COMPARA
Sentence alignment revision and markup

Ana Frankenberg-Garcia, Diana Santos and Rosário Silva - 18/04/2006

Note: COMPARA is a project developed under the scope of Linguateca, supported by grant POSI/PLP/43931/2001 from Fundação para a Ciência e Tecnologia (Portugal), co-financed by POSI. Ana Frankenberg-Garcia and Diana Santos are joint project leaders of COMPARA. For further information, please see http://www.linguateca.pt/COMPARA.

1. Criteria for text alignment

The basic unit of alignment in COMPARA is the source-text sentence. A sentence is defined as a word or sequence of words beginning with a capital letter and ending with a full-stop, ellipsis, exclamation mark or question mark, followed by a new sequence of words beginning with a capital letter, or by no text at all in the case of the end of a paragraph. The paragraph below illustrates the sentence separation criteria adopted (sentence beginnings are marked <s>) :

```
EURZ1 (five sentences)
<s>«You shouldn't listen to me,» Simon suddenly sighs. <s>«I'm an
old fool who no longer has any courage. <s>But for Master Abraham's
sake I will try to face the truth, if you like. <s>Now tell me, you
believe he was murdered by someone who knew him... a New Christian?»
<s>His questioning eyes seem almost hopeful, as if death by a Jew's
hand is preferable to Uncle having been murdered by a follower of
the Nazarene.
```

In cases of direct speech followed or preceded by reporting verbs (such as *say*, *tell*, *whisper*, *suggest* etc.), there can be words beginning with capital letters after the punctuation marks mentioned above without any resulting sentence separation. For example:

```
EBJT1 (one sentence)
<s>`You OK?´ Robin's daughter said, standing close to him, but not
touching.
```

Note that when direct speech is not followed or preceded by reporting verbs, sentence separation is maintained. In the example below, a new sentence begins after the second question mark because *realise* is not a reporting verb:

```
PBCB2 (three sentences)
<s>Then asks `What happened to Osbenio? <s>And to Clauir?´ <s> I
realise he was expecting someone else, a relative, someone or other.
```

The colon is only considered a sentence separator if it marks the end of a paragraph:

```
EBDL3 (two sentences)
<s>From long practice Philip was able to follow his drift pretty
well, and therefore answered confidently:
<s>«Oh, no, I couldn't leave Hilary behind to cope on her own.
```

If there is no end of paragraph, there is no sentence separation, no matter whether or not the word after the colon begins with a capital letter:

```
EBJB1 (one sentence)
<s>Flaubert wanted them to be: few writers believed more in the
objectivity of the written text and the insignificance of the
writer's personality; yet still we disobediently pursue.
```

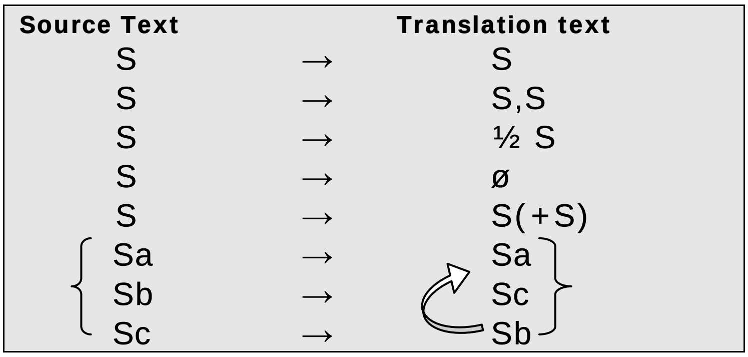
```
EUHJ1 (one sentence)
<s>But she did not commit herself, and in a moment she asked: «Now
that he has come back, will he stay here always?»
```

Source-text sentences are sometimes divided into two or more sentences in the process of translation. Translators may also join source-text sentences together, rendering them as a single translation sentence, or they may leave things out and insert elements that were not present in the source text. In addition to this, translators sometimes reorder elements so that the order in which they appear in the translation differs from that in which they appear in the source text.

In COMPARA, whenever there is not a one-to-one sentence correspondence between source-text and translation sentences, it is the translation sentences that are split or joined up to conform to the way sentences were originally divided in the source text. Thus an alignment unit is always one orthographic sentence[1] 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.[2] Sentences that have been reordered in the translation are "unreordered" in the alignment. So if source-text sentences ABC are translated into ACB, in the alignment procedure the translation sentences ACB change places so that they become ABC again. Figure 1 summarizes these alignment criteria.

