Simões, Barreiro, Santos, Sousa-Silva & Tagnin (eds.) *Linguística, Informática e Tradução: Mundos que se Cruzam*, Oslo Studies in Language 7(1), 2015. 359–377. (ISSN 1890-9639 / ISBN 978-82-91398-12-9)
http://www.journals.uio.no/osla

CORPUS-DRIVEN GLOSSARIES IN TRANSLATOR TRAINING COURSES

STELLA ESTHER ORTWEILER TAGNIN

RESUMO

A Linguística de Corpus tem-se mostrado um recurso valioso para a extração de candidatos a termos e unidades fraseológicas a partir de corpora especializados (Bowker & Pearson 2002). Na realidade, trata-se de uma abordagem relativamente nova já que a maioria dos glossários baseia-se, em geral, em material similar anteriormente existente. Embora haja muitos glossários no mercado, poucos foram compilados para atender às necessidades dos tradutores, cuja principal tarefa na tradução técnica é produzir um texto natural e fluente, seja na sua língua nativa, ou em uma língua estrangeira. Por essa razão, um glossário que consiste simplesmente de uma lista de termos e seus equivalentes não será satisfatório para o tradutor. Como produtores de texto, os tradutores precisam saber como a palavra é usada, ou seja, com quais palavras combina (Firth 1957; Sinclair 1991). Além disso, a linguagem técnica abriga termos que consistem de várias palavras assim como unidades fraseológicas ainda mais longas. A compilação de glossários era abordada no Curso de Especialização em Tradução na Universidade de São Paulo como metodologia para melhorar o conhecimento especializado dos alunos. Após algumas experiências, verificou-se que a abordagem condizia com o que Shreve (2006) chamou de "prática deliberada", metodologia que contribui para o desenvolvimento das habilidades de pesquisa e de tradução dos alunos, levando à aquisição de conhecimento e de técnicas especializados (Maia 1997, 2002; Tagnin 2002), de que os aprendizes poderão se valer em qualquer área na qual venham a trabalhar. Este artigo descreverá como isso foi realizado em várias ocasiões, ou seja, com o recurso a uma abordagem baseada em corpus, e ilustrará, com exemplos de vários projetos, os passos seguidos.

[1] INTRODUCTION

Corpus Linguistics, an empirical approach to language studies (McEnery & Hardie 2011; McEnery & Wilson 1997), has proved to be a valuable tool for the extraction of candidates for technical terms and phraseological units (Bowker & Pearson 2002). We understand terms as words or multiword units characteristic of specialized contexts. So, for instance, *cup* is a word that belongs to the general vocabulary of the English language. However, in a culinary context, *cup* is considered a term as it refers to a measurement, not to the utensil proper. In the same vein, we consider longer phraseological units, even without a term, as terminological units when they are typical of a certain domain. For example, *roll out the pastry on a lightly floured working surface.*

Even though a methodology that uses corpora has been used in various academic studies (Teixeira 2008; Perrotti-Garcia & Rebechi 2007; Tagnin & Bevilacqua 2013), many current glossaries, mostly commercial ones, are still based on existing ones, either editing previous editions or adding to them. In contrast, corpus-driven terminology derives all its data from a specialized corpus compiled for that specific purpose.

Although there are many glossaries available on the market, few meet the needs of technical translators (Teixeira 2008), who are expected to produce a natural and fluent text, either in their mother tongue, or in a foreign language, depending on the direction they are working in. For this reason, a simple list of terms and their equivalents will not suffice. A text producer needs to know how a word is used, that is, the words it combines with (Firth 1957; Sinclair 1991). In addition, technical language may have multi-word terms and even longer phraseological units which may also enjoy the status of terms and as such should feature as stand-alone entries in reference works. For instance, this would be the case of *freshly ground black pepper* in a glossary of culinary terms.

A corpus-driven compilation of glossaries was one of the main foci of Technical Translation, one of the disciplines of the two-year Translation diploma courses[1] at the University of São Paulo. Students were required to participate in projects envisaging the construction of specialized corpora and the extraction of relevant terminology. To that end they were introduced to the methodology and tools used by Corpus Linguistics. Thus students mastered corpus-related skills, such as defining criteria to build a reliable corpus, investigating a corpus with specific computational programs, designing criteria to select examples to include in a glossary entry, developing techniques to identify equivalent terms in two different languages and, finally, building appropriate glossary entries. This methodology produced, in general, good works, some of which have already been published (Perrotti-Garcia & Rebechi 2007; Teixeira & Tagnin 2008; Tagnin 2013).

From the perspective of translator training, this "deliberate practice" (Shreve 2006) — a well-defined motivating task with an adequate level of difficulty so as to promote students' improvement, and with appropriate feedback from the teacher — certainly contributed to the development of research and translation skills, leading to specialized knowledge which students would be able to put to use in any area they might come to work in.

[1] These courses were discontinued in 2005, that is, the last group completed the course in 2007.

This paper reports on the decisions made regarding what to teach in a translator training course and describes how Corpus Linguistics can be used for terminological works.

[2] CORPORA IN THE TRANSLATION CLASSROOM

The use of corpora in translator training courses has been a fact for over two decades (Maia 1997, 2002; Tagnin 2002). In Brazil it was introduced as a methodology for the compilation of technical glossaries in the Specialization Course in Translation at the University of São Paulo in 2001. During a course on Technical Translation students were divided into thematic groups and instructed to build an English–Portuguese comparable corpus in a specialized area, that is, a corpus with original texts in both languages. They should then extract the technical terms, identify equivalents and collect examples in both languages. Glossaries resulting from this activity were made available at the course's site[2] under "Trabalhos de alunos" - "Glossário" (Student works – Glossary). In 2005, students were asked to build a bilingual glossary along the lines of a series of technical glossaries brought out by a local publisher. Each group could choose one field of study, and the best works would be submitted to the publisher for possible publication. In 2008, as part of a similar course[3], it was suggested that the whole class engage in one collective project for the construction of a Photography glossary. This project is discussed in detail in Section [4].

