



USING SMALL PARALLEL CORPORA TO DEVELOP COLLOCATION-CENTRED ACTIVITIES IN SPECIALIZED TRANSLATION CLASSES

SORINA POSTOLEA*
TEODORA GHIVIRIGĂ**

Alexandru Ioan Cuza University of Iași, Romania

Abstract

The research devoted to special languages as well as the activities carried out in specialized translation classes tend to focus primarily on one-word or multi-word terminological units. However, a very important part in the making of specialist registers and texts is played by specialised collocations, i.e. relatively stable word combinations that do not designate concepts but are nevertheless of frequent use in a given field of activity. This is why helping students acquire competences relative to the identification and processing of collocations should become an important objective in specialised translation classes. An easily accessible and dependable resource that may be successfully used to this purpose is represented by corpora and corpus analysis tools, whose usefulness in translator training has been highlighted by numerous studies. This article proposes a series of practical, task-based activities—developed with the help of a small-size parallel corpus of specialised texts—that aim to raise the translation trainees' awareness of the collocations present in specialised texts and to provide suggestions about their processing in translation.

Keywords: *collocation, phraseology, specialised translation, translator training, parallel corpora, phraseology, classroom activities for translators, translation pedagogy*

Preliminary considerations

The research devoted to special languages as well as the activities carried out in the process of specialised translation (hence in the training of translators

* sorinapostolea@gmail.com

** teoghivi@yahoo.com

working with specialized texts) tend to focus primarily on one-word or multi-word terminological units, which are considered to be not only carriers of specialized conceptual content, but also markers of domain specificity. However, a very important part in the texture of the specialized text is played by collocations—a buffer zone between the vast area of general language and the finite set of terms pertaining to a particular domain. Irrespective of the definitions they have been given or the various classifications proposed, they have been regarded as one of the most salient indicators of native-like quality in a text (Bernardini, 2007, 3), so much so that they have been used to compare source-text (ST) to target-text (TT). It has been found, on comparison, that the collocational element in language produced by natives exceeds that in non-native texts, be they in task-based composition activities (Fan) or translations (Marco), in both number and complexity. While they raise no or few problems in comprehension, at production level they tend to be a source of frustration for non-natives (Laufer and Waldman). L2 learners—and translator trainees are no exception—have difficulty in producing correct combinations at all levels of proficiency, and it has been pointed out that errors in this area continue to persist even with advanced learners who are in possession of a wide range of vocabulary (*ibid*). Indeed, the discrepancy between knowledge of lexical material and the ability to organize it on the syntagmatic axis covers both general language and specialized language: according to Picht (qtd in Antia), even when the terminology has been correctly researched and appropriated, errors still occur on “incorporating the terms in stretches of target language” (120).

The first step in finding adequate ways to address this issue in language production is understanding the underlying difference in the acquisition of collocations between native speakers and L2 learners. Natives acquire them naturally, gradually and subconsciously, also effortlessly, while non-natives are exposed to them only in artificial class-room environment (Fan)—for the former, production is largely automated, collocational patterns emerge as a subliminal feature of language (Gledhill), which also enhances fluency, while for the latter it is a slow and sometimes painful process, a “time-consuming, error-laden decision” (Antia 119). Moreover, in the case of the non-natives, the co-occurring items are felt to represent arbitrary associations, more difficult to acquire and to master at production level; learners may fail to identify a pattern, entirely ignoring a governing rule, or they may produce deviant forms due to L2 interference (Fan). Hence the necessity of systematically including activities focused on the collocational element at every stage of the learning process; some studies (Webb and Kagimoto, Laufer, 2011, Peters 2016) also provide the ground for methodological guidelines and indicate factors and even types of

activities that can assist learners in the successful acquisition of such combinations.

Needless to say, the same holds true in the case of translators' training. Trainees' output has been shown to be deficient in this respect and, while textbooks aimed at non-professional, general language learners do include exercises and testing material focused on collocations, no consistent and systematic training on this topic is provided in either L1 or L2 (in Romania at least) in terms of either theoretical background (given pre-existing knowledge of linguistics concepts), or of acquisition and production. This is why helping students acquire competences relative to the identification and processing of collocations should become an important objective in specialised translation classes and aim at both the non-specialized and specialized language.

Collocations in the Literature

Word combinations that emerge in language in a recurrent manner and in a more or less stable form are known in the literature under many names: *collocations* (Sinclair 2004; Stubbs 2002), *set phrases*, *phrasemes* (Mel'čuk 1998), *phraseological units*, *phraseologisms* (Granger & Paquot 2008; Gläser 1998), *prefabricated patterns* or *chunks* (Granger 1998; Schmid 2003), or, more recently, especially in the field of computational linguistics and natural language processing, *multiword expressions* or *MWEs* (Wanner *et al.* 2006). They are instantiations and products of the "idiom principle", i.e. the so-called phraseological tendency of language according to which "words tend to go together and make meanings by their combinations" (Sinclair 29) and which, in the Firthian tradition, has ultimately led to the conclusion that "people speak in set phrases, rather than in separate words" (Mel'čuk 24).

