

COMPARA – Verificação da etiquetagem automática

Ana Frankenberg-Garcia (18/02/2004)

Na fase anterior, de alinhamento por frase, teve de inserir etiquetas especiais para as frases unidas, adicionadas e reordenadas na tradução. Por outro lado, para além do <p><s> inicial, não teve de fazer nada para as frases suprimidas e divididas na tradução. Isso porque a etiquetagem desses casos é automática. Contudo, depois de feita a etiquetagem automática desses casos, é necessário verificar manualmente se as frases divididas na tradução foram bem contadas. É comum haver problemas nos casos de discurso directo, em que o procedimento automático considera exemplos como o abaixo como sendo duas, ao invés de uma única, frase:

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

Como a palavra a seguir ao ponto de interrogação começa com letra maiúscula, o programa interpreta a frase acima como se fossem duas frases. O seu trabalho será corrigir isto através da Pesquisa Avançada do Compara, imediatamente após o par de textos em causa ter sido integrado ao corpus. Eis como:

1. Vá para <http://www.linguateca.pt/COMPARA/BuscaAvancada.html>
2. No primeiro passo, seleccione a direcção português-inglês se o texto de partida for em português, e a direcção inglês português se o texto de partida for em inglês.
3. No segundo passo, seleccione apenas "Separação de frases na tradução".
4. No passo 3.4, seleccione apenas o texto que vai verificar.
5. No passo 4, seleccione "concordâncias" e "mostrar propriedades de alinhamento".
6. Carregue no botão "procurar".

Nos resultados da sua pesquisa, irá encontrar **uma frase do texto de partida** no lado esquerdo do ecrã, e **mais de uma frase da tradução** no lado direito. Na coluna com o código do texto, encontrará o número correspondente à unidade de alinhamento em causa entre parênteses e a seguir o tip de alinhamento:

1-2 = uma frase dividida em duas
1-3 = uma frase dividida em três
1-1+1/2 = uma frase dividida em uma e meia, etc

A sua função será verificar se o tipo de alinhamento fornecido está correcto. Note que a etiquetagem automática considera o exemplo abaixo um alinhamento tipo 1-2, quando na verdade se trata de um alinhamento tipo 1-1:

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

Se encontrar casos deste género, deve abrir um documento novo em Word ou Wordpad e escrever apenas:

756: 1-1

Isso significa que na unidade de alinhamento 756, o tipo de alinhamento deve ser corrigido para 1-1, ou seja, uma frase no texto de partida corresponde a uma frase na tradução. Não se esqueça de deixar um espaço após os dois pontos.

Se detectar mais problemas, registre cada um numa linha separada. O documento final deverá ter o seguinte aspecto:

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

Em alguns pares de textos verificam-se muito poucos problemas, e em outros, às vezes, não se encontra nada de errado. Se não houver nenhum problema, deve apenas avisar-me que não encontrou nada. Se tiver encontrado algum caso a corrigir, mesmo que apenas um, guarde o ficheiro em formato texto, com o código do texto em questão seguido de *div*. Não é necessário pôr nenhuma extensão especial. Por exemplo:

PBMA5div.txt

Envie-o para Ana Frankenberg.