Translator-oriented, corpus-driven technical glossaries: the case of cooking terms

Stella E.O. Tagnin[1] and Elisa Duarte Teixeira[1]

Abstract

Although there are many English–Portuguese technical glossaries on the market, very few are designed to meet the specific needs of translators, whose main task is the production of an idiomatic text either in the mother tongue or in the foreign language. For that purpose, a mere list of monolexical terms (usually based on previous compilations or abstract conceptual schemata, and their equivalents) will not suffice. Most importantly, the producer of a text needs to know how a word is used in context, and this can be inferred from the observation of authentic examples. Corpus Linguistics has proved to be an invaluable tool in retrieving technical terms and phraseologies from corpora. In this paper, we employ a corpus-based methodology or, more precisely, the 'corpus-driven approach', to compile a bilingual, monodirectional English–Portuguese glossary of cooking terms (Teixeira and Tagnin, 2008), covering the various stages of the project, with an emphasis on the identification of candidate terms, and their subsequent validation through the generation of frequency and keyword lists using the lexical analysis software, WordSmith Tools.

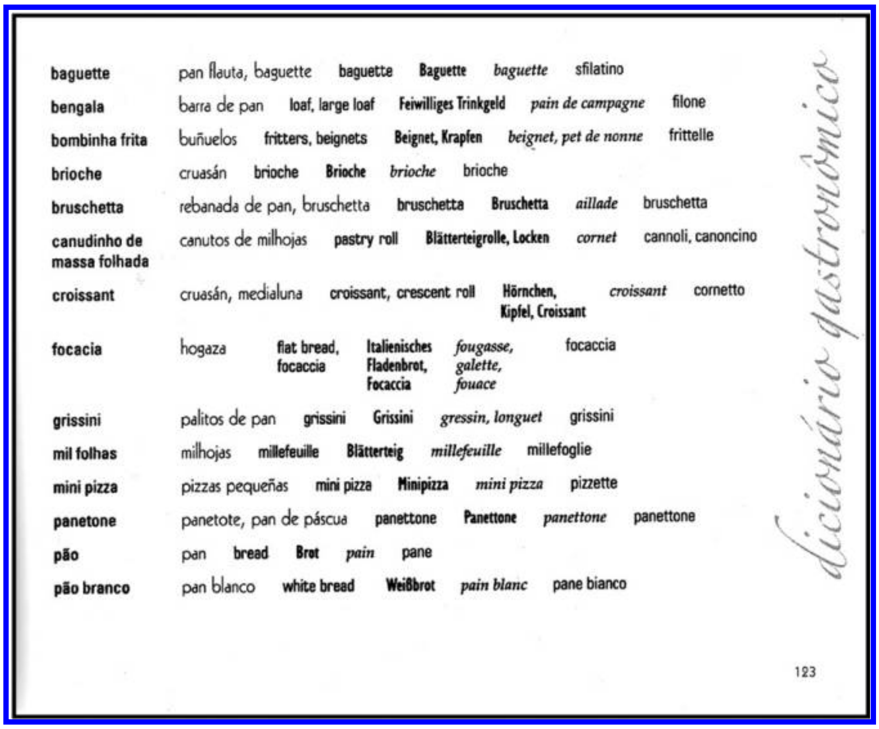
1. Technical translation, Terminology and Corpus Linguistics

It is a known fact that translators need a variety of reference sources to produce a good quality translation. The more specialised the text to be translated, the greater the need for specific terminological sources. However, most existing reference sources for the English–Portuguese pair pose two problems: first, they do not meet the needs of translators as regards usage, (i.e., they do not provide the context in which the term occurs), and, secondly,

[1] Projeto CoMET – Multilingual Corpus for Teaching and Translation, Modern Languages Department, University of São Paulo, Av. Prof. Luciano Gualberto, 403 – sala 14 (3o. andar), 05508-900 São Paulo, Brazil.
Correspondence to: Stella E.O. Tagnin, *e-mail*: seotagni@usp.br

Corpora 2012 Vol. 7 (1): 51–67
DOI: 10.3366/corp.2012.0017

www.eupjournals.com/cor



baguette	pan flauta, baguette	**baguette**	**Baguette**	*baguette*	sfilatino
bengala	barra de pan	**loaf, large loaf**	**Feiwilliges Trinkgeld**	*pain de campagne*	filone
bombinha frita	buñuelos	**fritters, beignets**	**Beignet, Krapfen**	*beignet, pet de nonne*	frittelle
brioche	cruasán	**brioche**	**Brioche**	*brioche*	brioche
bruschetta	rebanada de pan, bruschetta	**bruschetta**	**Bruschetta**	*aillade*	bruschetta
canudinho de massa folhada	canutos de milhojas	**pastry roll**	**Blätterteigrolle, Locken**	*cornet*	cannoli, canoncino
croissant	cruasán, medialuna	**croissant, crescent roll**	**Hörnchen, Kipfel, Croissant**	*croissant*	cornetto
focacia	hogaza	**flat bread, focaccia**	**Italienisches Fladenbrot, Focaccia**	*fougasse, galette, fouace*	focaccia
grissini	palitos de pan	**grissini**	**Grissini**	*gressin, longuet*	grissini
mil folhas	milhojas	**millefeuille**	**Blätterteig**	*millefeuille*	millefoglie
mini pizza	pizzas pequeñas	**mini pizza**	**Minipizza**	*mini pizza*	pizzette
panetone	panetote, pan de páscua	**panettone**	**Panettone**	*panettone*	panettone
pão	pan	**bread**	**Brot**	*pain*	pane
pão branco	pan blanco	**white bread**	**Weißbrot**	*pain blanc*	pane bianco

dicionário gastronômico

123

Figure 1: Example of multilingual dictionary (Carli and Klotz, 2006)

they are rarely up-to-date, because technology and knowledge now advance at such a pace that no published work can keep up with them. Thus, most bi- and multilingual technical glossaries tend to focus on the comprehension of technical texts by apprentices of specialised areas, rather than on the production of texts by translators. Indeed, they merely list monolexical terms and their 'equivalents', without examples from which the translator might glean some information as to their usage, as can be seen under Figure 1.

Many misconceptions regarding the practice of technical translation have been pointed out by various authors (Byrne, 2006: 2–7; and Korning Zethsen, 1999; among others), and all of them can be subsumed to one: the idea that technical texts can be reduced to the terminological content they convey. In technical translation, the correct use of terminology does not guarantee, alone, that the resulting text will be a good translation (Azenha Jr, 1999; Byrne, 2006; Korning Zethsen, 1999; and Wright, 1993) – that is to say, one that re-textualises the original message in an accurate, fluent and natural-sounding way. Wright (1993: 70, original emphasis) notes:

> Documents must speak 'the language' of the target audience and should resemble other texts produced within that particular language community and subject domain [. . .] These considerations frequently

> require that translators move beyond *merely correct* strategies in terms of lexical and grammatical content in order to account for *stylistically appropriate* solutions.

One must bear in mind that a translator is a text producer and, therefore, needs to know how a word is used – how it associates with other words (Firth, 1957); and this information can be inferred from the observation of authentic examples.

In this sense, a further handicap of the majority of technical reference materials is that entries generally comprise only nouns and adjectives while little or no attention is devoted to words with other grammatical functions (Finatto, 2007). In addition, these reference materials seem to ignore the fact that technical language does not consist solely of monolexical terms; multiword expressions, as Estopá (1999) and Bevilacqua (2001) claim, may also enjoy terminological status and should, therefore, be registered as entries in their own right. That would be the case, for example, of including not only 'pepper' or 'black pepper' but also 'freshly ground black pepper' in a glossary of cooking terms.[2] Moreover, the selection of which terms will be included in the glossary is usually the compiler's choice, and the compiler sometimes relies on previous works, abstract schemata or the discretion of one or a few specialists, irrespective of the term's relevance and frequency in actual texts from the field in question (Teixeira, 2008).

In short, most reference materials that are available are not translator-oriented: they do not provide context, collocations, phraseologies, translation suggestions in cases when there are no 'direct equivalents', or even information about the usage of a term in different genres and textual types (Fromm, 2008; Fuentes Morán and García Palacios, 2002; Gómez and Vargas, 2004; Salgado, 2006; Teixeira, 2004, 2008; and Varantola, 1998).

