

Using Parallel Corpora for Designing ESP Teaching Materials in Higher Education

Jelena Anđelković, Marija Novaković, Gordana Jakić

University of Belgrade, Serbia

Abstract

Language corpora have been used for the design of foreign language teaching materials and tests for several decades, especially in teaching languages for specific purposes. Using parallel corpora offers a number of possibilities in designing LSP teaching materials.

Aiming to create teaching materials for an *English for Specific Purposes* course for BA students of the Faculty of Organizational Sciences, University of Belgrade, a parallel English-Serbian corpus consisting of research articles from the journal *Management* has been designed. Based on this corpus, an overview of possibilities for using parallel corpora for designing ESP tasks and tests has been provided in this paper. The corpus allowed for an extraction of frequent terms, collocations, lexical patterns, and word chunks in the subject field of management. This was later used to design teaching materials and tests intended to improve the ESP skills of undergraduate students. In addition, the paper explores the possibilities of direct interaction between students and the corpus in the process of data-driven learning with the purpose of encouraging students' learning independence.

Key words: parallel corpus, specialized corpus, LSP teaching material design, data-driven learning



1 Introduction

The word *corpus* is used to describe “a collection of naturally occurring examples of language which have been collected for linguistic study” (Hunston, 2002, p. 2). Such text collections are nowadays machine-readable, i.e. they are stored and accessed electronically. Corpora can be used for a variety of purposes, such as comparative and contrastive language studies, translation studies, terminology extraction and analysis, second language teaching and learning, and many others.

There are different types of corpora available, such as general, specialized, written, spoken, comparable, parallel, reference, monitoring, learner’s corpora etc.

A general language corpus is a collection of spoken and /or written language data that is (or should be) balanced with regards to genres and domains that it covers. The most well-known general corpora for the English language are the *British national corpus*¹, the *American national corpus*², and the *Corpus of Contemporary American English*³. The largest general corpora for the Serbian language is the *Contemporary Serbian language corpus*⁴ (Serb. *Korpus savremenog srpskog jezika SrpKor2013*) created at the Faculty of Mathematics, University of Belgrade. This written corpus contains approximately 122 million words. It is fairly balanced (it consists of literary texts, research papers, newspaper articles, online news, and some literary and general text translations into the Serbian language), diachronic (time span: 19th century literary text–2011 newspaper articles), morphologically annotated, and available with permission granted by the authors. Another important Serbian corpus is the *Serbian Web Corpus SrWac*⁵ (Serb. *Srpski mrežni korpus*), created as a part of the *ReLDI*⁶ (*Regional Linguistic Data Initiative*) project containing approximately 555 million words. Even though *SrWac* has exceeded *SrpKor* in size, it is less balanced since all its texts have been collected from the .rs internet domain and therefore cannot be representative of the Serbian language as a whole.

A specialized language corpus, also known as LSP or LAP corpus, on the other hand, is a collection of language samples representing a language for specific purposes. For the most part, specialized corpora are *domain-specific* (thus representing the language used in particular domains, e.g. in the domain of medicine, law, civil engineering, etc.) or *genre-specific* (consists of texts or transcriptions of spoken discourse that belong to a single genre, e.g. research papers, newspaper articles, academic discussions, etc.). In some cases, specialized corpora can be extracted from general language corpora by focusing on a particular domain of language use. The most well-known English language corpus for the domain of business is the *Wolverhampton Business English corpus* (Gong, 2005, p. 16), now available with a license through meta-share⁷. Specialized language corpora are not as frequent and as widely exploited as general language corpora and are often made by researchers for a particular purpose, especially in languages of lesser diffusion (cf. Linn, 2006), such as Serbian.

A parallel or aligned corpus is a collection of texts in the source language and its translations into a foreign language. Generally speaking, the development of parallel corpora lags behind the progress made in monolingual corpora, since the compilation and processing of parallel texts is more complicated and usually involves sentence-level or word-level alignment.

