

Tutorial introdutório do Port4NooJ

Anabela Barreiro

[última actualização 21 de Novembro de 2008]

1 INTRODUÇÃO

O Port4NooJ é um conjunto de recursos desenvolvido no ambiente linguístico [NooJ](#) para o processamento automático do português. É constituído por dicionários e gramáticas, que podem também ser utilizados em aplicações como a tradução automática do português para o inglês. O dicionário resulta da conversão e adaptação dos recursos lexicais do sistema de tradução automática [OpenLogos](#). Para informações mais detalhadas acerca destes recursos, recomenda-se a leitura do artigo [[Barreiro, 2008](#)] ou do documento **Port4NooJ Linguistic Resources Overview**, que pode ser consultado na página da internet do Port4NooJ.

2 RECURSOS

Antes de utilizar os recursos do Port4NooJ, é necessário instalar a plataforma linguística do NooJ. Tanto a plataforma, como o pacote de recursos do Port4NooJ são descarregáveis a partir da página do [NooJ](#). Para descarregar a plataforma, clique em **Download**. Para descarregar o pacote de recursos, clique em **Resources** e seguidamente seleccione a língua, **Portuguese**. O pacote de recursos pode ser igualmente obtido a partir do sítio da [Linguateca](#).

O **Pacote de Recursos do Port4NooJ** inclui os seguintes ficheiros:

1. *PT-Dict.dic* – um dicionário de palavras da linguagem corrente que contém mais de 40.000 entradas simples (lemas) com as seguintes propriedades:
 - Informação acerca da categoria gramatical (nome/substantivo, verbo, adjectivo, advérbio, determinante, pronome, preposição, conjunção e expressão numérica);
 - Informação acerca do paradigma flexional;

- Informação acerca do paradigma derivacional;
 - Classe semântica como “nome comum, animal, mamífero (*cão*); “nome concreto, roupa” (*vestido*); “nome comum, agente, entidade geográfica” (*cidade*); informação sintáctica “verbo intransitivo de movimento” (*sair*); “adjectivo, país” (*português*); “advérbio de modo” (*adequadamente*), etc.;
 - Tradução em inglês de cada entrada lexical.
2. *PT-Morph.nof* – uma amostra das regras morfológicas do português (regras de flexão e de derivação);
 3. *PT-Contr.nom* – uma gramática morfológica para o processamento de contracções, tais como *das* = Prep(*de*) + Det(*as*) ou *neste* = Prep(*em*) + DemPron(*este*);
 4. *PT-Disamb-V-Pron.nog* – uma gramática para a anotação e desambiguação de pronomes a seguir a verbos e antes de verbos que ocorrem a seguir ao pronome relativo *que*;
 5. *PT-NE-PersonType.nog* – uma gramática para o reconhecimento e anotação de entidades mencionadas do tipo PESSOA.
 6. *PT2EN-Dates.nog* – uma gramática sintáctica para o reconhecimento e tradução de datas de português para inglês, tais como “*terça-feira, 27 de Fevereiro*”;
 7. Duas amostras de corpos:
 - A versão portuguesa da *Declaração Universal dos Direitos Humanos*
 - A obra *Viagens na Minha Terra* de Almeida Garret

3 INSTALAÇÃO DO NOOJ E APLICAÇÃO DOS RECURSOS

Abra a aplicação do NooJ em **NooJ/_App**, e comece a explorar a plataforma. É aconselhável a leitura do **manual do NooJ** como complemento deste tutorial introdutório antes de começar a usar o Port4NooJ. Este manual vem com a versão instalável do NooJ e ensina-lhe como utilizar a ferramenta, efectuar pesquisas, etc.

Pode também consultar os tutoriais e artigos de referência localizados no mesmo sítio onde descarrega o software NooJ.