All of the shortcomings of reference materials we have mentioned above are evident in the translation of most technical texts and become even more evident in the translation of cooking recipes (Colina, 1997; and Teixeira, 2004, 2008). Contrary to popular belief, cooking is a highly technical area, not only in its practice – an inadequate procedure may put the dish to waste – but especially when it comes to translating texts. First, recipes may convey culture-specific habits, ingredients or even procedures which may be strange to the target audience. Secondly, some ingredients or appliances may not exist in the target culture, which raises the question of how they should be translated. A literal translation may produce awkward and sometimes hilarious renderings such as *1/2 xíc. de chá de suco de lima* for '1/2 cup lime juice'. In Portuguese, 'lime' is *limão* (the most common variety, Tahiti) whereas 'lemon' is *limão siciliano*; and *lima* is a type of orange with a slightly bitter taste. Another example of a strange rendering would

[2] The fact that it can be rendered as 'pimenta-do-reino moída na hora' in Brazilian Portuguese and 'pimenta preta moída na altura' in European Portuguese is good evidence that this is a cohesive translation unit.

be *chocolate semi-doce* ('semi-sweet chocolate') instead of the conventional phraseology *chocolate meio-amargo* ('?half-bitter chocolate');[3] a cook would probably understand what *chocolate semi-doce* means, but would realise, nevertheless, that 'that is just not the way we say it'.

Corpus Linguistics is an empiricist approach to language (McEnery and Wilson, 1996), and has been playing an important role in Translation Studies and, more specifically, in the practice and teaching of technical translation (Maia, 1997; Tagnin, 2002; and Varantola, 2002). In addition, it has also proved to be a valuable tool for Terminology in that it allows for the retrieval of technical terms and phraseologies, and their equivalents, from corpora (Bowker, 1996; Bowker and Pearson, 2002; Pearson, 1998; and Tognini-Bonelli, 2002) by means of special software that identifies recurrent patterns of language.

Corpus Linguistics views language as a probabilistic system, since it is concerned not only with what it is 'possible' to produce in a language, but mainly with what will 'probably' occur. It relies on the observation of linguistic phenomena in large quantities of text – called 'corpora' – in order to produce generalisations about that language. Thus, it seems natural to assume that a glossary based on a corpus may provide the answer to some of the needs of a translator, mentioned above, because: (*a*) a corpus can be built according to the translator's needs; (*b*) it can be constantly updated; and (*c*) when carefully compiled, it offers authentic examples from natural-sounding and accurate texts – all of which help to reassure the translator that the term chosen is the most appropriate one.

In this paper, we will present the corpus-driven approach used to compile a bilingual monodirectional English–Portuguese glossary of cooking terms that we published in Brazil (Teixeira and Tagnin, 2008). We will discuss the various stages of the project, with an emphasis on the identification of candidate terms using frequency and keyword lists, and their subsequent validation through concordance lines, with the lexical analysis software, WordSmith Tools 3.0 (Scott, 1999).

2. Corpus-driven Terminology

Corpus Linguistics can be viewed from two perspectives: as a methodology and as an approach. In Brazil, it has mostly been used by terminologists as a methodology – that is, electronic corpora are compiled and then perused to retrieve definitions, examples and/or equivalents for selected terms (see, for example, Alves, 1998; and Krieger *et al.*, 2006). Although some glossaries are a step forward in the sense that they use Corpus Linguistics tools to generate Keyword lists from which some of the entries for the glossary are selected (e.g., Perrotti-Garcia and Rebechi, 2007), the corpora they are

[3] In this paper, the interrogation mark preceding a combination of words indicates that this is a doubtful rendering, (i.e., a non-natural sounding combination).

based on contain mainly didactic texts, in which defining contexts abound, rather than state-of-the-art materials that circulate among the specialists of a particular field of knowledge.

As stated before, traditional Terminology focusses mainly on nouns and adjectives; verbs have only recently received some attention (Bevilacqua, 2001; Estopá, 1999; and Finatto, 2007). Corpus-based terminology does the same: emphasis is usually given to the conceptual schema of the area under investigation, and these are defined by specialists who sometimes disagree about them. The meaning of terms is not, thus, deduced from the contexts in which they actually occur in authentic texts, but from the conceptual relations they have in the schema's hierarchy.

Last but not least, corpus-based glossaries, like traditional glossaries, will include phraseologies only if they contain a term (nouns or adjectives, but very rarely verbs). For example, *enquanto isso* ('meanwhile'), which is a recurring Brazilian Portuguese phrase in cooking recipes, (and would be rendered as *entretanto* in European Portuguese recipes), would not be considered a valid entry because it does not include a content word. This is evidence that such theoretical constraints can do a disservice to the translator.

In previous works (Teixeira, 2008; and Teixeira and Tagnin, 2008), we proposed the use of Corpus Linguistics as an *approach* to bilingual Terminology, as opposed to the corpus-based *methodology* used by some terminologists. By corpus-based *methodology* we mean that, although these researchers use a corpus, they do so to speed up traditional methods of terminology compilation. In such instances, the corpus is merely used as a 'fish pond' (Hanks, 2008: 220, citing Sinclair, 1987) from which the terminologist yields examples and definitions to fill in a list of terms that has been compiled previously; the list is usually based on a domain area schema devised and/or approved by a specialist. For instance, Farias and Bezerra (2008) resort to magazines, newspapers and catalogues to extract equivalents for their trilingual glossary of fashion terms based on a previously compiled monolingual *Portuguese Glossary of Fashion Terms* (Farias, 2003). The authors explicitly state that they used Farias (2003) as the 'corpus' from which they extracted the list of source language terms to be used as the basis for their trilingual glossary. Besides, any doubts resulting from this kind of procedure are, again, submitted to one or more specialists who will, ultimately, decide what enters and what does not enter the dictionary. In contrast, by corpus-driven *approach* we mean that even the list of terms should be retrieved from the corpus; in this sense, the corpus will 'tell' us which are the terms most commonly used in the area being investigated, not the specialist.

If we were to draw an analogy, we could say that in the same way an 'armchair linguist' (Fillmore, 1992) relies on a native speaker to validate intuitions, a corpus-based terminologist relies on a specialist to validate the information to be entered into a technical dictionary. We propose, instead, that a specialised corpus (and not one or a few specialists) be the main source of information used to build the dictionary entries and, especially, that the corpus be used to generate the list of headwords for the dictionary. The specialist can, then, have a more peripheral – and perhaps more effective – role in the compilation of translator-oriented technical dictionaries, such as helping the terminologist to choose the best sources of technical texts to build a sound, up-to-date and representative corpus.

Our *Vocabulário para Culinária Inglês–Português* ('English–Portuguese Glossary of Cooking Terms') was built according to this approach. We departed from the information contained in a carefully built comparable corpus of cooking recipes to select our entries, and we used a corpus-driven approach to extract information from the corpus using WordSmith Tools. In what follows, after briefly describing the corpus contents, we detail the methodology we adopted to build the list of entries in English and to find equivalents in the Portuguese comparable subcorpus. We then give some examples of the resulting 'translator-oriented entries', which are followed by our concluding remarks.

3. Compiling a corpus-driven glossary

3.1 Corpus building

In order to produce a bilingual corpus-driven glossary, we first built a comparable corpus – two corpora, one in each language, consisting of authentic texts covering the same domain, the same specific areas, and consisting of similar texts (in terms of genre, text typology, content, time span, extension, *etc*.). This would guarantee that all the resulting data were based on authentic language use. Parallel corpora (original texts and their translations) could also have been used as a source of *prima facie* equivalents for the entries identified in a corpus of authentic texts, but such corpora are rarely available on the Internet and, besides, require extra time to process, as they need to be aligned. So, our hypothesis was that a novice or even an experienced translator/terminologist working in a new subject area would be able to build a reliable bilingual glossary using well-built comparable corpora.

For the purpose of our cooking glossary, we gathered two subcorpora, one in English and one in Brazilian Portuguese, consisting of cooking recipes collected from reliable sources on the Internet, (i.e., websites containing recipes originally written, collected and/or edited by respected professionals in both languages). For this corpus in particular, we did not consult a specialist to help to locate representative and accurate texts on the Internet, since, being a corpus of home cooking recipes, the authors themselves – as home cooks and as home cookbook translators – agreed they should be able to evaluate which websites contained adequate texts. In the English language, some of the websites we used were *The Great British*

	Portuguese	English
Number of recipes	*c.* 8,300	*c.* 7,400
Tokens	1,520,864	1,578,125
Types	14,635	28,903
Type/token ratio	0.96	0.93

Table 1: Composition of the comparable corpus

Kitchen by Helen Watson, and the *Food* section of the BBC website; for the Portuguese, the cooking section of the *Terra* portal, as well as the *Basilico* website, and the *Gula* magazine website were some of the main sources.

