

Documentação do ELLE

Isabel Marcelino
29 de Setembro de 2005

O ELLE surgiu a partir de uma iniciativa de avaliação conjunta para o reconhecimento de entidades mencionadas (Mota et al., 2005). A avaliação conjunta (Santos & Rocha, 2003) é definida como um processo em que os participantes estabelecem entre si o conjunto de tarefas e os critérios de avaliação para a execução dessas mesmas tarefas numa determinada área de aplicação. A avaliação conjunta efectuada pela Linguateca (o HAREM, Santos & Cardoso, 2005) relativamente ao reconhecimento de entidades consistiu na identificação e classificação de expressões linguísticas incluindo, na sua maioria, os tradicionalmente conhecidos "nomes próprios". Essas expressões são constituídas por um elemento, geralmente um nome, que aponta para um referente específico e o identifica.

Esta documentação apresenta e descreve as funcionalidades do ELLE, os recursos que lhe servem de base e os módulos que integra. Para além disso, inclui uma apresentação dos principais problemas encontrados na concepção do sistema e sugere alternativas para os resolver. Finalmente, mostra como foi feita a integração do ELLE no anotador AnELL e explica como se invoca o ELLE em análises de REM.

1. Descrição do que é o ELLE e o que faz

ELLE significa Etiquetador LaBEL-Lex de Entidades (mencionadas). O ELLE é um anotador de entidades mencionadas, ou seja, dado um texto em português, reconhece e classifica as expressões consideradas com entidades mencionadas. Este sistema foi concebido no âmbito do HAREM, a primeira Avaliação conjunta de sistemas de Reconhecimento de Entidades Mencionadas para o português.

O ELLE anota entidades mencionadas referentes a Pessoas, Locais, Acontecimentos e Organizações.

Estas quatro categorias e também as subcategorias (tipos) são um subconjunto das que foram estabelecidas no âmbito do HAREM. Apresentamos em seguida exemplos reais de cada tipo na colecção HAREM.

Na categoria **Pessoa**, o ELLE anota os tipos:

-Individual:

Exemplo: <PESSOA TIPO="INDIVIDUAL">Bush</PESSOA> endureceu na atitude de conservação a todo o preço de uma ordem histórica condenada

-Cargo:

Exemplo: E o<PESSOA TIPO="CARGO"> Capitão</PESSOA> mandou em terra a Nicolau Coelho para ver aquele rio.

-GrupoCargo:

Exemplo: depois da visita por convite seu que nos fizeram os<PESSOA TIPO="GRUPOCARGO"> Presidentes</PESSOA> dos parlamentos israelita

-GrupoInd:

Exemplo: o nome de amores famosos, vamos dizer assim<PESSOA TIPO="GRUPOIND">Romeu e Julieta</PESSOA>,<PESSOA TIPO="GRUPOIND"> Otelo e Desdêmona</PESSOA>, que mais que tem?

- **GrupoMembro:**

Exemplo: No morro da Mogueira há vestígios evidentes da presença dos<PESSOA TIPO="GRUPOMEMBRO"> Celtas</PESSOA>,

Na categoria **Organização**, o ELLE anota os tipos:

- **Administração:**

Exemplo: como a<ORGANIZACAO TIPO="ADMINISTRACAO"> Câmara</ORGANIZACAO> não dá transporte, não vou.

- **Empresa:**

Exemplo: Pagar anuidade de R\$ 60,00, na conta da ANCIB no<ORGANIZACAO TIPO="EMPRESA"> Banco do Brasil</ORGANIZACAO>,

- **Instituição:**

Exemplo: Em vésperas de Carnaval, a<ORGANIZACAO TIPO="INSTITUICAO"> Associação Portuguesa para a Defesa do Consumidor</ORGANIZACAO> (DECO) mostra-se preocupada

- **Sub:**

Exemplo: quer através da actividade do recém criado<ORGANIZACAO TIPO="SUB"> Departamento Científico</ORGANIZACAO>.