1 <http://www.natcorp.ox.ac.uk/>

2 <http://www.anc.org/>

3 <http://corpus.byu.edu/coca/>

4 <http://www.korpus.matf.bg.ac.rs/prezentacija/korpus.html>

5 <http://nlp.ffzg.hr/resources/corpora/srwac/>

6 <https://reldi.spur.uzh.ch/>

7 <http://metashare.elda.org/repository/browse/wolverhampton-business-english-corpus/73439812de6811e2b1e400259011f6ea-7ba3c52b2b5648cfa1551c8009590db4/>

When it comes to parallel corpora available for the Serbian language, two aligned text collections, a French–Serbian *Aligned Corpus* (Serb. *SrpFranKor*) and a Serbian–English *Aligned Corpus* (Serb. *SrpEngKor*) have been created by the Faculty of Mathematics, University of Belgrade and now available through meta-share⁸. In addition, the aligned text collection search tool *Bibliša*⁹ contains several aligned text collections in Serbian and English (Stanković et al., 2012, p. 1710). Each collection is at the same time an aligned and a specialized corpus and is both genre-specific (all the texts are aligned research papers) and domain-specific (each collection represents a single domain: librarianship, geology, architecture, dentistry, management, etc.).

So far, corpora have been applied in comparative and contrastive language studies, lexicographical, terminology, translation studies, studies of grammar, sociolinguistics, pragmatics, semantics, discourse analysis and, most recently, in the field of second-language teaching and learning. In addition, each of the abovementioned types of corpora seems to be more suitable for a particular purpose, i.e. for the empirical research the researcher intends to conduct. Thus, parallel corpora are most frequently used in translation studies (for the extraction of concordance sentence pairs for translation) and for comparative and contrastive language studies, while specialized corpora found its most important application in terminology studies (for terminology research and the extraction of term equivalents in another language). In addition, corpora that are both specialized and aligned are very frequently used for automatic or semi-automatic terminology extraction and compilation of term bases and other terminographic products.

The following paragraphs will deal with a relatively new application of corpora, especially of specialized parallel corpora in LSP teaching and learning and, more precisely, in LSP teaching material design.

2 Use of corpora in LSP teaching and learning

While discussing the use of corpora in language teaching and learning in general, Römer (2011) mentions both its direct and indirect applications. According to Römer (2011), researchers and material writers can use language corpora *indirectly* by inserting corpora findings in their teaching syllabi design, reference works (e.g. vocabularies and grammars) and teaching materials. When designing syllabi and teaching materials based on corpora, authors generally rely on frequency list information for the creation of “lexical syllabi” (McEnery, Xiao, 2010). Perhaps even more important than the direct use of language corpora is that it can be used *directly* by both learners and teachers in the process of *data-driven learning* (DDL), which allows for direct interaction between the teacher and the corpus on the one hand, and the learner and the corpus, on the other hand. However, possible obstacles to the active (direct) use of corpora by language students are, according to Aston (1996), the lack of adequate training and a relatively high level of language proficiency that is required for independent use of corpora for language learning.

In recent years, corpus-based studies have recently become the core areas of LSP research. LSP and professional communication have benefited from both direct and indirect applications of specialized corpora. For LSP teachers and learners, the most obvious advantage of the use of specialized corpora is that they provide authentic examples of specialized language, which are, in turn, beneficial in developing reading and writing skills and in understanding and producing particular texts and types of texts (Aston, 2001; Flowerdew, 2001; Connor & Upton, 2004; Römer, 2011). Ball (2001) reports on the benefits of using corpora in ESP testing, i.e. in compilation and standardization of language tests,

8 <http://metashare.elda.org/repository/browse/french-serbian-aligned-corpus/29a1e8168bdf11e28ea3001517144592d79e-355941ff4503b26966fc958a94f7/>

9 <http://jerteh.rs/bibliša/>



and illustrates this by saying that the BEC Preliminary Test is based on view by wordlists extracted from *Cambridge Business English Texts Corpus*.

Several studies have also reported the effectiveness of using corpora for creating LSP teaching materials (indirect application of corpora in LSP teaching and learning), especially those aimed at teaching and learning collocations (Howarth, 1998), lexical acquisition (Cobb, 1999), writing and grammar (Tribble, 2001). Before the use of tailor-made specialized corpora for teaching material design, it seems that LSP teachers used to rely heavily on their own intuition and/or written discourse, which did not always seem to result in pragmatically and socio-linguistically appropriate teaching materials.