The recipes were copied manually and/or using offline browsers, such as HTTrack.[4] The html files were then converted into txt files using the Text Converter utility of WordSmith Tools (Scott, 1999). This corpus is now part of the CorTec, the Technical Comparable Corpus of the CoMET project, and is available online.[5] For detailed information about its compilation and its contents, see Teixeira (2008: 206–18). Information concerning its size can be seen under Table 1.

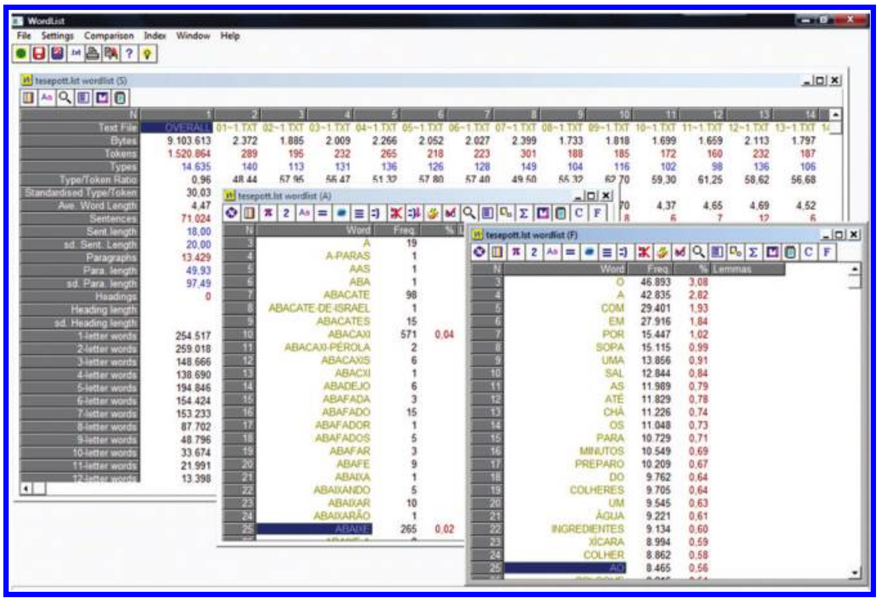
3.2 Corpus-driven approach

Wordsmith Tools was also the software used in our analysis. Since our aim was to produce a bilingual glossary in the English→Portuguese direction, we began by using the WordList function to generate a wordlist of the English corpus, (i.e., a list of all the words in all the texts of the corpus with their frequencies). Figure 2 shows the three screens produced by the Wordsmith's WordList tool with the results: statistics (S), words in alphabetical order (A) and words in frequency order (F).

In order to have a more accurate picture of the vocabulary specific to the cooking area, we then generated a KeyWord list. This is done by comparing the WordList of the English subcorpus with a WordList of a larger and more general corpus. This procedure removes the vocabulary that is commonly used in the texts of both corpora and highlights the vocabulary peculiar to the area under investigation – in our case, cooking recipes. We

[4] See: http://www.httrack.com/. These browsers automatically extract all files from a chosen Internet address and store them in the user's computer. This allows the user to surf the website even when the Internet connection is off, and means that they end up with a copy of the website in their computer (something that, unfortunately, most websites do not allow nowadays). The only problem is that one usually has to scan the files after the download is complete in order to delete unwanted ones (those that do not contain the type of text you are trying to collect). Even so, it is a much quicker process than copying each file manually.

[5] See: http://www.fflch.usp.br/dlm/comet/consulta_cortec.html



WordList
File Settings Comparison Index Window Help
tesepott.lst wordlist (S)
tesepott.lst wordlist (A)
tesepott.lst wordlist (F)

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Text File	OVERALL	01~1.TXT	02~1.TXT	03~1.TXT	04~1.TXT	05~1.TXT	06~1.TXT	07~1.TXT	08~1.TXT	09~1.TXT	10~1.TXT	11~1.TXT	12~1.TXT	13~1.TXT
Bytes	9.103.613	2.372	1.885	2.009	2.266	2.052	2.027	2.399	1.733	1.818	1.699	1.659	2.113	1.797
Tokens	1.520.864	289	195	232	265	218	223	301	188	185	172	160	232	187
Types	14.635	140	113	131	136	126	128	149	104	116	102	98	136	106
Type/Token Ratio	0,96	48,44	57,95	56,47	51,32	57,80	57,40	49,50	55,32	62,70	59,30	61,25	58,62	56,68
Standardised Type/Token	30,03													
Ave. Word Length	4,47									70	4,37	4,65	4,69	4,52
Sentences	71.024									8	6	7	12	6
Sent. length	18,00													
sd. Sent. Length	20,00													
Paragraphs	13.429													
Para. length	49,93													
sd. Para. length	97,49													
Headings	0													
Heading length														
sd. Heading length														
1-letter words	254.517													
2-letter words	259.018													
3-letter words	148.666													
4-letter words	138.690													
5-letter words	194.846													
6-letter words	154.424													
7-letter words	153.233													
8-letter words	87.702													
9-letter words	48.796													
10-letter words	33.674													
11-letter words	21.991													
12-letter words	13.398													

tesepott.lst wordlist (A)

N	Word	Freq.	%
3	A	19	
4	A-PARAS	1	
5	AAS	1	
6	ABA	1	
7	ABACATE	98	
8	ABACATE-DE-ISRAEL	1	
9	ABACATES	15	
10	ABACAXI	571	0,04
11	ABACAXI-PÉROLA	2	
12	ABACAXIS	6	
13	ABACXI	1	
14	ABADEJO	6	
15	ABAFADA	3	
16	ABAFADO	15	
17	ABAFADOR	1	
18	ABAFADOS	5	
19	ABAFAR	3	
20	ABAFE	9	
21	ABAIXA	1	
22	ABAIXANDO	5	
23	ABAIXAR	10	
24	ABAIXARÃO	1	
25	ABAIXE	265	0,02

tesepott.lst wordlist (F)

N	Word	Freq.	%	Lemmas
3	O	46.893	3,08	
4	A	42.835	2,82	
5	COM	29.401	1,93	
6	EM	27.916	1,84	
7	POR	15.447	1,02	
8	SOPA	15.115	0,99	
9	UMA	13.856	0,91	
10	SAL	12.844	0,84	
11	AS	11.989	0,79	
12	ATÉ	11.829	0,78	
13	CHÁ	11.226	0,74	
14	OS	11.048	0,73	
15	PARA	10.729	0,71	
16	MINUTOS	10.549	0,69	
17	PREPARO	10.209	0,67	
18	DO	9.762	0,64	
19	COLHERES	9.705	0,64	
20	UM	9.545	0,63	
21	ÁGUA	9.221	0,61	
22	INGREDIENTES	9.134	0,60	
23	XÍCARA	8.994	0,59	
24	COLHER	8.862	0,58	
25	AO	8.465	0,56	

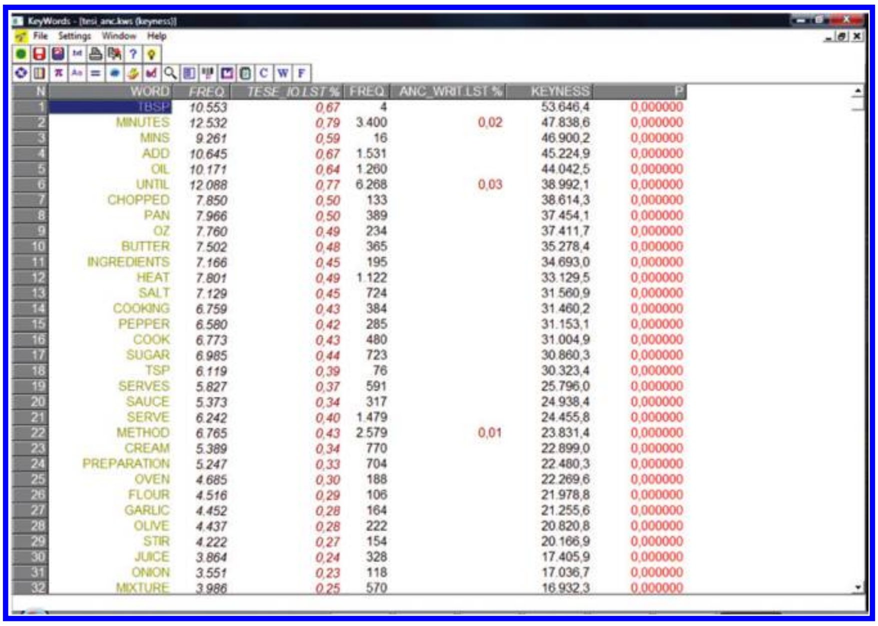
Figure 2: Results generated by the WordList tool

used the written part of the American National Corpus (ANC)[6] as a reference corpus. This produced over 3,670 keywords – words which occur in the English corpus of cooking recipes more often than we would expect them to occur by chance alone; the first thirty-two words can be seen under Figure 3.