Figure 1. COMPARA criteria for text alignment



Source Text		Translation text
S	→	S
S	→	S,S
S	→	½ S
S	→	ø
S	→	S(+S)
Sa	→	Sa
Sb	→	Sc
Sc	→	Sb

2. Editing sentence alignment

After you've aligned a pair of texts by paragraph, the texts are automatically aligned at the level of the sentence by EasyAlign. The *par* endings of the files that have been submitted to this procedure are renamed *easy* at this phase. (e.g. PBMA3easy.pt & PBMA3easy.en).

The automatic alignment obtained through Easyalign does not follow COMPARA's alignment criteria to the letter, and your job is to edit the results so that they match the criteria in figure 1. Here is how:

2.1 Preparing

a. Open source text "easy" file in Word.

b. Open translation "easy" file in Word.

Note that the divisions made by EasyAlign are marked <s> and that each <s> segment begins on a new line. If the sentence also represents the beginning of a paragraph, then it is marked <s><p>.

[1] Except in texts by José Saramago, as explained further on in 2.3.3.

[2] Except when the added sentence is the first one of a paragraph, in which case the extra sentence should come before.

c. Go to Window and select Arrange all. You should be able to see both source and translation, one on top of the other.

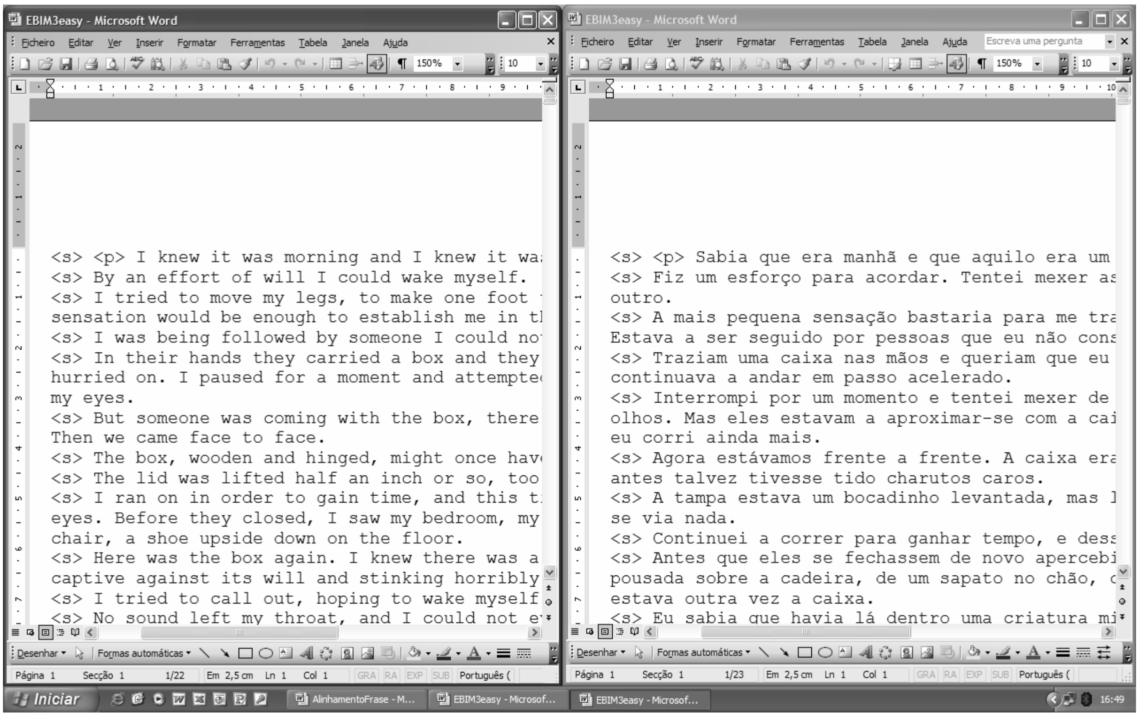
d. Drag the right margin of the source text to the centre of the screen.

e. Drag the left margin of the translation to the centre of the screen.

f. Drag the top margin of the bottom text to the top, so that it occupies half the screen.

g. Drag the bottom margin of the top text to the bottom, so that it occupies the other half of the screen. You should now be able to see the source text occupying the whole left-hand side of your screen and the translation occupying the whole right-hand side, as shown in fugure 2.