The research devoted to these linguistic associations in recent years seems to have followed two interrelated directions. First, although the collocational nature of language has been documented for several decades now (Gries 2013), many studies continue to tackle the issue of the specific criteria to be used in order to differentiate between more or less predetermined combinations and free associations of words. As several authors have noted (Nesselhauf 2005; Postolea 2014), the definitions provided so far fall into one of two main approaches or somewhere in-between: on the one hand, the *phraseological approach* focuses mainly on the semantic/syntactic-syntagmatic properties of fixed or semi-fixed word combinations (Gläser 1998; Nesselhauf 2005; Granger & Paquot 2008; Martin 2008). For instance, in Mel'čuk's view, a *set phrase* or *phraseme* is simply a phrase which is not *free*, i.e. in its formation either the selection of its constituent members or their combination, or both, are restricted to a

more or less limited number of choices (24). In the *corpus linguistics/semantics* approach, collocations are usually identified based on the criteria of frequency and co-occurrence. As Stubbs puts it, if *phrase* is a “neutral term for a string of words” which constitutes an “extended lexical unit”, *collocation* refers to “a lexical relation between two or more words which have a tendency to co-occur within a few words of each other in running text” (24). In other words, the difference between free word combinations—understood as phrases in this approach—and collocations lies in that the latter occur together more frequently “than expected on the basis of chance” (Gries, “Phraseology and linguistic theory” 6). This is why Sinclair simply defines collocation as “frequent co-occurrence of words” (28). Other definitions try to find a middle ground between these two approaches, integrating both semantic/syntactic and frequency criteria (e.g. see Gries, “Phraseology and linguistic theory” or Martin).

The second research trend within the many studies devoted to collocational phenomena has attempted to “disentangle the phraseological web” (Granger & Paquot 2008) and put forth various taxonomies of word combination types. This is a point where the two main approaches described so far diverge. Covering a very wide range of recurrent word combinations in the “frequency-based approach” (Nesselhauf 2005), the notion of *collocation* refers to just one type of phraseological units in the field of phraseology, where it is classified alongside many other entities, such as idioms and quasi-idioms, irreversible bi-and trinomials, similes, compounds, phrasal verbs, colligations, pragmatic formulae, textual phrasemes, etc. (Mel'čuk 1998; Granger & Paquot 2008). More specifically, phraseologists describe collocations as:

- “arbitrarily restricted lexeme combinations”, e.g. *make a decision* (Nesselhauf 1),
- semantic phrasemes “in which the meaning of one phrase member remains intact”, e.g. *launch an attack* (Mel'čuk 29), or
- “combinations of lexemes exhibiting a medium degree of observable recurrence, mutual expectancy and idiomaticity”, e.g. *keep one's distance* (Schmid 249).

It is this particular type of frequent word combinations—also known as *restricted collocations*—that the present paper focuses on. As shown by Nesselhauf (21-24), restricted collocations have been classified according to various syntactic, semantic, and combinatorial criteria, of which structural-syntactic taxonomies are the most common, including part-of-speech combinations such as N+N, V+N, A+N, Adv+V, Adv+A, etc. (Martin 56).

Even if they are usually discussed in relation to language in general, alongside terms, collocations also emerge in specialized discourse. Cabré talks about “combinations that occur frequently in the discourse of a special

subject field (which makes them similar to terminological phrases)” that “do not correspond to established concepts” (91), whereas Bowker & Pearson mention that collocations are “one type of word behaviour that can be identified with the help of a corpus”, emphasizing that “they feature quite prominently in LSP” (32). In spite of a series of studies devoted to collocations in specific areas of activity, such as Gledhill’s analysis of these entities in science writing (2000), Luzon Marco’s study about collocations in medical papers (2000), or Postolea’s discussion of the phraseology of general use ICT genres (2014), this linguistic phenomenon has been relatively less investigated in relation to specialised languages; the criteria to be used in order to differentiate specialised collocations from multi-word terms are still a topic of debate. Aguado de Cea proposes seven defining features for what she calls “specialised phraseological units”: “1) they are syntagmatic structures that include a term or a terminological unit; 2) they include a verb or a deverbal element; 3) they have a high degree of fixedness, although sometimes they allow the insertion of other elements, such as a quantifier or a pronoun; 4) they demonstrate regularity in their composition; 5) they admit the commutation of their constituents, but to a certain degree; 6) they maintain a specific meaning within a given field, and thus they acquire semantic stability; and 7) they are of frequent use in a specialised field of subfield of activity” (57-58). Just like the collocations emerging in everyday language, from a structural viewpoint, specialised word combinations may take various forms: V+N (or PrepP), e.g. *browse the Internet*; N+N, e.g. *editing tools*; Adj+N, e.g. *default settings*; N+Prep+N, e.g. *frames per second*; V+Adv, e.g. *briefly press*; Adv+Adj, e.g. *fully compatible*, etc. This type of frequent word combinations are all the more important in specialised discourse because they are closely linked to the notion of register/style. As Trosborg emphasizes, “it is often the collocation of two or more lexical items, rather than the occurrence of isolated terms, that determines the identity of a given register” (6). For Baker, “register-specific collocations are not simply the set of terms that go with a discipline”, and “they extend far beyond the list of terms that one normally finds in specialized dictionaries and glossaries” (52).