Due to time constraints, we decided to investigate the first 300 of these keywords: one author focussed on nouns and adjectives, the other on verbs and adverbs. Our glossary would, therefore, consist of the 300 most probable words (and the multiword combinations they are part of) to occur in recipes – rather than an ad hoc collection of terms based on pre-conceived assumptions or abstract conceptual schemata. A more comprehensive glossary, though, should certainly have a much higher threshold (which has still to be defined in the literature), and should also accommodate the same process of making wordlists and keywords for multiword units, such as bigrams, trigrams, quadrigrams, *etc*., because there is no guarantee that any key combination of two or more words will contain at least one that is key in a keyword list of unigrams. Take, for instance, *enquanto isso*: neither *enquanto* nor *isso* are in the unigram keyword list in the Portuguese corpus of recipes when it is compared to Lácio-Ref,[7] but the

[6] See: http://americannationalcorpus.org/. The ANC project aims to create a massive corpus of American English, 'including texts of all genres and transcripts of spoken data produced from 1990 onward.' The version used for our glossary was the second release which contained about 18.5 million words of written American English.

[7] See: www.nilc.icmc.usp.br/lacioweb. This corpus contains almost 10 million written words from various domains, such as Biology, Hard Sciences, Humanities, Social Sciences, Religion, *etc*. It was used as the reference corpus for our Portuguese keywords.



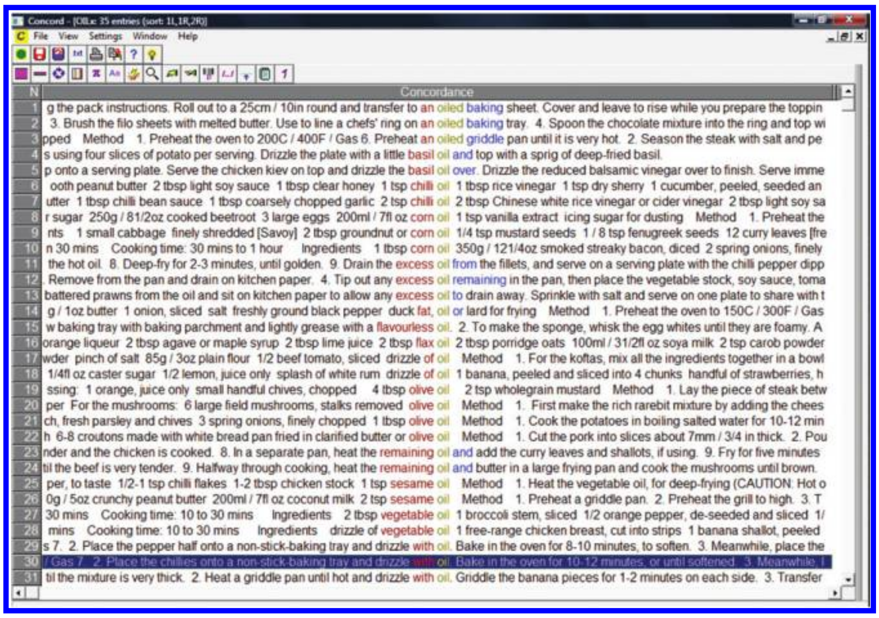
N	WORD	FREQ.	TESE_IO.LST %	FREQ.	ANC_WRIT.LST %	KEYNESS	P
1	TBSP	10.553	0,67	4		53.646,4	0,000000
2	MINUTES	12.532	0,79	3.400	0.02	47.838,6	0,000000
3	MINS	9.261	0,59	16		46.900,2	0,000000
4	ADD	10.645	0,67	1.531		45.224,9	0,000000
5	OIL	10.171	0,64	1.260		44.042,5	0,000000
6	UNTIL	12.088	0,77	6.268	0,03	38.992,1	0,000000
7	CHOPPED	7.850	0,50	133		38.614,3	0,000000
8	PAN	7.966	0,50	389		37.454,1	0,000000
9	OZ	7.760	0,49	234		37.411,7	0,000000
10	BUTTER	7.502	0,48	365		35.278,4	0,000000
11	INGREDIENTS	7.166	0,45	195		34.693,0	0,000000
12	HEAT	7.801	0,49	1.122		33.129,5	0,000000
13	SALT	7.129	0,45	724		31.560,9	0,000000
14	COOKING	6.759	0,43	384		31.460,2	0,000000
15	PEPPER	6.580	0,42	285		31.153,1	0,000000
16	COOK	6.773	0,43	480		31.004,9	0,000000
17	SUGAR	6.985	0,44	723		30.860,3	0,000000
18	TSP	6.119	0,39	76		30.323,4	0,000000
19	SERVES	5.827	0,37	591		25.796,0	0,000000
20	SAUCE	5.373	0,34	317		24.938,4	0,000000
21	SERVE	6.242	0,40	1.479		24.455,8	0,000000
22	METHOD	6.765	0,43	2.579	0,01	23.831,4	0,000000
23	CREAM	5.389	0,34	770		22.899,0	0,000000
24	PREPARATION	5.247	0,33	704		22.480,3	0,000000
25	OVEN	4.685	0,30	188		22.269,6	0,000000
26	FLOUR	4.516	0,29	106		21.978,8	0,000000
27	GARLIC	4.452	0,28	164		21.255,6	0,000000
28	OLIVE	4.437	0,28	222		20.820,8	0,000000
29	STIR	4.222	0,27	154		20.166,9	0,000000
30	JUICE	3.864	0,24	328		17.405,9	0,000000
31	ONION	3.551	0,23	118		17.036,7	0,000000
32	MIXTURE	3.986	0,25	570		16.932,3	0,000000

Figure 3: KeyWords of the English corpus compared to the ANC corpus

bigram *enquanto isso* occupies the 658th position in the bigram keyword list (it occurs 352 times in the corpus), in spite of the fact that it is also a rather common expression in the language overall (there are sixty occurrences in the Lácio-Ref corpus).

The next step was to extract concordance lines for each of these words and then analyse the co-text – words to the left and right of the target – with which they co-occurred in order to identify the multiword units they were part of. For this purpose, we used Wordsmith's Concord function, which allows the lines containing the search word in the centre to be sorted according to the words to its left and/or right. This procedure makes it quite easy to visualise recurrent patterns. Figure 4 illustrates some of the collocations found for the search word 'oil*', such as *oiled* [SURFACE], *basil oil*, *chilli oil*, *corn oil*, *excess oil*, *drizzle of oil*, *olive oil*, *sesame oil*, *vegetable oil* and *drizzle with oil*. All recurrent collocations and phraseologies that contained one of the 300 keywords analysed were included in the glossary.