Após a instalação do NooJ, os recursos do Port4NooJ podem ser copiados para a directoria **NooJ/pt**. Esta directoria contém duas pastas: **Lexical Analysis** e **Syntactic Analysis**. A pasta intitulada Lexical Analysis destina-se a dicionários e ficheiros de regras de flexão e derivação. A pasta intitulada Syntactic Analysis destina-se a gramáticas sintácticas. O Port4NooJ já traz os ficheiros de regras flexionais e derivacionais incorporados, assim como as regras de contracção de preposições com artigos e pronomes. Estes recursos podem ser aplicados no tratamento de vários tipos de fenómenos linguísticos. Por exemplo, o NooJ permite descrever e analisar aspectos relacionados com a ortografia, morfologia, e sintaxe local, estrutural e transformacional, mas também criar dicionários especializados, terminologias e fazer tradução automática.

3.1 Como visualizar os recursos do Port4NooJ no NooJ

3.1.1 Visualização de dicionários

Para visualizar um dicionário, vá a **File > Open > Dictionary** e clique no dicionário *PT-Dict.dic* que descarregou com o Pacote de Recursos do Port4NooJ e que está localizado na pasta dos recursos lexicais: **Lexical analysis/**. O dicionário aparece visualizado como uma lista, tal como pode ver na [Figura 1](#).

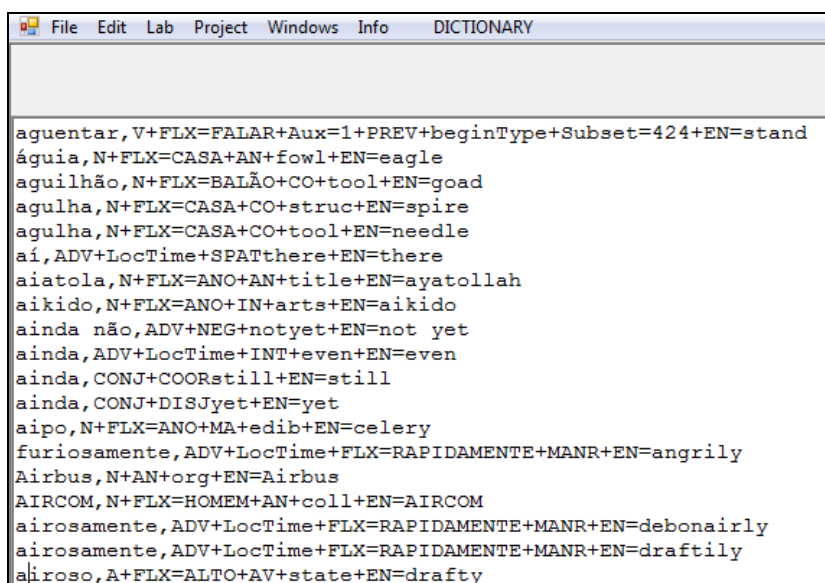
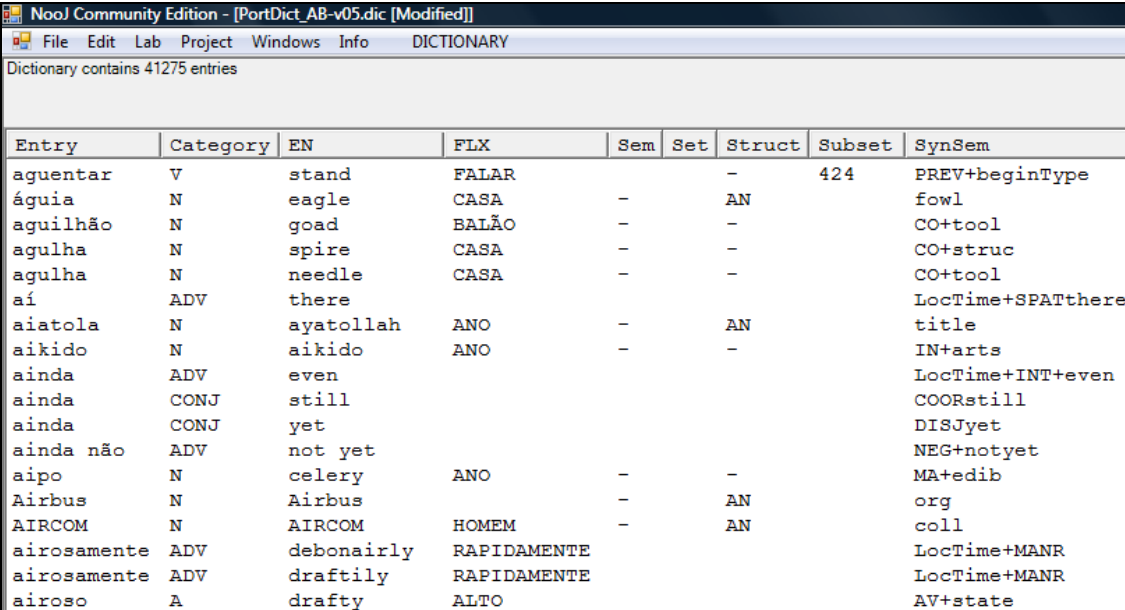


