Introducing COMPARA, the Portuguese-English parallel[1] corpus

Ana Frankenberg-Garcia (ISLA, Lisbon) & Diana Santos (SINTEF, Oslo)

This paper is an introduction to COMPARA. COMPARA is a machine-searchable and open-ended collection of Portuguese-English and English-Portuguese source texts and translations. It was made for people who have never used corpora before as well as for experienced corpus users. COMPARA's encoding and alignment criteria allow users to inspect translators' notes and investigate when and where translators have chosen to join, separate, delete, add and reorder sentences. Also, the corpus has been specifically designed to accommodate more than one translation per source text. COMPARA is freely accessible on the WWW.

Introduction

Modelling itself on the core structure of the English-Norwegian Parallel Corpus (Johansson et al. 1999), COMPARA is a machine-readable and searchable collection of source texts originally written in Portuguese and in English that have been aligned with their respective English and Portuguese translations. For the present, COMPARA contains only published fiction, but other genres are expected to be added to the corpus at a later stage. COMPARA is encoded according to the IMS Corpus Workbench system, developed at the University of Stuttgart (Christ et al. 1999), and can be searched online via the DISPARA interface, which has been developed under the broader framework of the Computational Processing of Portuguese project[2]. COMPARA has been conceived for people who are not necessarily corpus-literate as well as for experienced corpus users, and should thus be useful for arbitrarily complex research purposes.

Corpus Structure

COMPARA is an open-ended corpus. This means that it can grow in whichever direction proves to become important to users, and that the texts incorporated in the corpus can be put to use as soon as they are processed. Initially, COMPARA will not lend itself to studies that require large and representative language samples. However, as not all analyses depend on large corpora[3] and as there are no other publicly available corpora for the Portuguese-English language pair, it was felt that it was important to provide access to whatever was available as soon as it became available. An additional advantage of allowing users to access COMPARA at an early stage is that it is simpler to act on feedback from users when the corpus is small.

No a priori decisions were made regarding the kind of source texts and translations to be included in COMPARA, apart from the choice to limit the languages of the corpus to original English and English translated from Portuguese, and original Portuguese and Portuguese translated from English[4]. All varieties of Portuguese and English were considered, and no priority was given to any particular variety. The corpus was also structured so as to allow for the possibility of a source text being aligned with more than one translation.

Given this configuration, it is possible to use COMPARA to:

1. Study Portuguese translated from English and English translated from Portuguese.
2. Study Portuguese and English independently from one another.
3. Compare Portuguese and English source texts and translations from the six perspectives depicted in figure 1[5].
4. Compare different translations of the same source text, as examples A and B in figure 2 illustrate.
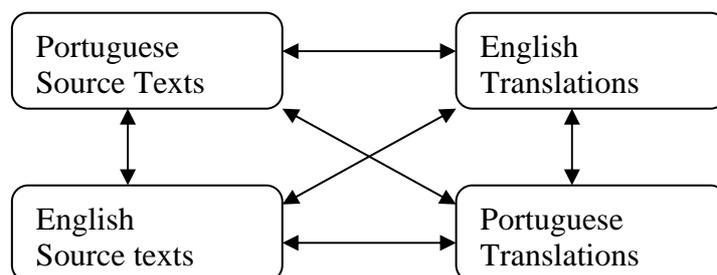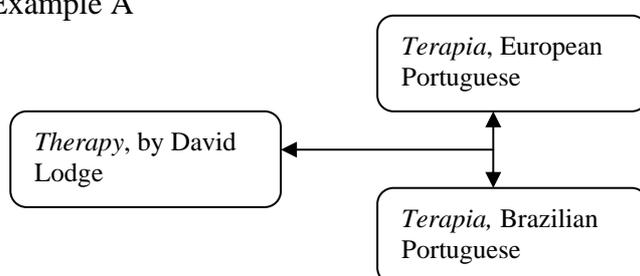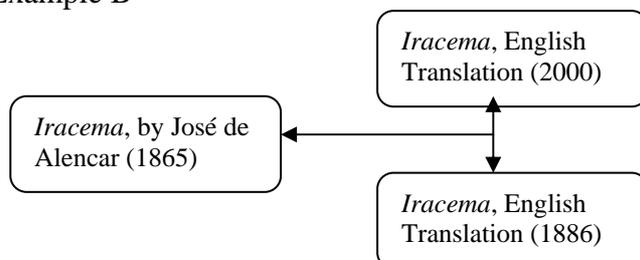
Figure 1

```
┌─────────────────┐            ┌─────────────────┐
│ Portuguese      │ ◄────────► │ English         │
│ Source Texts    │            │ Translations    │
└─────────────────┘ ╲        ╱ └─────────────────┘
        ▲            ╲      ╱            ▲
        │             ╲    ╱             │
        ▼              ╳                 ▼
┌─────────────────┐  ╱    ╲  ┌─────────────────┐
│ English         │ ◄────────► │ Portuguese      │
│ Source texts    │            │ Translations    │
└─────────────────┘            └─────────────────┘
```