Once all English entries were selected (words, collocations and phraseologies) it was then necessary to establish their translations in Portuguese. The first step we took was to look for equivalents in the Portuguese KeyWord list, which was obtained by comparing the Portuguese WordList with a WordList of Lácio-Ref, which was used as our Portuguese reference corpus. This showed several candidates for translation, such as *minutes* ('minutos'), *cream* ('creme') and *garlic* ('alho'). But it also showed variants like *heat* ('fogo', 'quente'), *stir* ('coloque', 'junte', 'acrescente') and so on (see Figure 5). In the case of *oil*, for instance, we identified *azeite*



```
N   Concordance
1   g the pack instructions. Roll out to a 25cm / 10in round and transfer to an oiled baking sheet. Cover and leave to rise while you prepare the toppin
2    3. Brush the filo sheets with melted butter. Use to line a chefs' ring on an oiled baking tray.  4. Spoon the chocolate mixture into the ring and top wi
3   pped    Method    1. Preheat the oven to 200C / 400F / Gas 6. Preheat an oiled griddle pan until it is very hot.  2. Season the steak with salt and pe
4   s using four slices of potato per serving. Drizzle the plate with a little basil oil and top with a sprig of deep-fried basil.
5   p onto a serving plate. Serve the chicken kiev on top and drizzle the basil oil over. Drizzle the reduced balsamic vinegar over to finish. Serve imme
6    ooth peanut butter  2 tbsp light soy sauce  1 tbsp clear honey  1 tsp chilli oil  1 tbsp rice vinegar  1 tsp dry sherry  1 cucumber, peeled, seeded an
7   utter  1 tbsp chilli bean sauce  1 tbsp coarsely chopped garlic  2 tsp chilli oil  2 tbsp Chinese white rice vinegar or cider vinegar  2 tbsp light soy sa
8   r sugar  250g / 81/2oz cooked beetroot  3 large eggs  200ml / 7fl oz corn oil  1 tsp vanilla extract  icing sugar for dusting    Method    1. Preheat the
9   nts    1 small cabbage  finely shredded [Savoy]  2 tbsp groundnut or corn oil  1/4 tsp mustard seeds  1 / 8 tsp fenugreek seeds  12 curry leaves [fre
10  n 30 mins    Cooking time: 30 mins to 1 hour     Ingredients     1 tbsp corn oil  350g / 121/4oz smoked streaky bacon, diced  2 spring onions, finely
11   the hot oil.  8. Deep-fry for 2-3 minutes, until golden.  9. Drain the excess oil from the fillets, and serve on a serving plate with the chilli pepper dipp
12  . Remove from the pan and drain on kitchen paper.  4. Tip out any excess oil remaining in the pan, then place the vegetable stock, soy sauce, toma
13  battered prawns from the oil and sit on kitchen paper to allow any excess oil to drain away. Sprinkle with salt and serve on one plate to share with t
14   g / 1oz butter  1 onion, sliced  salt  freshly ground black pepper  duck fat, oil or lard for frying    Method    1. Preheat the oven to 150C / 300F / Gas
15   w baking tray with baking parchment and lightly grease with a flavourless oil.  2. To make the sponge, whisk the egg whites until they are foamy. A
16  orange liqueur  2 tbsp agave or maple syrup  2 tbsp lime juice  2 tbsp flax oil  2 tbsp porridge oats  100ml / 31/2fl oz soya milk  2 tsp carob powder
17  wder  pinch of salt  85g / 3oz plain flour  1/2 beef tomato, sliced  drizzle of oil    Method    1. For the koftas, mix all the ingredients together in a bowl
18   1/4fl oz caster sugar  1/2 lemon, juice only  splash of white rum  drizzle of oil  1 banana, peeled and sliced into 4 chunks  handful of strawberries, h
19   ssing:  1 orange, juice only  small handful chives, chopped     4 tbsp olive oil    2 tsp wholegrain mustard    Method    1. Lay the piece of steak betw
20   per  For the mushrooms:  6 large field mushrooms, stalks removed  olive oil    Method    1.  First make the rich rarebit mixture by adding the chees
21  ch, fresh parsley and chives  3 spring onions, finely chopped  1 tbsp olive oil    Method    1. Cook the potatoes in boiling salted water for 10-12 min
22  h  6-8 croutons made with white bread pan fried in clarified butter or olive oil    Method    1. Cut the pork into slices about 7mm / 3/4 in thick.  2. Pou
23  nder and the chicken is cooked.  8. In a separate pan, heat the remaining oil and add the curry leaves and shallots, if using.  9. Fry for five minutes
24  til the beef is very tender.  9. Halfway through cooking, heat the remaining oil and butter in a large frying pan and cook the mushrooms until brown.
25   per, to taste  1/2-1 tsp chilli flakes  1-2 tbsp chicken stock  1 tsp sesame oil    Method    1. Heat the vegetable oil, for deep-frying (CAUTION: Hot o
26   0g / 5oz crunchy peanut butter  200ml / 7fl oz coconut milk  2 tsp sesame oil    Method    1. Preheat a griddle pan.  2. Preheat the grill to high.  3. T
27   30 mins    Cooking time: 10 to 30 mins     Ingredients     2 tbsp vegetable oil  1 broccoli stem, sliced  1/2 orange pepper, de-seeded and sliced  1/
28   mins    Cooking time: 10 to 30 mins     Ingredients     drizzle of vegetable oil  1 free-range chicken breast, cut into strips  1 banana shallot, peeled
29  s 7.  2. Place the pepper half onto a non-stick-baking tray and drizzle with oil. Bake in the oven for 8-10 minutes, to soften.  3. Meanwhile, place the
30  / Gas 7.  2. Place the chillies onto a non-stick-baking tray and drizzle with oil. Bake in the oven for 10-12 minutes, or until softened.  3. Meanwhile, I
31   til the mixture is very thick.  2. Heat a griddle pan until hot and drizzle with oil. Griddle the banana pieces for 1-2 minutes on each side.  3. Transfer
```

Figure 4: Selected, sorted concordance lines for *oil**

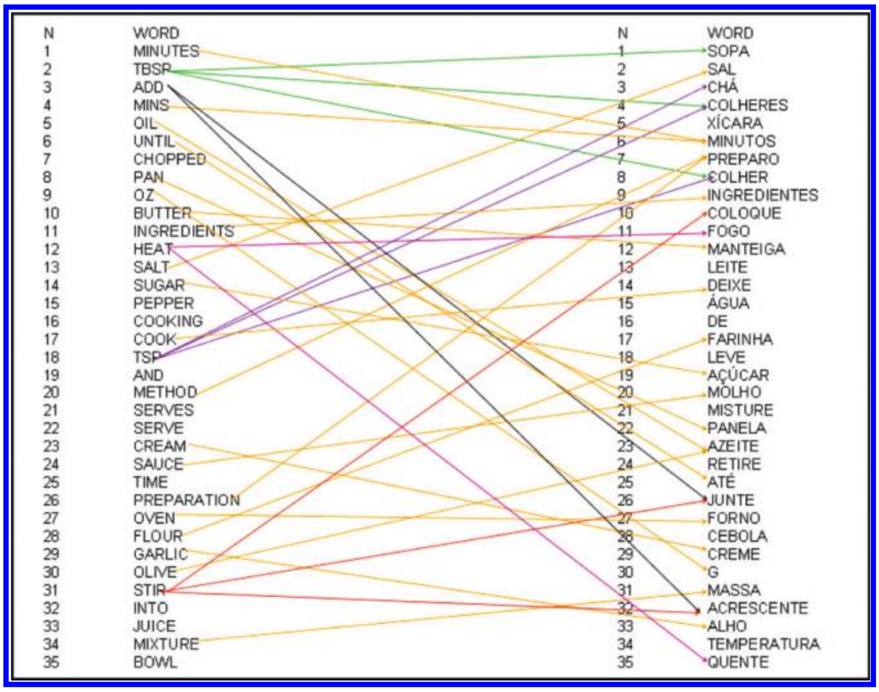


Figure 5: Translation candidates in the English–Portuguese comparable corpus of recipes

Collocate of *finely*	No. of occurrences
chopped	2,528
sliced	443
diced	379
grated	280
chop	167
shredded	85

