

# **Corpus representativeness in the selection of medical terms to be used in translation memory tools**

Paula Tavares Pinto Paiva

São Paulo State University – UNESP/ FCLAr

CAPES/Fundunesp

**Abstract** : Scientific journals have become one of the most important sources of information in medical areas. In Brazil, editors have published papers in two, sometimes, three languages, so that study results from Brazilian researchers will become known worldwide. This paper aims at presenting a Corpus-based study on bilingual medical terminology found on journals of Cardiology and Cardiac Surgery (Paiva, 2007, 2008). After using the software *WordSmith Tools* (1999) to identify representative terms in Portuguese and in English, we compiled two glossaries which were tested in a translation memory tool called *Wordfast* (Champollion, 1999). This experience showed that carefully elaborated glossaries containing representative terms proved to be a very helpful aid for translators who work or want to work with medical articles. We hope this study may illustrate how research and practice are well connected to each other and may produce significant research for the area of Corpus-based Translation Studies.

## **Introduction**

The use of constantly updated terms due to new discoveries and technologies in the medical field is evidenced by the reading of scientific journals which have become an important source of updating for medical professionals and researchers. These publications allow the diffusion of recent studies developed in several countries, including Brazil. A noteworthy aspect from these studies is the common use of new terminology connected to newly discovered findings.

Regarding this issue, many professional and future translators have looked for literature and special courses in order to learn how to use it proficiently since they know it might take years of experience to achieve this proficiency. Also, in recent years, an increasing number of translators have used translation memory tools to speed up the translating process as these tools provide previously used segments in the target language available for this professional. With this approach, the translator may decide to reuse the segment or words offered by the tool or not.

Bearing that in mind, this paper aims at describing a corpus-based research on medical terms which may help the work of specialized translators. The focus is given to the creation of a medical corpus composed of journals in two fields so that it would be possible to compile glossaries that could be inserted into translation memory tools.

In this article we will, firstly, point out the use of translation memory tools in Brazil and the relevant role of computerized corpora for the selection of terms to be included in bilingual glossaries; in part 2, we will introduce the types of corpora used in this study; section 3 will discuss, in four subsections, the whole process for the selection of medical terms used in our research, which was based on keyword lists; section 4 will introduce the definition of glossary used in the compilation of the medical glossaries of this research; in part 5 we will discuss the process of glossary insertion in a memory tool; finally, in section 6, we highlight the advantages of using corpus linguistics methodology in a terminological research such as this one and present the project which is currently being developed by our research team at the São Paulo State University/FCLAr, in Brazil .

## **1. Translation memory tools and corpora**

In Brazil, papers on the advantages and disadvantages of using translation memories are still scarce, although one can notice an increase of researchers interested in the role of these tools which influence the work of translators. One of the publications dedicated to discussions on the

use of translation memory tools and machine translation software is *Cadernos de Tradução* 14 (2004), entirely devoted to the subject. In this publication the article written by Nogueira & Nogueira (2004), two well-known professional Brazilian translators, show how enthusiastic they are about using these kinds of software. The authors state that some “translators tend to technological conservatism” (2004:18) but they add that these tools are useful, if not indispensable, for all kinds of translators and translation. They also say that for highly repetitive texts there are gains in quality and comfort for the translator regardless the type of text he is used to working with.

Considering the use of glossaries, the authors point out that most tools allow you to use more than one glossary at a time and they show all translations in every glossary, indicating that a certain terminology is preferred by a specific client, even if it conflicts with the content of glossaries.

The same publication also brings some perspective to the use of lexical analysis tools such as the software package *WordSmith Tools* (Scott, 1999), by companies that deal with a great number of translations. Such studies show how these tools can help in the preparation of bilingual glossaries which will be useful for translators. According to Ribeiro (2004), *WordSmith Tools* aims to help lexical analysis and the study of collocations, clusters and statistics related to the texts within a corpus. Some of its tools are *WordList*, used to generate lists of words from the corpus under study, *KeyWords* responsible for generating lists of keywords from the list generated by the *Wordlist*, and *Concord*, which generates lists of concordances or the occurrence of a specific item.

The fact that *Wordsmith Tools* deals with large amount of information and can be quick in providing information for specific analyses, gives it a feature of impartiality and representativeness for the analyst’s or researcher’s study, something that would be difficult to be done manually.

Regarding the use of tools developed for lexical analysis, as highlighted earlier, and the use of translation memory tools by professionals, Ribeiro (2004:161) states that “there is a clear division between the tools intended for scholars and those aimed at the professional market”. In general, according to the author “the tools reflect the priority of each sector: research and teaching, for universities, and productivity for the market”.

Ribeiro's opinion reflects a gap between recent theoretical studies carried out in universities and the needs of professional translators who deal with the current job market. Therefore, this research was carried out in order to develop a methodology for compiling bilingual glossaries containing medical terms which could be tested by students and professional translators in a translation memory tool called *Wordfast* (Champollion, 1999). These glossaries were compiled with the criteria set by Corpus Linguistics which are, among others, authenticity, content and representativeness.

## **2. Corpora in this study**

Several researchers have emphasized the importance of computerized corpora in the studies of Linguistics, Lexicography, Translation Studies and related areas (Sinclair, 1991; Stubbs, 1995; Biber *et al* 1998; Baker, 1993, 1995, 1996; Laviosa, 1997, 2002; Tognini-Bonelli, 2001; Tagnin, 2002; Berber Sardinha, 2004) precisely because of the possibility of working with a large amount of data which could be a guide into the search for lexical patterns. As many scholars have talked about computerized corpus, there have been several definitions for it. Summarizing, Carter and McCarthy define *corpus* as “a collection of texts, usually stored in computer-readable form” (2006: 11). Sinclair (2004) adds that “a corpus is made for the study of language” and emphasizes that some of the criteria to build a corpus are: representativeness, size, balance and homogeneity.

In this research we have compiled two kinds of corpora, that is, a *parallel corpus*, composed of original texts and respective translated texts and a *comparable corpus* composed of texts originally written in two languages (Baker, 1993, 1995, 1996; 2000; Olohan, 2004, 2007) which, in this case, are Portuguese and English. These texts were taken from two major Brazilian bilingual medical journals as well as six International journals to compose the non-translated comparable corpus. Both corpora are part of a larger corpus called *Brazilian MedCorp*, which currently contains 710.322 tokens of translated and non-translated papers in Portuguese and in English in the medical areas of Anesthesiology, Cardiology, Cardiac Surgery. In this paper, we are going to discuss the compilation of terms from the papers of Cardiology and Cardiac Surgery only.

The *Translated Brazilian Medical Corpus* (TBMC) is composed of four subcorpora, two in each language, divided into the two areas of medicine analysed in this study. The TBMC of Cardiology, in Portuguese, contains 45.788 tokens and its correspondent in English contains 46.661 tokens. The articles were all extracted from the *Brazilian Journal of Cardiology* (different editions during 2003). The TBMC of Cardiovascular Surgery contains 48.710 tokens and its correspondent in English contains 47.789 tokens. Articles were extracted from the *Brazilian Journal of Cardiovascular Surgery* (different editions during 2007).