[2.1] *What to teach: translators' needs*

Before deciding on the format of the glossaries to be produced, it was deemed necessary to determine the translator's terminological needs (Teixeira 2008; Fromm 2008). When one reflects about this, what immediately comes to mind is that a translator needs equivalents, which is actually only partially true. As González-Jover & Sierra (2004) have already pointed out, terminology materials should help translators make decisions that are part of their daily practice. And their daily practice involves much more than just finding an equivalent.

A survey carried out by Fromm (2008) with professional translators on the features of the bilingual dictionaries they mostly use showed (see Table 1) that the dictionaries translators find more valuable, apart from the ones that present "all of the above", are the ones the results that provide a translation as well as examples. And it is this preference that has been the basis on which the template for our entries was built.

[2] http://citrat.fflch.usp.br/node/18

[3] This was a single extracurricular discipline, also called Technical Translation, but not part of a full-fledged course anymore.

Features	Respondents	Percentage
only translation of word	14	8%
translation and examples in L2	34	19%
translation and definition in L2	19	11%
translation, definition in L2, examples in L1	23	13%
translation, definition, examples and cross-references	22	12%
definition in L1 and equivalent and examples in L2	22	12%
all of the above	41	23%

TABLE 1: Preferences of professional translators as to bilingual dictionaries (adapted from Fromm (2008, pg. 65)).

[2.2] *What kind of glossary?*

Given that translators are, above all, text producers and that their goal in technical translation is to produce a natural text, they need, in addition to equivalents, examples that contextualize a certain term found in the source text as well as information about its textual and linguistic patterns. In other words, they need to know the term's collocations and phraseologies (Tagnin 2002). For terms which do not have equivalents in the target language, translators would need other translation possibilities or even suggestions for adaptation. On such occasions, cultural information may help them to choose adequate substitutions.

Let us illustrate this with an example taken from the area of Cooking. If a translator needs to translate *1 large onion, finely chopped* into Portuguese, he/she would find it useful to have a glossary which would specify that the Portuguese cognate for *finely* (finamente) does not usually occur in this context. Rather, the most natural translation for *finely* into Portuguese would be the adverb *bem* (= *well*), which renders *bem picada* (*well chopped). Another option would be the diminutive *picadinha*, with or without the adverb *bem*. Thus, the glossary would specify that the best translation options are *1 cebola grande, bem picada* or *1 cebola grande (bem) picadinha*. In the case of *finely grated Parmesan cheese*, the glossary should provide the information that the usual translation is simply *queijo parmesão* (= *parmesan cheese*), since in Brazil this kind of cheese is customarily *finely grated*. Thus, the texture is only specified when the cheese should be coarsely grated, which would be *ralado grosso* in Portuguese. The cultural gap becomes even more evident when the translator encounters the term *buttermilk*. Although the Portuguese language has a corresponding term, *leitelho*, it is not used, mainly because this product does not exist in our country. Thus, the glossary could add an explanatory note or even suggest that *buttermilk* can be replaced by "a mixture of equal parts of milk and plain yogurt" (Teixeira & Tagnin 2008).

However, much of the material available on the market does not meet these needs and is often limited to a mere list of monolexical terms and their equivalents in the target language, without providing examples or other linguistic information that can help the translator to make adequate decisions and create a text in which naturalness (Sinclair 1984) prevails. Thus, as mentioned before, it is necessary to create a model for a glossary that meets the needs of the translator. In this sense, as Krieger & Finatto (2004) have suggested, translators can be instrumental in creating new methodologies for the production of reliable terminological sources of information.

In this paper we claim that a methodology relying on the premises of Corpus Linguistics can provide this so much needed "reliable terminological source of information" for translators.

[3] CORPUS LINGUISTICS

As we know, Corpus Linguistics is an empirical approach based on the observation of a large number of texts. These texts, always authentic, constitute a corpus, which can be investigated by means of specific computational programs that produce, among other data, concordance lines (see Figure 1). Concordance lines show the search word with its surrounding co-text, and allow investigators to identify recurrent patterns, terms and phraseological units. Concordance lines can also be sorted alphabetically by the words to the right or to the left of the search word, which makes identifying recurrent patterns even easier by grouping them together. The first example (Figure 1) is a selection of concordance lines for the Portuguese word *imagem* (= *image*), taken from the Photography corpus.

```
 1    está na posição centralizada.) É exibida uma imagem ampliada. 7Pressione
 2   Á e, em seguida, carregue em•. 1 Para ver uma imagem ampliada (zoom de
 3   ando os itens de regulação Corte: Girando uma imagem ampliada Utilizando a
 4   ne [ Sair] e depois pressione•. QPara ver uma imagem ampliada (zoom de
 5    está na posição centralizada.) É exibida uma imagem ampliada. 7Pressione
 6   a 22 do Manual da Cyber-shot.) 1 Para ver uma imagem ampliada (zoom de
 7   ne [ Sair] e depois pressione•. QPara ver uma imagem ampliada (zoom de
 8    imagem vista através do visor e o tamanho da imagem captada pelas lentes
 9   visualizar, editar. modificar ou imprimir uma imagem captada no modo Adobe
10   de pixels). Quando se intenciona imprimir uma imagem captada por uma
11   ls, que são poucos, não afetam a qualidade da imagem capturada. Além disso
12   antástica, porém nada se compara a tratar uma imagem capturada com
13   sde há algum tempo trabalhamos digitalmente a imagem capturada em película
14   ls, que são poucos, não afetam a qualidade da imagem capturada. Além disso
15   ilidade da câmera para uma nova foto porque a imagem capturada primeiro
```

FIGURE 1: A selection of concordance lines for *imagem*, sorted by 1st word to the right.