Table 2: Collocates of finely in the English corpus

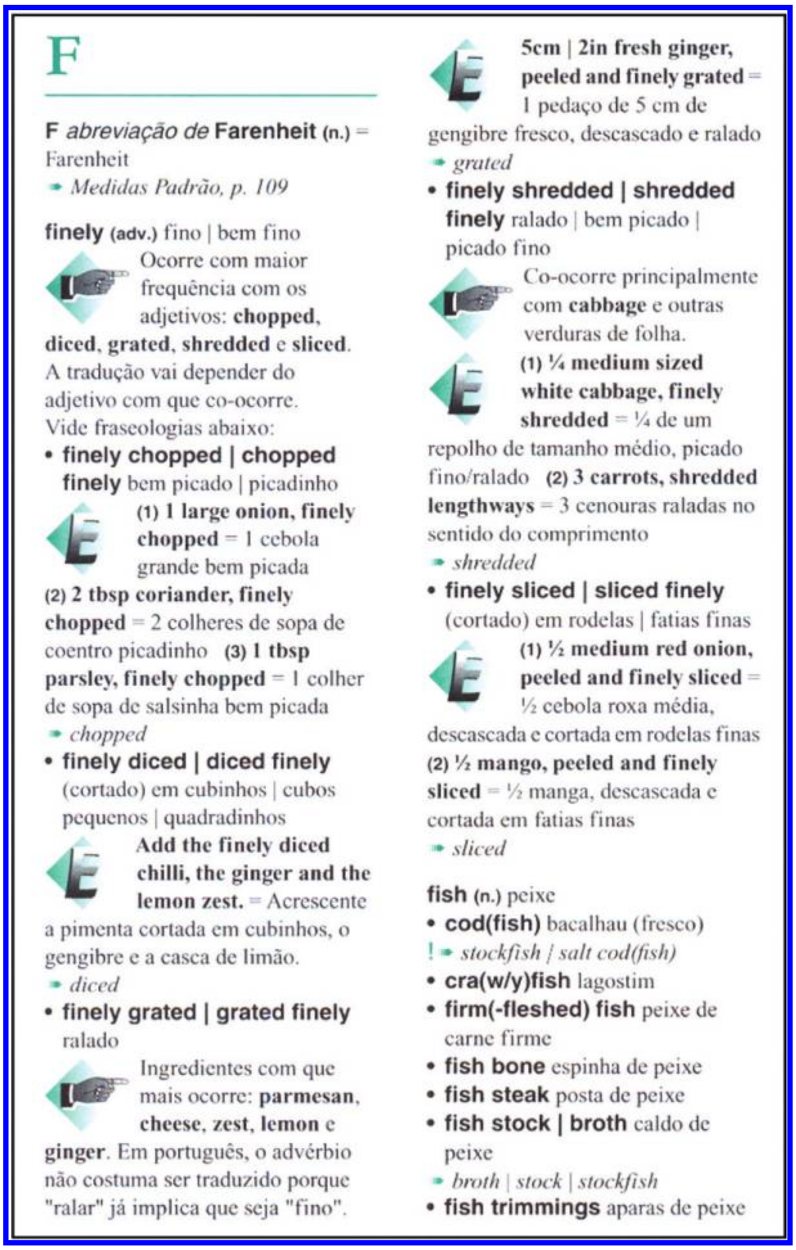
(‘olive oil’) as the twenty-third keyword and, further down, *óleo* (‘oil’) and the adjective *untado/a*, which can be an equivalent for ‘oiled’ in contexts such as ‘Place skin-side up on an oiled baking tray and grill for 4–5 minutes’.

We then generated concordance lines in Portuguese to make sure the candidates for translation occurred in similar contexts and with a similar frequency in the corpus. When the frequencies of equivalents were clearly different, we used the methodology devised by Tagnin (2007) – we looked for the collocates of the English term and generated concordance lists for them. For instance, *finely* occurred over 3,000 times in the English corpus while its Portuguese cognate, *finamente*, occurred only 236 times. Since the corpora were built to be comparable, it was expected that most corresponding terms would have similar frequencies. So, we looked for the most frequent collocates of ‘finely’ in the English corpus, which are shown under Table 2.

We then generated concordances for their candidate translations in Portuguese (for example, *picado* for ‘chopped’), to determine their collocates. This, in the case of *picado*, led us to *bem picado* (‘?well chopped’) or (*bem*) *picadinho* (‘?very well chopped’) as the best Portuguese equivalents for ‘finely chop(ped)’. Thus, although the adverbs ‘finely’ and *bem* would probably never be regarded as equivalents in other contexts, they do seem to play the same role as qualifiers of the corresponding participles *chopped* and *picado* in cooking recipes.

The second most frequent collocate of *finely* in the English corpus – the verb *slice* – could be translated as ‘fatiar’, but this possibility did not occur many times in the Portuguese corpus. So we used a wild card and searched for *fatia**, which gave us ‘cort* em fatias finas’ (‘?cut into fine slices’) as a better equivalent for *finely sliced*. A further search for *cortar* (‘cut’) produced ‘cortar em rodelas finas’ (‘?cut into fine round slices’), usually used with *onions*, *carrots* and other round-shaped ingredients.

Another collocate of *finely* in English is *grate*. Common collocates for this verb are ‘parmesan’, ‘cheese’ and ‘chocolate’. The corresponding verb in Portuguese, *ralar*, occurs with similar ingredients, but a noticeable difference is that while in English it is usual to say ‘finely grated parmesan cheese’ (fourteen occurrences in the corpus), in Portuguese *parmesão* (‘parmesan cheese’) does co-occur with *ralado* (‘grated’) but usually with no



F

F *abreviação de* **Farenheit** (n.) = Farenheit
➔ *Medidas Padrão, p. 109*

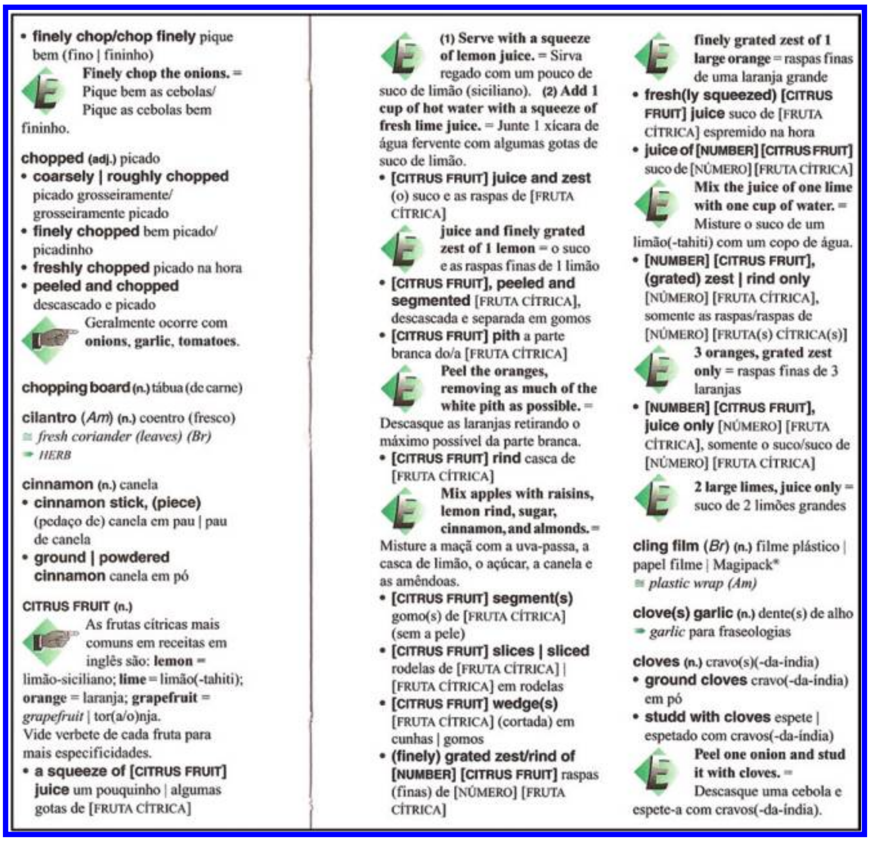
finely (adv.) fino | bem fino
☞ Ocorre com maior frequência com os adjetivos: **chopped, diced, grated, shredded** e **sliced**. A tradução vai depender do adjetivo com que co-ocorre. Vide fraseologias abaixo:

- **finely chopped | chopped finely** bem picado | picadinho
  (1) **1 large onion, finely chopped** = 1 cebola grande bem picada (2) **2 tbsp coriander, finely chopped** = 2 colheres de sopa de coentro picadinho (3) **1 tbsp parsley, finely chopped** = 1 colher de sopa de salsinha bem picada
  ➔ *chopped*
- **finely diced | diced finely** (cortado) em cubinhos | cubos pequenos | quadradinhos
  Add the finely diced chilli, the ginger and the lemon zest. = Acrescente a pimenta cortada em cubinhos, o gengibre e a casca de limão.
  ➔ *diced*
- **finely grated | grated finely** ralado
  ☞ Ingredientes com que mais ocorre: **parmesan, cheese, zest, lemon** e **ginger**. Em português, o advérbio não costuma ser traduzido porque "ralar" já implica que seja "fino".
  5cm | 2in fresh ginger, peeled and finely grated = 1 pedaço de 5 cm de gengibre fresco, descascado e ralado
  ➔ *grated*
- **finely shredded | shredded finely** ralado | bem picado | picado fino
  ☞ Co-ocorre principalmente com **cabbage** e outras verduras de folha.
  (1) **¼ medium sized white cabbage, finely shredded** = ¼ de um repolho de tamanho médio, picado fino/ralado (2) **3 carrots, shredded lengthways** = 3 cenouras raladas no sentido do comprimento
  ➔ *shredded*
- **finely sliced | sliced finely** (cortado) em rodelas | fatias finas
  (1) **½ medium red onion, peeled and finely sliced** = ½ cebola roxa média, descascada e cortada em rodelas finas (2) **½ mango, peeled and finely sliced** = ½ manga, descascada e cortada em fatias finas
  ➔ *sliced*

fish (n.) peixe
- **cod(fish)** bacalhau (fresco)
  ! ➔ *stockfish | salt cod(fish)*
- **cra(w/y)fish** lagostim
- **firm(-fleshed) fish** peixe de carne firme
- **fish bone** espinha de peixe
- **fish steak** posta de peixe
- **fish stock | broth** caldo de peixe
  ➔ *broth | stock | stockfish*
- **fish trimmings** aparas de peixe

Figure 6: Published entry for *finely* in Teixeira and Tagnin (2008)

modifier. This may point to a relevant cultural difference – parmesan cheese is ‘finely grated’ by default in Brazil, so there is no need, apparently, to specify that. However, when the cheese is to be grated differently, this fact is made clear: *ralado grosso* (‘coarsely grated’). In other words, for the Portuguese text to sound natural, *finely* should be omitted in the translation.



- **finely chop/chop finely** pique bem (fino | fininho)

  Finely chop the onions. = Pique bem as cebolas/ Pique as cebolas bem fininho.

chopped (adj.) picado
- **coarsely | roughly chopped** picado grosseiramente/ grosseiramente picado
- **finely chopped** bem picado/ picadinho
- **freshly chopped** picado na hora
- **peeled and chopped** descascado e picado

  Geralmente ocorre com **onions, garlic, tomatoes.**

chopping board (n.) tábua (de carne)

cilantro (*Am*) (n.) coentro (fresco)
≅ *fresh coriander (leaves) (Br)*
→ *HERB*

cinnamon (n.) canela
- **cinnamon stick, (piece)** (pedaço de) canela em pau | pau de canela
- **ground | powdered cinnamon** canela em pó

CITRUS FRUIT (n.)

As frutas cítricas mais comuns em receitas em inglês são: **lemon** = limão-siciliano; **lime** = limão(-tahiti); **orange** = laranja; **grapefruit** = *grapefruit* | tor(a/o)nja.
Vide verbete de cada fruta para mais especificidades.

- **a squeeze of [CITRUS FRUIT] juice** um pouquinho | algumas gotas de [FRUTA CÍTRICA]

  (1) Serve with a squeeze of lemon juice. = Sirva regado com um pouco de suco de limão (siciliano). **(2) Add 1 cup of hot water with a squeeze of fresh lime juice.** = Junte 1 xícara de água fervente com algumas gotas de suco de limão.
- **[CITRUS FRUIT] juice and zest** (o) suco e as raspas de [FRUTA CÍTRICA]

  juice and finely grated zest of 1 lemon = o suco e as raspas finas de 1 limão
- **[CITRUS FRUIT], peeled and segmented** [FRUTA CÍTRICA], descascada e separada em gomos
- **[CITRUS FRUIT] pith** a parte branca do/a [FRUTA CÍTRICA]

  Peel the oranges, removing as much of the white pith as possible. = Descasque as laranjas retirando o máximo possível da parte branca.
- **[CITRUS FRUIT] rind** casca de [FRUTA CÍTRICA]

  Mix apples with raisins, lemon rind, sugar, cinnamon, and almonds. = Misture a maçã com a uva-passa, a casca de limão, o açúcar, a canela e as amêndoas.
- **[CITRUS FRUIT] segment(s)** gomo(s) de [FRUTA CÍTRICA] (sem a pele)
- **[CITRUS FRUIT] slices | sliced** rodelas de [FRUTA CÍTRICA] | [FRUTA CÍTRICA] em rodelas
- **[CITRUS FRUIT] wedge(s)** [FRUTA CÍTRICA] (cortada) em cunhas | gomos
- **(finely) grated zest/rind of [NUMBER] [CITRUS FRUIT]** raspas (finas) de [NÚMERO] [FRUTA CÍTRICA]

  finely grated zest of 1 large orange = raspas finas de uma laranja grande
- **fresh(ly squeezed) [CITRUS FRUIT] juice** suco de [FRUTA CÍTRICA] espremido na hora
- **juice of [NUMBER] [CITRUS FRUIT]** suco de [NÚMERO] [FRUTA CÍTRICA]

  Mix the juice of one lime with one cup of water. = Misture o suco de um limão(-tahiti) com um copo de água.
- **[NUMBER] [CITRUS FRUIT], (grated) zest | rind only** [NÚMERO] [FRUTA CÍTRICA], somente as raspas/raspas de [NÚMERO] [FRUTA(s) CÍTRICA(s)]

  3 oranges, grated zest only = raspas finas de 3 laranjas
- **[NUMBER] [CITRUS FRUIT], juice only** [NÚMERO] [FRUTA CÍTRICA], somente o suco/suco de [NÚMERO] [FRUTA CÍTRICA]

  2 large limes, juice only = suco de 2 limões grandes

cling film (*Br*) (n.) filme plástico | papel filme | Magipack*
≅ *plastic wrap (Am)*

clove(s) garlic (n.) dente(s) de alho
→ *garlic* para fraseologias

cloves (n.) cravo(s)(-da-índia)
- **ground cloves** cravo(-da-índia) em pó
- **studd with cloves** espete | espetado com cravos(-da-índia)

  Peel one onion and stud it with cloves. = Descasque uma cebola e espete-a com cravos(-da-índia).

Figure 7: Published entry for CITRUS FRUIT in Teixeira and Tagnin (2008)

Shredded, another collocate of *finely* in English, co-occurs with *cabbage* while the Portuguese equivalent for this vegetable, *repolho*, co-occurs with *picar* in phraseologies such as 'picado bem fino', 'picado bem fininho' ('finely chopped'). All this information appears in the final entry for *finely* in our published glossary (see Figure 6).

3.3 Translator-oriented entries

If we go back to the translator's needs, as mentioned above, we will see that they are met in our glossary. With very few exceptions, all entries feature examples, all of which are extracted from the corpus, and their corresponding translations into Portuguese, as can be seen under Figure 5. Information on usage is also provided by listing common collocates and the type of recipe or the part of the recipe the words are more commonly associated with; for instance, we highlighted the fact that *fold* usually co-occurs with *egg whites* in the procedure part of recipes for cakes, soufflés, *etc*.

Culture-specific information is given in notes, as in the case of 'buttermilk', an ingredient that is not available in Brazil. In such instances, an explanation is given, as are suggestions on how to replace the ingredient in the recipes (specialists were consulted to provide a substitute for the ingredient). Several illustrations and charts throughout the glossary also add useful information, such as charts showing the differences between cuts of beef, pork and lamb meat, which were included in the corresponding entries. Conversion charts for the basic measuring units, and a colour-illustrated table of herbs and spices, appear at the end of the book.

Another feature of the glossary is the use of lemmas, or categories, as entries. For instance, after analysing the concordance lines for *lemon*, *lime*, *orange* and *grapefruit*, we realised that many collocates and multiword expressions were common to all or some of these words, which are all citrus fruits, such as 'candied [CITRUS FRUIT] rind / peel', '[CITRUS FRUIT] wedges', 'freshly squeezed [CITRUS FRUIT]', *etc*. So, we created an entry, with capital letters to distinguish it from the other entries, for CITRUS FRUIT, listing all the phraseologies commonly associated with citrus fruits in the corpus (see Figure 6). This entry is cross-referenced in all entries of the actual citrus fruits (lemon, lime, orange, *etc*.).