Figura 2. Preparing your computer screen for alignment revision



```
<s> <p> I knew it was morning and I knew it wa
<s> By an effort of will I could wake myself.
<s> I tried to move my legs, to make one foot
sensation would be enough to establish me in t
<s> I was being followed by someone I could no
<s> In their hands they carried a box and they
hurried on. I paused for a moment and attempte
my eyes.
<s> But someone was coming with the box, there
Then we came face to face.
<s> The box, wooden and hinged, might once hav
<s> The lid was lifted half an inch or so, too
<s> I ran on in order to gain time, and this t
eyes. Before they closed, I saw my bedroom, my
chair, a shoe upside down on the floor.
<s> Here was the box again. I knew there was a
captive against its will and stinking horribly
<s> I tried to call out, hoping to wake myself
<s> No sound left my throat, and I could not e
```

Página 1 Secção 1 1/22 Em 2,5 cm Ln 1 Col 1 Português (

```
<s> <p> Sabia que era manhã e que aquilo era um
<s> Fiz um esforço para acordar. Tentei mexer as
outro.
<s> A mais pequena sensação bastaria para me tra
Estava a ser seguido por pessoas que eu não cons
<s> Traziam uma caixa nas mãos e queriam que eu
continuava a andar em passo acelerado.
<s> Interrompi por um momento e tentei mexer de
olhos. Mas eles estavam a aproximar-se com a cai
eu corri ainda mais.
<s> Agora estávamos frente a frente. A caixa era
antes talvez tivesse tido charutos caros.
<s> A tampa estava um bocadinho levantada, mas l
se via nada.
<s> Continuei a correr para ganhar tempo, e dess
<s> Antes que eles se fechassem de novo apercebi
pousada sobre a cadeira, de um sapato no chão, c
estava outra vez a caixa.
<s> Eu sabia que havia lá dentro uma criatura mi
```

Página 1 Secção 1 1/23 Em 2,5 cm Ln 1 Col 1 Português (

h. Click your mouse on the source text and Select all (Control A), change font to Times New Roman, and change font size to 10.

i. Go to View and select Normal.

j. Go to File and Page setup. Set left margin at 0.5 cm and right margin at 13.5 cm and apply to Whole document. The source text should now be thin enough so that you can see a full line on only the left half the screen.

k. Select all (control A) again. Go to Format, Bullets and numbering and then Numbered. Select the “1. 2. 3.” format. All the sentences in the source text should now be numbered.

l. Save the source text in Word format (at this stage, it is important not to save it in text format) and then click the mouse on the translation text.

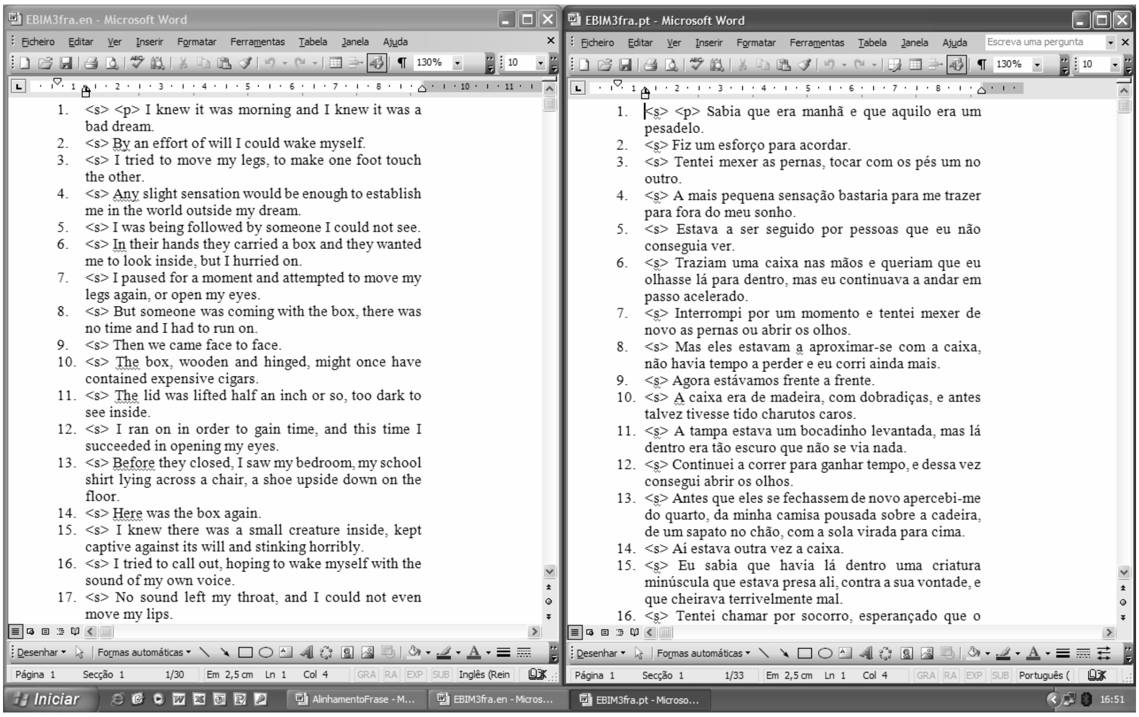
m. Select all (control A), change font to Times New Roman, and change font size to 10.