Collocations in Lexicography

One of the most obvious aids in the process of acquiring both specialized and non-specialized collocations would be the extensive use of lexicographic work that includes information in this area along with other types of linguistic information. However, the quantity of information to be found in learners’ dictionaries (Keller, Siepmann) or specialised dictionaries (L’Homme, Campos & Castells) is insufficient or may not be organized in a

user-friendly manner. Ideal treatment of collocational material should consist of one entry for each combination (Keller 16), however, given space limitations, this is rarely the case, which hampers the search. While recent instruments in electronic format may make retrieval faster and easier (based on key word searches), the problem of word order still remains: should such collocations be entered under the collocator or the base? According to Antia, dictionaries approach the issue with a focus on comprehension and start from the former; however, it is at production level that most problems appear (as shown *supra*), so a production oriented approach should favour entries that start from the base, since the producer of discourse proceeds from the base to the collocator (118). The problem is compounded by the fact that, given the very nature of this linguistic entity, it presents formal variations, especially when one element is a verb. (Keller)

This is an issue not only in general language or learner dictionaries, but in specialized lexicographic sources, terminology databases included. Few such sources provide information regarding word combinations (L'Homme and Leroyer 260) and those which do exist are domain sensitive, focusing mainly on business, legal language, and IT; however, their treatment of this type of word combinations is not consistent. In databases all the collocations of a base are usually associated to one entry (in TermiumPlus, for instance, as shown by L'Homme) and retrieval is done by selecting from a list of combinations—a difficult task for a novice, or are classified according to the parts of speech the components belong to, which relies heavily on the user's expert knowledge of the language, since it “assumes that a user already knows the part of speech of the collocate he or she is looking for” (*ibid* 240). While efforts are made to improve access to collocational material in the various resources, and recent instruments such as DiCoInfo are created with the needs of translators and technical writers in mind, providing rich and well structured information on lexical relations (Pimentel 434), organizing such combinations in lexicographic works remains a sensitive matter.

The translator trainer also needs to satisfactorily answer the question of providing resources not only for L2, but also for the native language. Most of the times, knowledge of the collocational stock of L1 is not only considered a prerequisite, but is also taken for granted, as part of the native speakers' competence; however, it rests largely on their individual background and linguistic expertise, and Wray (*qtd. in* Fan 121) considers that it is closely related to the speakers' identity. At present, Romanian secondary education does not provide any specific forms of teaching in this area, so students have no previous experience from L1 when they start training as translators and therefore lack the theoretical instruments to address the issue. The existing Romanian literature focuses on transferring

such groupings from L2 (English mainly) into L1 with little regard for what the student can resort to in L1 in the translation process. The few existing bilingual dictionaries of collocations focus on specialized languages (e.g. domain: Business¹) or on restricted lexical fields (e.g. *the body*²). Even monolingual resources are somewhat scarce and of varying value; *Dicționarul limbii române*, the leading work³ on the market, and the others that have evolved round it are admittedly deficient in presenting collocational information due to technical reasons, as well as the need to preserve its initial structural unity⁴. The need for such tools is even more stringent in L1 to L2 translation tasks; while they may not reflect the sense of translation in an official or working environment, they are an important area of practice that can be an indicator of the student's language production competence.

In these circumstances, the translator trainer is faced with the task of providing students with reliable lexicographic instruments and suitable tasks for practice. Corpus and corpus based materials are an immediate, obvious answer in that they are affordable, open to customization and opportunistic use, adaptable in terms of domain, reusable. However, they come at a cost: when using corpora, students may extract incorrect information or extract information incorrectly, they may have difficulty in using or interpreting the data. Therefore, prior to corpus use, they need to be in possession of a methodology that can adequately assist them in tackling such linguistic entities.

Corpora in Translation Education

Helping students acquire collocation-related skills should therefore represent a consistent part of foreign language teaching as well as translator education. Trainees should be encouraged to learn how to identify these word combinations and distinguish them from other syntagmatic entities of frequent use in particular genres, such as pragmatic/ textual phrasemes, colligations, or compounds and multi-word terms. For future translators, learning how to deal with collocations in LSPs is very important since,

¹ Teodora Popescu, Mariana Toma, *Dictionary of Business Collocations: with Romanian Translation and Practice Session*, revised edition, Alba Iulia: Aeternitas, 2009

² Pia Brînzeu, Aba-Carina Pârlog, Hortensia Pârlog, *Translating the Body*, Lincom Europa, 2007

³ *Dicționarul limbii române* (DLR), Editura Academiei Române, 2010

⁴ Bogdan Harhătă, "Collocations and dictionaries: A Case-Study Applied to Romanian". *DACOROMANIA*, 18.2 (2012): 147–153. Web. Dec 2016.

unlike in general language translation, where the use of collocations may lead to normalization, inhibiting the students' creativity (Stewart qtd. in Bernardini 2004), knowing how to identify and use the right collocations in specialized translation is a key competence when the aim of the translation task at hand is to preserve the style/register features of the source text in the target language. Moreover, since word combinations of frequent use in specialized contexts may sometimes go against the usual norms of everyday/colloquial speech and thus prove to be a source of difficulty in translation, students should be trained and provided with the tools they need in order to identify quickly and reliably equivalent or matching collocations in the target LSP. Corpora, used by the students on their own or as a resource in the translation classroom, offer an easy and dependable way of achieving this desideratum.

