COMPARA - Checking automatic alignment

Ana Frankenberg-Garcia (18/02/2004)

In the previous, sentence alignment and alignment markup phase, you were asked to insert alignment markup for sentences that had been joined together in the translation, for sentences that had been and added to the translation, and for sentences that had been reordered in the translation. But, apart from leaving initial <p><s> marks, you were not asked to do anything special for the sentences that had been deleted from the translation and the ones that had been split into more than one sentence in the translation. This is because this is done automatically. However, after the automatic markup is inserted, the sentences that were split in the translation need to be inspected manually, to confirm that everything, especially certain alignment units involving direct speech, are have been counted well. The problem is the automatic procedure is not able to interpret that

```
`You OK?´ Robin's daughter said, standing close to him, but not
touching.
```

is just one sentence. Because the next word after the question mark begins with a capital letter, the program counts it as two sentences. Your job is to correct this, and you do it immediately after the pair of texts in question is made available online, using Compara's Complex Search facility. Here is how:

1.Go to http://www.linguateca.pt/COMPARA/ComplexSearch.html
2. In step 1, if the source text is in Portuguese, select the Portuguese to English direction; if it is in English select the English to Portuguese direction.
3.In step 2, check the box saying "sentences split in translation" and leave everything else unchecked.
4. In step 3.4, check the box pertaining to the text pair you are working on and leave all else unchecked.
5. In step 4, check the box for "concordance" and check "show alignment properties" as well.
6. Submit your query.

When you get your results, you should see **one source text sentence** on the left-hand side of your screen and **more than one translation sentence** on the right-hand side. On the column with the text code, you should see the number of the alignment unit in brackets and underneath it the type of alignment:

1-2 = one sentence split into two
1-3 = one sentence split into three
1-1+1/1/2 = one sentence split into one and a half, etc

You are to check whether the type of alignment given is in fact right. Remember that the automatic alignment markup will consider the example below to be a 1-2 alignment, when it is really a 1-1 alignment:

PBMA1(756): 1-2	-- Que frutas são? perguntou Rubião fechando a carta.	«What kind of fruit is it?» Rubião asked, folding the letter.

You are therefore to open a new document in Word or Wordpad, and write just this:

756: 1-1

Which means that in alignment unit 756, the alignment type is one source text sentence to one translation sentence. Don't forget to leave a space after the colon.

If you detect more problems, record them on a new line. Your document should look something like this:

```
407: 1-1
751: 1-1
756: 1-1
862: 1-2
```

Some text pairs have very few problems, and some have none at all. If there weren't any problems, simply let me know that everything was fine. If there were problems (even if just one), save the file in text format. The file should be named with the text code, followed by *div*. No special extension is necessary. For example:

PBMA5div.txt

Send it over to Ana Frankenberg.