Na categoria **Local**, o ELLE anota os tipos:

- **Correio:**

Exemplo: Olhe aqui no<LOCAL TIPO="CORREIO"> Largo do Paço</LOCAL> começaram as aulas,

- **Administrativo:**

Exemplo: Tenho agora um convite para ir à<LOCAL TIPO="ADMINISTRATIVO"> Maia</LOCAL>

- **Geográfico:**

Exemplo: A elevada pluviosidade conduz, por vezes, àformção de lagoas em zonas elevadas, como éo caso da<LOCAL TIPO="GEOGRAFICO"> Lagoa do Marinho</LOCAL>

- **Virtual:**

Exemplo: Uma boa série de televisão: a que a<LOCAL TIPO="VIRTUAL"> RTP</LOCAL> estreou ontem à noite, na<LOCAL TIPO="VIRTUAL"> TV2</LOCAL>,

- **Alargado:**

Exemplo: o<LOCAL TIPO="ALARGADO"> Museu Paraense Emílio Goeldi</LOCAL>, que reúne exemplares da flora e da fauna

Na categoria **Acontecimento**, o ELLE anota os tipos:

- **Efeméride:**

Exemplo: Após o<ACONTECIMENTO TIPO="EFEMERIDE"> 25 de Abril</ACONTECIMENTO>, houve uma certa instabilidade no país,

- **Organizado:**

Exemplo: Temos de o ultrapassar ...a fim de restaurar a confiança do<ACONTECIMENTO TIPO="ORGANIZADO"> Congresso</ACONTECIMENTO> e do público

- **Evento:**

Exemplo: O fracasso dos países ocidentais em converter suas indústrias militares para finalidades civis, depois do fim da<ACONTECIMENTO TIPO="EVENTO"> Guerra Fria</ACONTECIMENTO>,

O ELLE baseia-se em gramáticas, descritas na forma de grafos, que descrevem contextos locais. Este sistema usa o INTEX (Silberstein, 1993) versão 4.33 para a descrição e aplicação das gramáticas e dos dicionários.

2. Descrição dos recursos em que se apoia

O ELLE utiliza os recursos do LABEL: para anotar os textos, o ELLE utilizou os dicionários usados pelo AnELL (ver Marcelino (2005) para uma documentação dos recursos linguísticos subjacentes ao AnELL):

	Recursos	Tamanho
Dicionários de palavras simples e grafos	DelafNpr_V2.dic	136 KB
	DelafNprDemoHarem.dic	3 KB
	Delaf_V3.bin	1.011 KB
	Romanos.fst	1 KB
Dicionários de palavras compostas	DelacfNpr_V2.dic	175 KB
	DelacfNprDemoHarem.dic	4 KB
	DelacPrep.dic	1 KB

3. Descrição do módulo ELLE

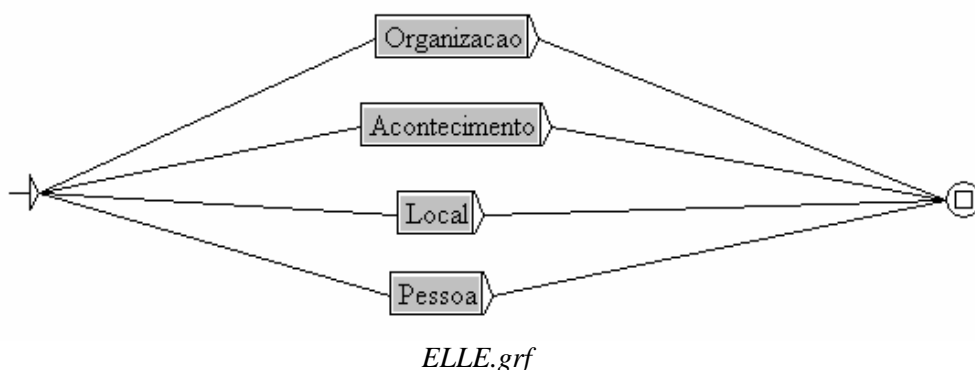
a. As gramáticas

Para o HAREM, foram utilizadas 24 gramáticas, escritas inicialmente por Elisabete Ranchhod e alargadas pela autora.

Cada gramática corresponde a um tipo de uma categoria mas, às vezes, quando uma só gramática é muito complicada, usam-se duas gramáticas para um só tipo:

Por exemplo, para a Categoria ORGANIZACAO Tipo INSTITUICAO, o ELLE utiliza duas gramáticas: InstAdj.grf (entidade mencionada é começada por Nome + Adjectivo) e InstPrepNome.grf (entidade mencionada é começada por Nome + Preposição).