Unlike monolingual corpora, the use of parallel corpora in language teaching and learning and the creation of teaching materials seems to be still in its infancy. For Danielsson and Mahlberg (2003), the main reason why a parallel corpus can be used in language teaching and learning is managing problems that occur in students' writing due to the interference of their native language. In data-driven learning (DDL), students are encouraged to make use of parallel corpora on their own, in order to discover how words, idioms, phrases, and grammatical phenomena are used in the target language, while teachers only act as facilitators in the learning process by indicating regularities in the parallel texts and thus making learners aware of grammatical, lexical, or discourse features (Granger, 1998). In addition, parallel texts can be used by learners to find potential equivalent expressions in the given languages and investigate semantic, syntactic and collocational differences. Such investigations can lead to comparative language studies (Johansson, 2007).

3 ESP corpus compilation and use: our case

3.1 ESP teaching and learning at the Faculty of Organizational Sciences

At the Faculty of Organizational Sciences, University of Belgrade, *English for Specific Purposes* is taught as part of both undergraduate and graduate studies. There are two mandatory, two-semester courses in the first and the second year (ESP1 and ESP 2), and an elective one-semester course (ESP 3) in the fourth year of undergraduate studies, as well as an elective graduate course (ESP 4).

As teachers of ESP courses at Organizational Sciences, University of Belgrade, we face large, mixed-ability classes consisting of students with varied majors (Management, Quality Management and Standardization, Operations Management, and Information Systems and Technologies) attending the same course. There is no option to group the students according to their foreknowledge, nor according to their major. In theory, all students should have B2 English language skills in order to select an ESP1 course, but, in practice, since there are no possibilities for organizing placement testing, our students' English language skills differ significantly.

Since there are neither ready-made course books nor teaching materials suitable for our heterogeneous classes, we have to create our own teaching materials that could respond to the needs of our students. So far, we have relied on our intuition and other courses' syllabi and materials in the process of selecting and creating our own teaching materials. We have also used authentic texts in their adapted or original form. These methods are, for the most part, time-consuming (especially the selection and adaptation of authentic materials) and often do not bring satisfactory results.

Inspired by research papers proving the increasing use of specialized and parallel corpora in LSP teaching, learning, and material design, we have decided to investigate the potential of a parallel specialized corpus in the creation of materials for teaching and learning ESP at the Faculty of Organizational Sciences. Starting from the assumption that a parallel corpus would be useful for vocabulary acquisition and writing exercises on the one hand, and for comparative and translation

exercises on the other, we conducted an analysis based on a domain- and genre- specific parallel corpus that had initially been compiled for the purpose of terminology analysis and extraction as a part of a PhD study.

3.2 General information on parallel corpus of management

The Serbian-English parallel corpus of management consists of research articles taken from the journal *Management*, published by the Faculty of Organizational Sciences in both Serbian and English (Serbian texts and their English translations) between 2009 and 2012. The corpus consists of 17 issues in each language, with the total of 362 documents / texts (181 per language). The Serbian part of the corpus contains 668,681 tokens and 29,326 sentences. The management corpus is both domain-specific (all the texts belong to the domain of management) and genre-specific (all the texts are research articles), i.e. it is designed to represent a sub-language of general Serbian and English as used in the subject-field of management.

3.2.1 Corpus compilation and processing

The creation of the management corpus was a demanding and time-consuming task that was performed in a sequence of steps:

a) Text selection

The main criteria for the selection of texts to be included in the management corpus were their *representativeness* and *availability*.

First, all the chosen texts belong to the research paper genre and represent empirical studies conducted in the domain of management. Therefore, we believe that the corpus is highly representative of both research paper genre and the specialized academic language used in the domain of management. In addition, research papers represent advanced writings and a complex type of discourse that exemplify an advanced level of difficulty and therefore can be a useful corpus for the design of management-specific language teaching materials, especially those aimed at vocabulary acquisition and writing.