n. Go to View and select Normal.

o. Go to File and Page setup. Set left margin at 13.5 cm and right margin at 0.5 cm and apply to Whole document. The translation text should now be thin enough so that you can see a full line on only the right half the screen.

p. Select all (control A) again. Go to Format, Bullets and numbering and then Numbered. Select the “1. 2. 3.” format. All the sentences in the translation should now be numbered too as shown in figure 3. Save the file in Word format (again, remember not to save it in text format at this phase).

Figure 3. Numbered texts for alignment revision


EBIM3fra.en - Microsoft Word
Ficheiro Editar Ver Inserir Formatar Ferramentas Tabela Janela Ajuda
130%
10
Página 1 Secção 1 1/30 Em 2,5 cm Ln 1 Col 4 GRA RA EXP SUB Inglês (Rein
EBIM3fra.pt - Microsoft Word
Ficheiro Editar Ver Inserir Formatar Ferramentas Tabela Janela Ajuda
Escreva uma pergunta
130%
10
Desenhar
Formas automáticas
Página 1 Secção 1 1/33 Em 2,5 cm Ln 1 Col 4 GRA RA EXP SUB Português (
Iniciar
AlinhamentoFrase - M...
EBIM3fra.en - Micros...
EBIM3fra.pt - Microso...
16:51

1. <s> <p> I knew it was morning and I knew it was a bad dream.
2. <s> By an effort of will I could wake myself.
3. <s> I tried to move my legs, to make one foot touch the other.
4. <s> Any slight sensation would be enough to establish me in the world outside my dream.
5. <s> I was being followed by someone I could not see.
6. <s> In their hands they carried a box and they wanted me to look inside, but I hurried on.
7. <s> I paused for a moment and attempted to move my legs again, or open my eyes.
8. <s> But someone was coming with the box, there was no time and I had to run on.
9. <s> Then we came face to face.
10. <s> The box, wooden and hinged, might once have contained expensive cigars.
11. <s> The lid was lifted half an inch or so, too dark to see inside.
12. <s> I ran on in order to gain time, and this time I succeeded in opening my eyes.
13. <s> Before they closed, I saw my bedroom, my school shirt lying across a chair, a shoe upside down on the floor.
14. <s> Here was the box again.
15. <s> I knew there was a small creature inside, kept captive against its will and stinking horribly.
16. <s> I tried to call out, hoping to wake myself with the sound of my own voice.
17. <s> No sound left my throat, and I could not even move my lips.