The advantages of corpora in translator training have been discussed in many works. Zanettin *et al.* describe various ways in which electronic collections of texts of various sizes and forms may be used as translation/documentation aids. Bernardini (2004) proposes exploring corpora to support and promote "corpus-aided discovery learning" whereas Bowker shows how these tools may be used in translation assessment. Bowker & Pearson discuss the benefits of monolingual corpora as translation resources that may serve several purposes: to verify one's own intuitions, to search for unknown equivalents, to choose between multiple options, to investigate the usage of particular items, or to search for explanatory contexts (198-206).

However, most of the studies dealing with corpora in translator training have mainly focused on general issues and only a few have put forth practical, task-based activities developed with the help of these tools to be used in translation classes, such as, for instance, the activities discussed by Marco & van Lawick. In fact, besides helping students with their translation and documentation process, corpora may also be used as a resource for trainers, seeing that "in the translation classroom it would be the teacher's role to plan and carry out tasks leading progressively to the achievement of a given learning objective" (Marco & van Lawick 13). Drawing on various types of searchable collections of texts, from general-monolingual to small-size parallel specialised corpora, translator trainers may create a wide range of custom-made, task-based materials tailored to meet the specific needs of their trainees. In this type of approach to translator training, rather than acting as the sole problem-solver, the teacher works as a cartographer who provides the students with maps that help them deal with and overcome given difficulties.

Corpus-based Collocation-centred Activities

In the following lines we propose a series of practical classroom activities that put to use the advantages of corpora and corpus tools in order to raise the students' awareness of specialized collocations, to make them discover the pivotal yet discrete role played by these preset word combinations in specialist registers, and to offer them suggestions about the way in which these linguistic entities should be dealt with in translation. Seeing that corpus linguistics is a branch of studies still underdeveloped in Romania, another goal was to introduce corpora as a resource and tool that the trainees themselves might use in their future work.

Methodology

The activities developed drew on a small-size specialized parallel corpus built by one of the authors for her doctoral research. It comprised 15 pairs of information and communication technology (ICT) user manuals translated from English into Romanian, aligned at sentence level, and amounting to a total number of about 150,000 words for each language involved. The texts referred to 15 different ICT products (e.g. a laptop, a printer, a tablet, a smartphone) and were produced by 15 different international ICT companies. Since the wide array of collocational phenomena present in specialised texts cannot be covered in a single class, for our study we focused on just two types of such formations, namely V+N / N+V and Adj+N specialised (restricted) collocations retrieved from the corpus: e.g. *press and hold a/the (X) button/key* (77 occurrences), *enter the password/address/name* (69 tokens), or *default settings* (21 occurrences). To test how these activities may be actually used in the classroom we asked a small working group of 12 students for help, i.e. 6 undergrads from the translation programme at our university as our target group (TG) and 6 undergrads from a language and literature programme as a control group (CG). The students in the TG had already been exposed to specialised translation training, whereas the trainees in the CG had no previous experience with translation in general. Our activity comprised two steps: 1) a preliminary test that the students had to solve on their own and 2) 4 activities that the students solved in the translation class.

Preliminary test

The main aim of the preliminary test was to establish the students' baseline level in terms of overall sensitivity to SL/TL collocations, i.e. their

ability to spot these word groupings and process them in translation, and, in relation to this, the overall consistency of the translation solutions they would provide, since, in theory, once identified as such, a specialised collocation should be translated using its established, most frequent, equivalent in the target language. For the preliminary test, the students received a *homework assignment* made up of two parts:

A) they were asked to translate 10 small excerpts from the corpus according to the following brief, which offered no clue about the treatment of specialised collocations:

“You are a freelance translator and one of your clients has just asked you to translate the following excerpts from several user manuals (IT devices). Translate them into Romanian to the best of your knowledge and abilities.”

B) they were asked to fill in a small survey comprising the following questions about the task performed:

- 1) Did you use any additional materials to help you in your translation? If your answer is YES, what materials did you use? List them and, if possible, specify the addresses of the websites you used.
- 2) What were the elements the most difficult to translate for you? Give examples and explain the difficulties you encountered.
- 3) How long did it take for you to translate these excerpts? (Think of the time you actually spent translating them and looking up unknown items or pieces of information.)

The results of this preliminary test showed that both groups of students demonstrated a rather low degree of sensitivity to the issue of specialised collocations and an even lower degree of consistency in the translation solutions they provided for each source language collocation. For instance, the 3 collocations of interest in the first sample excerpt they had to translate received many different translation solutions:

Sample excerpt:

Printing a demo page

Print a demo page to make sure that the machine is operating correctly.

To print a demo page:

In ready mode, press and hold the OK button for about 2 seconds.