All the *Management* research articles that were included in the corpus are publicly available in *.pdf* format on the journal's website¹⁰. The availability of journal articles online significantly simplified the process of text pre-processing.

b) Text processing

Each text was downloaded in its original *.pdf* format, converted into *.txt* format, and cleaned up of material irrelevant for our linguistic study (e.g. references, tables, graphs, tables of contents, figures, headers and footers, page numbers, reference numbers, image captions, etc.).

After the initial preparation of individual texts, Serbian-English pairs of texts (e.g. issue 60, no.1 in Serbian and issue 60, no.1 in English) had to be aligned on the paragraph level so that comparative and contrastive studies can be performed. A number of problems were encountered during this time-consuming process: such as the differences in layout of two parallel texts, missing or misplaced sentences, or sentences that had not been translated or had been inadequately translated. When the alignment process was complete, it needed to be verified so that there is a minimum amount of noise during corpus analysis.

¹⁰ <http://management.fon.bg.ac.rs/index.php/mng/about>

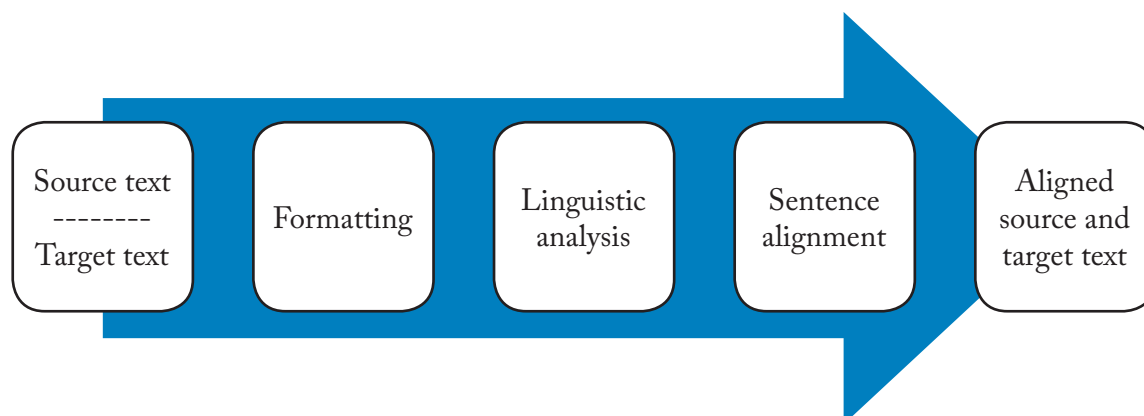


Figure 1: *Stages of parallel corpus processing*

Parallel corpus prepared in this way was ready to be analyzed by using freely available software tools and concordancers such as *AntConc* and *AntPConc*¹¹. An advantage of such tools is that they can easily generate wordlists and key word lists, n-grams and clusters, extract collocates, etc. for any given language, but a possible drawback is that they are not ideal for the analysis of highly flexive languages such as Serbian, since working with Serbian in these tools requires a lot of manual “cleaning”.

For this reason, the Serbian part of the corpus was further processed. This process included character encoding, tokenization, morphological analysis, PoS tagging, determination of sentence boundaries etc. performed by members of *JerTeh – Serbian Society for Language Resources and Technologies*¹². The prepared parallel management corpus was incorporated into *Bibliša*¹³, an aligned text collection tool developed by this society, and is now available online, together with eight other domain-specific aligned collections.

3.3 Parallel Management corpus and ESP teaching and learning at Faculty of Organizational Sciences, University of Belgrade

The parallel management corpus, though initially created for the purpose of terminology extraction and analysis, proved to be a valuable addition to the teaching and learning process and materials design at ESP courses at the Faculty of Organizational Sciences, especially for ESP 1 course taught as part of Year One of undergraduate studies to students with lower English language proficiency level. Several ways of using this corpus both directly and indirectly, some of which were tested during our classes, will be presented below.

3.3.1 Core vocabulary and teaching materials design

One of the most important applications of corpora in ESP teaching is the extraction of words and word chunks that constitute core linguistic material for our ESP setting because of their recurrent use in the corpus. We believe that such a semi-technical list of subject-matter words represents the core of specialized vocabulary that our students need to acquire at lower proficiency levels, i.e. during the ESP1 course at the Faculty of Organizational Sciences, and therefore greatly influences syllabus and teaching material design for this course.