1. <s> <p> Sabia que era manhã e que aquilo era um pesadelo.
2. <s> Fiz um esforço para acordar.
3. <s> Tentei mexer as pernas, tocar com os pés um no outro.
4. <s> A mais pequena sensação bastaria para me trazer para fora do meu sonho.
5. <s> Estava a ser seguido por pessoas que eu não conseguia ver.
6. <s> Traziam uma caixa nas mãos e queriam que eu olhasse lá para dentro, mas eu continuava a andar em passo acelerado.
7. <s> Interrompi por um momento e tentei mexer de novo as pernas ou abrir os olhos.
8. <s> Mas eles estavam a aproximar-se com a caixa, não havia tempo a perder e eu corri ainda mais.
9. <s> Agora estávamos frente a frente.
10. <s> A caixa era de madeira, com dobradiças, e antes talvez tivesse tido charutos caros.
11. <s> A tampa estava um bocadinho levantada, mas lá dentro era tão escuro que não se via nada.
12. <s> Continuei a correr para ganhar tempo, e dessa vez consegui abrir os olhos.
13. <s> Antes que eles se fechassem de novo apercebi-me do quarto, da minha camisa pousada sobre a cadeira, de um sapato no chão, com a sola virada para cima.
14. <s> Aí estava outra vez a caixa.
15. <s> Eu sabia que havia lá dentro uma criatura minúscula que estava presa ali, contra a sua vontade, e que cheirava terrivelmente mal.
16. <s> Tentei chamar por socorro, esperançado que o

2.2 Basic sentence alignment editing

According to COMPARA’s alignment criteria, each <s> or <s><p> segment of the source text should consist of just one sentence. Therefore:

2.2.1 Source text alignment

Detect all the <s> or <s><p> segments in the source text that contain more than one sentence and all the ones which are empty. For example:

```
1. <s> I was bewildered, utterly lost in amazement -- but I could
not forget the peculiar whine of my Newfoundland dog Tiger, and the
odd manner of his caresses I well knew. It was he.
2. <s>
3. <s> I experienced a sudden rush of blood to my temples -- a giddy
and overpowering sense of deliverance and reanimation.
```

To detect empty <s> or <s><p> segments or ones which have more than one sentence in the source text in a semi-automatic way, without having to read through the text all over again, press *Control F* and search first for all fullstops (.), then all question marks (?), and then all exclamation marks (!).

> Note that the colon (:) is *not* considered to be a sentence separator, unless it is followed by a paragraph break. This means you do not have to worry about them now, because they have already been dealt with in the paragraph alignment phase.

Whenever you find segments with more than one sentence in the source text, press *Enter* to separate them and insert a new <s> mark at the beginning of the sentence you have separated. Word will renumber the new <s> segments accordingly. For example:

Change

```
1. <s> Did you know about the time he had the ass keel-hauled? Is
that in your archives?
```

Into

```
1. <s> Did you know about the time he had the ass keel-hauled?
2. <s> Is that in your archives?
```

> Don’t worry about <s><p> tags, for all necessary <p> for paragraph marks have been dealt with in the paragraph alignment phase

Whenever you find an empty segments, remove the empty line and the <s> or <s><p> mark completely with the *Backspace* key (Word will renumber the remaining source text segments accordingly). For example:

Change

```
1. <s> It was he.
2. <s>
3. <s> I experienced a sudden rush of blood to my temples -- a giddy
and overpowering sense of deliverance and reanimation.
```

Into

```
1. <s> It was he.
2. <s> I experienced a sudden rush of blood to my temples -- a giddy
and overpowering sense of deliverance and reanimation.
```

Once you have finished the source text alignment revision, save the file and don’t change anything else in it.

2.2.2 Translation alignment revision

Scroll down the translation text side by side with the source text. You will notice the text and the numbered segments of the translation do not always match the ones in the source text. In this part of the aligment revision, you have to make them match again, changing only the translation text.

When you come across unaligned segments, you are to join or separate any extra or missing segments in the translation in order to realign them with those of the source text.

To join <s> segments of the translation, press *Backspace* to erase line breaks until the numbers of the segments match the ones on the source text. The new translation segment should now contain more than one sentence and more than one <s> mark. Leave only the first <s> or <s><p> mark (the one immediately next to the number) and delete the other(s).