The above concordance lines show the recurrence of three collocations: *imagem ampliada*, *imagem captada* and *imagem capturada*, which might indicate that they are candidate terms. Besides, one notices that *imagem capturada* occurs five times while *imagem captada*, which has the same meaning, only occurs three times. This seems to indicate that the first one is probably more common and thus a more natural choice. It is important to point out that Corpus Linguistics looks at language as a probabilistic system, that is, it observes which patterns have a higher **probability** of occurring to the detriment of those that just feature a grammatical **possibility** of occurrence (Kennedy 1998). Therefore, if a technical translator seeks to produce a natural-sounding text he/she should use the terms that are more likely to occur in the specialized area he/she is working in.

Recurring patterns in the English counterpart of the Photography corpus can be seen in Figure 2. These concordance lines show mainly verbal collocations such as *capture an image*, *copy an image*, *delete an image*, *display an image* and *edit an image*.

```
 1   video monitor. When you capture an image, it automatically appears on the
 2   imum depth of field. 2. Capture an image of a plain white object, such as
 3    hoices are: Off-If you capture an image using long exposure while this f
 4   n list box. continued Capturing an image 43 Saving the camera images as a
 5   to save the image. 44 capturing an image Saving the selected camera image
 6   e to another album You can copy an image of an album to another album.
 7   l file data is deleted. To copy an image file to the computer without ove
 8   ge and the otherimages. Copying an image to another album You cancopyan i
 9   pening image as another Copying an image 115 an album of the hard file fo
10    position in the album. Copying an image to another album You can copy an
11   mages Your camera cannot delete an image. • Cancel the protection (page R
12   monitor. When you should delete an image, select IDeletel in the Edit men
13    language. Accidentally deleted an image. Use the Recover function to rec
14   istake. • Once you have deleted an image. you cannot restore it. We recom
15   ORED AS DIGITAL DATA To display an image that is digitally stored on your
16    is displayed Tool Bar Displays an Image Information dialog box of the im
17   t at a time. Opens and displays an image file from the hard disk or MO di
18    Open File . Opens and displays an image file from the hard diskorMO disk
19    Camera (page 113) You can edit an image: changing image comments, moving
20    image file tobe opened. 4 Edit an image / Print an image/Transfer an ima
21          ed. m to 3 to (D Editing an image 113 Sorting the images 114 H
22    ky001 Sky002 Sky003... Editing an image The images in album of the hard
```

FIGURE 2: Selection of concordance lines for *image*, sorted by 1st and 2nd word on the left.

Another method to extract terminological units is by using a list of n-grams (Guinovart & Simões 2009; Maia et al. 2008). These lists show all combinations of two words (bigrams), three words (trigrams) or even longer combinations, depending on how the researcher adjusts the settings of the program being used. Again, however, these lists need to be examined by the researcher in order to decide which combinations are, in fact, terminological units.

Corpus Linguistics can be used in two ways to compile glossaries: as a methodology or as an approach. In the first case, we refer to it as corpus-based Terminology; in the second, as corpus-driven Terminology. It is the latter that was used in our courses.

[3.1] *Corpus-based Terminology*

A terminological reference source is said to be corpus-based when texts are selected because they offer a variety of defining contexts, which will be used to build the definitions for its entries. Besides, work is usually based on a pre-selected list of nouns — and only more recently of verbs — derived from an ontology, which shows the structure of the area being addressed and all of its subareas. This allows the terminologist to decide which areas to address in the glossary to be built. Once the list has been compiled, definitions and examples are extracted from the corpus built for that purpose. Basically, only pre-established **terms** and phraseological units which contain **these terms** will make up the entries of such a reference work. In short, the corpus is seen as a repository of definitions and examples (Teixeira 2008).

[3.2] *Corpus-driven Terminology*

In contrast, Corpus Linguistics is used as an approach when all entries that will make up the glossary are extracted directly from the corpus. In other words, only terms present in the texts that make up the corpus will be included in the glossary. Also, corpora are composed of the texts most commonly written or referred to by specialists, such as articles published in journals, textbooks, manuals, articles in newspapers, etc. The type of texts to be collected will depend on the area being addressed but they are expected to feature the actual and updated terminology used in that area. Whether these texts have defining contexts or not is not relevant.

N	Word	Freq.	%	Texts	%
1	THE	13,665	7.91	10	100
2	#	13,197	7.64	10	100
3	TO	4,173	2.42	10	100
4	AND	2,705	1.57	10	100
5	IN	2,621	1.52	10	100
6	A	2,560	1.48	10	100
7	CAMERA	2,216	1.28	10	100
8	IS	2,168	1.26	10	100
9	OR	2,164	1.25	10	100
10	OF	2,111	1.22	10	100
11	IMAGE	1,697	0.98	10	100
12	ON	1,643	0.95	10	100
13	YOU	1,576	0.91	10	100
14	WITH	1,309	0.76	10	100
15	FOR	1,284	0.74	10	100
16	BUTTON	1,187	0.69	10	100
17	IMAGES	1,156	0.67	10	100
18	MODE	1,043	0.60	10	100
19	YOUR	973	0.56	10	100
20	WHEN	946	0.55	10	100

TABLE 2: WordList – 20 most frequent words in the *Camera* subcorpus of the *Photography* project.