Trainees' translations:

printing a page → 3 solutions: *imprimarea unei pagini* (6) / *printarea unei pagini* (5) / *listarea unei pagini* (1)

ready mode → 7 solutions: *modul Gata* (3) / *modul ready* (3) / *modul automat* (1) / *modul imediat* (1) / *modul disponibil* (1) / *când aparatul este pregătit* (1) / *când este pornită* (1)

press and hold the button → 5 solutions: *apăsați și mențineți apăsat butonul* (4) / *țineți apăsat butonul* (3) / *ține(ți) apăsată tasta* (3) / *mențineți apăsat butonul* (1) / *apăsați butonul* (1)

Another indicator of the students' low degree of sensitivity with regard to the collocations present in the excerpts was represented by their answers to the question about the difficulties they encountered in the process of translation. Most of the trainees in both groups identified only a series of terms as sources of difficulty (e.g. *touchpad*, *hotspot*, *cellular data connection*) and only two people mentioned collocations—without using this term (e.g. *run a utility* and *linking/unlinking a contact*). Some differences between the two groups emerged in their answers to the other two questions in the mini-survey. Even if none of the students mentioned corpora as a translation aid, the students in the TG, who had previously had translation classes, listed a wider variety of resources and materials used in translation (bilingual dictionaries, parallel texts, Romanian dictionaries, documentation resources, translator forums) than the students in the CG, who listed only Romanian online dictionaries, the Internet, Google translate or even no additional resources. This may be explained by the fact that the trainees in the TG were more aware of the additional documentation steps that a specialised translation requires, while the students in the CG used only their general knowledge about the target language, paying less attention to the accuracy of their output. This difference is also visible in the answers to the last question in the survey: the students in the translation programme (TG) spent much more time on the translation of the excerpts (~300 words) because they tried to document their choices more thoroughly. The average time declared by the trainees in the CG was of ~ 80 minutes (spanning from 1h to 2½h), whereas the average time of the TG was of ~ 120 minutes (spanning from 50min to 3h).

Classroom activities

Since the preliminary test revealed that the students in both groups were unaware of the existence of collocations in the specialised language at hand and lacked information about how these units should be dealt with in translation, with the help of the corpus, we designed a course unit meant to overcome these lacunae. In the opening part of the class (100 min), the trainer offered some theoretical background about collocations: definitions, examples of collocations in general as well as in specialised languages (both in L1 and L2), useful criteria to be used so as to distinguish set phrases from compounds or multi-word terms, etc.

COLLOCATION SPOTTING (ACTIVITY 1)

After this introductory sequence, the first activity proposed was a collocation spotting exercise. The students were asked to identify possible collocations in a series of excerpts retrieved from the corpus and then to suggest their own translations for these units.

ACTIVITY 1 (EXAMPLE)

Identify the specialized ICT collocations present in the three excerpts from user manuals shown below then try to translate them into Romanian:

If you want to keep your configuration settings, you must save changes before exiting the BIOS setup utility.
If you want to restore default settings, choose Load Setup Defaults.

Possible collocation	Your translation

This type of exercise aims to stimulate the trainees' practice and acquisition of the target specialised language as well as to capitalize on their knowledge and increase their sensitivity to specialist registers in their native language (L1). After having completed the exercise, the trainer and the students discussed the collocations present in the excerpts, the solutions proposed by the trainees as well as the "official", published translations of the sample fragments retrieved on the spot from the corpus.

CLOZE TEST (ACTIVITY 2)

The second collocation-centred activity developed for the class was a cloze test adapted from Marco & van Lawick. As these authors emphasize, "the main advantage of this kind of exercise is that it allows the class to focus on a specific translation problem, leaving aside all other aspects of a text which, interesting as they may be, are perceived at a given moment as peripheral to the issue in hand" (13-14). The trainees were presented with several parallel fragments retrieved from the corpus in which they were asked to fill the gaps with missing collocation elements. This exercise was meant as both an L2 and L1 practice opportunity. Moreover, since the source and target fragments were approached in parallel, this exercise was also meant to raise the students' awareness of the differences that may or may not emerge between the source language collocations and their corresponding units in the target text.

ACTIVITY 2 (EXAMPLE)

Fill in the blanks with a suitable word to form collocations in both the source and the target texts:

<p>To _____ the computer, simply press and release the power button below the LCD screen.</p>	<p>Pentru a porni computerul, _____ și eliberați butonul de pornire aflat sub ecranul LCD.</p>
<p>To turn the power off, do any of the following:</p> <p>- Use the Windows _____ command</p> <p>Click on Start then click on Shut Down.</p>	<p>Pentru _____ alimentarea, faceți oricare dintre următoarele:</p> <p>- Folosiți comanda de oprire din Windows _____ pe Start, apoi _____ pe Opreire.</p>
<p>You can also put the computer in sleep mode by _____ the “sleep” hotkey <Fn> + <F4>.</p>	<p>Puteți pune computerul și în modul _____, apăsând tasta rapidă “_____” <Fn> + <F4>.</p>

With the corpus and a parallel concordancer at hand, this exercise was also followed by a discussion on the difficulties encountered by the students when they tried to solve the exercise as well as on the solutions provided in the published translations of the fragments.

MULTIPLE CHOICE EXERCISE/LEARNER CORPUS (ACTIVITY 3)

The third activity carried out and discussed in class was also based on Marco & van Lawick’s ideas. In this exercise, the students had to rate, on a scale from 1 to 3, 5 different translations of a single fragment retrieved from the corpus. The activity actually included three of the fragments that the students had to translate as a homework assignment for the preliminary test. Out of the 5 different translations for each fragment, 1 was the “official” translation published in the corpus, and the remaining 4 were translations produced by fellow students. The trainees in the TG were asked to rate translations proposed by the students in the CG and vice versa.