On the other hand, the *Comparable Medical Corpus* (CMC) of Cardiology, in Portuguese, contains 113.508 tokens, all articles taken from the journal *Arquivos Brasileiros de Cardiologia* (2003, 2004). Its correspondent in English contains 137.905 tokens, taken from the journals *BMC Cardiovascular Disorders* (2004); *Current Interventional Cardiology Reports* (2001); *Journal of the American College of Cardiology* (2003); and *Hellenic Journal of Cardiology* (2004).

The CMC of Cardiovascular Surgery, in Portuguese, contains 107.620 tokens, extracted from the journal *Revista Brasileira de Cirurgia Cardiovascular* (2006 and 2007) whereas its

correspondent in English contains 162.345 tokens, all articles taken from the journals *The Journal of Thoracic and Cardiovascular Surgery* and *Interactive and CardioVascular Thoracic Surgery* ( 2006 and 2007).

Discussion on the process of keyword selection will be shown in the next section of this paper.

### **3. Keywords and the selection of simple and complex terms with the aid of *WordSmith Tools***

As discussed previously, the glossaries from this study were drawn up based on the approach of Corpus Linguistics. The theoretical details can also be found in Paiva & Camargo (2007) and Paiva *et al* (2008). In order to generate lists of keywords for each subcorpus in Portuguese we used two reference corpora : the corpus *Lácio-Ref* (4.156.816 tokens) and the corpus *Folha de São Paulo 97* (235.036 tokens). Thus, we obtained eight lists: a) two lists of keywords of Cardiology from the TBMC; b) two lists of keywords from the TBMC of Cardiac Surgery; c) two lists of keywords from the CMC of Cardiology and d) two lists of keywords from the CMC of Cardiac Surgery. The same process was used to obtain the keyword lists in English, this time, using as reference corpora the BNC and TEC. Next we will discuss how each list of keywords was generated and how we selected the words to guide the compilation of the glossaries.

#### **3.1. Selection of terms of Cardiology in the TBMC**

When generating the list of keywords from the subcorpus of Cardiology in Portuguese, and, afterwards, showing the list to a specialist in this medical area we obtained a list of the top twenty noun and adjective keywords using the *Lacio-Ref* corpus as reference:

<b>Ranking</b>	<b>Keywords in Portuguese</b>	<b>Frequency</b>	<b>%</b>	<b>Keyness</b>
1 <sup>st</sup>	ARTERIAL	197	0,43	1.097,70
2 <sup>nd</sup>	PACIENTES	216	0,47	1.005,30
3 <sup>rd</sup>	HIPERTENSÃO	164	0,36	901,7
4 <sup>th</sup>	VENTRÍCULO	119	0,26	662,9
5 <sup>th</sup>	ESQUERDO	105	0,23	573,7
6 <sup>th</sup>	MORTALIDADE	111	0,24	566,1
7 <sup>th</sup>	COLS	91	0,2	506,9
8 <sup>th</sup>	ARTÉRIA	86	0,19	479
9 <sup>th</sup>	MITRAL	86	0,19	479
10 <sup>th</sup>	SISTÓLICA	84	0,18	467,9
11 <sup>th</sup>	ATRIAL	74	0,16	412,2
12 <sup>th</sup>	DOENÇAS	97	0,21	408,1
13 <sup>th</sup>	CARDÍACA	72	0,16	401
14 <sup>th</sup>	ARTÉRIAS	71	0,16	395,4
15 <sup>th</sup>	INSUFICIÊNCIA	77	0,17	386,6
16 <sup>th</sup>	PULMONAR	68	0,15	378,7
17 <sup>th</sup>	SISTÊMICA	73	0,16	375,6
18 <sup>th</sup>	CARDIOVASCULAR	70	0,15	371,9
19 <sup>th</sup>	DOENÇA	90	0,2	365
20 <sup>th</sup>	ESTUDO	146	0,32	350,8

Likewise, and now, to confirm the presence of the same keywords, we generated a new list using another reference corpus, *Folha de Sao Paulo 97*, also in Portuguese :

<b>Ranking</b>	<b>Keywords in Portuguese</b>	<b>Frequency</b>	<b>%</b>	<b>Keyness</b>
1 <sup>st</sup>	ARTERIAL	197	0,43	2.289,70
2 <sup>nd</sup>	HIPERTENSÃO	164	0,36	1.697,10
3 <sup>rd</sup>	VENTRÍCULO	119	0,26	1.596,50
4 <sup>th</sup>	PACIENTES	216	0,47	1.466,40
5 <sup>th</sup>	COLS	91	0,2	1.210,30
6 <sup>th</sup>	MITRAL	86	0,19	1.135,80
7 <sup>th</sup>	SISTÓLICA	84	0,18	1.085,70
8 <sup>th</sup>	P	149	0,33	1.039,50
9 <sup>th</sup>	ATRIAL	74	0,16	999,9
10 <sup>th</sup>	ARTÉRIA	86	0,19	981,3
11 <sup>th</sup>	MORTALIDADE	111	0,24	867,2
12 <sup>th</sup>	VENTRICULAR	62	0,14	837,7
13 <sup>th</sup>	SISTÊMICA	73	0,16	816,7

14 <sup>th</sup>	CARDIOVASCULAR	70	0,15	779,5
15 <sup>th</sup>	CORONÁRIA	62	0,14	758,6
16 <sup>th</sup>	ESQUERDO	105	0,23	717
17 <sup>th</sup>	ARTÉRIAS	71	0,16	709,4
18 <sup>th</sup>	TROMBO	55	0,12	699,4
19 <sup>th</sup>	VASCULAR	61	0,13	690,6
20 <sup>th</sup>	MIOCÁRDIO	59	0,13	690,2

As we may see, fourteen keywords are common to both lists of keywords from the subcorpus of Cardiology: *arterial* (arterial), *pacientes* (patients), *hipertensão* (hypertension), *ventrículo* (ventricle), *esquerdo* (left), *mortalidade* (mortality), *cols. (al)*, *artéria* (artery), *mitral* (mitral), *sistólica* (systolic), *atrial* (atrial), *artérias* (arteries), *sistêmica* (systemic) and *cardiovascular* (cardiovascular).

It was also observed that other keywords that were not among the top twenty, were part of the first one hundred in both lists : *doença* (disease), *cardíaca* (heart/cardiac), *insuficiência* (failure), *pulmonar* (lung/pulmonary), *doença* (disease), *estudo* (study), *p* (p), *ventricular* (ventricular), *coronária* (coronary), *artérias* (arteries), *trombo* (thrombus), *vascular* (vascular), and *miocárdio* (myocardium/myocardial)

Among the keywords in the lists we could notice the presence of terms which are not only part of Cardiology, but also part of all subareas of Medicine, such as *pacientes* (patients), *mortalidade* (mortality) and *doença* (disease). Once they are representative in the subcorpus we decided to keep them in the glossary that was being compiled at the time.

Both lists of keywords generated with two different reference corpora in Portuguese, the *Lacio-Ref* and *Folha de Sao Paulo 97*, proved the representativeness of the candidate of terms for the subarea of Cardiology. This result indicates the validity of a research in Terminology which uses the methodology of Corpus Linguistics.