In corpus-driven terminology, the first step is to extract a list of all the words in the corpus with their frequencies (Table 2).

It is interesting to notice that most highly frequent words are grammatical words; the first content word — *camera* — only appears in position seven, which gives an indication of the field the corpus covers.

In order to establish which of these words are typical of the area being addressed, a wordlist is usually compared to another wordlist extracted from a corpus of general language, usually three to five times larger than the study corpus, which is known as a "*reference corpus*" or "*comparison corpus*." This comparison yields a list of keywords (see Table 3), which are the words that show a statistically relevant frequency in the specialized corpus in relation to the reference corpus. In other words, these lexical items are relatively more frequent in the study corpus than in the reference corpus. For this reason, they are regarded as potential "*candidate terms*."

N	Key word	Freq.	%	RC. Freq.	RC. %	Keyness
1	CAMERA	2,216	3.71	46		9,941.70
2	IMAGE	1,697	2.84	220	0.01	6,623.78
3	BUTTON	1,187	1.98	37		5,230.30
4	IMAGES	1,156	1.93	49		5,009.45
5	MODE	1,043	1.74	11		4,759.94
6	SELECT	828	1.38	94		3,284.80
7	FLASH	703	1.18	17		3,130.42
8	PHOTOGRAPHS	478	0.80	29		2,968.63
9	OR	2,164	3.62	4,022	0.25	2,937.68
10	MENU	703	1.18	61		2,874.92
11	EXPOSURE	636	1.06	39		2,684.16
12	BATTERY	587	0.98	12		2,629.87
13	SHUTTER	554	0.93	2		2,564.53
14	PRESS	835	1.40	415	0.03	2,400.90
15	CARD	655	1.10	144		2,338.61
16	KODAK	485	0.81	1		2,253.63
17	FILM	446	0.75	433	0.03	2,201.97
18	PHOTOGRAPHIC	333	0.56	2		2,196.38
19	DIGITAL	508	0.85	49		2,053.56
20	LIGHT	370	0.62	238	0.01	2,017.40

TABLE 3: KeyWord List – First 20 keywords in the *Photography* corpus.

This list, entirely extracted from the corpus, will be used as the starting point for the selection of candidate terms. Each of these candidates is examined in its context in order to identify possible collocations and longer phraseological units. This is done by running concordance lines for the search word and then looking for recurrent patterns, which can be seen in Figure 3 for the word *camera*.

```
 1          built the CCD into the worlds first CCD camera. This
 2          state video camera. During 1975 the CCD camera with its
 3         if: • The card was formatted using a DCS camera. • The
 4          computer as they are captured. Once DCS Camera Manager
 5         rting the Capture KODAK PROFESSIONAL DCS Camera Manager is
 6          images using the KODAK PROFESSIONAL DCS Camera Manager
 7       s your ``digital negative.'') Refer to DCS Camera Manager
 8          Camera Manager software. If you use DCS Camera Manager's
 9         racketing was added. Even with a digital camera, bracketin
10         you won't get as wide angle on a digital camera as you
11         fe with Kodak Ni-MH rechargeable digital camera batteries.
12          le); 2 Kodak Ni-MH rechargeable digital camera batteries
13         eries 600-800 Ni-MH rechargeable digital camera battery
14         only a Kodak Li-Ion rechargeable digital camera battery
15          WEP Appendix EasyShare-One zoom digital camera
16         visory Kodak EasyShare C433 zoom digital camera This
17                Kodak EasyShare C433 zoom digital camera User's guid
18         col), via USB cable model U-8, EasyShare camera dock or
19         col), via model U-8 USB cable, EasyShare camera dock, or
20         r pictures and videos. • Kodak EasyShare camera dock, Koda
21         dapter included with the Kodak EasyShare camera dock or
22         hich gives a medium wide angle on a film camera acts as a
23          a digital camera as you would on a film camera using the
24         l Zoom capability. If you've used a film camera, you'll be
25         ver 1500 photos -- 40 rolls -- on a film camera that cost
```

FIGURE 3: A selection of concordance lines for *camera*.

The lines from Figure 3 show various collocations and phraseological units such as *CCD camera*, *DCS camera*, *DCS Camera Manager*, *digital camera*, *(Kodak) Ni-MH rechargeable digital camera battery*, *Kodak Li-Ion rechargeable digital camera battery*, *(Kodak) EasyShare camera dock* and *film camera*. In a corpus-driven terminological reference source each one of these recurrent combinations will be listed along with relevant examples extracted from the concordance lines.

[4] THE PHOTOGRAPHY GLOSSARY PROJECT

The above sequence of activities was followed on various occasions during Technical Translation courses at the University of São Paulo. The most recent ones took place in 2005 and 2008, as mentioned before. For the sake of illustration, we will concentrate on the 2008 project on Photography, but will resort to other areas from the 2005 project when they provide better examples to illustrate the procedures being discussed.

[4.1] *Class procedures*

The first step was to establish the subareas that would be addressed in the project. Examining instructional material on Photography, we determined the following six topics to be covered: history of photography[4], light, cameras, studio, storage and digital photography. The class was accordingly divided into six groups, each of which should build a comparable bilingual English-Portuguese corpus in the area assigned to them. They also had to select a one-page text from their English corpus to be translated into Portuguese by the whole class. Each group would be responsible for discussing their translation with the whole class. Besides, preliminary results for the glossary were also to be presented so that procedures and doubts could be discussed. The stages of the project are described below.