ACTIVITY 3 (EXAMPLE)

Paying attention to the way collocations were processed in translation, rate the following translations on a scale from 1 to 3, where:

- 1 = incorrect
- 2 = more or less correct, but inadequate
- 3 = correct and adequate

1. Print a demo page to make sure that the machine is operating correctly
In ready mode, press and hold the OK button for about 2 seconds.

a)	Printează o pagină demonstrativă pentru a te asigura că imprimanta operează corect. În modul automat, ține apăsată tasta OK timp de aproximativ 2 secunde.	
b)	Printare a unei pagini de probă pentru a verifica funcționarea corectă a dispozitivului. Când aparatul este pregătit, mențineți apăsat butonul OK timp de aproximativ două secunde.	
c)	Imprimați o pagină demo pentru a vă asigura că aparatul funcționează corect. În modul imediat, apăsați și mențineți apăsat butonul OK pentru aproximativ 2 secunde.	
d)	<i>*Imprimați o pagină demonstrativă pentru a vă asigura ca mașina funcționează corect. În modul disponibil, apăsați și mențineți butonul OK circa 2 secunde.</i>	
e)	Listați o pagină de probă pentru a vă asigura că imprimanta funcționează corect. În modul ready, țineți apăsată tasta OK timp de aproximativ 2 secunde.	

* Published (corpus) translation

This was by far the most interactive activity in the class and it raised many discussions about the manner in which collocations should be approached in translation as well as about translation assessment in general. Since they were given the possibility to play the role of the evaluator for a change, the trainees felt empowered and were more open. In fact, “this kind of activity is intended to enhance the translator trainees’ critical sense, as it forces them to consider different translation possibilities and give each one of them its due, always in a reasoned way” (Marco & van Lawick 17).

An interesting outcome was the fact that in some cases (as in the example shown below), the trainees considered that the solutions proposed by fellow students were, in fact, better than the “official” translations from the corpus.

RESULTS (SAMPLE)

a)	Printează o pagină demonstrativă pentru a te asigura că imprimanta operează corect. În modul automat, ține apăsată tasta OK timp de aproximativ 2 secunde.	3	2	2	2	1	1	11
b)	Printare a unei pagini de probă pentru a verifica funcționarea corectă a	2	3	1	1	2	2	11

	dispozitivului. Când aparatul este pregătit, mențineți apăsat butonul OK timp de aproximativ două secunde.							
c)	Imprimați o pagină demo pentru a vă asigura că aparatul funcționează corect. În modul imediat, apăsați și mențineți apăsat butonul OK pentru aproximativ 2 secunde.	1	1	1	1	2	2	8
d)	<i>*Imprimați o pagină demonstrativă pentru a vă asigura ca mașina funcționează corect. În modul disponibil, apăsați și mențineți butonul OK circa 2 secunde.</i>	2	1	3	1	2	1	10
e)	Listați o pagină de probă pentru a vă asigura că imprimanta funcționează corect. În modul ready, țineți apăsată tasta OK timp de aproximativ 2 secunde.	2	3	2	3	1	3	14

* Published (corpus) translation

This also led to a discussion about how the acceptability of collocation translations should be assessed and verified in a given specialist register, and the role of specialised parallel corpora in this process.

CLASSROOM TRANSLATION (ACTIVITY 4)

The last collocation-centred activity developed with the help of our small-size parallel corpus was a translation exercise. Using the corpus with the trainer, the students were asked to translate several excerpts from a text belonging to the same genre and specialised field but not included in the corpus.

ACTIVITY 4 (EXAMPLE)

Using some of the collocations discussed so far and the corpus, translate the following excerpt from a user manual:

- “1) Tap once on an item - Performs an action, such as starting an app, opening a link, or performing a command.
- 2) Swiping in from the left edge can:
 - Bring in and open an app currently running in the background. If there is more than one currently open app, you can...”

The purpose of this final activity was to provide the students with an opportunity to apply the knowledge acquired in the class and to demonstrate how the corpus can be used in actual practice. The translation was carried out checking the translations of the collocations (and the terms) in the source text available in the corpus. Thus the students were able to observe how this

resource can be used to help them in their everyday translation activities and how this kind of documentation may substantially reduce the time they usually spend doing research for their translations.

Classroom activities – conclusions

The corpus-based collocation-centred activities developed for this specialised translation class proved to be very useful from at least three different viewpoints: 1) they helped us assess our students' baseline level and training needs relative to the linguistic issue at hand using readily available authentic materials; 2) they provided a means to address the training needs identified in our trainees in a personalised, coherent, and task-based manner; 3) they served to approach the topic of specialised collocations within a contrastive framework that increased the students' awareness of this linguistic phenomenon both in L1 and L2.

On the other hand, at the end of the class some students still had trouble in identifying specialised collocations both in L1 and L2, and this means that other activities of this kind, centred on other types of word combinations, are still needed in the future. An advantage of the activities designed based on our parallel corpus was the fact that the class was very interactive, and many issues relative to collocations and their translation as well as other translation-related questions were discussed in group. The students felt empowered to express their own opinions and doubts and to recognise their own weaknesses as far as the translation process was concerned. A very important advantage of this approach was the fact that the class focused on the corpus and on the published translations included in it as reference, and not on the teacher as the sole problem-solver/holder of the truth.