Figure 2

Example A

```
                    ┌──────────────────────┐
                    │ Terapia, European    │
                    │ Portuguese           │
                    └──────────────────────┘
┌──────────────────────┐        ▲
│ Therapy, by David    │ ◄───────┤
│ Lodge                │         ▼
└──────────────────────┘ ┌──────────────────────┐
                         │ Terapia, Brazilian   │
                         │ Portuguese           │
                         └──────────────────────┘
```

Example B

```
                    ┌──────────────────────┐
                    │ Iracema, English     │
                    │ Translation (2000)   │
                    └──────────────────────┘
┌──────────────────────┐        ▲
│ Iracema, by José de  │ ◄───────┤
│ Alencar (1865)       │         ▼
└──────────────────────┘ ┌──────────────────────┐
                         │ Iracema, English     │
                         │ Translation (1886)   │
                         └──────────────────────┘
```

In example A, the source text can act as a common denominator for a study of similarities and differences between Brazilian and European Portuguese. In example B, the source text can act as a common denominator to a study of diachronic differences in translation.

A rather important corollary to this corpus structure is that conventional corpus design issues such as balance and representativeness are simply not addressed. The responsibility of achieving balance (if and when balance is an issue) and of judging the representativeness of the corpus for any given study is left in the hands of the user. To enable users to have control over this, it is possible for them to determine exactly which parts of COMPARA they wish to search.

The corpus can be automatically narrowed down so as to search only within specific varieties of Portuguese and English, and any language variety combination is possible. For example, users can search only Brazilian Portuguese and British English, or all varieties of Portuguese but only American English, and so on.

It is also possible to narrow down the corpus by date of publication. Users who wish to focus only on contemporary (or non-contemporary) language, for example, can automatically remove from their searches source texts and translations published before (or after) any particular date.

The third narrowing-down option available allows users to select exactly which texts they want to use in their searches, thus creating their own, tailor-made sub-corpora of COMPARA. For example, it is possible to search texts by only one particular author, group of authors, translator, and so on.

Eventually, when texts of different genres are included in the corpus, there will also be an option that allows users to select texts automatically by genre.

To help users assess the questions of balance and representativeness, access to information on the size of each separate text in the corpus is provided. The full reference of all the texts in the corpus plus extra-textual information on copyright and language variety are also provided.

Text selection
Although COMPARA is to remain open to different genres, the initial corpus contains only fiction. Obtaining copyright clearance for a corpus of fiction is notoriously difficult, and requesting permission to use the same texts in two different languages is twice as complicated. For the Portuguese-English language pair, however, there are several good reasons for taking fiction as a starting point.

To understand this, one should note that in the world there is far more English translated into Portuguese than Portuguese translated into English. When looking at how this reality applies to a few individual genres, it can be seen that:

- *Academic prose*. There are practically no translations of Portuguese into English. Portuguese-speaking intellectuals and scientists usually write and publish directly in either Portuguese or English, and it is only very rarely that academic books and articles originally written in Portuguese appear in English translation. Conversely, as most Portuguese-speaking academics are expected to be able to read in English, scientific articles in English translated into Portuguese are not very common.

- *Journalistic texts*. Although Portuguese-language magazines and newspapers often buy news articles written in English, the Portuguese version they publish is often an adaptation more than a translation[6]. Conversely, Portuguese-language newspaper articles in English translation are practically inexistent.

- *Instruction booklets.* There is far more English translated into Portuguese than Portuguese translated into English in the texts that accompany consumer goods. Also, these texts do not usually go through the editing and proofreading stages normally required of published texts[7], so their quality can be extremely variable. Finally, it is not always possible to determine whether Portuguese has been translated directly from English and English from

Portuguese: this type of translation is often relayed, with other languages standing in between.