Instruction in Corpus Linguistics

As most of the class had no previous knowledge in Corpus Linguistics, they were introduced to its basic notions in a series of three lectures, with special emphasis on the stages of building a specialized corpus and using linguistic software to investigate it, in that case, *WordSmith Tools* version 5 (Scott 1996), with its suite of tools: *WordLists, Keywords* and *Concord.*

Building a corpus

Students were required to build a bilingual comparable corpus with approximately 100,000 words in each language according to the following steps:

(i) search for texts on the Internet so as to avoid having to scan them. Although most texts were indeed retrieved from the Internet, some groups had to resort to written material and hence scan it;

(ii) clean the texts, eliminating figures, tables, charts, illustrations and any other non-linguistic material which the researcher believes will not contribute relevant material[5];

(iii) save texts in .txt format;

(iv) include a header with metatextual information such as: title of the text, place of publication, date of publication, subarea etc.

The final composition of the five subcorpora compiled by each group is presented in Table 4.

Extracting terms (Wordlist and Keywords)

Once the corpora were built, students generated WordLists for each of their corpora and then compared these lists with similar lists for general language corpora.

[4] This group was discontinued during the course.

[5] It is true that some tables exhibit terminological material, though not in context. It is up to the researcher, in those cases, to include the tables or not in the corpus.

Subcorpus	Number of words
Camera	72,665
Digital photography	72,864
Light	36,668
Storage	59,803
Studio	72,716
Total	**314,716**

TABLE 4: Final composition of the *Photography* corpus.

This comparison yielded words — keywords — that occurred at a statistically significant higher frequency in the study corpus (see Table 3).

These words were considered candidate terms as they were peculiar to the study corpus. In order to confirm whether they were actually terms or not, students ran concordance lines for each of the words to examine their context of occurrence (see Figure 3).

Extracting patterns

Let us remember that recurrent patterns in concordance lines may be candidate terms. Figure 4 shows some of these patterns for the word *photographs.*

The Figure 4 concordance lines allow us to identify nominal collocations such as *albumen photographs*, *colo[u]r photographs*, *digital photographs* and *family photographs*, as well as verbal collocations like *clean photographs*, *display photographs* and even longer phraseological units like *water-damaged photographs.*

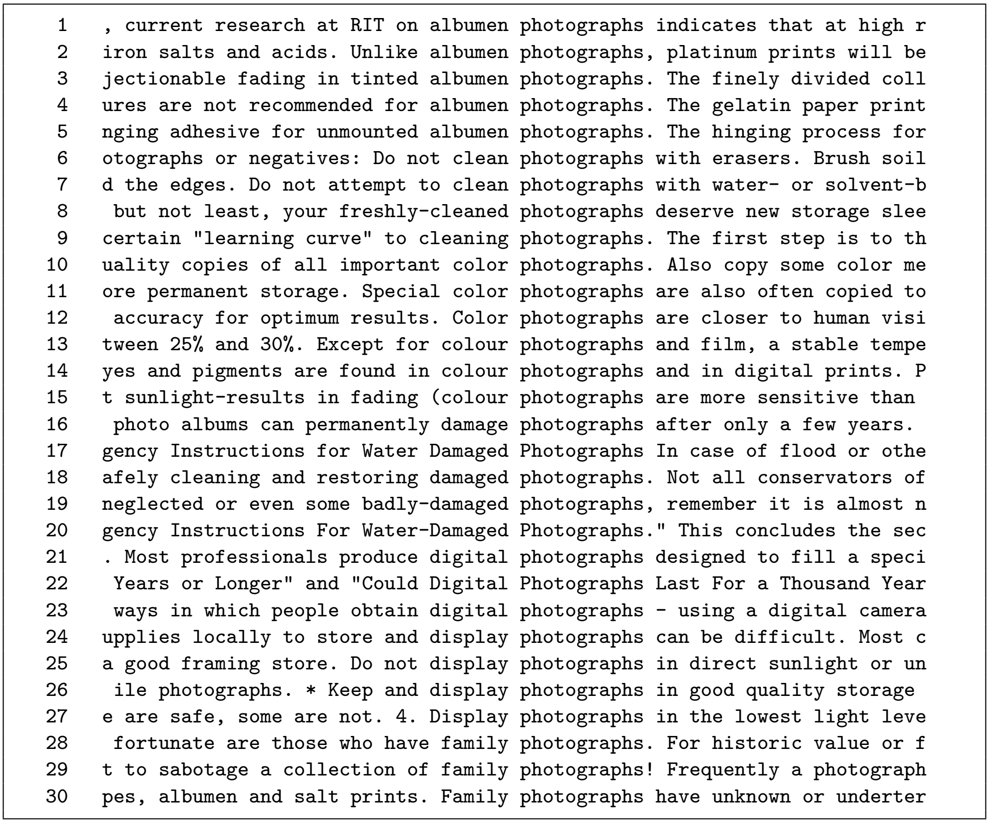
Extracting relevant context (examples)

Once all relevant terms and phraseologies had been identified, examples were retrieved from the concordance lines to be inserted in the entries. If the concordance line did not show the full context, a double click on it led to the full source text. Part of it is shown below for concordance line 25 in Figure 5.