After the process carried out with the subcorpora in Portuguese, we did the same procedure to generate the keywords in English using as reference corpora the *BNC* and *TEC*. Among the twenty words present in both lists, fifteen were the same: “patients”, “coronary”, “ventricular”, “hypertension”, “artery”, “mortality”, “study”, “cardiovascular”, “diseases”, “mitral”, “pressure”, “pulmonary”, “atrial”, “posterior” and “disease”. Five other words that were not in the top twenty but were among the first thirty words in the lists were “arteries”, “myocardial”, “valve”, “reported” and “systolic”. The word “heart”, which ranked as the thirteenth word in the list with BNC as reference corpus, ranked the seventy-fourth position in the list having TEC as reference corpus. This result, together with the analysis of keywords in Portuguese, showed, once again, that the representative words in a corpus of study appear in the list of keywords even if they have different reference corpora.

The keywords selected from the parallel subcorpora of Portuguese and English which were considered simple terms and guided the selection of complex terms were the following:

Table 1. Simple Cardiology terms present in the parallel subcorpus

1. <i>ARTÉRIA</i>	1. ARTERY
2. <i>ARTÉRIAS</i>	2. ARTERIES
3. <i>ARTERIAL</i>	3. ARTERIAL
4. <i>CARDÍACA</i>	4. CARDIAC
5. <i>CARDIOVASCULAR</i>	5. CARDIOVASCULAR
6. <i>CORONÁRIA</i>	6. CORONARY
7. <i>DOENÇA</i>	7. DISEASE
8. <i>DOENÇAS</i>	8. DISEASES
9. <i>ESQUERDO</i>	9. LEFT
10. <i>ESTUDO</i>	10. STUDY
11. <i>HIPERTENSÃO</i>	11. HYPERTENSION
12. <i>INSUFICIÊNCIA</i>	12. INSUFFICIENCY
13. <i>MIOCÁRDIO</i>	13. MYOCARDIUM/ MYOCARDIAL
14. <i>MITRAL</i>	14. MITRAL
15. <i>MORTALIDADE</i>	15. MORTALITY
16. <i>PACIENTES</i>	16. PATIENTS

<i>17. PULMONAR</i>	17. PULMONARY
<i>18. SISTÊMICA</i>	18. SYSTEMIC
<i>19. SISTÓLICA</i>	19. SYSTOLIC
<i>20. TROMBO</i>	20. THROMBUS
<i>21. VASCULAR</i>	21. VASCULAR
<i>22. VENTRÍCULO</i>	22. VENTRICLE

As it could be seen, the two lists of keywords originated a list of twenty-two simple terms, along with their equivalents in the English, which were shown to a specialist in Cardiology.

Although the abbreviation “cols” and “et al” were used frequently by the authors and revealed a characteristic of the scientific article on the use of abbreviations, they were not included in the glossary of terms because they did not have a noun or adjective base.

After observing keywords, concordance lines and word clusters we compiled a list of simple and complex terms such as the following one:

Table 2. List of bilingual simple and complex Cardiology terms

<i>ARTERIAL</i>	<b>ARTERIAL</b>
<i>DOENÇA ARTERIAL CORONARIANA</i>	CORONARY ARTERY DISEASE
<i>HIPERTENSÃO ARTERIAL</i>	ARTERIAL HYPERTENSION
<i>HIPERTENSÃO ARTERIAL SISTÊMICA</i>	SYSTEMIC (ARTERIAL) HYPERTENSION
<i>HIPERTENSÃO ARTERIAL SUSTENTADA</i>	SUSTAINED ARTERIAL HYPERTENSION
<i>HIPERTENSÃO ARTERIAL PULMONAR</i>	PULMONARY ARTERIAL HYPERTENSION
<i>HIPERTENSÃO ARTERIAL TRANSITÓRIA</i>	TRANSIENT ARTERIAL HYPERTENSION
<i>PRESSÃO ARTERIAL</i>	BLOOD PRESSURE/ARTERIAL PRESSURE

<i>PRESSÃO ARTERIAL DIASTÓLICA</i>	DIASTOLIC BLOOD PRESSURE
<i>PRESSÃO ARTERIAL PULMONAR</i>	PULMONARY ARTERIAL PRESSURE
<i>PRESSÃO SISTÓLICA ARTERIAL PULMONAR</i>	PULMONARY ARTERIAL SYSTOLIC PRESSURE
<i>PROTOCOLO HIPERTENSÃO ARTERIAL TRANSITÓRIA</i>	TRANSIENT ARTERIAL HYPERTENSION PROTOCOL
<b><i>HIPERTENSÃO</i></b>	<b>HYPERTENSION</b>
<i>HIPERTENSÃO</i>	HYPERTENSION
<i>HIPERTENSÃO ARTERIAL</i>	ARTERIAL HYPERTENSION
<i>HIPERTENSÃO ARTERIAL SISTÊMICA</i>	SYSTEMIC HYPERTENSION
<i>HIPERTENSÃO ARTERIAL SUSTENTADA</i>	SUSTAINED ARTERIAL HYPERTENSION
<i>HIPERTENSÃO ARTERIAL PULMONAR</i>	PULMONARY ARTERIAL HYPERTENSION
<i>HIPERTENSÃO ARTERIAL TRANSITÓRIA</i>	TRANSIENT ARTERIAL HYPERTENSION

As it can be seen, the use of terms formed by the syntagmatic composition process is very frequent. Because of that, the discussion about the terms with an specialist in the area was crucial to check the extension of each term.

In the next section we will present the procedures used in order to select the terms of Cardiac Surgery.

### **3.2. Selection of Cardiac Surgery terms in the TBMC**

A similar process was used to generate lists of keywords from the parallel subcorpus of Cardiac Surgery. By having the corpus *Lácio-Ref* as a reference, the top ranking of keywords was the following:

<b>Ranking</b>	<b>Keywords in Portuguese</b>	<b>Frequency</b>	<b>%</b>	<b>Keyness</b>
1 <sup>st</sup>	PACIENTES	303	0,62	1.433,50
2 <sup>nd</sup>	OR	193	0,4	1.040,60
3 <sup>rd</sup>	VD	169	0,35	922
4 <sup>th</sup>	OPERATÓRIO	154	0,32	840,1
5 <sup>th</sup>	CIRURGIA	145	0,3	729,3
6 <sup>th</sup>	AORTA	112	0,23	610,9
7 <sup>th</sup>	ARTÉRIA	109	0,22	594,5
8 <sup>th</sup>	DOR	114	0,23	563,8
9 <sup>th</sup>	PULMONAR	98	0,2	534,5
10 <sup>th</sup>	FORAM	283	0,58	462
11 <sup>th</sup>	ARTERIAL	78	0,16	425,4
12 <sup>th</sup>	CALCIFICAÇÃO	77	0,16	419,9
13 <sup>th</sup>	GRUPO	169	0,35	407,3
14 <sup>th</sup>	APROTEININA	74	0,15	403,6
15 <sup>th</sup>	ET	120	0,25	387,9
16 <sup>th</sup>	AL	121	0,25	387,4
17 <sup>th</sup>	TP	71	0,15	387,2
18 <sup>th</sup>	BYPASS	70	0,14	381,8
19 <sup>th</sup>	ENXERTO	67	0,14	365,4
20 <sup>th</sup>	PÓS	157	0,32	356

The same procedure was performed using the corpus of Cardiac Surgery and the corpus of the *Folha de Sao Paulo 97* as a reference. The list of keywords generated is the following one:

<b>Ranking</b>	<b>Keywords in Portuguese</b>	<b>Frequency</b>	<b>%</b>	<b>Keyness</b>
1 <sup>st</sup>	VD	169	0,35	2.262,90
2 <sup>nd</sup>	PACIENTES	303	0,62	2.212,30
3 <sup>rd</sup>	OR	193	0,4	2.164,20
4 <sup>th</sup>	OPERATÓRIO	154	0,32	1.942,30
5 <sup>th</sup>	AORTA	112	0,23	1.395,10
6 <sup>th</sup>	ARTÉRIA	109	0,22	1.259,20
7 <sup>th</sup>	ET	120	0,25	1.077,70
8 <sup>th</sup>	PULMONAR	98	0,2	1.006,50
9 <sup>th</sup>	APROTEININA	74	0,15	990,7
10 <sup>th</sup>	CALCIFICAÇÃO	77	0,16	982,7

11 <sup>th</sup>	BYPASS	70	0,14	926,7
12 <sup>th</sup>	PÓS	157	0,32	919,6
13 <sup>th</sup>	TP	71	0,15	919,3
14 <sup>th</sup>	CIRURGIA	145	0,3	908,8
15 <sup>th</sup>	ARTERIAL	78	0,16	804
16 <sup>th</sup>	ENXERTO	67	0,14	789,4
17 <sup>th</sup>	INTERMITENTE	65	0,13	786,6
18 <sup>th</sup>	TORÁCICA	65	0,13	783,5
19 <sup>th</sup>	DOR	114	0,23	709,7
20 <sup>th</sup>	CORONARY	52	0,11	696,2

After comparing the two lists shown above, we observed that seventeen keywords or abbreviations appeared in the two lists: *pacientes* (patients), *or* (or), *vd* (rv), *operatório* (operative), *cirurgia* (surgery), *aorta* (aorta), *artéria* (artery), *dor* (pain), *pulmonar* (lung), *arterial* (arterial), *calcificação* (calcification), *aprotinina* (aprotinin), *et* (et), *tp* (pt), *bypass* (bypass), *enxerto* (graft) e *pós* (post).

The words or abbreviations that appear in one list and were not present in the other one were : *grupo* (group), *foram* (were), *al* (al), *torácica* (chest), *intermitente* (intermittent) and *coronária* (coronary). As in the previous section, we generated two lists with two different corpora so that we could only select the most representative keywords.

After checking the concordance lines we observed that the abbreviation *tp* was used to refer to the term *tronco pulmonar* (pulmonary trunk) which had been found in English as “pulmonary artery trunk”. The same abbreviation could also be referring to the term *tempo de protrombina* (prothrombin time) which, according to the specialist, refers to a coagulation study. This last term, however, was not present in the analysed subcorpora analyzed, therefore, the term included in the glossary for the abbreviation “tp” was “pulmonary artery”.

The abbreviations *et* and *al* also presented a high frequency and representativeness in the subcorpus of Cardiac Surgery, as all the articles contained references to other authors. This,

again, shows one characteristic of the academic genre, however, as likewise the data from the sub-corpus of Cardiology, these abbreviations were not included in the glossary since they were not typical terms this sub-area.

Two English words that called our attention at first, were “bypass” and “coronary” , since they were in a list of Portuguese words. For this reason, we decided to verify the concordance lines with these two words. What we found out was that they had been used as search words by *Medline*. Besides these two words, the word “or” was also present in the search expression of *Medline* therefore, it would not be part of the article itself and it was not included in the glossary. The verb form *foram* (were) is also representative in the sub-corpus of Cardiac Surgery, however, the glossary which was being compiled should contain nouns and adjectives, so we decided not to include this verb form, but we considered studying the passive voice in these articles.

Similarly to the sub-corpus of Cardiology, general terms of medicine were also present in the sub-corpus of Cardiac Surgery. Terms such as *pacientes* (patients) and *dor* (pain) were kept in the glossary as they were also representative of the Cardiac Surgery domain. After a discussion with a specialist from the area he stated that “pain” was a common word which is frequently used to refer to a common clinical sign of surgical specialties which involve trauma. In the end of the process of selecting keywords we came up with a list of fifteen words. As the initial purpose was to select twenty keywords to serve as guidelines for the compilation of the glossary, also, as the list of Cardiology already had twenty-two words, we included the following words from the ranking in order to have a more complete list, therefore we also included the keywords *cardíaca* (cardiac), *sistólica* (sistólica), *hipotermia* (hipotermia), *aminoglisídeo* (aminoglycoside) and the abbreviation *ve* (left ventricle).

After verifying the keywords in Portuguese we followed the same process in order to generate keyword lists in English. For doing so, we used the reference corpora *BNC* and *TEC* .

Both lists of the top keywords had the following fifteen words in common : “patients”, “rv”, “surgery”, “postoperative”, “artery”, “aortic”, “pulmonary”, “coronary”, “surgical”, “arterial”, “et”, “calcification”, “study”, “pt” e “aprotinin”.

The next words which were among the top thirty words were: “pain”, “grafts”, “group”. The word “bypass” was present in the list generated with the *BNC*, but was not part of the list generated with *TEC*. Once again, we could confirm that, although not in the same ranking position the same words were representative in the sub-corpus of Cardiac Surgery in English. After comparing the keywords in Portuguese and in English the selected terms were the following :

Table 3. Simple terms of Cardiac Surgery present in the parallel subcorpus

1. AMINOGLICOSÍDEO	1. AMINOGLYCOSIDE
2. AORTA	2. AORTA/AORTIC/AO
3. APROTININA	3. APROTININ
4. ARTÉRIA	4. ARTERY
5. ARTERIAL	5. ARTERIAL
6. CALCIFICAÇÃO	6. CALCIFICATION
7. CARDÍACA	7. CARDIAC
8. CIRURGIA	8. SURGERY
9. DOR	9. PAIN
10. ENXERTO	10. GRAFT(ING)
11. HIPOTERMIA	11. HYPOTHERMIA
12. INTERMITENTE	12. INTERMITTENT
13. OPERATÓRIO	13. (PERI/POST/PRE/INTRA) OPERATIVE
14. PACIENTES	14. PATIENTS
15. PULMONAR	15. PULMONARY
16. SISTÓLICA	16. SISTOLIC
17. TORÁCICA	17. THORACIC
18. TP (TRONCO PULMONAR)	18. PT (PULMONARY TRUNK)
19. VD (VENTRÍCULO DIREITO)	19. RV (RIGHT VENTRICLE)
20. VE (VENTRÍCULO ESQUERDO)	20. LV (LEFT VENTRICLE)

The keywords selected from the sub-corpus in Portuguese were compared to the keywords extracted from the English sub-corpus. This comparison also showed that all the keywords had equivalents, sometimes, with more than one option in both languages. An example of this fact is the simple term *aorta* which had, as equivalents, three terms in English, “aortic”, as an adjective, “aorta” as a noun and the abbreviation “*ao*”.

The keyword “aminoglycoside” refers to an antibiotic of wide clinical use and is part of the representative list from the sub-corpus analyzed as one of the articles discusses the effectiveness of antibiotic treatment in hospital.

Among the limitations of the software used in this study the analyst must be aware of problems he might have with hyphenated words which may be found separately in the lists, such as *pós-operatório* in Portuguese whose equivalent in English is “postoperative”. In order to avoid problems like that we also observed the aligned original and translated texts so that we could spot words with prefixes and hyphens.