Figura 1: Dicionário visualizado como lista

Se clicar em **DICTIONARY > View as table**, pode visualizar o dicionário como uma tabela, como mostra a Figura 2. Não é possível modificar o dicionário quando este está em formato tabela.



| Entry | Category | EN | FLX | Sem | Set | Struct | Subset | SynSem |
|-------------|----------|------------|-------------|-----|-----|--------|--------|-------------------|
| aguentar | V | stand | FALAR | | | - | 424 | PREV+beginType |
| águia | N | eagle | CASA | - | | AN | | fowl |
| aguilhão | N | goad | BALÃO | | | - | | CO+tool |
| agulha | N | spire | CASA | - | | - | | CO+struc |
| agulha | N | needle | CASA | - | | - | | CO+tool |
| aí | ADV | there | | | | | | LocTime+SPATthere |
| aiatola | N | ayatollah | ANO | | | AN | | title |
| aikido | N | aikido | ANO | - | | - | | IN+arts |
| ainda | ADV | even | | | | | | LocTime+INT+even |
| ainda | CONJ | still | | | | | | COORstill |
| ainda | CONJ | yet | | | | | | DISJyet |
| ainda não | ADV | not yet | | | | | | NEG+notyet |
| aipo | N | celery | ANO | - | | - | | MA+edib |
| Airbus | N | Airbus | | | | AN | | org |
| AIRCOM | N | AIRCOM | HOMEM | - | | AN | | coll |
| airosamente | ADV | debonairly | RAPIDAMENTE | | | | | LocTime+MANR |
| airosamente | ADV | draftily | RAPIDAMENTE | | | | | LocTime+MANR |
| airoso | A | drafty | ALTO | | | | | AV+state |

Figura 2: Dicionário visualizado como tabela

Para voltar à visualização do dicionário como lista, clique em **DICTIONARY > View as list**. Só neste formato é que o dicionário pode ser modificado.

Se desejar ver o dicionário com todas as formas flexionadas, deverá gerar primeiro esse dicionário, que tem a extensão **.flx**. Clique em **Lab > Dictionary**. Surge um painel de controlo onde deve seleccionar o dicionário a ser compilado, como mostra a Figura 3.