- A gramática principal:

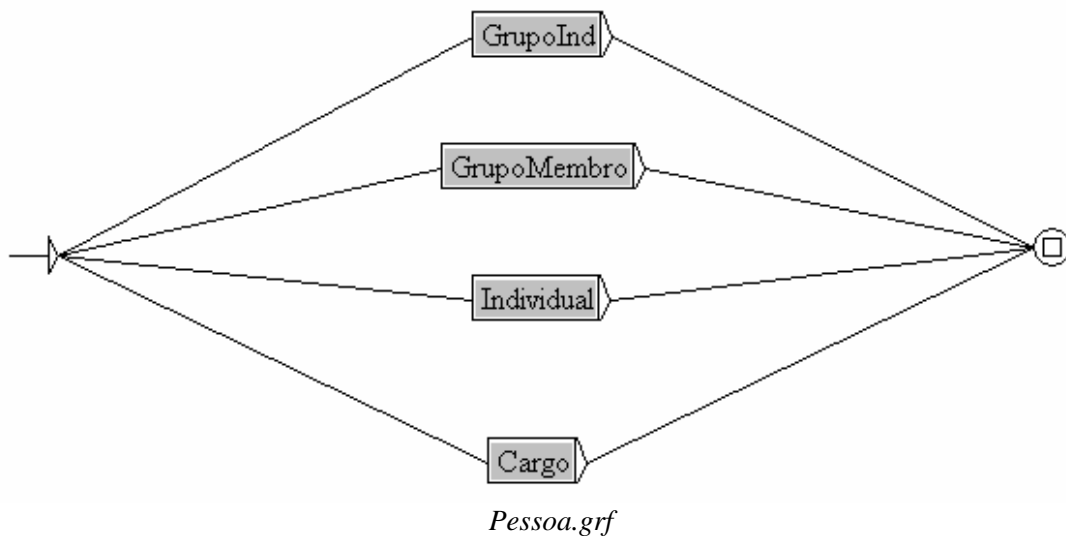


O grafo «ELLE.grf» é a gramática principal do ELLE.

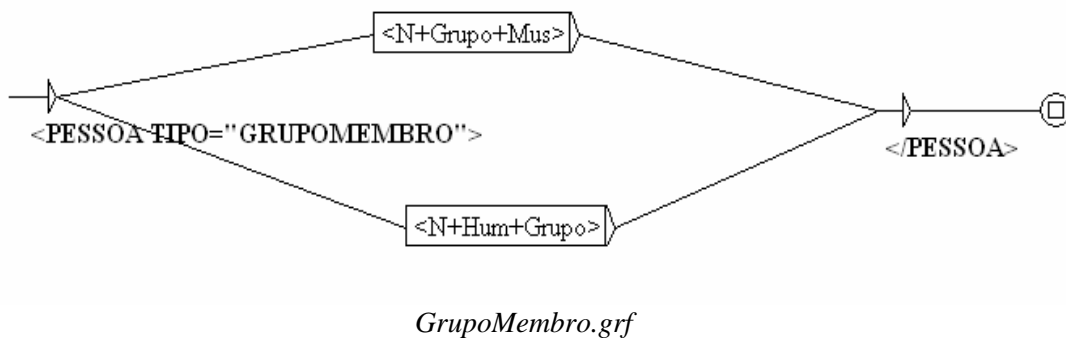
É esta gramática que chama todas as outras gramáticas.

Este grafo principal chama os grafos em paralelo para que nenhum tenha prioridade. Eles são todos aplicados ao mesmo tempo.

Por exemplo, quando chama o grafo: *Pessoa.grf*, obtemos uma gramática que chama os tipos dessa categoria.



E se a anotação utilizada será do tipo “GrupoMembro”, teremos o grafo seguinte:



Nesta gramática, temos dois nós:

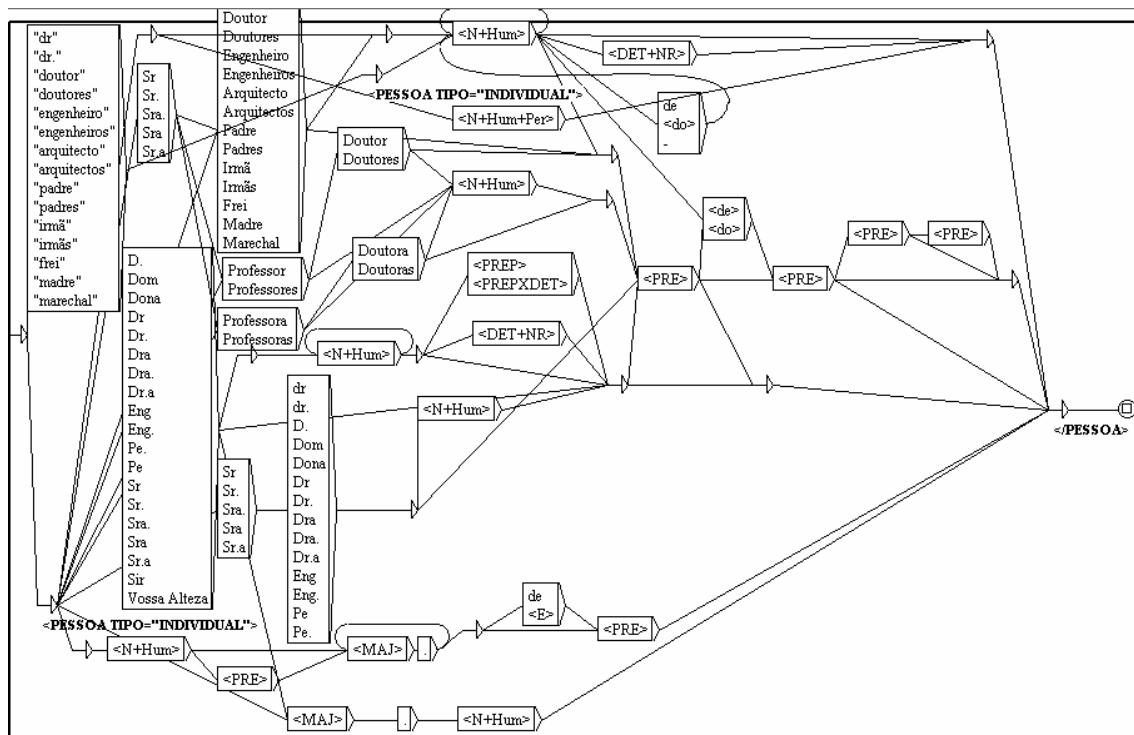
<N+Grupo+Mus>¹ e <N+Hum+Grupo>²

Todas as expressões que estiverem classificadas nos dicionários com esses atributos estarão anotadas como Pessoa, Tipo ”GRUPOMEMBRO”.

Mas o ELLE não é feito só de gramáticas simples, também tem várias muito mais complexas: por exemplo, o grafo individual.grf é muito mais complicado que as gramáticas anteriores:

¹ A etiqueta <N+Grupo+Mus> é usada na referência de um grupo de música, como por exemplo os “Delfins”.

² A etiqueta <N+Hum+Grupo> refere um conjunto de pessoas, referidas como sendo membros de uma organização.



Individual.grf

- Explicação de cada gramática:

- O grafo principal: **ELLE.grf**: é constituído por 4 sub-grafos que estão também estruturados (que chamam outros grafos). Este grafo principal tem 761 nós e 3957 transições.

- Como se vê acima no grafo ELLE.grf, este grafo é constituído por 4 sub-grafos:

- **Organização.grf**: tem 312 nós e 2057 transições.

Este grafo é também constituído por 4 sub-grafos:

- **Administração.grf**: tem 43 nós e 175 transições.

- **Instituição.grf**: tem 248 nós e 1778 transições.

Este grafo é constituído por 2 sub-grafos:

- **InstPrepNome.grf**: tem 108 nós e 638 transições.

- **InstAdj.grf**: tem 81 nós e 562 transições.

- **Empresa.grf**: tem 41 nós e 138 transições.

- **Sub.grf**: tem 12 nós e 20 transições.

- **Acontecimento.grf**: tem 105 nós e 345 transições

Este grafo é constituído por 3 sub-grafos:

- **Efeméride.grf**: tem 31 nós e 44 transições.
- **Organizado.grf**: tem 66 nós e 265 transições.
- **Evento.grf**: tem 14 nós e 38 transições.
- **Local.grf**: tem 249 nós e 990 transições.
Este grafo é constituído por 5 sub-grafos:
 - **Correio.grf**: tem 97 nós e 456 transições.
 - **Alargado.grf**: tem 26 nós e 108 transições.
 - **Geográfico.grf**: tem 46 nós e 161 transições.
 - **Virtual.grf**: tem 48 nós e 125 transições.
 - **Administrativo.grf**: tem 50 nós e 152 transições.
- **Pessoa.grf**: tem 104 nós e 567 transições.
Este grafo é constituído por 4 sub-grafos:
 - **GrupoInd.grf**: tem 6 nós e 7 transições.
 - **GrupoMembro.grf**: tem 4 nós e 4 transições.
 - **Individual.grf**: tem 39 nós e 248 transições.
 - **Cargo.grf**: tem 65 nós e 312 transições.

b. Utilização do ELLE no HAREM

- Número de padrões reconhecidos por cada gramática:

Na colecção HAREM, constituída por mais de 500 mil palavras de vários géneros textuais:

- A gramática **ELLE.grf** anotou 19319 padrões³.
- A gramática **Organização.grf** anotou 3842 padrões.
 - A gramática **Administração.grf** anotou 1431 padrões.
 - As gramáticas do tipo “**Instituição**” anotarão 1740 padrões.