Identifying equivalents

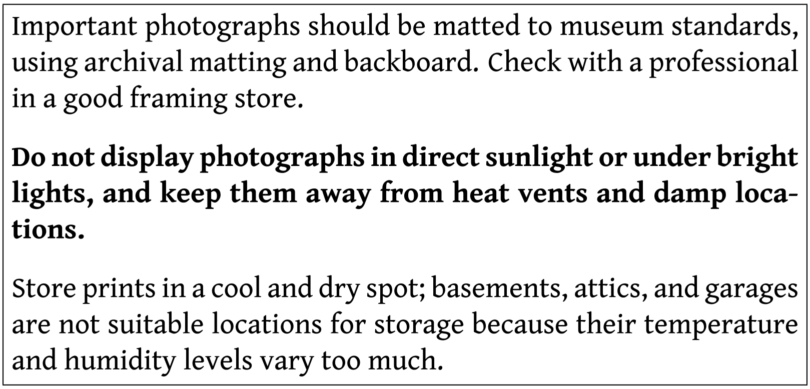
One way to identify possible equivalents is to compare the lists of keywords in both languages. Figure 6 illustrates this procedure for an English–Portuguese Cooking glossary (Teixeira & Tagnin 2008).

Once a pair is identified, concordance lines should be generated to check whether the selected equivalents occur in similar contexts. When there is no such *prima facie* (literal) equivalent, search can be pursued by the word's collocates or context (Tagnin 2007). For example, if we wish to find the equivalent for *finely* — the most frequent adverb in a Cooking corpus — we will realize that it is not *finamente*, the Portuguese cognate for *finely*, because this adverb displays a very



```
  1   , current research at RIT on albumen photographs indicates that at high r
  2   iron salts and acids. Unlike albumen photographs, platinum prints will be
  3   jectionable fading in tinted albumen photographs. The finely divided coll
  4   ures are not recommended for albumen photographs. The gelatin paper print
  5   nging adhesive for unmounted albumen photographs. The hinging process for
  6   otographs or negatives: Do not clean photographs with erasers. Brush soil
  7   d the edges. Do not attempt to clean photographs with water- or solvent-b
  8    but not least, your freshly-cleaned photographs deserve new storage slee
  9   certain "learning curve" to cleaning photographs. The first step is to th
 10   uality copies of all important color photographs. Also copy some color me
 11   ore permanent storage. Special color photographs are also often copied to
 12    accuracy for optimum results. Color photographs are closer to human visi
 13   tween 25% and 30%. Except for colour photographs and film, a stable tempe
 14   yes and pigments are found in colour photographs and in digital prints. P
 15   t sunlight-results in fading (colour photographs are more sensitive than
 16    photo albums can permanently damage photographs after only a few years.
 17   gency Instructions for Water Damaged Photographs In case of flood or othe
 18   afely cleaning and restoring damaged photographs. Not all conservators of
 19   neglected or even some badly-damaged photographs, remember it is almost n
 20   gency Instructions For Water-Damaged Photographs." This concludes the sec
 21   . Most professionals produce digital photographs designed to fill a speci
 22    Years or Longer" and "Could Digital Photographs Last For a Thousand Year
 23    ways in which people obtain digital photographs - using a digital camera
 24   upplies locally to store and display photographs can be difficult. Most c
 25   a good framing store. Do not display photographs in direct sunlight or un
 26    ile photographs. * Keep and display photographs in good quality storage
 27   e are safe, some are not. 4. Display photographs in the lowest light leve
 28    fortunate are those who have family photographs. For historic value or f
 29   t to sabotage a collection of family photographs! Frequently a photograph
 30   pes, albumen and salt prints. Family photographs have unknown or underter
```

FIGURE 4: Selection of concordance lines for *photographs* sorted by 1st word to the left.



Important photographs should be matted to museum standards, using archival matting and backboard. Check with a professional in a good framing store.

Do not display photographs in direct sunlight or under bright lights, and keep them away from heat vents and damp locations.

Store prints in a cool and dry spot; basements, attics, and garages are not suitable locations for storage because their temperature and humidity levels vary too much.

FIGURE 5: Expanded context in source text. Relevant concordance line highlighted by author.

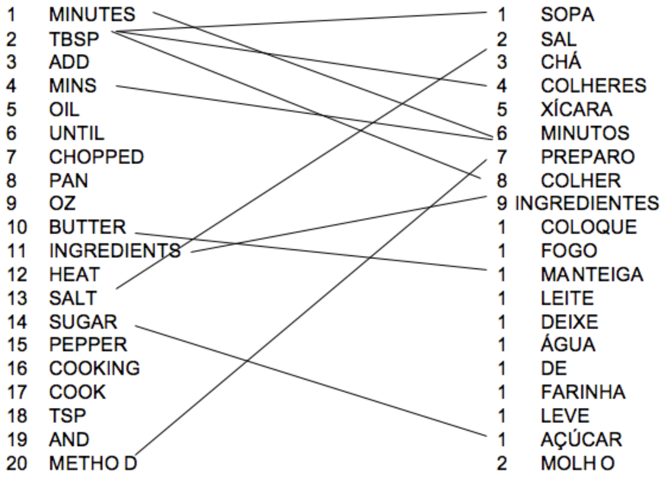


FIGURE 6: Matching candidate terms in bilingual keyword lists.

low frequency in the Portuguese Cooking corpus. So, we can look at the collocates of *finely* and see with which words they occur in the target language corpus. One of these collocates is *chopped*, *picado* in Portuguese. The concordance lines will show that *picado* co-occurs with *bem*, yielding the collocation *bem picado*, but they also show a typical Portuguese term *picadinho*, which may also occur with *bem*: *bem picadinho* (Figure 7).

2 cebolas médias **bem picadas**
½ dente de alho **bem picado**
junte os tomates pelados **bem picados**.
Calabresa **picadinha**
100 g de bacon **picadinho**
2 dentes de alho **picadinhos**
Polvilhar salsa **bem picadinha**

FIGURE 7: Selection of some concordance lines for *picad**, sorted by 1st word to the left.

