a Spanish *Libro de Registro de Declaraciones*, it will be necessary to type this title in a search string and click on the option *Pescudar no CONSUMER* (Table 8). As a result we can see retrieved parallel concordances with translation equivalents in four languages.

Table 8- The starting point for a search: typing *Libro de Registro de Declaraciones* in the *CONSUMER*.

Corpus CONSUMER español-galego-catalán-euskara de información sobre consumo (5.586.431 palabras)


Palabra ou expresión en español:

Palabra ou expresión en galego:

Palabra ou expresión en catalán:

Palabra ou expresión en euskara:

Table 9- A sample of retrieved parallel concordances from the *CONSUMER*.

1-CMI (10432)	través de las anotaciones que los mismos hayan podido realizar en el Libro de Registro de Declaraciones de Voluntad o en la historia clínica, junto a un examen de la documentación y pertenencias personales que el difunto portaba.	sanitario, a través das anotacións que os mesmos puidesen realizar no Libro de Rexistro de Declaracións de Vontade ou na historia clínica, xunto a un exame da documentación e pertenzas persoais que o defunto portaba.	centre sanitari, per mitjà de les anotacions que els mateixos haguessin pogut fer en el Llibre de Registre de Declaracions de Voluntat o en la història clínica, al costat d'un examen de la documentació i les pertinences personals que el difunt portava.	borondatea zein zen, eta, horretarako, erietxeko Borondate Deklarazioen Erregistro Liburuan edo historia klinikoan idatzi ahal izan duten edozein ohar erabil dezakete, eta hildakoak zeraman dokumentazioa eta gauza pertsonalak aztertu behar dira.	
------------------	--	--	--	---	---

EUROPARL (<<http://opus.lingfil.uu.se/bin/opuscqp.pl?corpus=Europarl>>) is another multidirectional parallel corpus that may help in the search for translation equivalents. It stores database of the proceedings extracted from the European Parliament web site and includes information in 21 European languages. The significant advantage of this corpus is the possibility to retrieve aligned concordances in any chosen combinations of proposed languages. The *EUROPARL* is a good source for identifying appropriate translation equivalents in the spheres of policy and economics.

5. FINAL REMARKS

In spite of the fact that Corpus linguistics is not a new field, nowadays we are still at the beginning of the use of electronic corpora not only as specific research subjects but also as language consultants or advisers, applicable in various areas and suitable for a large number of purposes. Development and production of various electronic corpora have been in process. It is quite possible that in the near future we will have access to more advanced, complex and rich monolingual and multilingual, parallel, multimedia or other types of corpora.

In fact, electronic corpora provide quality selected information by competent specialists. Corpora are available online any time and possess quick search functions. They allow identification of full contexts and references. Corpora may be considered as potential sources for improving language knowledge and communication skills. Their application is one of the ways to facilitate busy professional lives.

REFERENCES

- [1] Anderson, W. & Corbett, J. (2009). *Exploring English with Online Corpora*, Basingstoke: Palgrave MacMillan.
- [2] Aston, G. (2001). *Learning with corpora*, Houston: Athelstan.
- [3] Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and applications. In M. Baker, G. Francis & E. Tognini-Bonelli (Ed.), *Text and Technology* (pp. 233-250). Amsterdam and Philadelphia: John Benjamins,
- [4] Bowker, L. (2000). Towards a methodology for exploiting specialized target language corpora as translation resources, *International Journal of Corpus Linguistics* 5(1), 17-52.
- [5] Bowker, L. & Pearson J. (2002). *Working with Specialized language: A practical guide to using corpora*, London and New York: Routledge.
- [6] Johns, T. (1991). Should you be persuaded – two samples of data-driven learning materials. In T. Johns, & P. King (Ed.), *Classroom Concordancing* (pp. 1-16). Birmingham: The University of Birmingham Centre for English Language Studies.
- [7] Laviosa, S. (2002). *Corpus-based Translation Studies: Theory, Findings, Applications*, Amsterdam and New York: Rodopi.
- [8] *Oxford dictionaries* (2013). Retrieved November 15, 2013, from <http://www.oxforddictionaries.com>
- [9] Sinclair, J. (1992). The Automatic Analysis of Corpora, Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4 – 8 August, 1991. In J. Svartvik (Ed.) *Direction on Corpus Linguistics*. Berlin and New York: Mouton de Gruyter.
- [10] Zanettin, F., Bernardini, S. & Stewart, D. (2003). *Corpora in Translator Education*, Manchester/Northampton: St Jerome.
- [11] Online corpora: *Academia Sinica Balanced Corpus of Modern Chinese*. Retrieved November 20, 2013, from <http://rocling.iis.sinica.edu.tw/CKIP/engversion/20corpus.htm>, <http://db1x.sinica.edu.tw/kiwi/mkiwi/>.
- [12] *BACKBONE*, Retrieved November 20, 2013, from <http://projects.ael.uni-tuebingen.de/backbone/moodle/>.
- [13] Braun, S. & Kohn, K. *et al.* (2003-2004). *The English Language Interview Corpus as a Second-Language Application*, University of Tuebingen. Retrieved November 21, 2013, from http://www.uni-tuebingen.de/elisa/html/elisa_index.html#topic_keys.

- [14] *Business Letter Corpus*. Retrieved November 20, 2013, from <http://www.someya-net.com/concordancer>.
- [15] *CLUVI – Corpus Lingüístico da Universidade de Vigo*. Retrieved November 20, 2013, from <http://sli.uvigo.es/CLUVI/index.html>.
- [16] *COMPARA*. Retrieved November 23, 2013, from <http://www.linguateca.pt/COMPARA/>>.
- [17] *Corpus del Español*. Retrieved November 20, 2013, from <http://www.corpusdelespanol.org/>.
- [18] *EUROPARL*. Retrieved November 20, 2013, from <http://opus.lingfil.uu.se/bin/opuscqp.pl?corpus=Europarl3;lang=pt;adv=1>.
- [19] *National Corpus of Polish*, Retrieved November 20, 2013, from <http://nkjp.pl/poliqarp/>.nj>.
OPUS. Retrieved November 23, 2013, from <http://opus.lingfil.uu.se/>.
- [20] *PolyU Language Bank*. Retrieved November 23, 2013, from <http://langbank.engl.polyu.edu.hk/indexl.html>.
- [21] *Real Academia Española: Banco de datos (CORDE). Corpus diacrónico del español*. Retrieved November 23, 2013, from <http://www.rae.es>.
- [22] *Real Academia Española: Banco de datos. Corpus de Referencia del Español Actual (CREA)*. Retrieved November 23, 2013, from <http://corpus.rae.es/creanet.html>.
- [23] *Russian National Corpus*. Retrieved November 23, 2013, from <http://www.ruscorpora.ru/en/index.html>.