Cross-references like the one mentioned above, as well as synonyms, language variants (British *versus* American) and possible pitfalls in translation – such as translating *red onion* as '?cebola vermelha' instead of 'cebola roxa' ('?purple onion') – are all indicated by coloured icons which the translator can easily identify in the entry.

4. Conclusion

Our aim in this paper was to present the methodology used to compile a translator-oriented, bilingual glossary from a comparable corpus using a corpus-driven approach. We believe that such a glossary reflects the actual vocabulary of the area in a more reliable way, from the point of view of text production, than most glossaries that have been produced for the English–Brazilian Portuguese language pair using traditional methodologies in Terminology. As argued in the introduction, a reference source for this language professional should include context, collocates, as well as usage and culture-specific information so as to meet the needs of the translation task. The cooking glossary we have built complies with these requirements.

We hope other researchers may show an interest in testing this approach in other specialised areas, adjusting it to genre and text-specific features of other technical texts in other pairs of languages. This would help not only to improve the methodology but could also suggest ways to address other important issues in terminology compilation, such as writing definitions and creating conceptual representations using the corpus-driven approach.

References

Alves, I.M. (ed.). 1998. Glossário de termos neológicos da Economia. Cadernos de Terminologia, 3. São Paulo: Humanitas.

Azenha Jr, J. 1999. Tradução técnica e condicionantes culturais. Primeiros passos para um estudo integrado. São Paulo: Humanitas FFLCH/USP.

Bevilacqua, C.R. 2001. 'Unidades fraseológicas especializadas: novas perspectivas para sua identificação e tratamento' in M.G. Krieger and A.M.B. Maciel (eds) Temas de Terminologia, pp. 106–17. São Paulo/Porto Alegre: Humanitas/Editora da Universidade.

Bowker, L. 1996. Towards a Corpus-based Approach to Terminography. Terminology 3 (1), pp. 27–52.

Bowker, L. and J. Pearson. 2002. Working with Specialized Language. London and New York: Routledge.

Byrne, J. 2006. Technical Translation: Usability Strategies for Translating Technical Documentation. The Netherlands: Springer.

Carli, F. and E. Klotz. 2007. Dicionário gastronômico. São Paulo: Contorno Editora.

Colina, S. 1997. 'Contrastive rhetoric and text-typological conventions in translation teaching', Target 9 (2), pp. 335–53.

Estopá, R. 1999. Extració de terminologia: elements per la construció d'un SEACUSE (Sistema d'Extració Automàtica de Candidats de Significació Especializada). Ph.D. thesis. Barcelona, Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.

Farias, E.M.P. 2003. Glossário de Termos da Moda. Fortaleza: Edições da UFC/Edições SEBRAE.

Farias, E.M.P. and T.M.F. Bezerra. 2008. Glossário Trilíngue de Termos do Vestuário. Fortaleza: Edições UFC.

Fillmore, C.J. 1992. '"Corpus linguistics" or "computer-aided armchair linguistics"' in J. Svartvik (ed.) Directions in Corpus Linguistics, Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991, pp. 35–60. Berlin and New York: Mouton de Gruyter.

Finatto, M.J.B. 2007. 'Exploração terminológica com apoio informatizado: diálogos entre terminologia e lingüística de corpus' in M. Lorente, R. Estopà, J. Freixa, J. Martí and C. Tebé (ed.) Estudis de lingüística i de lingüística aplicada en honor de M. Teresa Cabré Castellví. Vol. 2: De deixebles. Barcelona: IULA.

Firth, J.R. 1957. Papers in Linguistics: 1934–1951. London: Oxford University Press.

Fromm, G. 2008. VoTec: a construção de vocabulários eletrônicos para aprendizes de tradução. Tese (Ph.D. thesis) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo. Accessed 1 June 2009, at: http://www.teses.usp.br/teses/disponiveis/8/8147/tde-08072008–150855/

Fuentes Morán, M.T. and J. García Palacios. 2002. 'Los diccionarios de especialidad y el traductor' in G. Guerrero Ramos and M.F. Pérez Lagos (eds) Panorama actual de la terminología, pp. 117–36. Málaga: Comares.

Gómez González-Jover, A. and Ch. Vargas Sierra. 2004. Aspectos metodológicos para la elaboración de diccionarios especializados bilingües destinados al traductor. El español, lengua de traducción. II Congreso Internacional pp. 365–98. Bruselas: ESLEtRA. Accessed 1 March 2012, at: http://cvc.cervantes.es/lengua/esletra/pdf/02/032_gomez-vargas.pdf

Hanks, P. 2008. 'The lexicographical legacy of John Sinclair', International Journal of Lexicography 21 (3), pp. 219–29.

Korning Zethsen, K. 1999. 'The dogmas of technical translation: are they still valid?', Hermes, Journal of Linguistics 23, pp. 65–75. Accessed 5 January 2008, at: http://hermes2.asb.dk/archive/FreeH/H23_05.pdf

Krieger, M.G., A.M.B. Maciel, C.R. Bevillaqua, M.J.B. Finatto and P.C.R. Reuillard. 2006. Glossário de Gestão Ambiental. São Paulo: Disal Editora.

Maia, B. 1997. 'Do-it-yourself corpora... with a little bit of help from your friends!' in B. Lewandowska-Tomaszczyk and P.J. Melia (eds) PALC '97 Practical Applications in Language Corpora. Łódź: Łódź University Press, pp. 403–10.

McEnery, A.M. and A. Wilson. 1996. Corpus Linguistics. Edinburgh: Edinburgh University Press.

Pearson, J. 1998. Terms in Context. Amsterdam: John Benjamins.

Perrotti-Garcia, A.J. and R.R. Rebechi. 2007. Vocabulário para Química. Série Mil e Um Termos. São Paulo: SBS.

Scott, M. 1999. WordSmith Tools 3.0. Oxford: Oxford University Press.

Salgado, A.R. 2006. Unidades Fraseológicas Especializadas na perspectiva da tradução. Dissertação. (MA thesis) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre.

Sinclair, J.M. 1987. 'The nature of the evidence' in J. Sinclair (ed.) Looking Up: An Account of the COBUILD Project in Lexical Computing, pp. 150–9. Collins ELT.

Tagnin, S.E.O. 2002. 'Os Corpora: Instrumentos de auto-ajuda para o tradutor' in S.E.O. Tagnin (ed.) Cadernos de Tradução no. 9, 2002/1. Florianópolis: Núcleo de Tradução da Universidade Federal de Santa Catarina.

Tagnin, S.E.O. 2007. 'A identificação de equivalentes tradutórios em corpora comparáveis' Anais do I Congresso Internacional da ABRAPUI: Belo Horizonte, 3 a 6 de junho de 2007. Accessed 4 June 2009, at: http://www.fflch.usp.br/dlm/comet/Novo/Stella_Abrapui%202007_artigo.pdf

Teixeira, E.D. 2004. Receita qualquer um traduz. Será? – a Culinária como área técnica de tradução. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2v. (MA dissertation.)

Teixeira, E.D. 2008. A lingüística de corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. Ph.D. thesis. Accessed 6 April 2010, at: www.teses.usp.br/teses/disponiveis/8/8147/tde-16022009–141747/

Teixeira, E.D. and S.E.O. Tagnin. 2008. Vocabulário para Culinária inglês/português. Série Mil e Um Termos. São Paulo: SBS.

Tognini-Bonelli, E. 2002. 'Functionally complete units of meaning across English and Italian: towards a corpus-driven approach' in B. Altenberg and S. Granger (eds) Lexis in Contrast: Corpus-based Approaches, pp. 73–95. Amsterdam: John Benjamins.

Varantola, K. 1998. 'Translators and their use of dictionaries. User needs and user habits' in B. T.S. Atkins (ed.) Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators, pp. 179–92. Tübingen: Niemeyer.

Varantola, K. 2002. 'Disposable corpora as intelligent tools in translation' in S.E.O. Tagnin (ed.) Cadernos de Tradução no. 9, 2002/1. Florianópolis: Núcleo de Tradução da Universidade Federal de Santa Catarina.

Wright, S.E. 1993. 'The inappropriateness of the merely correct: stylistic considerations in scientific and technical translation' in S.E. Wright and L.D. Wright Jr (eds) Scientific and Technical Translation, pp. 69–86. Amsterdan: John Benjamins.