³ Quando se diz que a gramática anotou 19319 padrões, isso significa que a dita gramática encontrou no texto 19319 Entidades Mencionadas.

- A gramática **Empresa.grf** anotou 663 padrões.
- A gramática **Sub.grf** anotou 343 padrões.
- A gramática **Acontecimento.grf** anotou 349 padrões.
- A gramática **Efeméride.grf** anotou 61 padrões.
- A gramática **Organizado.grf** anotou 258 padrões.
- A gramática **Evento.grf** anotou 30 padrões.
- A gramática **Local.grf** anotou 8560 padrões.
- A gramática **Correio.grf** anotou 221 padrões.
- A gramática **Alargado.grf** anotou 245 padrões.
- A gramática **Geográfico.grf** anotou 251 padrões.
- A gramática **Virtual.grf** anotou 1251 padrões.
- A gramática **Administrativo.grf** anotou 6592 padrões.
- A gramática **Pessoa.grf** anotou 6567 padrões.
- A gramática **GrupoInd.grf** anotou 57 padrões.
- A gramática **GrupoMembro.grf** anotou 16 padrões.
- A gramática **Individual.grf** anotou 5923 padrões.
- A gramática **Cargo.grf** anotou 526 padrões.
- A gramática **GrupoCargo** anotou 45 padrões.

O uso de gramáticas em paralelo torna o desempenho do ELLE não determinístico. Assim, no caso de mais do que uma gramática poder ser aplicada à mesma expressão, não é possível de momento controlar qual a classificação obtida.

As gramáticas do ELLE (o grafo ELLE.grf) foram aplicadas quatro vezes à Colecção HAREM antes dos resultados serem enviados para a organização do HAREM. O grafo ELLE.grf foi aplicado mais do que uma vez simplesmente para confirmar o resultado da primeira aplicação antes de enviar os resultados para a organização do HAREM.

4. Problemas encontrados e descrição de uma avaliação do ELLE

a) Problemas conhecidos

Nos dicionários do LABEL, e essa é uma opção tomada pelo LABEL, existem entradas que são marcadas sistematicamente de forma ambígua (aliás criadas antes sequer do HAREM existir).

Existe uma limitação do sistema INTEX, na aplicação dos grafos, em produzir mais do que uma informação para o mesmo emparelhamento. Isso quer dizer que a mesma sequência no texto emparelha com mais do que um caminho da gramática, que produzem informações diferentes. Mas o INTEX só pode dar uma informação e nós não controlamos isso. Ou seja, o INTEX introduz novas informações no texto (graças ao modo *MERGE*) mas só pode introduzir **uma** anotação por Entidade Mencionada. Por exemplo, nos dicionários, a palavra “Portugal” está marcada com três informações diferentes:

- Portugal,.N+Hum (nome de pessoa)
- Portugal,.N+Top (topónimo)
- Portugal,.N+HumCol (entidade geopolítica)

Ao ver “Portugal”, o INTEX faz o emparelhamento através de três gramáticas distintas: LOCAL, ORGANIZACAO e PESSOA.

Se não encontrar nenhum contexto nas gramáticas que coincida com o contexto do texto, vai produzir uma única informação que nem sequer se pode controlar qual vai ser. É conforme ele fizer a pesquisa, é por acaso. Em suma, da forma como ELLE foi escrito, não é possível controlar a saída que o sistema produz quando é possível produzir mais do que uma saída.

Esse problema aparece várias vezes com o grupo de categorias seguinte:

LOCAL tipo ADMINISTRATIVO	ORGANIZACAO tipo ADMINISTRACAO (ex: Lisboa)
PESSOA tipo INDIVIDUAL	ABSTRACAO tipo NOME
ORGANIZACAO tipo INSTITUICAO	ABSTRACAO tipo NOME (ex: Escola Normal Livre)
LOCAL tipo VIRTUAL	ORGANIZACAO tipo EMPRESA (ex: Expresso)
ORGANIZACAO tipo EMPRESA	LOCAL tipo ALARGADO (ex: Ler Devagar)
LOCAL tipo ADMINISTRATIVO	ABSTRACAO tipo IDEIA (ex: Portugal)
ORGANIZACAO tipo EMPRESA	ABSTRACAO tipo NOME (ex: Isotexsa)
ORGANIZACAO tipo EMPRESA	ABSTRACAO tipo MARCA (ex: Intelbras)
PESSOA tipo INDIVIDUAL	LOCAL tipo ADMINISTRATIVO (ex: São José)

O ELLE só anota com as etiquetas que estão na esquerda. A categoria ABSTRACCAO não foi tratada, então não a pode reconhecer, e para as outras, o problema pode vir da ordem pela qual o grafo *ELLE.grf* chama os grafos.