- *Tourist brochures.* Tourist brochures in Portuguese translation are practically inexistent: Portuguese-speaking tourists abroad are expected to get by in other, more widely known languages. In contrast, almost all material destined to be read by tourists in Portuguese-speaking countries comes with an English translation. Quality is also an issue. Like instruction booklets, tourist brochures do not seem to be as carefully edited and proofread as published texts.

- *Fiction.* Although the amount of Portuguese fiction translated into English cannot be compared with the amount of English fiction in Portuguese translation, there are quite a few Portuguese-speaking fiction writers who have been translated into English[8]. A survey carried out in the very beginning of this project revealed that there are over 100 titles currently in print. Although there is no absolute guarantee of quality in the translation of fiction, these translations usually have to go through several stages of editing and proofreading, and are normally undertaken by qualified and experienced translators, not to mention the fact that the quality of the source texts is also relevant – while original fiction has to merit publication in the first place, this is not always true of certain tourist brochures, instruction manuals, etc.

This is obviously not the place for an extensive analysis of the present Portuguese-English translation reality for each individual genre. The brief outline given above, however, should serve to justify the choice for an initial corpus of fiction.

As COMPARA is publicly accessible on the Internet, we anticipated that it would be extremely hard to obtain permission to use fiction texts that are still in copyright. This factor overruled all possible arguments in favour of requesting permission to use integral texts[9]. Despite the choice for using extracts rather than entire texts, an attempt was made not to replicate design problems encountered in other corpora based on text extracts. The English-Norwegian Parallel Corpus, for example, used similar length extracts taken only from the beginnings of novels, and this was found to be a disadvantage by Santos and Oksefjell (1999) when attempting to validate corpus-based contrastive work. In COMPARA, extracts of around 30% of the total of each work were selected at random from their beginning, middle or end.

Alignment
Aligning source texts and translations is not simple, for translators do not always translate texts in a predictable and linear manner. Translators may (and often do):

a. split source-text sentences into two or more sentences in the translation
b. join two or more source-text sentences together, rendering them as a single translation sentence
c. leave things out
d. insert elements that were not present in the source text
e. reorder elements so that the order in which they appear in the translation differs from that in which they appear in the source text.

Existing parallel corpora which take the sentence as the basic unit for text alignment assume that sentences are not very often split, joined and reordered in translation (Véronis, 2000). They also do not pay too much heed to the direction of translation. Although these assumptions might be necessary if one is to automatically align large amounts of text, they also prevent one from using aligned corpora to study factors which are in themselves interesting to translation studies: i.e., what makes translators join, separate, add, remove or reorder sentences.

In COMPARA, we felt it was important to align texts directionally always from source to translation. This would not only open the way to a better understanding of translators' decisions, but would also facilitate the process of analysing multiple translations of the same source text. The basic unit of alignment in COMPARA was therefore defined as the source-text sentence, which means the direction of translation is intrinsically taken into account.

Whenever there is not a one-to-one sentence correspondence between source and translation, the sentences in the translation are split or joined together to conform to the way sentences were originally divided in the source text. Thus an alignment unit is always one orthographic sentence in the source text and the corresponding text in the translation, whether it is one, more than one, or even only part of a sentence. Source-text sentences that have been left out of the translation are aligned with blank units. Sentences that have been added to the translation with no corresponding text in the original are fitted into the nearest preceding alignment unit. Figure 3 summarizes these alignment criteria.

Figure 3

SOURCE		TRANSLATION
S	•	S
S	•	S,S
S	•	½ S
S	•	ø
S	•	S(+S)

Apart from the above, if there are any sentences that have been reordered in the translation, they are aligned with the sentences that prompted them in the source texts, and the original published order is marked so that sentence reordering, a non-trivial aspect of translation studies, can be automatically retrieved.

None of the existing automatic aligners seem to take the direction of translation into account. The aligner used in COMPARA – EasyAlign tool (v.1.0) written for use with the IMS Corpus Workbench – does not single out source text sentence divisions as being more important than translation text divisions. Whenever source and translation sentence divisions do not correspond, EasyAlign either matches one (source sentence) to many (translation sentences) - which is fine according to COMPARA's alignment criteria - or it matches many (source sentences) to one (translation sentence) - which is not what COMPARA's alignment requires. When one-to-many cases occur, automatic sentence separators for Portuguese and English are applied in order to tag those translation units appropriately. When many-to-one matches are detected, the source text sentences that have been kept together have to be manually separated and matched to the corresponding text in the translation.