The term *aprotinina* (aprotinin) was only found as a simple term in the original and translated texts. After checking the concordance lines and discussing the keyword selection with the Cardiac Surgery specialist we found out that this term is commonly used during surgeries as it is a kind of drug used to improve blood coagulation, therefore, it should be included in the glossary of the sub-area .

After observing the concordance lines, word clusters and collocates on a list of we organized a list of simple and complex terms as the example below:

Table 4. List of bilingual simple and complex Cardiology terms

<i>AORTA</i>	<b>AORTA</b>
<i>AORTA ABDOMINAL</i>	ABDOMINAL AORTA



<i>AORTA ASCENDENTE</i>	ASCENDING AORTA
<i>AORTA TORÁCICA</i>	THORACIC AORTA
<i>COARCTAÇÃO DA AORTA</i>	AORTIC COARCTATION
<i>DISSECÇÃO DA AORTA</i>	AORTIC DISSECTION
<i>VD (VENTRÍCULO DIREITO)</i>	<b>RD (RIGHT VENTRICLE)</b>
<i>DILATAÇÃO DO VD</i>	DILATATION OF THE RD
<i>DISFUNÇÃO DO VD</i>	DYSFUNCTION OF THE RD
<i>HIPERTROFIA DO VD</i>	RV HYPERTROPHY
<i>INFUNDÍBULO DO VD</i>	RV INFUNDIBULUM
<i>SOBRECARGA SISTÓLICA DO VD</i>	RV SYSTOLIC OVERLOAD
<i>VOLUME DIASTÓLICO FINAL DO VD</i>	RV FINAL DIASTOLIC VOLUME

After the selection of simple and complex terms from the lists of the parallel corpus (TBMC) we started the selection of simple and complex terms from the comparable corpora (CMC) as discussed in the next section.

### 3.3. Selection of terms of Cardiology in the CMC

Similarly to the procedures previously performed for the observation of the TBMC, we generated lists of keywords from the non-translated articles in Portuguese using as reference corpora in Portuguese the *Lácio-Ref* and *Folha de Sao Paulo in 1997* for a double verification. The words on both lists were : *pacientes* (patients), *atrial* (atrial), *pulmonar* (pulmonary), *ventricular* (ventricular), *mitral* (mitral), *esquerdo* (left), *arterial* (arterial/ blood), *artéria* (artery), *insuficiência* (failure), *ventrículo* (ventricle), *doença* (disease), *estudo* (study), *vascular* (vascular), *cardíaca* (cardiac/heart), *trombo* (thrombus), *mortalidade* (mortality), *hipertensão* (hypertension) and *coronária* (coronary).

As noted, there were eighteen keywords extracted from the top one hundred words originated by the non-translated papers in Portuguese (CMC). These words would indicate a high degree of representativeness within the sub-area of Cardiology used by Brazilian authors.

The same procedure was performed using the non-translated papers of Cardiology in English, only, this time, using as reference corpus the frequency list from the *BNC*. In this phase we decided not to use the *TEC* list since the CMC is composed of non-translated texts, however we think it would be interesting to conduct a future study contrasting *TEC* to non-translated and translated texts.

The list of keywords in English from the CMC showed the following equivalents for the words that were used being used as guidelines for the selection of terms for the glossary: “patients”, “coronary”, “ventricular”, “cardiac”, “artery”, “study”, “disease”, “systolic”, “mortality”, “atrial” e “hypertension”.

After this procedure, we verified that there were eleven equivalent words to those ones that had been selected from the non-translated articles in Portuguese. This also points to a representation of words in the Cardiology sub-area which are commonly used by authors who write originally in English.

Once again, we must state that only the top one hundred keywords were observed in each list, however, the remaining words could be included to the total list of keywords from the comparable corpora. This was not done due to time and team constraint.

Having selected the keywords in Portuguese and in English from the CMC, which at this stage were considered simple terms, we started the process of analysing the clusters, collocates and concordance lines to be inserted into the Cardiology glossary. This glossary should contain the terms withing their context of use.

As an illustration, we present below some examples of words taken from the CMC :

Table 5. Simple and complex terms in Portuguese and in English extracted from the CMC

ARTÉRIA	ARTERY	
<i>Artéria</i>	Artery	<p><i>TOP:</i> O estresse cardiovascular causa isquemia miocárdica em regiões supridas por uma <b>artéria</b> com grau significativo de estenose, e este fenômeno é manifestado por alteração transitória da contração segmentar.</p> <p><i>TOI:</i> Using a femoral arterial approach, an angioplasty balloon was advanced to the mid-left anterior descending <b>artery</b>.</p>
<i>Artéria coronária</i>	Coronary artery	<p><i>TOP:</i> O ultra-som intracoronariano é também uma modalidade invasiva, que tem por base a aquisição de imagens tomográficas, a partir do interior da <b>artéria coronária</b>, utilizando-se um cateter que tem incorporado na sua extremidade um único transdutor.</p> <p><i>TOI:</i> After adjustment for established risk factors, the odds ratio for <b>coronary artery</b> disease was 1.45 (95% CI, 1.25 to 1.68) for participants in the highest third of the group with respect to those in the bottom third.</p>
<i>Artéria coronária esquerda</i>	Left coronary artery	<p><i>TOP:</i> A operação foi feita sem o emprego de circulação extracorpórea (4/7/2003) e realizados anastomose da artéria torácica interna esquerda com o ramo interventricular anterior da <b>artéria coronária esquerda</b> e enxerto de veia safena autóloga para o ramo diagonal da artéria coronária esquerda.</p> <p><i>TOI:</i> Anomalous origin of the <b>left coronary artery</b> from the right sinus of Valsalva.</p>
<i>Artéria coronária direita</i>	Right coronary artery	<p><i>TOP:</i> O grau de comprometimento arterial foi analisado pela angiografia convencional e os pacientes divididos em portadores de lesões de tronco de coronária esquerda se apresentassem lesões &gt; 50% neste local ou portadores de lesões uniaarteriais, biarteriais, triarteriais de acordo com o comprometimento, com lesões &gt; ou iguais a 70%, de um, dois ou três dos seguintes territórios: artéria descendente anterior e/ou ramos diagonais; artéria circunflexa e/ou ramos marginais; <b>artéria coronária direita</b>.</p> <p><i>TOI:</i> A resting myocardial perfusion acquisition is then performed using 0.05 mmol/kg body weight of dimeglumine gadopentetate. Next, a coronary scout image is acquired, followed by a high-resolution angiogram aligned along the course of the <b>right coronary artery (RCA)</b>.</p>
<i>Artéria pulmonar</i>	Pulmonary artery	<p><i>TOP:</i> A principal vantagem da técnica era que a <b>artéria pulmonar</b> direita, não sendo separada da confluência, permitia que o volume de sangue da veia cava superior fosse dividido para ambos os pulmões.</p> <p><i>TOI:</i> Spiral CT image confirming enlarged pulmonary artery and branches with thromboembolic material. Ao = Aorta, Pa =Pulmonary Artery, Rbpa = Right branch of <b>pulmonary artery</b>, Lbpa = Left branch of pulmonary artery.</p>

A result to be discussed is that there was a variety of complex terms found in CMC which were not included in the glossary for failing to have an equivalent in the other language. As an example, we highlight the terms from the keyword *artérias* (arteries), whose complex terms used in Portuguese were: *artérias naturais* (natural arteries), *artérias sistêmicas* (systemic arteries), *artérias coronárias humanas in vitro* (human coronary arteries *in vitro*) and *artérias pulmonares lobares* (lobar pulmonary arteries). The same was observed with the complex terms from the non-translated articles in English as for the terms from the keyword “arteries”: “native coronary arteries”, “totally occluded coronary arteries”, and “infarct-related arteries”.