Conclusions

The advent and subsequent rise to prominence and popularity of corpus instruments and methodologies has generated a large number of studies that explore and assess the value and benefits that can be extracted from their use in various areas of interest. Access to a variety of such collections of texts is easier than ever, inviting extensive use for theoretical research, as well as didactic and translation purposes; reflections upon this have identified multiple advantages derived from their use, some of which have also been demonstrated by the activities devised and tested for this paper. Zanettin (2002) discusses the role of corpora as valid translation aids and resources whose use may result in improved terminological and phraseological accuracy. McEnery and Xiao point out that use of parallel corpora can reveal

differences between source and target texts (18), thus offering the basis for a sound contrastive approach, and can provide “a useful and effective reference tool and a workbench for translators and trainees” (22), which adds an important didactic dimension to their use for translation purposes. In relation to this, Bernardini (2004) finds they are an important source of materials for learning activities and knowledge about language, and mentions that access to authentic language can stimulate the learners and foster genuine reasoning and inductive learning as they are engaged in problem solving. In using corpora they take a more active role being encouraged to interpret data rather than merely memorize them. Moreover, corpus data can be used to develop communication skills and, more specifically and in line with the topic at hand, to raise the learners’ “awareness of variation [...] at the phraseological level” (105-7).

Nevertheless, enthusiasm about the manifold purposes for which corpora can be used and the advantages they can bring in translation—and translation training—needs to be moderated by awareness of the dangers that may come along with such a versatile instrument. For one thing, much depends on the quality of the corpus: while in terms of size, no corpus is too small to render its utility—authors speak about small and even tiny corpora—the quality of the texts that constitute the collection is crucial: Melby (2012), for instance, discusses the danger of using source texts plagued by inconsistencies in parallel corpora, which are then replicated by translators and/as corpora users. Zanettin (2002) also warns against the possible unreliability of some documents used to compile collections of texts as one of the shortcomings of DIY (and, we may add, other types of) corpora. Bernardini (2004) considers that, while they are useful, they should not end up being the only and ultimate element to inform language pedagogy (101) and we add that a higher degree of awareness regarding a particular language phenomenon neither equals nor substitutes active engagement in the learning process.

The immediate and obvious consequence should be that corpora are to be considered in association with other types of training materials and methods. Compiled and used wisely, however, such instruments can prove invaluable in research, training, as they are in actual translation. That is why some authors (Zanettin 2002, Bernardini 2004) consider that learning how to use corpora should be an integral part of the translation students’ training and that they should be equipped with adequate know-how in the compilation of DIY collections and with a methodology to use them, as well as any other type of similar collections of texts.

The experimental integration of task-based activities created from an (ad-hoc)(mini)corpus in a class focused on teaching collocations to a group of translation students may have, we feel, a number of consequences for

further developments. Corpus-based activities can be used by Romanian trainers and trainees to raise awareness of linguistic phenomena such as collocations, which they have not been taught systematically; to produce meaningful and sound training materials that could be tailored to level, needs, learning objectives; to motivate trainees to perform better in the acquisition and use of collocations—and subsequently other language issues; to foster trainees' autonomy in researching collocational material; and, not least, to question the value of published existing translations, thus triggering the students' competitiveness and creativity with a view to improving them or offering better solutions; and, in the long run, to create a sound basis for Continuing Professional Development for translators.

Works Cited

- Aguado de Cea, Guadalupe. "La fraseología en las lenguas de especialidad". *Las lenguas profesionales y académicas*. Ed. Enrique Alcaraz-Varó, José Mateo Martínez and Francisco Yus Ramos. Barcelona: Editorial Ariel, 2007. Print.
- Antia Edem, Basse. *Terminology and Language Planning*. Amsterdam: John Benjamins, 2000. Print.
- Bernardini, Silvia. "Collocations in Translated Language: Combining Parallel, Comparable and Reference Corpora". Paper given at the fourth Corps Linguistics conference held at the University of Birmingham, 27 – 30 July 2007. Web. Oct. 2016.
- . "The theory behind the practice: Translator training or translator education?". *Translation in Undergraduate Degree Programmes*. Ed. Kirsten Malmkjær: John Benjamins, 2004.
- Bowker, Lynne. "A Corpus-based Approach to Evaluating Student Translations". *The Translator*, 6.2 (2000):183-210.
- Bowker, Lynne and Jennifer Pearson. *Working with Specialized Language: a Practical Guide to Using Corpora*. London: Routledge, 2002.
- Campos, Alonso A. and Sergi Torner Castells. "Adjectives and collocations in specialized texts: lexicographical implications". *Proceedings of the XIV EURALEX International Congress*. Ed. A. Dykstra & T. Schoonheim. Leeuwarden/Ljouwert: Fryske Akademy, 2010.
- Fan, May. "An exploratory study of collocational use by ESL students – A task based approach". *System*, 37 (2009): 110–123. Web. Jan. 2016.
- Gläser, Rosemarie. "The Stylistic Potential of Phraseological Units in the Light of Genre Analysis". *Phraseology. Theory, Analysis, and Applications*. Ed. A. P. Cowie. Oxford: Clarendon Press, 1998. pp. 125-144. Print.
- Gledhill, Christopher. *Collocations in Science Writing*, Tübingen: Gunter Narr Verlag, 2000. Print.
- Granger, Sylviane. "Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae". *Phraseology. Theory, Analysis, and Applications*. Ed. A. P. Cowie. Oxford: Clarendon Press, 1998. pp. 145-160. Print.