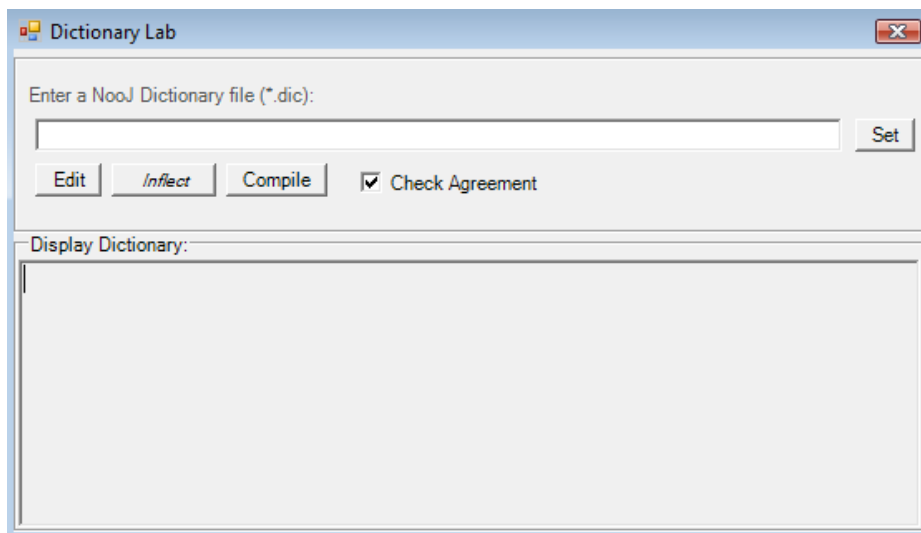


Figura 3: Painel de controlo do *Dictionary Lab*

Clique em **Set** e escolha o dicionário da pasta **Lexical Analysis**, a partir da janela que mostra onde se encontram os dicionários. O caminho para o dicionário surge na linha a seguir a “Enter a NooJ Dictionary file”. Seguidamente, clique no comando **Inflect**. Enquanto o processo de flexão está a ser realizado, pode acompanhá-lo no fundo da página (letras vermelhas e quadrados evolutivos azuis). Quando o processo estiver terminado pode fechar a janela do painel de controlo. Se desejar visualizar o dicionário de formas flexionadas deverá fazer **File > Open > Dictionary** e abrir o dicionário com a extensão **.flx**. Se vir formas erradas, pode corrigi-las no dicionário de lemas. Na **Secção 5**, será mencionada a forma como melhor pode contribuir para a correcção e aperfeiçoamento dos recursos públicos.

3.1.2 Visualização de gramáticas

Para visualizar uma gramática, vá a **File > Open > Grammar** e escolha uma das gramáticas que se encontra na pasta **Syntactic analysis/**. Experimente seleccionar a gramática *PT-NE-PersonType*. Para ver a estrutura completa da gramática, clique em **Grammar > Show structure**. Pode clicar em cada um dos subgrafos para visualizar o conteúdo de cada um deles. A janela que mostra a estrutura interna do grafo tem três comandos: **Expand**, **Collapse** e **Refresh**. O comando **Expand** permite visualizar a estrutura completa da gramática. O comando **Collapse** permite esconder todos os subgrafos. E o comando **Refresh** permite refrescar. É um comando útil quando se está

a alterar uma gramática. A **Figura 4** ilustra a gramática *PT-NE-PersonType* com a sua estrutura representada do lado direito.