Podíamos, é certo, mudar a ordem pela qual os grafos são chamados no grafo *ELLE.grf*, mas essa mudança não permite encontrar tantos padrões como está actualmente.

- Há imensos problemas para distinguir um nome humano e um topónimo que sejam homónimos.

Exemplo: «Santos» pode ser um apelido e um topónimo. (O ELLE só consegue fazer a diferença se estiver a proposição «em» antes de «Santos»; neste caso, ele sabe que é topónimo)

- A maior parte das siglas não estão a ser anotadas pelo ELLE.

- Outros exemplos de algumas EM mal anotadas:

internet no site:<LOCAL TIPO="VIRTUAL">http://www.brasil</LOCAL>2mil.com.br.

→ no grafo “Local tipo Virtual”, os números não estão inseridos num endereço electrónico. Para que este endereço seja bem anotado, é preciso inserir no grafo a possibilidade dos números fazerem parte do nome dum site e também se pode acrescentar essa informação no reconhecimento dos e-mails.

Andei em Braga, na <PESSOA TIPO="INDIVIDUAL">São João do Souto</PESSOA> talvez na primária e depois é que mudei para o colégio.

→ para que o nome desta escola (São João do Souto) fosse bem anotado, tinha que ser inserido no dicionário de nomes próprios compostos como instituição. Mas como “São João” está inserido no dicionário de nomes próprios compostos como “Nome Humano”, «São João do Souto» foi anotado como nome de pessoa.

b) Como melhorar o ELLE

Os problemas acima descritos podem resolver-se eventualmente produzindo uma saída do tipo <ORGANIZACAO|LOCAL>.

Mas também, por exemplo, se estivesse descrito no grafo ‘Local tipo Administração’:

"<deslocar>-se a <N+Top>"

seria possível neste contexto atribuir a etiqueta certa quando emparelhasse com "deslocou-se a Lisboa".

Para melhorar em geral o ELLE, pode-se actualizar os dicionários de nomes próprios, criar novas gramáticas para os outros Tipos das outras Categorias que ainda não foram tratadas e corrigir (melhorar) as gramáticas existentes.

c) Diferenças entre as opções do ELLE e as do HAREM

As anotações do ELLE têm em conta as decisões do LABEL sobre certas formas de anotação que já foram discutidas anteriormente por Elisabete Ranchhod.

- Para o LabEL, as vírgulas não são sempre um critério linguístico para separar EMs: Por exemplo, «Presidente da República, Jorge Sampaio» é UMA SÓ EM.

Para o LabEL, «Presidente da República» é uma entrada lexical com os atributos 'Hum+Cargo' (ocupação tipicamente humana), Jorge Sampaio é um nome de pessoa (atributo 'Hum'). Do ponto de vista sintáctico, trata-se de dois SN em aposição: a vírgula é obrigatória!

5. Integração do ELLE no AnELL

O AnELL (**A**notador **E**lectrónico **L**abEL-**L**inguateca) é um serviço on-line de anotação linguística de corpora cujo principal objectivo é produzir corpora anotados com informação gramatical. Foi desenvolvido em parceria entre o **LabEL**, que elaborou os recursos linguísticos, e a **Linguateca**, que implementou a interface web. Foi apresentado em (Mota & Moura, 2003). Encontra-se neste momento instalado na máquina acdc da Linguateca, acessível através do endereço [www.linguateca.pt/ANELL](http://www.linguat<u>e</u>ca.pt/ANELL). O motor de processamento de linguagem natural é o sistema **INTEX**.

Para poder inserir o ELLE no AnELL foi necessário modificar alguns grafos e criar outros.

No AnELL, quando o utilizador desejar ter a anotação de entidades mencionados, terá que seleccionar “Entidades Mencionadas” na janelita *Anotações*.