Because COMPARA's alignment criteria requires that source-texts should always be divided in the same way no matter which translation they are aligned with, source texts can act as common denominators to multiple translations, as previously shown in figure 2. Also, because the alignment is directional, it is possible to search automatically for translational discourse changes such as where and when translators have decided to join, split, delete, add or reorder sentences[10].

Additional markup
Apart from alignment markup, COMPARA also marks highlighted text whenever it is used to set off titles, foreign words or within-sentence emphasis, thus facilitating the automatic retrieval of these features.

Categorizing highlighted text in this way, however, is not as straightforward and unambiguous as it seems. In Julian Barnes´s *Flaubert's Parrot*, for example, titles like *Un coeur simple* and *Madame Bovary* can be classified as both title and foreign. However, in the case of *Madame Bovary,* since both the English and Portuguese translations of this title are also *Madame Bovary,* and given that *Madame Bovary* is a proper name, it could be reasonably argued that it should be marked just title, but not foreign.

Marking foreign words *per se* is also not simple. The boundaries dividing what an author or translator (not to mention a corpus maker) considers or not to be foreign is by no means clearcut. For example, in Margaret Jull Costa's English translation of Sá-Carneiro's *Lúcio's Confession,* words like *coupé* and *décolletage* are not in italics but *manqué* and *passé* are.

There may also be a certain amount of difficulty in discriminating between text that has been highlighted because it is foreign and text that has been highlighted for emphasis. In David Lodge's *Therapy,* at one point it is hard to tell whether *Au contraire* has been italicised just because it is foreign or both because it is foreign and for emphasis.

For the sake of consistency, then, the criteria outlined in figure 4 have been adopted when inserting title, foreign and emphasis marks. According to these criteria:
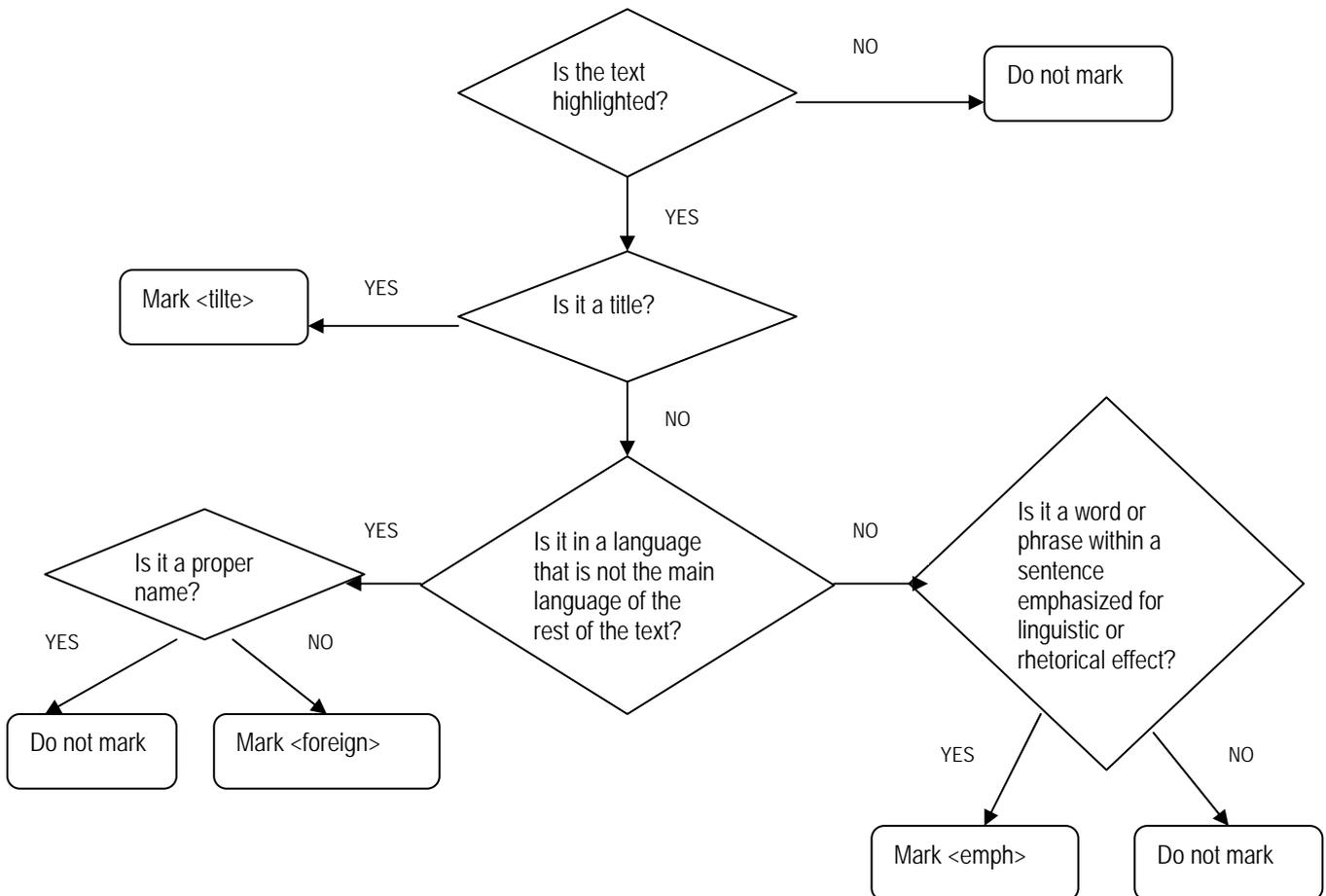
a. Titles and foreign and emphatic words and phrases that have not been highlighted by the author or the translator have not been marked in the corpus.