11 Created by Laurence Anthony and downloadable from <http://www.laurenceanthony.net/software.html>

12 <http://jerteh.rs/>

13 <http://jerteh.rs/biblisha/>

To provide the core vocabulary list, we generated a general word list initially consisting of 19,643 word types by using the untagged version of the English part of the management corpus and the *AntConc* software tool. The initial list was reduced to 19,492 word types by eliminating some Serbian words and the noise resulting from the imperfect conversion of documents from *.pdf* into *.txt* format. Fifty most common words in the English sub-corpus are presented in Table 1.

Table 1: *English sub-corpus: 50 word types with highest frequency*

rank	frequency	word type
1	60356	the
2	33735	of
3	22332	and
4	18177	in
5	15924	to
6	12378	a
7	11360	is
8	7659	that
9	6405	for
10	5992	as
11	5700	are
12	5006	be
13	4958	on
14	3911	it
15	3577	with
16	3407	this
17	3276	by
18	2942	management
19	2880	an
20	2869	which

As expected, words with the highest frequencies belong to the group of function words, i.e. articles, prepositions, auxiliary words, pronouns, etc., whereas the only content word is the key word of the subject-field in question – the term *management* itself. As the frequencies drop, however, the number of content words, and especially the number of key terms, increases.

Our focus, however, was not on the general word list, but rather on the list of most frequent nouns found in the corpus. To generate this list, we expanded and reviewed the initial list of frequent words to include only the most frequent nouns, which proved to be either semi-technical words, i.e. the ones used in more than one subject-field (*system, project, development, information, and organization*),¹⁴ or words of general language (*company, business and market*) or technical terms (*management*).

The ten most frequent nouns in the English part of the management corpus are shown in Table 2.

¹⁴ By exploring the presence and the frequency of these words in the 9 text collections for various domains (e.g. geology, architecture, library and information technologies, etc.) available in *Bibliša*, it was proven that they appear relatively frequently in at least 3 of the text collections, and therefore can be considered semi-technical.

Table 2: *English sub-corpus: 10 nouns with highest frequency*

rank	frequency	word type
1	2942	<i>management</i>
2	2720	<i>project</i>
3	2139	<i>company</i>
4	2076	<i>business</i>
5	1831	<i>system</i>
6	1827	<i>process</i>
7	1674	<i>development</i>
8	1583	<i>information</i>
9	1532	<i>organization</i>
10	1505	<i>market</i>

Starting from the hypothesis that new specialized vocabulary is acquired better when taught in chunks or word clusters, we used the *AntConc Cluster / N-gram tool* to generate lists of the most common word clusters around the ten most frequent nouns related to the term *management* so that we can use the list for creating our teaching materials. We believe that word clusters, compounds, and phrases, and especially technical collocations in nominal compounds, are of extreme importance in scientific and technical language. Words acquire fixed positions in compounds and phrases, and thus form characteristic language bonds. In other words, the most frequent nouns, together with their most frequent collocations, compounds, and phrases, represent the core vocabulary that students need to acquire during their ESP1 course.

To illustrate, 30 most frequent *business*¹⁵ word chunks that appear in at least three of the research papers of the *Management* journal (range) are presented in Table 3.

Table 3: *Business word chunks*

No.	<i>Business - clusters</i>	frequency	range
1	business operations	129	51
2	electronic business	129	10
3	business processes	74	25
4	business environment	63	35
5	do business	37	26
6	business strategy	35	15
7	business activities	34	24
8	business intelligence	32	4
9	business decisions	31	12
10	e-business	30	8
11	responsible business	30	5
12	business units	28	11
13	business partners	20	13
14	business performance	20	15

¹⁵ It is interesting to note that the plural form of the word *business*, i.e. *businesses*, appears only 84 times in the entire corpus, and that it rarely appears in recognizable word chunks, as it cannot be found in the table above. The most commonly found clusters are nominal phrases (e.g. *business operations*, *electronic business*), while the only verb-noun collocation with high frequency in the corpus is *to do business*.