- Granger, Sylviane, and Magali Paquot. "Disentangling the phraseological web". *Phraseology: an interdisciplinary perspective*. Ed. Sylviane Granger and Fanny Meunier. Amsterdam: John Benjamins, 2008. pp. 27-50. Print.
- Gries, Stefan Tn. "50-something years of work on collocations. What is or should be next...". *International Journal of Corpus Linguistics*, 18.1 (2013): 137-165. Web. Sept. 2016.
- . "Phraseology and linguistic theory. A brief survey". *Phraseology: an interdisciplinary perspective*. Ed. Sylviane Granger and Fanny Meunier. Amsterdam: John Benjamins, 2008. pp. 3-26. Print.
- Keller, Nicole. *Integrating Multi-Word Terms in Terminology Management Systems: A Case Study*. Trier: Wissenschaftlicher Verlag, 2010. Print.
- Laufer, Batia and Tina Waldman. "Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English". *Language Learning* 61.2 (2011): 647-672. Web. Jan. 2016.
- L'Homme, Marie-Claude. "A methodology for describing collocations in a specialised dictionary". *Lexicography in the 21st century*. S. Ed. Nielsen & S. Tarp. Amsterdam: John Benjamins, 2009. Print.
- L'Homme, Marie-Claude and Patrick Leroyer. "Combining the semantics of collocations with situation-driven search paths in specialized dictionaries". *Terminology* 15.2 (2009):258-283. Print.
- Marco, Josep and Heike van Lawick.. "Using corpora and retrieval software as a source of materials for the translation classroom". *Corpus Use and Translating. Corpus use for learning to translate and learning corpus use to translate*. Ed. A. Beeby, P. Rodríguez Inés and P. Sánchez-Gijón. Amsterdam: John Benjamins, 2009. pp. 9-28. Print.
- Martin, Willy. "A unified approach to semantic frames and collocational patterns". *Phraseology: an interdisciplinary perspective*. Ed. Sylviane Granger and Fanny Meunier. Amsterdam: John Benjamins, 2008. pp. 51-66. Print.
- McEnery, Tony and Richard Xiao. "Parallel and Comparable Corpora: What Is Happening". *Incorporating Corpora. The Linguist and the Translator*. Ed. Gunilla Anderman and Margaret Rogers. Clevedon: Multilingual Matters, 2007. Print.
- Melby, Alan. K. "Terminology in the age of multilingual corpora". *Journal of Specialised Translation*, 18, July 2012. Web. Nov. 2016.
- Mel'čuk, Igor. "Collocations and Lexical Functions". *Phraseology. Theory, Analysis, and Applications*. Ed. A.P. Cowie. Oxford: Clarendon Press, 1998. pp. 23-54. Print.
- Nesselhauf, Nadja. *Collocations in a Learner Corpus*. Amsterdam: John Benjamins, 2005. Print.
- Peters, Elke. "The learning burden of collocations: The role of interlexical and intralexical factors". *Language Teaching Research*, 26.1 (2016):113-138. Web. Dec. 2016.
- Pimentel, Janine. "Using Frame Semantics to build a bilingual lexical resource on legal terminology". *Handbook of Terminology*. Eds.: Hendrik J. Kockaert & Frieda Steurs. Amsterdam: John Benjamins, 2015. 425-450. Print.

- Postolea, Sorina. *Terminology and Translation of General Use ICT Texts: A Romanian Perspective*. Iași: PhD Thesis, 2014.
- Schmid, Hans-Jörg. "Collocation: hard to pin down, but bloody useful". *Z A A*, 51. 3 (2003): 235-258.
- Siepmann, D. "Phraseology in learners' dictionaries: What, where and how?". *Phraseology in Foreign Language Learning and Teaching*. Ed. F. Meunier & S. Granger. Amsterdam: John Benjamins, 2008.
- Sinclair, John. *Trust the Text. Language, Corpus and Discourse*. Ed. R. Carter. London: Routledge, Taylor & Francis e-Library, 2004.
- Stubbs, Michael. *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishing, 2002.
- Wanner, Leo, Bernd Bohnet, and Giereth, Mark. "Making sense of collocations". *Computer Speech & Language* 20.4,(2006):609–624.
- Webb, Stuart and Eve Kagimoto. "Learning Collocations: Do the Number of Collocates, Position of the Node Word, and Synonymy Affect Learning?" *Applied Linguistics* 32.3 (2011): 259–276.
- Zanettin, Federico, Silvia Bernardini, and Dominic Stewart, (Eds.). *Corpora in Translator Education*. Manchester: St. Jerome Publishing, 2003. Print.
- Zanettin, Federico. "Corpora in Translation Practice". ResearchGate. 2002. https://www.researchgate.net/publication/228806527_Corpora_in_translation_practice Web. dec 2016.