b. Highlighted text indicating features other than titles, foreign words and phrases, and within-sentence emphasis (like changes of voice, for example) have not been marked.

c. Whenever the categories for title, foreign and emphasis overlap, title overrules foreign, which in turn overrules emphasis. Thus emphasis is only marked if it is neither foreign nor title, and foreign is only marked if it is not title.

In addition to the above, translators' notes have also been marked and the body of the notes have been inserted in the text at the point which their identifiers appear. This allows users to study exactly when, where and why translators' notes were thought to be appropriate. The implementation of this particular feature involved the initial removal of the notes in order not to damage automatic alignment, and then putting them automatically back in place.

Obvious typographical errors that escaped proofreading in the editions from which the texts were extracted have been corrected and marked. However, because we assumed that the

retrieval of these errors was not important for translation studies, there is no corresponding query in the interface.

Figure 4

Is the text highlighted?

NO → Do not mark

YES

Is it a title?

YES → Mark <tilte>

NO

Is it in a language that is not the main language of the rest of the text?

YES → Is it a proper name?

YES → Do not mark

NO → Mark <foreign>

NO → Is it a word or phrase within a sentence emphasized for linguistic or rhetorical effect?

YES → Mark <emph>

NO → Do not mark

Accessibility
One of our major concerns in creating COMPARA was to make it widely accessible to anyone interested in comparing and contrasting English and Portuguese. The most logical way of providing widespread access to COMPARA seemed to be to enable the corpus to be searched via the Internet[11].

A particularly delicate matter for a corpus that is to be made available online is the copyright issue. In COMPARA, users only have access to the results of their search queries, and are not be able to read or copy the corpus excerpts in their entirety[12].

The groundwork for actually making COMPARA available online had been previously laid by the Computational Processing of Portuguese project, whose objective is to create, evaluate, catalogue and distribute Portuguese language computational resources. Thanks to their collaboration, providing free online access to COMPARA was mainly a matter of developing the DISPARA interface, which serves as a bridge between the IMS Corpus Workbench software and the specific requirements of COMPARA[13].

COMPARA can be accessed free of charge at http://www.portugues.mct.pt/COMPARA/. No registering or password is necessary. Every page in the COMPARA website is available in both Portuguese and English, so that people with very little Portuguese or very little English can still access them. Two search options are available. The Simple Search was made for people who have never used corpora before. It allows users to search the entire corpus either in the Portuguese-English or in the English-Portuguese direction. The Complex Search was made for those who want to conduct more sophisticated queries. It allows users to choose the search direction (an option to consider if the directionality of translation is relevant to a particular query), to narrow down the corpus (and so control which texts they are going to use if their queries require a balanced corpus or a specific subset or other of the corpus), and to select how the results are to be presented. Users can view concordances, distribution of forms (for queries involving more than one form, for example, *untrue/false*), distribution of sources (how a search expression is distributed in the texts within the corpus) and a quantitative wrap up (the distribution of the search expression in the two languages, for searches that involve alignment constraints - see below). When defining a query, the IMS Corpus Workbench syntax (Christ et al. 1999) can be used to include in a single query different spellings of a word (for example, *analyse* and *analyze*), different morphological variants of a word (for example, *walk*, *walked*, *walks*, etc.), a word and a collocate with any number of elements in between (for example *make* and *decision*), and so on. It is also possible to enter alignment constraints. For example, users searching for the Portuguese translation of *yes*, which is often rendered as *sim*, can retrieve just the cases in which *yes* is translated into *sim* or just the cases in which *yes* is translated into something other than *sim*[14].