No.	Business - clusters	frequency	range
15	business success	17	12
16	strategic business	17	5
17	business results	16	13
18	of business doing	16	11
19	business development	14	7
20	business model	14	8
21	banking business	14	5
22	strategic business units	12	4
23	business and management	11	5
24	business decision making	10	7
25	efficient business	8	7
26	global business	8	8
27	company business	7	4
28	does business	6	6
29	retail business	5	3
30	do business with	4	4

Each word cluster can also be observed in the context by using *AntConc* concordance tool.

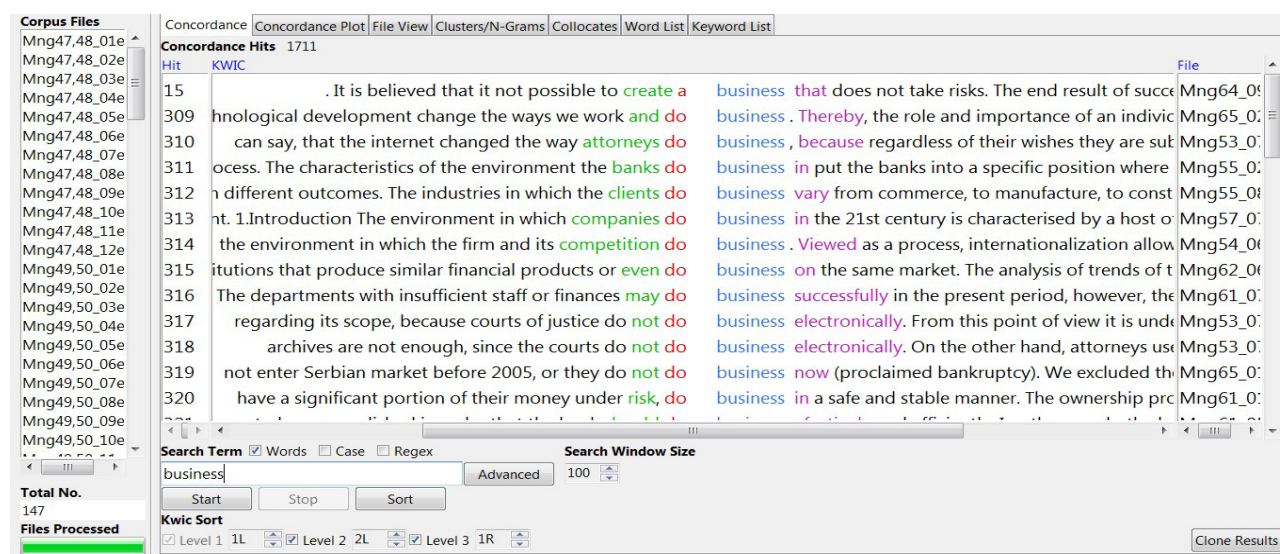


Figure 2: Printscreen of AntConc 3.4.4w concordance tool interface

3.3.2 Parallel concordances and translation practice

Since our students sometimes struggle to find an adequate translation of English lexical word chunks, especially the ones that consist of more than two noun elements (e.g. *strategic business unit*, *business decision making*), the Serbian sub-corpus of the parallel management corpus proved to be an invaluable source of translation equivalents. The *Bibliša* concordance tool enables teachers to create translation-based teaching materials, while students can be trained to use it for translation practice or when searching for translation of the most frequent clusters.

BIBLIŠA: ALIGNED COLLECTION SEARCH TOOL

Home Metadata browse Metadata search Mongo search Manage data Help Tutorial About

SEARCH DOCUMENT

Type words: Language:

Number of concordances (en): 2409		Broj konkordansi (sr): 2409
Knežević et al., 2011, vol. XVI:58, ID: 7.2011.58.6 metadata	Usual in the business practice are deferred payments in a defined term whose time span is conditioned by a usual business practice, by the stability of economic conditions, by the business stability of a company, etc.	U poslovnoj praksi uobicajena su odložena placanja u određenom roku čija je dužina uslovljena uobicajenom poslovnom praksom, stabilnosti ekonomskih prilika, poslovne stabilnosti jednog preduzeća i drugo.
Tošović Stevanović A., 2010, vol. XV:55, ID: 7.2010.55.9 metadata	Strategic alliances function as a complex and interrelated innovation system, reaching by far further into business than the traditional implementation of technology through imitation.	Strateške alijanse funkcionišu kao složen i međusobno povezan inovacioni sistem, koji u poslovanju ide mnogo dalje od tradicionalnog korišćenja tehnologije preko imitacija.
Suknović et al., 2010, vol. XV:54, ID: 7.2010.54.1 metadata	In conclusion, the data quality can seriously affect the quality of the business intelligence reporting, therefore it is necessary that as high as possible a quality should be achieved.	Zaključno, kvalitet podataka je pojava koja znatno može uticati na kvalitet izveštavanja poslovne inteligencije, pa je zbog toga potrebno obezbediti što veći kvalitet.