Another fact to be mentioned is the occurrence of complex terms in the Portuguese and in English from the non-translated articles in the CMC which were not exactly the same as in the TBMC, but whose context indicated that these were equivalent terms referring to the same medical procedures. One example was the term *necrose cardíaca* (cardiac necrosis), from the original articles in the TBMC. According to the specialist in Cardiology, this would be a generic term whose specific terms were found in the non-translated articles from the CMC, such as, *necrose miocárdica* and “myocardial necrosis”. A similar case may be found with terms which have the same referent, but may have one or more element in its equivalent or may have one omitted element in its equivalent. One example of this phenomenon is the term *artéria coronária direita* (right coronary artery), found on translated and non-translated articles in Portuguese, and had as an equivalent in English “right coronary artery” or just “coronary artery” being used in the same contexts.

After the selection of terms in the sub-area of Cardiology, the glossary was prepared to be inserted in the *Wordfast*, a translation memory tool. This will be discussed in sections 4 and 5 of this paper.

### 3.4 Selection of terms of Cardiac Surgery in the CMC

The same procedures were previously used for the verification of keywords, simple and complex terms in the sub-area of Cardiac Surgery from the TBMC and the CMC. The twenty keywords in Portuguese, which consisted of nouns, adjectives, conjunctions and abbreviations, found in both corpora were: *pacientes* (patients), *artéria* (artery), *cirurgia* (cirurgia), *aorta* (aorta), *foram* (were), *or* (or), *mortality* (mortalidade), *operatório* (operative), *grupo* (group), *arterial* (arterial), *cardíaca* (cardiac), *et* (et), *al* (al), *pulmonar* (pulmonary), *graft* (enxerto), *cirúrgica* (surgical), *calcificação* (calcification), *bypass* (bypass), *pós* (post) and *torácica* (thoracic).

In English, the keywords which were part of the list were: “patients”, “surgery”, “cardiac”, “lung”, “pulmonary”, “artery”, “aortic”, “postoperative”, “bypass”, “mortality”, “preoperative”, “systolic”, “were”, “heart”, “coronary”, “aprotinin” and “et”.

The words “lung” e “heart”, which would be the *prima face* equivalent of *coração* and *pulmão* were considered key in this list because, in English, they could function as an adjective, for example, “lung transplantation” and “lung cancer”. As for the terms “heart chamber” and “congestive heart failure”, in English, the equivalent terms found in Portuguese were *câmara cardíaca* (cardiac chamber) and *insuficiência cardíaca congestiva* (congestive cardiac failure).

It is important to highlight that, although the keywords discussed appeared on the lists from the CMC in Portuguese and in English, , words and abbreviations that had been excluded from the lists of TBMC of Cardiology and Cardiac Surgery were not included in glossaries for the same reasons discussed earlier.

Another interesting phenomenon observed was the change of syntact roles of certain terms such as the noun *intraoperatório* (intraoperative), in Portuguese, referring to the

“intraoperative period” which is only used as an adjective in English, as in “intraoperative morbidity”, “intraoperative complications” or “intraoperative arrhythmia”. This may indicate that differently from the English usage, Brazilian authors omit the noun “period” from the complex term and have used the adjective as a noun, not informally, but in formal language papers. Similarly, we found that the professional translator of Cardiac Surgery is aware of the use of this term as a noun since the lines of concordance from the TBMC also contain sentences with the adjective as a noun. Likewise, terms in Portuguese such as *pós-operatório* (postoperative) and *perioperatório* (perioperative) are used as nouns and also refer to the “period” in which the patient is. On the other hand, these terms were only used by the authors of non-translated English papers as adjectives.

A term that stood out for its frequency in Portuguese was the term *canal arterial* (arterial canal) in the TBMC which was also found in the CMC. Its equivalent used by the translator of Cardiac Surgery was “patent ductus arteriosus”. According to the specialist in Cardiac Surgery, by adding the word “patent” was the term would be referring to the *canal arterial aberto* (open arterial canal), but could also be used as an equivalent to *canal arterial* (arterial canal) as it was used by the Brazilian authors. This term was not found among the keywords in English from the CMC, which may indicate two things, either authors who write originally in English refer to this term in a different way, or the CMC did not contain any paper referring to this term.

Some terms used with abbreviations in Portuguese were not found with the same abbreviations in English. One example was the term *hipertrofia do ventrículo VD* (ventricular hypertrophy RV) whose equivalent used by the translator was “RV hypertrophy”. This same term was not found in the list of CMC in the same way, but among other compound terms such as “reverse eccentric hypertrophy”.

Having made the comparison between the keywords in English and in Portuguese in the sub-area of Cardiac Surgery, we, once again, observed the concordance lines, word clusters and collocatesso that we could select the simple and complex terms within their cotext. After that, we enlarged the cotext so they would become contexts of use which could be part of the glossaries such as the following illustration :

Table 6. Simple and complex terms in Portuguese and in English extracted from the CMC

<i>AORTA</i>	AORTA	
<i>Aorta</i>	Aorta	<p><i>TOP</i>: Coarctação da <i>aorta</i> ocorreu em cinco pacientes. A comparação das curvas de sobrevida (Kaplan-Meier) mostrou melhor resultado com o Tubo VD-AP (p=0,003).</p> <p><i>TOE</i>: As previously described, each dog underwent CABG with a reversed SVG from the <i>aorta</i> to the left anterior descending artery (LAD).</p>
<i>ARTERIAL</i>	ARTERIAL	
<i>parede arterial</i>	arterial wall	<p><i>TOP</i>: No caso da reoperação por degeneração tecidual, havia significativa calcificação da <i>parede arterial</i> do homoenxerto, obrigando sua ressecção completa e implante de novo homoenxerto valvar.</p> <p><i>TOE</i>: Chronic inflammation results in endothelial dysfunction and facilitates the interactions between modified lipoproteins, monocyte-derived macrophages, T cells, and the normal cellular elements of the <i>arterial wall</i>.</p>

The next step was to compile the glossary in a way that the simple and complex terms could be read, within their contexts, by the translator who used glossaries as memories. We will discuss how this was done in the next section.