Modo Automático

Insira um pequeno texto na janela "Texto" e prima "Anotar Texto". Obterá o texto anotado após uns breves instantes. Para alterar os recursos linguísticos, o conteúdo das etiquetas e/ou o formato de visualização do texto anotado prima "Mostrar Opções".

Seleccione [Anotações](#) para ver o valor dos códigos.
Seleccione [Ajuda](#) para mais informações.

The screenshot displays the AnELL web interface in automatic annotation mode. It features a main text input area labeled 'Texto' containing the sentence 'O João foi ao Hospital.'. To the right, there are two panels: 'Recursos' and 'Anotações'. The 'Recursos' panel has five checked options: 'Dicionários de palavras simples', 'Dicionários de palavras compostas', 'Dicionários de siglas', 'Dicionários de nomes próprios', and 'Gramáticas de resolução de ambiguidades'. The 'Anotações' panel has six checked options: 'Lema', 'Categoria gramatical', 'Subcategorias gramaticais', 'Atributos semânticos', 'Atributos morfológicos', and 'Entidades Mencionadas'. Below the text input, there are three buttons: 'Anotar Texto', 'Mostrar Opções', and 'Limpar Formulário'.

E para o exemplo acima escrito, o resultado será:

```


Texto Anotado



Para uma descrição da informação contida nas etiquetas carregue aqui  

  Resultados em ficheiro txt.



```

<S>
{o(o.ms):PRO:Dem} {o(o.ms):DET:Art, Def} {o(me:3ms[o],2"ms[o]):PRO:Pes, A}
<PESSOA TIPO=INDIVIDUAL>
{João(João.ms):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}
<AMB_ATOM>
{ao(ao.ms):PREPXPRO:Dem} {ao(ao.ms):PREPXDET:Art, Def}
<OR>
{a(a):PREP}
{o(o.ms):PRO:Dem} {o(o.ms):DET:Art, Def}
</AMB_ATOM>
<ORGANIZACAO TIPO=INSTITUICAO>
{hospital(hospital.ms):N}
</ORGANIZACAO>
.
</S>

```


```

6. Como se invoca o ELLE

Para fazer uma anotação com o ELLE, é preciso ter o INTEX instalado⁴, assim como os recursos linguísticos do LABEL e o ELLE numa directoria do próprio computador.

- Abrir o INTEX.
- Seleccionar «**Português**» no Language Resource Directory.
- Seleccionar **Text** → **Open**
- Escolher o texto pretendido na directoria certa. (se o texto estiver em .txt, tem que escolher «**Windows ANSI text**» para que ele apareça na directoria escolhida)
- Responder **Sim** à pergunta «Do you want to pré-process the text?»
- Seleccionar só «**Apply FST em MERGE mode**» e «**Apply All Default Dictionaries**» e clique em **GO!**.
- Depois de ter feito isto (pre-processar o texto e aplicar os dicionários), seleccionar **Text** → **Locate Pattern**.

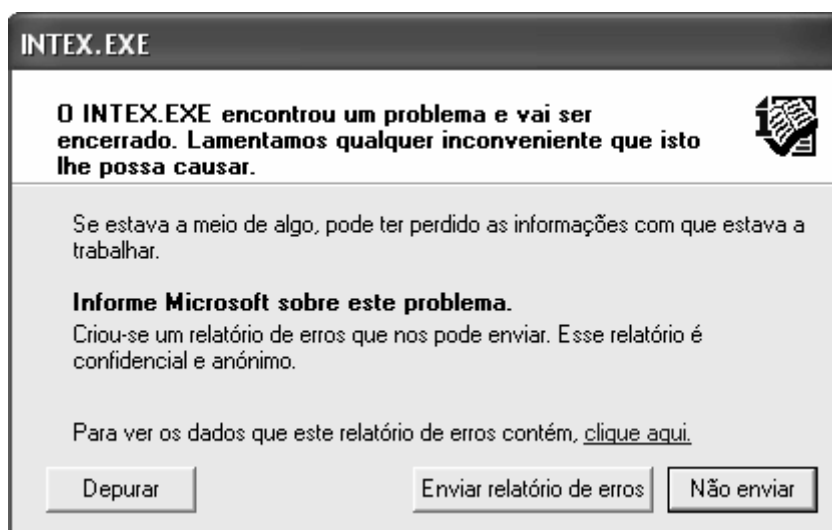
⁴ Para instalar o INTEX, dirija-se ao endereço: <http://www.nyu.edu/pages/linguistics/intex/> e siga as indicações de instalação.