When joining <s> segments in the translation, you may notice that the translator added one or more than one whole sentence to the translated text without there being any equivalent text in the original source text. If this is the case, surround the additional sentence(s) with <add> in the beginning and </add> at the end. For example:

```
1. <s> What would God            1. <s> Que pensará Deus
think?                           disto? <add>Que pensará
                                 Deus daquilo? Que pensará
                                 Deus de tudo?</add>
```

In order to split <s> segments of the translation into two (or three, etc.), press *Enter* so that you split a translation sentence into two (or three, etc) and insert a new <s> mark at the beginning of each new segment. If you split a sentence into two, change the <s> mark of each half into <s2>. If you split a sentence in three, change the <s> mark at the beginning of each third into <s3>, etc. You must be careful to make sure you split the translation segment at the point of closest correspondence to the source text[3]. For example:

```
88. <s> - Pois devias             88. <s2> " Well, you
rir, meu querido.                 should be laughing, my
                                  dear fellow,
89. <s> Porque a                  89. <s2> because
imortalidade é o meu lote         immortality is my lot or
ou o meu dote, ou como            my spot or whatever name
melhor nome haja.                 you can come up with for
                                  it.
```

When dividing <s> segments in the translation, you may find the translator has left out a whole sentence. If this happens, mark the translation segment just with <s>, leaving the rest blank. For example:

```
1. <s> A cara                     1. <s> Zito's face was
impenetrável, os olhos            inscrutable, his eyes
não diziam nada.                  said nothing.
2. <s> Não estava mais            2. <s>
ali quem falou.
3. <s> Ele agora atendia          3. <s> Now he was serving
uma freguesa que queria           a customer who wanted
três metros de morim.             three metres of cambric.
```

Correctly aligned source text and translation text segments (i.e., one sentence in the source text and the corresponding text in the translation both of which identified by the same number) are called *alignment units*.

2.2.3 Authors' and translators' notes alignment

Authors' notes <anote> should be placed in separate alignment units. If the note is made up of more than one sentence, each sentence must be on a separate alignment unit.

Translators' notes <tnote> should remain inside the alignment unit where the note belongs (immediately after the note mark) even when the note contains more than one sentence.

[3] It is not always possible to achieve 100% correspondence.

2.3 Complicated cases of sentence alignment

2.3.1 Alignment with whole sentences and fractions

If you come across a source text sentence that matches 1½ translation sentences, align them normally, as described earlier. The only thing that changes are the alignment tags used on the translation text. Normally, <s> is used for full sentences (no matter how many) and <s2> for half a sentence, <s3> for one third of a sentence, an so on. When there is a combination of a whole sentence with a fraction of a sentence, our programs cannot make automatic counts, so we must use explicit tags:

a. Instead of identifying half sentences with <s2>, use <s1/2>. Similarly, instead of using <s3> for a third of a sentence, use <s1/3>, an so on.

b. Instead of identifying whole sentences just with <s>, count exactly how many whole sentences there are and write it down: <s1> = one whole sentence, <s2> = two whole sentences, etc.

c. Use a plus sign (+) to indicate that there is a comination of a whole sentence with a fraction. Thus <s1/2+1> = one and a half sentences, <s1/3+2> = two and a third sentences, <s1/2+3> = three and a half sentences, etc.

d. The fraction must always come before the whole number within the tag, even if in the actual text the whole sentence appears first. For example:

```
347. <s> Augustus called          347. <s1/2+1> Augusto
to me at first in a low           chamou-me. Primeiro, em
voice and without closing         voz baixa e sem fechar o
the trap -- but I made            alçapão, mas eu não dei
him no reply.                     qualquer resposta;
348. <s> He then shut the         348. <s2> fechou então o
trap, and spoke to me in          alçapão e falou-me num
a louder, and finally in          tom de voz mais alto;
a very loud tone -- still         depois aos berros, mas eu
I continued to snore.             continuava a ressonar.
```

So in the above example, alignment unit 347 containis one full source text sentence and 1½ translation sentences; alignment unit 348 contains one full source text sentence and ½ a translation sentence.

Note that this rule also applies when the alignment unit includes sentences that have been added to the translation. For example:

```
611. <s> Parece até que           611. <s1/2+1> there were
chorou.                           even tears...» <add>«Who
                                  cried?</add>
```

Alignment unit 611 contains one full source-text sentence and 1½ translation sentences.

2.3.2. Reorderings

You may also come across source-text sentences that have been reordered in the translation. In the alignment you have to change the place of the corresponding text in the translation (which can be a full sentence or half a sentence, etc.) so that it follows the order of the source text. Whenever you do that, you have to surround the parts you are moving with <reord 1>bla, bla,

bla </reord>. After you've done that, you have to insert <place 1> in the place where the bit you’ve moved used to be, i.e., the place of the published order. If there is a second reordering in the translation, you should mark it <reord 2> bla, bla, bla </reord>, and then mark the place where the translator originally put the bit you’ve moved with <place 2>. If there is a third one, use <reord 3> bla, bla, bla </reord> and <place 3>, and so on, until the end of the translation text.