#### 4. Preparation and definition of glossary in this study

After selecting the simple and complex terms and their context of use, the glossaries were prepared in *Excel* tables to be transformed into texts to be used in translation tools as in the following figure:

Table 7. Simple and complex terms in Portuguese and in English in the glossary

<p><i>Artéria Coronária Esquerda</i></p>	<p>Left Coronary Artery</p>	<p><i>OT</i>: No tipo balanceado, a artéria coronária direita irrigava somente o ventrículo direito e parte posterior do septo interventricular, não fornecendo ramos significantes para o ventrículo esquerdo, enquanto este era irrigado pela <i>artéria coronária esquerda</i>.</p> <p><i>TT</i>: In the balanced type, the right coronary artery irrigated only the right ventricle and the posterior part of the interventricular septum, and did not provide significant branches for the left ventricle, which was irrigated by the <i>left coronary artery</i>.</p> <p><i>TOP</i>: A operação foi feita sem o emprego de circulação extracorpórea (4/7/2003) e realizados anastomose da artéria torácica interna esquerda com o ramo interventricular anterior da <i>artéria coronária esquerda</i> e enxerto de veia safena autóloga para o ramo diagonal da artéria coronária esquerda.</p> <p><i>TOE</i>: Anomalous origin of the <i>left coronary artery</i> from the right sinus of Valsalva.</p>
--	-----------------------------	--

Since the aim of compiling these glossaries was to submit lists of simple and complex terms in two medical areas and not define those terms, Barros's (2004:144) definition for glossaries was the one used in this study: "its main feature is not to present definitions, but simply a list of lexical units and terminology accompanied by their equivalents in other languages".

It is believed that most glossaries prepared by and for translators follow the same format since they facilitate the search for terms and equivalents during the translation process.

Regarding the type and macrostructure organization, the glossaries of this study are defined as bilingual and unidirectional: Portuguese → English. They also bring the terms containing keywords generated from the TBMC and the CMC, in Portuguese and in English, in



two medical subareas, that is, Cardiology and Cardiac Surgery. The glossaries also contain the term context of use as it was suggested by Castanho (2004).

The table displays the microstructure in the glossaries:

Table 8. Microstructure of glossaries prepared in this study

KEYWORD IN PORTUGUESE	EQUIVALENT KEYWORD IN ENGLISH	
SIMPLE OR COMPLEX TERM IN PORTUGUESE	SIMPLE OR COMPLEX TERM IN ENGLISH	Term inserted in the context of use from the TBMC in Portuguese + Term inserted in the context of use from the TBMC in English + Term inserted in the context of use from the CMC in Portuguese + Term inserted in the context of use from the CMC in English

In the next section, discuss the insertion of the compiled glossaries in the translation memory tool used in this study.

### 5. The insertion of glossaries in the tool *Wordfast*

After completing the tables on *Excel*, they were converted and saved as plain text (txt) to be read by the software *Wordfast* as glossaries. They were saved in a folder called “Glossaries”, within the program, and named according to their direction and translational medical subarea, in this case, Brazilian-Portuguese (PT-BR) to American-English (EN-US). Once the glossaries had been activated as memories, the translator should, first, choose the glossary in the software by clicking *f > Terminology > Glossaries*.

This way, during the translation process, the medical term, that is part of the glossary, will be highlighted on the screen as below:

angioplastia. Após ressuspensão em 20 ml de salina albuminada (5%), cerca de 2,4 X 10<sup>8</sup> células foram injetadas no sistema coronariano direito e esquerdo, através de cateter de angioplastia.

Foram injetados 10ml da suspensão na artéria coronária descendente anterior, lentamente em 10min, 5 ml circunflexa e 5 ml na artéria coronária direita.

Figure 1. Screen showing the text to be translated with the words *artéria coronária* and *artéria coronária direita* highlighted by the program *Wordfast*

Then, the software offers the option found in the glossary, as seen in the following illustration:

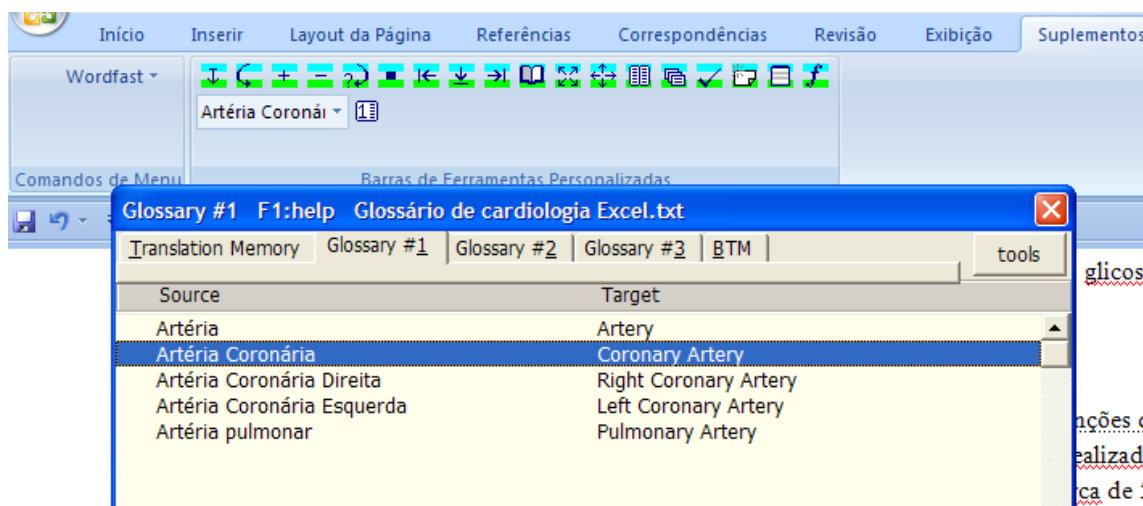


Figure 2. Term *artéria coronária* in the glossary of the *Wordfast*

As it can be seen in the next figure, by using the glossaries proposed in this study, the translator can also see how the terms were previously used by the authors of texts originally written either in Portuguese or in English. The information from the third column of the glossary is that in the “Comment” window in the next figure:

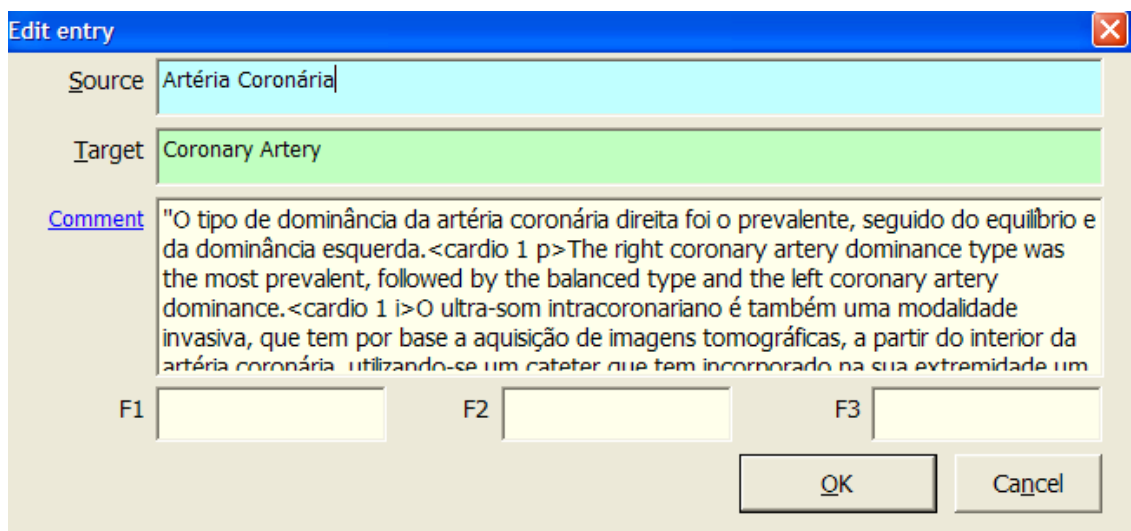


Figure 3. Screen with *Wordfast* showing the term in its context of use

Part of the glossaries compiled in this study was made available to be tested by three students who wrote about their experience (Felice, 2008; Garcia, 2008; Santos, 2008). Also, at a professional office, there was a translator who tested them. After two months, the students and the professional translator considered the glossaries useful and also said that instead of using two separate glossaries, they preferred to join them into a single file and name it as “Cardiology and Cardiovascular Surgery Glossary” since the two sub-reas have medical terms in common.

