

AJUDA À REVISÃO do AnELL

Isabel Marcelino
30 de Setembro de 2005

Há pelo menos dois grandes tipos de problemas envolvidos na revisão da anotação do AnELL: ambiguidade e palavras desconhecidas.

AJUDA PARA AS AMBIGUIDADES

Para resolver os problemas das ambiguidades, pode-se utilizar o DiaTag (*DIALogic TAGging*) <http://www.nyu.edu/pages/linguistics/intex/> (Viprey, 2004)

Adaptado ao INTEX, o DiaTag pode completar esse dispositivo porque tem uma interface onde se pode escolher a boa informação e assim resolver a ambiguidade da anotação.

Por exemplo, na frase: «A Maria foi ao médico.», a anotação automática feita pelo AnELL é:

```
<S>
{a(o:fs):PRO:Dem} {a(o:fs):DET:Art, Def} {a(me:3fs[o],2"fs[o]):PRO:Pes, A}
{a(a):PREP}
<PESSOA TIPO=INDIVIDUAL>
{Maria(Maria:fs):N:Hum}
</PESSOA>
{foi(ser:J3s[icqo],J2"s[icqo]):V} {foi(ser:J3s[icqo],J2"s[icqo]):V:Vaux}
{foi(ir:J3s[icqo],J2"s[icqo]):V} {foi(ir:J3s[icqo],J2"s[icqo]):V:Vaux}
{ao(ao:ms):PREPXPRO:Dem} {ao(ao:ms):PREPXDET:Art, Def}
{médico(médico:ms):N} {médico(médico:ms):A:Rel}
</S>
```

Neste caso o DiaTag vai automaticamente seleccionar todas as entradas com ambiguidade, e a interface propõe ao utilizador de escolher qual a anotação que quer guardar.

E depois podemos obter o resultado:

```
<S>
{a(o:fs):DET:Art, Def}
<PESSOA TIPO=INDIVIDUAL>
{Maria(Maria:fs):N:Hum}
</PESSOA>
{foi(ir:J3s[icqo],J2"s[icqo]):V}
{ao(ao:ms):PREPXDET:Art, Def}
{médico(médico:ms):N}
</S>
```

AJUDA PARA AS PALAVRAS DESCONHECIDAS

Há pelo menos duas abordagens principais para tratar o problema das palavras desconhecidas:

- Acrescentar as formas canónicas aos dicionários.
- Introduzir essas entradas directamente já flexionadas.

A primeira abordagem tem a vantagem de que no futuro pode-se encontrar não só a forma que era desconhecida como também as relacionadas com o mesmo paradigma flexional.

Por exemplo, se "arrobas" não é encontrado, e se adicionar ao dicionário arroba,N301 e depois gerar as formas flexionadas, então da próxima vez se poderá encontrar não só "arrobas" como também "arroba".

Explicação do esquema de interface que poderá ter a Ajuda à revisão do AnELL

1. Uma coluna com todas as palavras desconhecidas do texto.

2. Um "menu descendente" chamado «Ordenação».

Este menu contém: - *Alfabética* (para que as palavras estejam ordenadas pela ordem alfabética)
- *Alfabética inversa* (para que as palavras estejam ordenadas pela ordem alfabética inversa)
- *Frequência* (para que as palavras estejam ordenadas pela frequência em que ocorrem)

→ *A mudança deve ser feita na coluna onde estão as palavras desconhecidas*

3. Um "menu descendente" chamado «Tipo».

Este menu contém: - *Analizador Morfológico*: aplica os autómatos aumentativos, diminutivos e superlativos.
(*seleccionar toda a lista e verificar quais são as palavras que foram anotadas pelos autómatos → essas palavras anotadas devem aparecer numa lista para que possa ser validada ou não essa selecção e correcção no texto*)
- *Estrangeiras*: aplica os dicionários do brasileiro, espanhol, francês e inglês.
(*seleccionar toda a lista e verificar quais são as palavras que foram anotadas pelos dicionários → essas palavras anotadas devem aparecer numa lista para que possa ser validada ou não essa selecção e correcção no texto*)

→ Deve aparecer uma pequena janela que indica se a/as palavra(s) foi/foram bem analisada(s) e corrigida(s) na anotação do AnELL.

Estes dois tipos de análises ajudem a filtrar a lista das palavras desconhecidas.

4. Um “menu descendente” chamado «Pesquisa».

Este menu contém: - *Net*: para que a pesquisa da palavra desconhecida se faça na Internet, no Google, por exemplo!

- *Texto (concordância)*: para que a pesquisa da palavra desconhecida se faça em concordância no texto onde foi encontrada.

(*Só se pode seleccionar uma palavra na coluna das palavras desconhecidas.*)

→ Deve aparecer uma pequena janela com a concordância da palavra ou com o Google.

5. Um botão chamado «Acrescentar lema»

Este botão serve para dar a anotação correcta no pedido de anotação do AnELL e ao mesmo tempo para acrescentar as formas canónicas aos dicionários do AnELL e do LabEL.

Deve aparecer uma pequena janela com a palavra que foi seleccionada na lista, para que lhe seja atribuída a informação correcta. Depois de aceitar essa inserção na anotação do AnELL, aparece outra janela que permita acrescentar as formas canónicas aos dicionários.

6. Um botão chamado «Apagar»

Este botão serve para apagar as palavras “sem sentido” que vão ficar marcadas como XERR na anotação do pedido.

Agradecimento

O AnELL foi (parcialmente) financiado pela Fundação para a Ciência e Tecnologia, co-financiada pelo POSI, através do projecto POSI/PLP/43931/2001 (Linguatca).

REFERÊNCIAS

Viprey, J-M, «DIATAG_INTEX : Convivialité d'étiquetage des ambigus résiduels, Applications pour les formes composées». 5th Intex Workshop, Marseille (France), 2004.

ANEXO 1 :

Ajuda à Revisão

A vertical sidebar menu with a scroll bar on the right. At the bottom, there are two items: "número de palavras" with a dropdown arrow and a checkbox labeled "seleccionar tudo".

- Ordenação das palavras desconhecidas

"Ordenação" ▼

- Classificação das palavras:

"Tipo" ▼

→ aparecer uma janela que diz se foi bem analisada e corrigida.

- Pesquisa das palavras:

"Pesquisa" ▼

→ aparecer uma janela com a concordância ou com o google.

Acrescentar lema

Apagar