If even this procedure does not reveal an equivalent, it may be because there is no equivalent in the target language. Thus, in such instances, it would be useful to suggest an adaptation or insert an explanatory note, as was the case for *buttermilk*, mentioned earlier in this paper. Because we are dealing with a comparable corpus, with original texts in both languages, this kind of information may be retrieved from the corpus itself.

Building entries

To meet translators' needs, as discussed above, entries portrayed the following information:

(1) **head word** (*part-of-speech*)
(2) Example in English
(3) **equivalent**
(4) example in Portuguese
(5) *Comments* (*if necessary*)
(6) → *cross-reference*

Here are a few sample entries from the Photography glossary:

(1) **acid-free** (*adj.*)
(2) For added protection, **acid-free** envelopes and boxes are availabe from conservation suppliers.
(3) **de pH neutro**
(4) Só são aceitáveis para embalagens de arquivo de fotografias papéis **de pH neutro** ou próximo de neutro, isentos de lignina e sem corantes.
(5) *Termo usado quando um produto contém nível de pH acima de 7.0. Indica que em sua composição não foi utilizado nenhum componente com reação ácida ou que, com o passar do tempo se decomponha produzindo resíduos ácidos que causam sérios danos às fotografias.*

(1) **adapter card** (*n.*)
(2) The **adapter card** may have multiple ports.
(3) **cartão adaptador**
(4) Conecte a extremidade de 6 pinos do cabo em qualquer port disponível ao **cartão adaptador** IEEE 1394 do computador.

(1) **additional development** (*n.*)
(6) → development, additional

At the end of this process, students had built their bilingual glossaries, which were examined by the instructor and returned with comments and suggestions. This way, students had the opportunity to revise their work and make any necessary changes, adjustments or additions. Only the final version was evaluated.

[5] RESULTING PRODUCTS

As mentioned above, this procedure was carried out on two occasions, 2005 and 2008. From the glossaries produced by the 2005 class, one on Chemistry was published in 2007 (Perrotti-Garcia & Rebechi 2007).

A Cooking glossary built along the same lines was produced by a former translation student and co-authored by me (Teixeira & Tagnin 2008). Although not part of either the 2005 or the 2008 project, it is an offspring of a glossary on Cooking spices and condiments compiled in the 2001 course. After finishing the Translation course, Teixeira pursued her master's degree with a thesis on the translation of cooking recipes (Teixeira 2004) and her PhD with a dissertation on a proposal for a Cooking dictionary aimed at a translator's textual production (Teixeira 2008)[6].

The results of the Photography project, unsurprisingly, were a bit uneven. One group excelled and one presented very poor material. The work of the other groups was good but needed some improvement. As the aim was to submit high quality material to a publisher and only one glossary met this requirement, after grades had been assigned, the instructor called a meeting of those who would be interested in pursuing the project on their own time and making all necessary adjustments for the work to be suitable for submission to the publisher. A group of five students[7] decided to embrace the project and the final material was submitted in early 2009. As it is the publisher's policy to have all technical glossaries revised by a professional in the area, the material was examined by a professional photographer who returned it with a few comments and suggestions. These were worked on by the group and the *Vocabulário para fotografia* was eventually published in 2013 (Tagnin 2013).

[6] AN INTERESTING OUTCOME

A couple of years ago I participated in a round table on the teaching of translation. One of my colleagues, Fabio Alves, from the Federal University of Minas Gerais, presented the concept of "deliberate practice." It goes something like this: for students to acquire translation competence, their training should aim at developing specific skills that will contribute to their optimal learning and expert performance in a certain field (Ericsson & Charness 1997). This requires certain conditions to be met, among which the most mentioned one is "subjects' motivation to attend to the task and exert effort to improve their performance" (Ericsson et al. 1993, pg. 367) .This is developed by Shreve (2006, pg. 29) who states that for deliberate practice to occur, the following requirements must be met:

[6] Both works were done under the author's supervision.

[7] Angelica Royo, Eliana C. R. Antonopoulos, Helena Akemi Misumi, Moira Martins de Andrade and Veridiana Rocha Schwenck.

(i) tasks should be well-defined;

(ii) they should involve appropriate difficulty for the student;

(iii) there should be possibility for informative feedback;

(iv) there should be opportunities for repetition and correction.

Although the methodology described did not follow – at least consciously – any learning theory, it has been pointed out (Alves & Tagnin 2010) that it met the conditions for Shreve's "deliberate practice." First of all, the task was highly motivating because, quality permitting, the final outcome would be submitted to a publisher who brings out a collection of technical glossaries. Besides,

(i) the task was **well-defined**: students knew what was expected of them and they were instructed in the stages to be followed to complete the project;

(ii) it involved **an adequate degree of difficulty** as most of the class had no previous training in Corpus Linguistics;

(iii) students received **informative feedback** by means of comments and suggestions provided by the instructor, both throughout the course and on the pre-final glossary;

(iv) students had the chance to **repeat any of the stages**, if needed, and **make necessary corrections** and only then turn in their final version.

Further evidence was obtained in a questionnaire aimed at checking whether the above conditions had actually been met. The questionnaire was answered by both authors of the Chemistry glossary (Perrotti-Garcia & Rebechi 2007) independently – two years later. They remarked that a) they learned a lot in their work with corpora; b) they realized that, in retrospective, they could have produced a more complete glossary, which attests to the fact that they had incorporated the methodology into their professional practice; c) the feedback they received from the technical reviser (as part of the publisher's preparation of the final version of the material for publication) helped them to improve the glossary. One of the authors underscored that "the methodology really worked and that the use of a corpus can help overcome difficulties which are inherent to working in an area in which one is not an expert" (Alves & Tagnin 2010). It must be remembered that they were both students and not experts in chemistry, although one of them had studied a bit of Biochemistry as part of her training as a dentist. They also mentioned that the reviser, a translator and chemical engineer, commented that she "would never have been able to collect the terms as [the authors] did".