In addition to this, users can inspect translators' notes, emphasis, foreign words and expressions, and titles. Finally, users can inspect alignment properties and see when and where translators have decided to join, separate, delete and add sentences to the translation. The possibility of looking at reordered sentences was not yet operational at the time this paper was written.

Output
Thinking of language learning and translation training in particular, it was important that the results should be presented in a way that would be easily transposable to the classroom. The concordances in COMPARA are displayed in two columns, with the Portuguese or English search item appearing in bold on the left-hand side, and the corresponding text in English or Portuguese on the right-hand side. Displaying the results in columns rather than in rows makes it easier for the user to compare different translations of the same search string and scroll up and down the results on screen.

Instead of a key-word-in-context (KWIC) concordance with a fixed number of characters to the left and to the right, the context is always one full source-text sentence and the corresponding text in the translation. Full source text sentences can help one understand the choices underlying a given translation better than a set number of characters to the left and to the right of the search string, although in some cases more co-text may be necessary. There are plans to allow the user to expand the amount of co-text given within the limits of fair-use, but this feature was not yet operational at the time this paper was written.

Next to each parallel concordance displayed, there is a link to the full reference of the pair of texts from where it was retrieved. This makes it easier for users to identify the source text and the translation in question, should additional, extra-textual information (like authorship, date of publication, language variety, and so on) be important to help one understand a translation.

Language learners can be trained to use COMPARA to look things up for themselves, and there are many ways the corpus can be used in the language or translation classroom. As proposed in Frankenberg-Garcia (2000; forthcoming), it is not unduly complicated to edit the results obtained in COMPARA so as to convert them into teaching materials. Appendix 1 contains a sample cloze exercise based on the output of COMPARA, in which students are required to give the Portuguese translation of *even,* a word susceptible of creating confusion among native speakers of Portuguese given its different meanings and translations. To prepare the exercise, a search for *even* was carried out in COMPARA, and the results were saved as an HTML file. The file was then opened from within a word processor and edited. The table menu was used to delete the row where corpus reference links appear (not necessary for the exercise), and the replace function was used to change the Portuguese translations of *even* (*até, mesmo, sequer* and *ainda*) into blank spaces.

Appendix 2 contains an extract of a worksheet prepared to help students understand different uses of negative prefixes in English and Portuguese. Based on the principle of data-driven learning (Johns 1991), students are asked to look at the concordances extracted from COMPARA and underline the Portuguese words that correspond to English words beginning with the negative prefix *un*. The exercise helps students realize that negative prefixes seem to be used much more sparingly in Portuguese, and that translators use different strategies to deal with them. Out of the 57 occurences of English words beginning with the negative prefix *un* contained in the total exercise, less than half were translated into a word containing a Portuguese negative prefix. Negative particles and root antonyms (for example, *false* instead of *untrue*) were frequently used as alternative translation strategies.

Current corpus and concluding remarks
COMPARA was first announced in January 2001. At the time this paper was written, COMPARA had permission to include extracts of 60 different Portuguese-English text-pairs by authors and translators from Angola, Brazil, Mozambique, Portugal, South Africa, the United Kingdom and the United States. These texts represent the combined product of the work of 33 authors and 31 translators[15]. The part of the corpus available for research presently adds up to around 190 K words in each language. In less than six months, COMPARA has had over 3000 visits from more than 400 different computers. It is hoped that feedback from users will contribute towards the development of COMPARA, and that this might take place alongside a growing interest in the use of corpora for research and education.

Notes

[1] Parallel is being used here to refer to a bilingual collection of source texts and their translations. In the contrastive linguistics tradition, this would have been referred to as a translation corpus.
[2] For further information on the Computational Processing of Portuguese project, see http://www.portugues.mct.pt.
[3] For Biber et al.(1998), different analyses require different sized corpora, and the optimum size of a corpus for an analysis to be reliable is something that can be investigated empirically. It has been shown that lexicographic studies, for example, require larger corpora than studies investigating grammar alone.
[4] Following Baker (1998), it is believed that there are distinctive differences between texts written originally in one language and texts translated into that same language.
[5] As in the English-Norwegian Parallel Corpus.
[6] R.Garcia, journalist at the Portuguese daily PUBLICO, personal communication. Pearson (2000) reached a similar conclusion after analysing French translations of *Scientific American.*
[7] For Biber (1993), "published" means not only printed in multiple copies, but also copyright registered or recorded by a major indexing service.