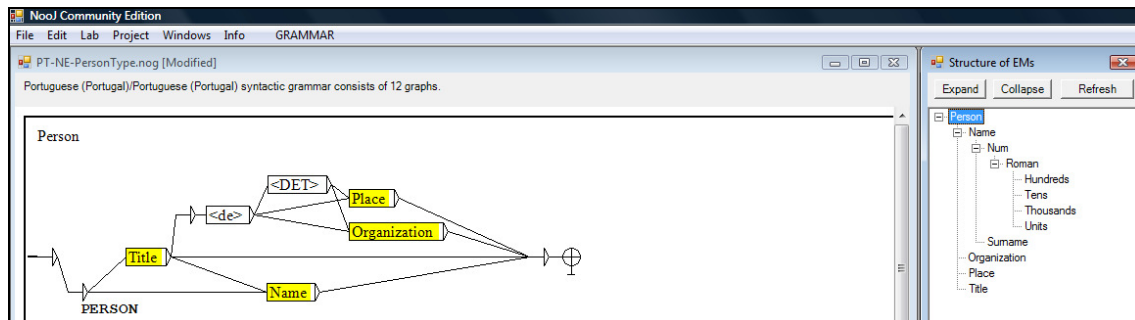


Figura 4: Visualização de uma gramática e da sua estrutura

3.1.3 Visualização de regras

Para visualizar regras e gramáticas morfológicas, vá a **File > Open > Grammar**. A janela para a escolha da gramática abre-se automaticamente na pasta **Syntactic analysis/**. Recue para a directoria **pt/** e seleccione a pasta **Lexical analysis/**, onde se encontram todas as regras e gramáticas morfológicas. Todos os ficheiros relacionados com a morfologia têm a extensão **.nof**, à excepção da gramática de contracções, que tem a extensão **.nom**. Pode visualizar o ficheiro com a amostra das regras flexionais e derivacionais, tal como ilustra a **Figura 5**. No entanto, o Port4NooJ vem já com toda a flexão incorporada. O ficheiro serve apenas para fins didácticos.

```

NooJ Community Edition - [PTSample.nof]
File Edit Lab Project Windows Info

# Example of a noun inflection paradigm

# plural adding -s
MENINO = <E>/m+s + s/m+p ;

# Example of a verb inflection paradigm

# regular ending in -ar (1st conjugation), like "falar"
FALAR = <E>/INF +
  <E>/INFI+1+s + es/INFI+2+s + <E>/INFI+3+s + mos/INFI+1+p + des/INFI+2+p + em/INFI+3+p +
  <B>do/PP+m+s + <B>dos/PP+m+p + <B>da/PP+f+s + <B>das/PP+f+p +
  <B>ndo/G +
  <B2> (o/PR+1+s + as/PR+2+s + a/PR+3+s + amos/PR+1+p + ais/PR+2+p + am/PR+3+p) +
  <B2> (ei/PS+1+s + aste/PS+2+s + ou/PS+3+s + ámos/PS+1+p + astes/PS+2+p + aram/PS+3+p) +
  <B> (va/PI+1+s + vas/PI+2+s + va/PI+3+s + <B>ávamos/PI+1+p + <B>áveis/PI+2+p + vam/PI+3+p) +
  <B2> (ara/PMP+1+s + aras/PMP+2+s + ara/PMP+3+s + áramos/EMP+1+p + áreis/PMP+2+p + aram/PMP+3+p) +
  ei/F+1+s + ás/F+2+s + á/F+3+s + emos/F+1+p + eis/F+2+p + ão/F+3+p +
  ia/C+1+s + ias/C+2+s + ia/C+3+s + iamos/C+1+p + íeis/C+2+p + iam/C+3+p +
  <B2> (e/PRS+1+s + es/PRS+2+s + e/PRS+3+s + emos/PRS+1+p + eis/PRS+2+p + em/PRS+3+p) +
  <B2> (asse/PIS+1+s + asse/PIS+2+s + asse/PIS+3+s + ássemos/PIS+1+p + ásseis/PIS+2+p + assem/PIS+3+p) +
  <B2> (ar/FS+1+s + ares/FS+2+s + ar/FS+3+s + armos/FS+1+p + ardes/FS+2+p + arem/FS+3+p) +
  <B2> (a/IMP+2+s + e/IMP+2+s + emos/IMP+1+p + ai/IMP+2+p + em/IMP+3+p) ;

```

Figura 5: Exemplos de regras flexionais

Pode também experimentar visualizar a gramática de contrações que se encontra na mesma pasta, *PT-Contr.nom*. Visualize a estrutura completa da gramática da mesma forma que visualiza as restantes gramáticas, conforme representado na [Figura 6](#).