- Seleccionar **Graph** e clicar em «**Set**» e depois procure a pasta “ELLE” na directoria onde estiver.
- Seleccionar a pasta Grafos e depois **ELLE.grf**.
- No “Locate Panel”, seleccionar «Merge with input text» no *FST outputs*; seleccionar «**Index all utterances in text**» no *Search limitation* e clicar em **START**.
- Para ver a concordância, na janela *Display indexed sequences...* que aparece, pôr 115 chars no *Right Col* (em vez de 55) e clicar em **Build concordance**.
- Para ver o texto anotado, fechar a janela *Display indexed sequences...* e fechar o grafo *ELLE.grf*.
- Se preferir ver a anotação no texto em formato rtf (sem concordância), ir à directoria onde está o texto que foi anotado, abrir a pasta **nome-do-texto_snt** e seleccionar o ficheiro **result**.

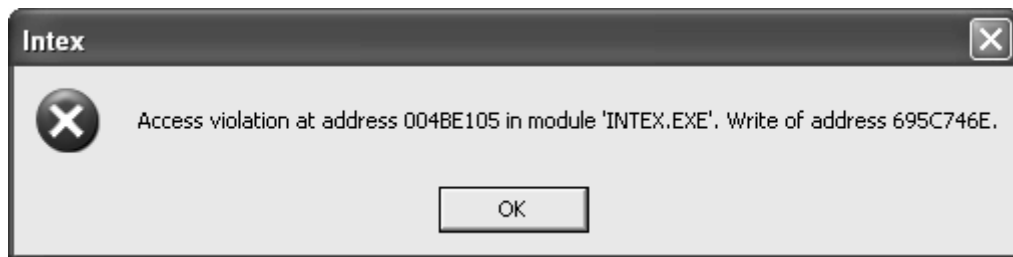
- PROBLEMAS QUE PODEM OCORRER

O principal problema que pode ocorrer, é o Intex bloquear quando o caminho para encontrar o texto for muito comprido:

Por exemplo, se o texto se encontrar dentro duma pasta no ambiente de trabalho, o Intex pode dar uma mensagem de erro:



ou a mensagem:



porque para ele o caminho é muito comprido.

Para resolver este problema, é preciso entrar no C:\, abrir a pasta *INTEX*, depois a pasta *PORTUGUÊS* e colocar o texto (em .txt) na pasta *Corpus*.

Mas este problema também pode ocorrer se os dicionários não estiveram bem situados. Os dicionários devem estar na directoria do Intex nas pastas Delacf e Delaf.

Para aplicar os grafos, não há este tipo de problemas.

Agradecimento

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

REFERÊNCIAS

- Marcelino, Isabel, «Documentação dos recursos linguísticos subjacentes ao AnELL», 2005.
- Mota, Cristina; P. Moura, «ANELL: A Web System for Portuguese Corpora Annotation». In Nuno Mamede; J. Baptista; I. Trancoso; G. Volpes Nunes (eds.), *Computational Processing of the Portuguese Language*, LNAI 2721, pp. 184-188, Berlin: Springer, 2003.
- Mota, Cristina; D. Santos; E. Ranchhod, «Avaliação de reconhecimento de entidades mencionadas: princípio de AREM». In Diana Santos (org.), *Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa*, Lisboa, no prelo.
- Ranchhod, Elisabete M., «O uso de dicionários e de autómatos finitos na representação lexical das línguas naturais». In Elisabete M. Ranchhod (org.), *Tratamento das Línguas por Computador. Uma Introdução à Linguística Computacional e suas Aplicações*, pp. 13-47, Lisboa: Caminho, 2001.
- Ranchhod, Elisabete; P. Carvalho; C. Mota; A. Barreiro, «Portuguese Large-scale Language Resources for NLP Applications». In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of the 4th Language Resources and Evaluation Conference, LREC'04*, pp. 1755-1758, Lisbon, 2004.
- Santos, Diana; N. Cardoso, «Assumptions and evaluation resources of HAREM - Evaluation Contest for Named Entity Recognizers for Portuguese», em apreciação.
- Santos, Diana; P. Rocha, «AvalON: uma iniciativa de avaliação conjunta para o português». In Amália Mendes & Tiago Freitas (orgs.), *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística (APL 2002)* (Porto, 2-4 de Outubro de 2002), Lisboa: APL, pp. 693-704, 2003.
- Silberztein, Max, *Dictionnaires électroniques et analyse lexicale du français. Le système INTEX*. Paris, Masson, 1993.