In the example that follows, the translator split the first source-text sentence into two, and moved the second half of it to the sentence immediately after: *But there was still another and very different source of disquietude, and one, indeed, whose harassing terrors had been the chief means of arousing me to exertion from my stupor on the mattress. It arose from the demeanor of the dog. → Mas, havia ainda uma outra razão completamente diferente para me inquietar. Esta inquietação, em consequência da qual os extenuantes terrores me tinham arrancado ao torpor e me haviam obrigado a soerguer-me do colchão, provinha do comportamento do cão.* This should be marked as follows:

286. <s> But there was still another and very different source of disquietude, and one, indeed, whose harassing terrors had been the chief means of arousing me to exertion from my stupor on the mattress.	286. <s1/2+1> Mas, havia ainda uma outra razão completamente diferente para me inquietar. <reord 1> em consequência da qual os extenuantes terrores me tinham arrancado ao torpor e me haviam obrigado a soerguer-me do colchão, </reord>
287. <s> It arose from the demeanor of the dog.	287. <s2> Esta inquietação, <place 1> provinha do comportamento do cão.

You should only mark <reord> when translators change the order of whole sentences or of sentences that they themselves decided to split. In other words, you needn’t bother with <reord> when translators change the order of words or of clauses within the same sentence.

It is important that <reord> and <place> tags always appear in this order, first <reord>, and then <place>, no matter in what direction you've interpreted the reordering. This means that the following encoding is wrong because the <place> tag was inserted before the <reord> tag:

14. <s> There was a pause, and then Judy said, 'Mum liked it, though. '	14. <s2> Houve uma pausa e depois Judy disse <place 1> : -- No entanto, a mãe gostava disto.
15. <s> Her voice shook.	15. <s2> <reord 1> em voz trémula </reord>

The right way to do it, or rather, the way to make it work in our system, is:

<s> There was a pause, and then Judy said, 'Mum liked it, though. '	14. <s2> Houve uma pausa e depois Judy disse <reord 1> : -- No entanto, a mãe gostava disto. </reord>
15. <s> Her voice shook.	15. <s2> em voz trémula <place 1>

Note that <reord> and <place> tags needn’t be in adjacent alignment units as long as the number that identifies them is the same. For example, take *Robin said sharply, 'Caro was Judy's mother. And my wife. Not Joe's.' → -- A Caro era mãe de Judy. E minha mulher. Não era nada ao Joe -- disse Robin com brusquidão.* This should be encoded as follows:

```
<s> Robin said sharply, 'Caro              72. <s1/2+1> -- A Caro era mãe
was Judy's mother.                         de Judy. <reord 3> -- disse
                                           Robin com brusquidão. </reord>
73. <s> And my wife.                       73. <s>  E minha mulher.
74. <s> Not Joe's. '                       74. <s2> Não era nada ao Joe
                                           <place 3>
```

Last of all, it is important to note that it is *not* possible to (a) have one reordering within another neither (b) open a <reord> tag in one alignment unit and close it </reord> in another one. The example below is therefore wrong, because the <reord> tag opens in alignment unit 97 and closes in alignment unit 98.

There was a sundial in the centre of the sweep, with an engraved metal plate bolted to its surface. 'Onlie count,' the engraving ran, 'the sunny houres.' Caro had put it there. It had been her first Christmas present to Robin.→ No centro do terreiro, Caro pusera um relógio de sol, o primeiro presente de Natal que dera a Robin, com uma placa de metal gravado aparafusada à superfície que dizia "Conta apenas as horas de sol".

```
97. <s> There was a sundial in             97. <s4> No centro do
the centre of the sweep, with              terreiro,  <reord 4> com uma
an engraved metal plate bolted             placa de metal gravado
to its surface.                            aparafusado à superfície
98. <s> 'Onlie count,' the                 98. <s4> que dizia "Conta
engraving ran, 'the sunny                  apenas as horas de sol".
houres.'                                   </reord>
99. <s> Caro had put it there.             99. <s4> Caro pusera um
                                           relógio de sol,
100. <s> It had been her first             100. <s4> o primeiro presente
Christmas present to Robin.                de Natal  que dera a Robin
                                           <place 4>
```