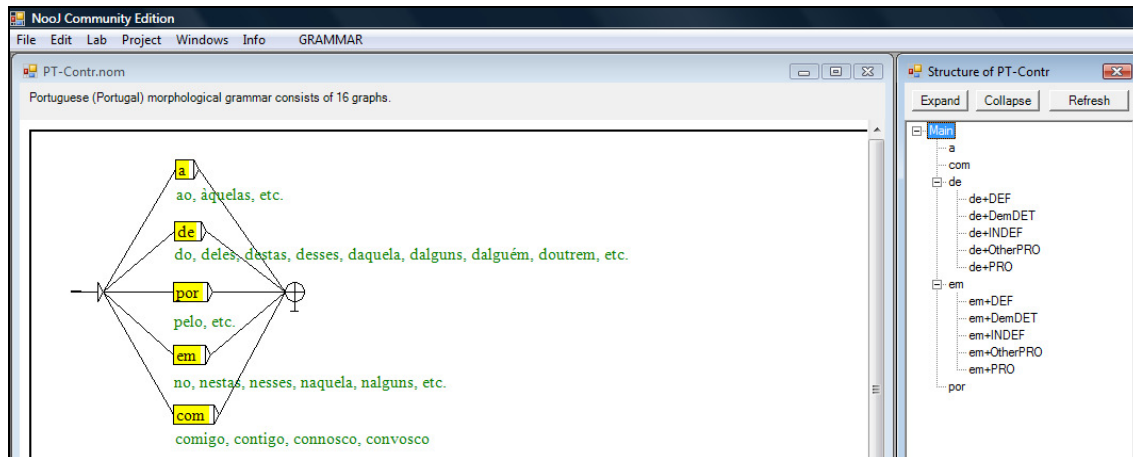


Figura 6: Visualização da gramática de contrações com a sua estrutura

3.1.4 Visualização de textos com informação pré-processada

Experimente abrir o texto da Declaração Universal dos Direitos Humanos, um dos textos que vem no pacote de recursos do Port4NooJ. Clique em **Open > Text**. O sistema apresenta-lhe a janela da pasta **_Projects**, onde todos os textos devem ser guardados. Selecione o texto. A [Figura 7](#) apresenta o texto seleccionado.

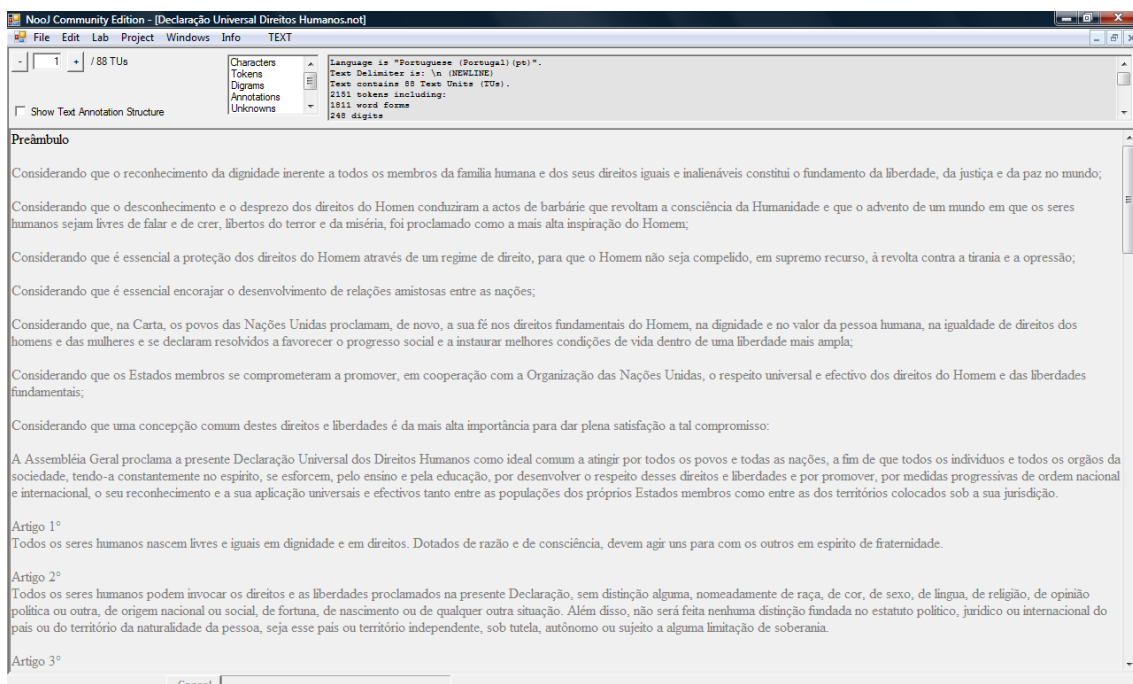


Figura 7: Texto com as suas informações gerais

Na parte superior do ecrã, pode encontrar vários tipos de informação acerca deste texto. Na janela cinzenta encontra as características principais do texto. Na janela branca, pode clicar em cada um dos elementos da lista e ver os detalhes de cada um deles. Pode ver o número de *tokens* ou digramas, por ordem alfabética ou por frequência no texto. Pode ver todas as anotações criadas para esse texto, criar um ficheiro com todas as palavras desconhecidas (ou seja, palavras que não se encontram nos dicionários locais). Pode visualizar as palavras ambíguas e as não ambíguas. Se clicar em **Show Text Annotation Structure**, no canto superior esquerdo, pode ver as anotações vindas do dicionário, frase a frase. A Figura 8 apresenta as anotações para a frase seleccionada (TU 21).