[8] Similarly, there is more English fiction translated into Norwegian than Norwegian fiction translated into English (Johansson, S. , Ebeling, J. and Oskefell, S 1999).

[9] See Baker (1995), for example.

[10] Ideally, it would have been desirable to retrieve the addition, deletion and reordering of clauses, for, as Malmkjaer (1997) put it, it is at the level of the clause that translation sense for sense and translation structure for structure usually meet. Empirical evidence of this for the Portuguese-English pair can be found in Santos (1994). Clause alignment was not an option, however, because of the extreme difficulty in detecting clause boundaries automatically.

[11] Several other advantages of making corpora available online are outlined in Santos (1998).

[12] The maximum number of concordances shown is 500. If the user chooses to search only a subset of the corpus, then this number is further limited to 200. Whenever the results exceed these limits, the concordances shown are selected at random. Even if for certain queries it is not possible to *show* all the concordances, the total number of solutions found is always provided, so that the user can at least have a quantitative measure of the results.

[13] DISPARA is a general system for DIStributing PARAlell corpora on the Web. Although the system was conceived to cater for the specific needs of COMPARA, it can also be very easily adapted to other parallel corpora encoded according to the IMS Corpus Workbench.

[14] There is no underlying word alignment. The search is based on finding out whether or not *sim* is present in the Portuguese alignment unit corresponding to the English alignment unit containing *yes*.

[15] For a full, regularly updated list of the texts in COMPARA, see:
http://www.portugues.mct.pt/COMPARA/CorpusContents.html

References

Baker, Mona (1995) "Corpora in translation studies: an overview and some suggestions for future research" *Target* 7, 2, pp 223-243.

Baker, Mona (1998) "Reexplorer la langue de la traduction: une approche par corpus" in Laviosa, S. (ed.) *The corpus-based approach: a new paradigm in translation studies*. Special Issue of META, XLIII, 4.

Biber, Douglas (1993) "Representativeness in corpus design", in *Literary and Linguistic Computing,* 8, pp 243-257.

------, S. Conrad & R. Reppen (1998) *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Christ, Oliver, B. Schulze, A. Hofmann & E. Koenig (1999) "The IMS Corpus Workbench: Corpus Query Processor (CQP): User's Manual", Institute for Natural Language Processing, University of Stuttgart, March 8, 1999 (CQP V2.2).

Frankenberg-Garcia, Ana (2000) "Using a translation corpus to teach English to native speakers of Portuguese" in *Op.Cit.* Vol. 3, 65-78.

------ (*forthcoming)* "COMPARA, language leraning and translation training", in *Proceedings of Training the Language Services Provider for the New Millennium.* Porto, 25-26 May 2001.

IMS Corpus Workbench. Available at http://www.ims.uni-stutgart.de/CorpusWorkbench/

Johansson, Stig, J. Ebeling & S. Oksefjell (1999) English-Norwegian Parallel Corpus: Manual http://www.hf.uio.no/iba/prosjekt/ENPCmanual.html [Access Date 7/7/2000]

Johns, Tim (1991) "Should you be persuaded: two examples of data-driven learning" in *ELR Journal* Vol.4, 1-16.

Malmkjaer, Kirsten (1997) "Unit of translation" in Baker, M. (ed.)(1997) *The Routledge Encyclopedia of Translation Studies*. London: Routledge, pp 286-288.

Pearson, Jennifer (2000) "Translating popular science: implications for translator training" Paper presented at CULT2K, the Second International conference on Corpus Use and Learning to Translate. Bertinoro, 3-4 November 2000.

Santos, Diana. (1994) "Bilingual alignment and tense" in *Proceedings of the Second Annual Workshop on Very Large Corpora* (Kyoto, 4 August 1994), ACL, pp.129-141.

------ (1998) "Providing access to language resources through the World Wide Web: the Oslo Corpus of Bosnian Texts". In Antonio Rubio, Natividad Gallardo, Rosa Castro and Antonio Tejada (eds.), *Proceedings of The First International Conference on Language Resources and Evaluation* (Granada, 28-30 May 1998), Vol. 1, pp.475-481.