Figure 3: Printscreen of Bibliša parallel concordance search tool interface

3.3.3 Direct and indirect application of parallel corpus in ESP courses at the Faculty of Organizational Sciences

The findings presented above can be implemented both indirectly and directly throughout our ESP 1 course. Indirectly, a core vocabulary list and parallel concordances are useful additions to designing “lexical” syllabi (McEnery, Xiao, 2010) and teaching materials whose main purpose is the acquisition of the core vocabulary for the subject-field / domain in question and developing translation competence. These targeted teaching materials can include, for example, gap-filling and cloze exercises for students to explore specific vocabulary, key phrases, and specialized language word chunks (patterns), reading texts, writing and translation tasks that revolve around the targeted core vocabulary, test design, etc. For students with higher levels of language proficiency, corpus-based teaching material does not have to be focused only on core vocabulary, but rather on other discursive patterns, meta-linguistic markers, or on the analysis of common mistakes found in the corpus (having that the texts in the corpus were originally written in Serbian and then translated into English).

For direct application of our corpus in the ESP learning process (i.e. in data-driven learning), we believe that students need to be trained in using both *AntConc* and *Bibliša* tools for corpus exploitation. Students can be encouraged to use the *AntConc* word list and concordance tool to generate their own lists of words, collocations, and word chunks to independently look for authentic examples of their use in the context, as well as to disambiguate the meaning of two similar terms, words, or partial synonyms by observing them in context. This type of exercise seems to be more suitable for students with higher levels of language proficiency. By using *Bibliša*, students can look for translation equivalents, make their own glossaries, look for contextual examples, or use concordance pairs as a point of reference when practicing translation.

4 Discussion

The focus of this paper was on potential ways of exploiting a parallel, specialized, genre-specific Serbian-English corpus in teaching material design for ESP courses at the Faculty of Organizational Sciences, University of Belgrade. After presenting the potential and documented benefits of using corpora (in particular specialized and parallel corpora) in the teaching and learning of both general

language and language for specific purposes, the paper focuses on the process of management corpus compilation and processing. The central part of the paper is dedicated to its application in the ESP teaching and learning process, and especially in ESP teaching material design.

After pilot testing the possibilities of using the corpus in the ESP teaching and learning process both directly and indirectly, we have drawn several conclusions.

First, the application of parallel corpora can be considered beneficial to both teachers and students of ESP in higher education. As this is a relatively new medium, it may contribute significantly to teaching and learning diversification and thus improve students' motivation for ESP learning. In DDL, for example, corpora use greatly contributes to the development of students' independence by encouraging them to independently acquire terms, term phrases, collocations, grammatical patterns, and other structures characteristic of specialized discourse. In addition, they can compile their own word and term lists with equivalents in another language and learn to rely on context when seeking an adequate word and word chunk translations. Indirectly, corpus frequency wordlists enable teaching material writers to decide on the core vocabulary to build their teaching materials, tests, and syllabi, especially for courses aimed at students with a lower level of ESP competence.

However, despite all the benefits of both direct and indirect parallel corpora applications, we believe that our ESP teaching and learning cannot be based exclusively on corpora exploitation but should rather be an addition to the traditional approach. The reason for this lies in the limitations and drawbacks of both the corpus itself, and the drawbacks of corpora application in ESP teaching and learning in general.