The way to go about it is to mark <reord> twice:

```
97. <s> There was a sundial in             97. <s4> No centro do
the centre of the sweep, with              terreiro,  <reord 4> com uma
an engraved metal plate bolted             placa de metal gravado
to its surface.                            aparafusado à superfície
                                           </reord>
98. <s> 'Onlie count,' the                 98. <s4> <reord 5> que dizia
engraving ran, 'the sunny                  "Conta apenas as horas de
houres.'                                   sol". </reord>
99. <s> Caro had put it there.             99. <s4> Caro pusera um
                                           relógio de sol,
100. <s> It had been her first             100. <s4> o primeiro presente
Christmas present to Robin.                de Natal  que dera a Robin
                                           <place 4> <place 5>
```

Another possible option, that would cut down on the number of reorderings, is:

```
97. <s> There was a sundial in             97. <s3> No centro do
the centre of the sweep, with              terreiro, Caro pusera um
an engraved metal plate bolted             relógio de sol, <reord 4> com
to its surface.                            uma placa de metal gravado
                                           aparafusado à superfície
98. <s> 'Onlie count,' the                 98. <s3> que dizia "Conta
engraving ran, 'the sunny                  apenas as horas de sol".
houres.'                                   </reord>
99. <s> Caro had put it there.             99. <s>
100. <s> It had been her first             100. <s3> o primeiro presente
Christmas present to Robin.                de Natal  que dera a Robin
                                           <place 4>
```

This example illustrates the subjectiveness underlying part of the markup process and the need for non-trivial decisions when complicated cases arise. In the present case, there is a choice between a 4-1 and a 3-1 plus 1-0 alignment. Fortunately, however, reorderings are not very common!

2.3.3 Texts by José Saramago

The texts by the Portuguese author José Saramago are considered a special case and therefore COMPARA treats them differently from other texts. We chose to consider as separate sentences, i.e. independent alignment units, the parts of text where the author uses commas followed by direct speech beginning with a capital letter. For example:

231. <s> O cego não o ouviu, já iam a entrar no gabinete do médico, e a mulher dizia,	231. <s> The blind man did not hear him, they were already going into the doctor's consulting room, and the wife was saying,
232. <s> Muito obrigada pela sua bondade, senhor doutor, é que o meu marido, e tendo dito interrompeu-se, em verdade ela não sabia o que realmente sucedera, sabia apenas que o marido estava **cego** e lhes tinham roubado o carro.	232. <s> Many thanks for being so kind, doctor, it's just that my husband, and that said, she paused, because frankly she did not know what had really happened, she only knew that her husband was blind and that their car had been stolen.

In these cases, the <s> or <s><p> tags do not contain any special indication, for we have already stated that these texts are a special case. However, it is possible that the translator has split, joined, etc. sentences within the scope of the criteria that COMPARA established for the alignment units of José Saramago's texts. For example, below is a case in which the translator joined two sentences of the original text (according to Saramago's writing criteria):

537. <s> Sim, sou eu, disse, ouviu com atenção o que estava a ser lhe comunicado e só acenou ligeiramente a cabeça antes de desligar.	537. <s2> Yes, speaking, he said, listened attentively to what he was being told and merely nodded his head slightly before ringing off,
538. <s> Quem era, perguntou a mulher,	538. <s2> Who was that, his wife asked,

And a case in which the translator split a sentence of the original (according to Saramago's writing criteria):

1251. <s> O estrondear sacudido das detonações fez surgir quase imediatamente de dentro das tendas, meio vestidos, os soldados que compunham o piquete encarregado da guarda do manicómio e de quem lá fora posto dentro.	1251. <s2> The noise of the blast immediately brought the soldiers, half dressed, from their tents. These were the soldiers from the detachment entrusted with guarding the mental asylum and its inmates.

4. Conclusion

When you’ve finished, save the source text and the translation file in text format (now that the numbers of the alignment units have been fully revised, we do not need to use Word anymore). The files should be renamed with a *fra* for "frase" ending (instead of *easy*) and extension *pt* for texts in Portuguese and *en* for texts in English. For example:

PBMA3fra.pt (for a text in Portuguese)

PBMA3fra.en (for the same text in English)

Send your text in this format to Ana Frankenberg.