------ & S. Oksefjell (1999) "Using a parallel corpus to validate independent claims", *Languages in Contrast*, Vol. 2(1):117-132.

Véronis, Jean (ed) (2000) *Parallel Text Processing*, Dordrecht: Kluwer Academic Publishers.

Exercise based on output of *COMPARA http://www.portugues.mct.pt/COMPARA/ [15-May-2001]*

Fill in the gaps with an appropriate Portuguese translation for *even*:

I had the ideas; I **even** made notes.	Tinha as ideias; _____ coligi notas.
They had planned the trip in detail, had their hair specially curled for the occasion, and had **even** stolen flowers for the girls.	Tinham planeado a visita em pormenor, tinham ondulado especialmente o cabelo para a ocasião, e _____ tinham roubado flores para as raparigas.
Memories came out of hiding, but not emotions; not **even** the memories of emotions.	Surgiam as recordações, mas não as emoções; nem _____ recordações de emoções.
The other rooms contained medical instruments of the eighteenth and nineteenth centuries: heavy metal relics coming to sharp points, and enema pumps of a calibre which surprised **even** me.	As outras salas tinham instrumentos médicos dos séculos XVIII e XIX: pesadas relíquias de metal que terminavam em pontas agudas e irrigadores de um calibre que _____ a mim me surpreendia.
She keeps the adored relic beside her, and **even** takes to saying her prayers while kneeling before him.	Guarda junto de si a relíquia adorada e começa _____ a dizer as suas orações ajoelhada na sua frente.
A cheeky bird, inducing affection, **even** reverence.	Um pássaro atrevido, que suscitava afecto, respeito _____.
All that remains of Flaubert's residence is a small one-storey pavilion a few hundred yards down the road: a summer house to which the writer would retire when needing **even** more solitude than usual.	Tudo o que ficou da residência de Flaubert é um pequeno pavilhão a poucas centenas de metros da estrada: uma casa de Verão para onde o autor se retirava quando precisava de _____ maior solidão do que a habitual.
Then I realised the fallacy in this: Flaubert, after all, hadn't been given a choice of parrots; and **even** this second one, which looked the calmer company, might well get on your nerves after a couple of weeks.	Depois descobri a ironia disto: Flaubert, apesar de tudo, não tinha podido escolher o papagaio; e _____ o segundo, que parecia uma companhia mais calma, podia muito bem tornar-se irritante depois de umas semanas.

Exercise based on output of *COMPARA http://www.portugues.mct.pt/COMPARA/ [10-May-2001]*

Read the extracts below and underline the Portuguese word or words that correspond to English words beginning with the negative prefix *un*:

The thrower remained a stylish, temporary statue: knees not quite **unbent**, and the right hand ecstatically spread.	O jogador ficou como uma estilizada estátua temporária: os joelhos um pouco dobrados e a mão direita erguida e estática.
Let me start with the statue: the one above, the permanent, **unstylish** one, the one crying cupreous tears, the floppy-tied, square waistcoated, baggy-trousered, straggle-moustached, wary, aloof bequeathed image of the man.	Vou começar pela estátua: a de cima, a permanente, a sem estilo, a que chora lágrimas de cobre, a que lega à posteridade a imagem circunspecta de um homem com um laço desajeitado, colete quadrado, calças largas como sacos, bigode em desalinho.
If so, then how tantalising are the **unfinished** books.	Se assim é, então que excitantes são os livros inacabados.
The **unwritten** books?	Os livros que não se escreveram?
Dot, dash, dash, dash went: the concrete caissons, with the **unhurried** water between them.	Ponto, traço, traço, traço, faziam as caixas de cimento, separadas umas das outras pela água calma.
I was close to where friends had died - the sudden friends those years produced - and yet I felt **unmoved**.	Ali perto amigos meus tinham morrido - os amigos inesperados que esses anos nos dão - mas não me sentia comovido.
But here, in this **unexceptional** green parrot, preserved in a routine yet mysterious fashion, was something which made me feel I had almost known the writer.	Mas aqui, neste vulgar papagaio verde, perservado de uma maneira vulgar e no entanto misteriosa, havia algo que me fazia sentir que quase tinha conhecido o escritor.
It's about a poor, **uneducated** servant-woman called Félicité, who serves the same mistress for half a century, unresentfully sacrificing her own life to those of others.	É acerca de uma pobre criada ignorante chamada Félicité, que serve a mesma patroa durante meio século, sacrificando sem ressentimentos a sua vida à dos outros.