[7] FINAL REMARKS

This article was intended to demonstrate how a corpus-driven methodology can produce glossaries that meet the translator's needs and how this practice can enhance students' translation competence.

The methodology described showed that building corpus-driven glossaries can be an adequate practice to enhance students' performance towards achieving translation competence. First, because Corpus Linguistics has shown to be an effective approach to build technical glossaries that meet the translator's needs. Second, because, as it was later discovered, the methodology was considered an adequate practice in helping students to achieve specialized knowledge and master translation techniques which they will be able to put to use in any technical area they may come to work in (Alves & Tagnin 2010).

REFERENCES

Alves, Fábio & Stella Esther Ortweiler Tagnin. 2010. Corpora e ensino de tradução: o papel do auto-monitoramento e da conscientização cognitivo-discursiva no processo de aprendizagem de tradutores novatos. In Vander Viana, Stella Esther Ortweiler Tagnin & Fábio Alves (eds.), *Corpora no ensino de línguas estrangeiras*, 189–203. HUB Editorial.

Bowker, Lynne & Jennifer Pearson. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora.* Routledge.

Ericsson, Anders, Ralf Th. Krampe & Clemens Tesch-Romer. 1993. The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review* 100. 363–406.

Ericsson, K. Anders & Neil Charness. 1997. Cognitive and developmental factors in expert performance. In P. J. Feltovich, K. M. Ford & R. R. Hoffman (eds.), *Expertise in context: Human and machine*, 3–41. MIT Press.

Firth, John Rupert. 1957. *Papers in linguistics 1934-1951.* Oxford University Press.

Fromm, Guilherme. 2008. *Votec: A construção de vocabulários eletrônicos para aprendizes de tradução.* São Paulo: Universidade de São Paulo PhD dissertation.

González-Jover, Adelina Gómez & Chelo Vargas Sierra. 2004. Aspectos metodológicos para la elaboración de diccionarios especializados bilingües destinados al traductor. In L. González & P. Hernuñez (eds.), *Las palabras del traductor: Actas del II Congreso El español, lengua de traducción*, 365–398.

Guinovart, Xavier Gomez & Alberto Simões. 2009. Parallel corpus-based bilingual terminology extraction. In Marie-Claude L'Homme & Sylvie Szulman (eds.), *8th international conference on terminology and artificial intelligence*, .

Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics.* Longman.

Krieger, Maria da Graça & Maria José Bocorny Finatto. 2004. *Introdução à Terminologia: Teoria e prática.* Contexto.

Maia, Belinda. 1997. Do it yourself corpora... with a little bit of help from your friends! In B. Lewandowska-Tomaszczyk & P. J. Melia (eds.), *Practical applications in language corpora*, 403–410. Lodz University Press.

Maia, Belinda. 2002. Do-it-yourself, disposable, specialised mini corpora - where next? Reflections on teaching translation and terminology through corpora. *Cadernos de Tradução* 1(9). 221–235.

Maia, Belinda, Rui Silva, Anabela Barreiro & Cecília Fróis. 2008. N-grams in search of theories. In Barbaba Lewandowska-Tomaszczyk (ed.), *Corpus linguistics, computer tools, and applications: State of the art*, vol. 17 Lodz Studies in Language, 71–84.

McEnery, Tony & Andrew Hardie. 2011. *Corpus Linguistics: Method, Theory and Practice.* Cambridge University Press.

McEnery, Tony & Andrew Wilson. 1997. *Corpus Linguistics.* Edinburgh University Press.

Perrotti-Garcia, Ana Júlia & Rozane Rodrigues Rebechi. 2007. *Vocabulário para química - Português-Inglês / Inglês-Português* Série Mil & Um Termos. SBS.

Scott, Mike. 1996. *Wordsmith tools.* Oxford University Press.

Shreve, Gregory. 2006. The deliberate practice: translation and expertise. *Journal of Translation Studies* 9(1). 27–42.

Sinclair, John McHardy. 1984. Naturalness in language. In Jan Aarts & W. Weijs (eds.), *Corpus Linguistics*, Rodopi.

Sinclair, John McHardy. 1991. *Corpus, concordance, collocation.* Oxford University Press.

Tagnin, Stella Esther Ortweiler. 2002. Os corpora: instrumentos de auto-ajuda para o tradutor. *Cadernos de Tradução* 1(9). 191–219.

Tagnin, Stella Esther Ortweiler. 2007. A identificação de equivalentes tradutórios em corpora comparáveis. In *I Congresso Internacional da ABRAPUI*, s/pp.

Tagnin, Stella Esther Ortweiler. 2013. *Vocabulário para fotografia.* SBS.

Tagnin, Stella Esther Ortweiler & Cleci Regina Bevilacqua. 2013. *Corpora na terminologia*. HUB Editorial.

Teixeira, Elisa Duarte. 2004. *Receitas qualquer um traduz. Será? - a Culinária como área técnica de tradução*. Universidade de São Paulo MSc thesis.

Teixeira, Elisa Duarte. 2008. *A Linguística de Corpus a serviço do tradutor: Proposta de um dicionário de Culinária voltado para a produção textual*: Universidade de São Paulo PhD dissertation.

Teixeira, Elisa Duarte & Stella Esther Ortweiler Tagnin. 2008. *Vocabulário para Culinária inglês-português* Série Mil & Um Termos. SBS.

CONTACTS

Stella Esther Ortweiler Tagnin
Universidade de São Paulo
seotagni@usp.br