In addition, these users reported they have added other simple and common terms so as not to have to type them again whenever they appeared in a translation as, with only one command on the computer keyboard, the tool enters the word into the text that is being translated. They added that due to the short deadline for delivering the translations, the presence of the contexts of use for the confirmation of terms has been very helpful.

Despite the advantages offered by translation memories, Azzan (2004) also highlights the importance of managing the memories and glossaries in these programs so that the translator does not “get lost” in his/her work. The author adds that good glossaries are crucial and valuable.

It is noteworthy that this procedure has started a project for compiling glossaries of other technical areas to be included in translation memory tools which will be carried out at the university and will be used by other translation offices.

## **6. Final Comments**

The main objective in this study was to unite theory and practice through a methodology that makes it possible to do so. The glossaries compiled on a scientific basis, although not having the traditional printed format, could be tested by students and professional translators who stated their opinion about them and could also suggest some changes.

Regarding the research of terms itself we may state that computerized corpora and corpus-based approach play important roles in a study such as the one carried out by us since they empirically guide the analyst into the most representative terms of the area. Needless to say that following the basic criteria during the compilation of a study corpus is essential for the success of this kind of research.

We should also add that this methodology is useful to raise learners' awareness in understanding that there might be more than one option to be used in their translations and that context will be important in determining which choice should be made.

This research is now being developed together with the Chemistry Institute from the São Paulo State University as three students have compiled a parallel and a comparable corpus of Chemistry and Biotechnology. They have also started using the software *WordSmith Tools* in order to observe the keywords of these areas and are testing the use of glossaries within other translation memory tools such as *Trados* and *Google Toolkit*.

Finally, we believe that, whether in teaching or in professional context, glossaries compiled with the help of *WordSmith Tools* and, afterwards, inserted into translation memory

tools such as *Wordfast* may put students and translators in contact with a bilingual terminology in a more dynamic and useful way.

## Acknowledgments

I am grateful to Professors Diva Cardoso de Camargo (UNESP/IBILCE) and Maeve Olohan (MANCHESTER UNIVERSITY) for the valuable suggestions in this study. I am also grateful to CAPES and *Fundunesp* for the financial support.

## References

Azzan, F. (2004). “Gerenciamento de memórias de tradução e de glossários”. *Cadernos de Tradução*, 2004, 14, 2, 87- 119.

Barros, L. A. (2004). *Curso Básico de Terminologia*. São Paulo: USP.

Baker, M. (1993). “Corpus linguistics and translation studies: implications and applications”. In Baker, M.; Francis, G.; Tognini-Bonelli, E. (ed.). *Text and Technology: in Honour of John Sinclair*. Amsterdam: John Benjamins, 233-250.

Berber Sardinha, A. P. (2003). “Uso de corpora na formação de tradutores”. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, São Paulo, 9, 323-367.

Berber Sardinha, T. (2004). *Lingüística de Corpus*. Barueri, SP: Manole.

Biber, D.; Conrad, S.; Reppen, R. (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: CUP.

British National Corpus (BNC). Available at <http://www.natcorp.ox.ac.uk/> (accessed: 06 April 2006)

Camargo, D. C. de. (2007). *Metodologia de pesquisa em tradução e lingüística de corpus*. São Paulo: Cultura Acadêmica; São José do Rio Preto, SP: Laboratório Editorial do IBILCE, UNESP.

Castanho, R. (2004). *Proposta para a elaboração de um glossário de colocações na área médica – subárea hipertensão arterial*. Dissertation (Master’s in Linguistics and English Literature)-Departamento de Letras Modernas da Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.

Champollion, Y. (1999) *Wordfast*.

Felice, A. A. (2008). *A utilização de glossários de medicina no programa de memórias de tradução Wordfast*. Final research report. Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP.

Garcia, R. C. P. (2008). *Tradução e arte: comercial, graffiti e games*. Final research report. Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP.

Lacio-Ref. Available at <http://www.nilc.icmc.usp.br/lacioweb/corpora.htm> (accessed: 09 July 2008)

Laviosa, S. (1997). "How Comparable Can 'Comparable Corpora Be?'. *Target*, Amsterdam, 9, 2, 289-319.

Laviosa, S. (2002). *Corpus-based translation studies: theory, findings, applications*. Amsterdam, Rodopi.

Carter, R.; McCarthy, M. (2006). *Cambridge Grammar of English : a Comprehensive Guide. Spoken and Written English Grammar and Usage*. Cambridge : CUP.

Nogueira, D.; Nogueira, V. M. C. (2004). "Por que usar programas de apoio à tradução?" *Cadernos de Tradução*, 2, p. 17-35.

Olohan, M. (2004). *Introducing corpora in translation studies*. New York: Routledge.

Olohan, M. (2002). "Leave it out! Using a comparable corpus to investigate aspects of explicitation in translation". *Cadernos de Tradução: Corpora e Tradução*. Available at <http://www.cadernos.ufsc.br/> (accessed : 15 August 2007)

Paiva, P. T. P.; Camargo, D. C. (2007). "Estudos da tradução baseados em corpus e lingüística de corpus: levantamento de termos médicos para a elaboração de um glossário bilíngüe". *Estudos Lingüísticos*, São Paulo), 2, 428-436.

Paiva, P. T. P.; Camargo, D. C.; Xatara, C. M. (2008). "Uma reflexão sobre a elaboração de um léxico bilíngüe preliminar na subárea de cardiologia a partir do uso de termos encontrados em um corpus paralelo e em dois corpora comparáveis". *DELTA*. Documentação de Estudos em Lingüística Teórica e Aplicada, São Paulo, 24, 1, 1-22.

Ribeiro, G. C. B. (2004). "Tradução técnica, terminologia e lingüística de corpus: ferramenta WordSmith Tools". *Cadernos de Tradução*, 2, 159 - 174.

Santos, S. F. (2008). *Memórias de tradução como auxílio ao tradutor de textos especializados*. 2008. Final Essay. (Graduated in Letters and Translation) - União das Faculdades dos Grandes Lagos.

Scott, M. *WordSmith Tools*: version 3.0. (1999) Oxford: Oxford University Press.

Sinclair, J. (1991) *Corpus, concordance and collocation*. Oxford: Oxford University Press.

Sinclair, J. "Developing Linguistic Corpora: A Guide to Good Practice Corpus and Text – Basic Principles"(2004).

Available at: <http://ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm> (accessed: 10 January 2009)

Stubbs, M. (1995). "Corpus evidence for norms of lexical collocation". In: cook, G; Seidlhofer, B. (ed). *Principle and practice in Applied Linguistics: Studies in Honour of H Widdowson*. Oxford: Oxford University Press.

Tognini-Bonelli, E. (2001) *Corpus linguistics at work*. Amsterdam :Jonh Benjamins.