One of the limitations of the corpus is its representativeness and balance. Namely, given that this corpus is genre-specific (all the texts are research papers) and domain-specific (all the texts belong to a single domain of management), we cannot regard it as a fully balanced and full representation of the subject-field of management as a whole. Therefore, we believe that it should be further expanded to include other text types and genres, i.e. conference papers, introductory textbooks, technical reports, online discussions, and similar. This view is supported by Conrad's (1996, p. 302) claim that "different genres must be coped with by learners in order for them to widen their knowledge and enhance their linguistic competence."

Another drawback of parallel corpus use is related to our class size and the training of students. We believe that using a parallel corpus in the ESP teaching and learning process is much more effective with small classes, since this enables the teacher to truly observe and facilitate students' corpus work. Large classes make it difficult to monitor students' corpus work and progress, and to train them how to properly exploit the corpus.

To conclude, we believe that our Serbian-English parallel corpus on management represents an invaluable source for indirect use in our ESP classes, since it can enable designing teaching materials and other resources based on frequency wordlist, frequencies of phrases, collocations and grammar patterns found in the corpus. The main obstacle to its direct application in our ESP courses seems to be prompted by the size of our classes, since large, mixed-ability classes make the corpus use training and corpus exploitation by the students themselves significantly more difficult.

References

- Aston, Guy (1996). The British National Corpus as a language learner resource. In S. Botley, J. Glass, T. McEnery & A. Wilson (Eds.), *Proceedings of Teaching and Language Corpora* (pp. 178–191), Lancaster: UCREL.
- Ball, F. (2001). Using corpora in language testing. *Research notes*, 6, 6–8.



- Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12(4), 345–360.
- Connor, U., Upton, T.A. (2004). *Discourse in the professions: Perspectives from corpus linguistics*. Amsterdam: John Benjamins.
- Conrad, S. M. (1996). Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and education*, 8(3), 299–326.
- Danielsson, P, Mahlberg, M. (2003). There is more to knowing a language than knowing its words: Using parallel corpora in the bilingual classroom. *English for Specific Purposes World* (available at http://www.esp-world.info/articles_6/DanielssonMahlberg2003.htm).
- Flowerdew, L. (2001). EAP Materials Design. *Small corpus studies and ELT: Theory and practice*, 5, 363.
- Gong, W. (2005). English in computer-mediated environments: a neglected dimension in large English corpus compilation. In *Proceedings of Corpus linguistics conference series*. (pp. 14–17). Birmingham.
- Granger, S. (1998). The computer learner corpus: a testbed for electronic EFL tools. *Linguistic databases*, 175–88.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied linguistics*, 19(1), 24–44.
- Hunston, S. (2002). *Corpora in applied linguistics*. Ernst Klett Sprachen.
- Johansson, S. (2007). Seeing through multilingual corpora. *Language and Computers*. Retrieved Sept 30 2017 from ingentaconnect.com.
- John, E. S. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language learning & technology*, 5(3), 185–203.
- Linn, S. (2006). Trends in the translation of a minority language: the case of Dutch. *Benjamins Translation Library*, 67, 35.
- McEnery, T., & Xiao, R. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning*, 2 (pp. 364–380). London & New York: Routledge.
- Romer, U. (2011). Corpus Research Applications in Second Language Teaching. *Annual Review of Applied Linguistics*, 31, 205–225.
- Sinclair, J. (2005). Corpus and Text – Basic Principles. In M. Wynne (Ed.), *Developing Linguistic Corpora: a Guide to Good Practice* (pp. 1–16). Oxford: Oxbow Books. Retrieved April 13, 2017 from <http://ahds.ac.uk/linguistic-corpora/>.
- Stanković, R., Krstev, C., Obradović, I., Trtovac, A., & Utvić, M. (2012). A tool for enhanced search of multilingual digital libraries of e-journals. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, (Istanbul, Turkey).
- Tribble C. (1997). Improving corpora for ELT: Quick and dirty ways of developing corpora for language teaching. In B. Lewandowska-Tomaszczyk & P. Melia (Eds.), *Practical Applications in Language Corpora – Proceedings of PALC '97* (pp. 107–117). Lodz: Lodz University Press.
- Tribble, C. (2001). Small corpora and teaching writing. *Small corpus studies and ELT: theory and practice* (pp. 381–408). Amsterdam: John Benjamins.