| Token | Annotation |
|----------------|--------------------------------------|
| todos | o,DET+m+p+DEFART+the+EN=the |
| os | o,N+m+p+UNalnum+EN=o |
| os | os,PRO+PERS+OD+3+p+m+EN=them |
| indivíduos | indivíduo,N+m+p+AN+Hum+EN=bloke |
| indivíduos | indivíduo,N+m+p+AN+Hum+EN=fellow |
| indivíduos | indivíduo,N+m+p+AN+Hum+EN=individual |
| têm | direito,A+m+s+AV+loc+EN=righthand |
| têm | direito,A+m+s+AV+state+EN=blunt |
| têm | direito,A+m+s+AV+state+EN=straight |
| têm | direito,A+m+s+MA+EN=right |
| têm | direito,A+m+s+MA+EN=rightmost |
| têm | direito,A+m+s+NAVpred+EN=right-hand |
| têm | direito,N+m+s+AB+cause+EN=right |
| ao | o,DET+m+s |
| reconhecimento | reconhecer,N+N |
| reconhecimento | reconhecer,N+N |
| reconhecimento | reconhecer,N+N |
| reconhecimento | reconhecer,N+N |
| reconhecimento | reconhecer,N+N |
| reconhecimento | reconhecer,N+N |
| reconhecimento | reconhecer,N+N |
| reconhecimento | reconhecimento |

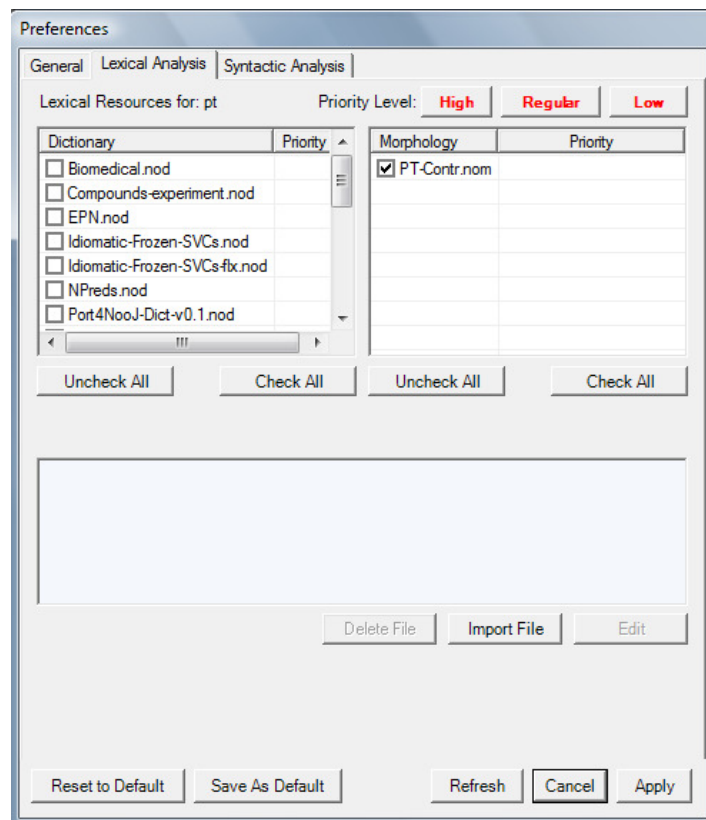
Figura 8: Texto com anotações do dicionário

Como não foi aplicada nenhuma gramática de desambiguação ao texto, todas as análises do dicionário associadas a uma dada forma encontram-se representadas nas anotações. Na [Secção 3.3](#), será possível ver a aplicação de gramáticas de desambiguação ao texto.

Se pretender, pode alterar o texto ou exportá-lo. Para alterar o texto, clique em **TEXT > Modify text**. Para exportar o texto, clique em **TEXT > Export annotated text as an XML document**. Consulte o manual do NooJ.

3.2 Aplicação de recursos aos textos

O NooJ funciona de acordo com um número de parâmetros que se estabelecem no painel de controlo **Info > Preferences**. Nesse painel podem ser especificadas as preferências por omissão, como a língua de trabalho, o tipo de fonte, que recursos lexicais e sintáticos são aplicados para cada língua, etc. A [Figura 9](#) mostra o painel de controlo **Preferences**.



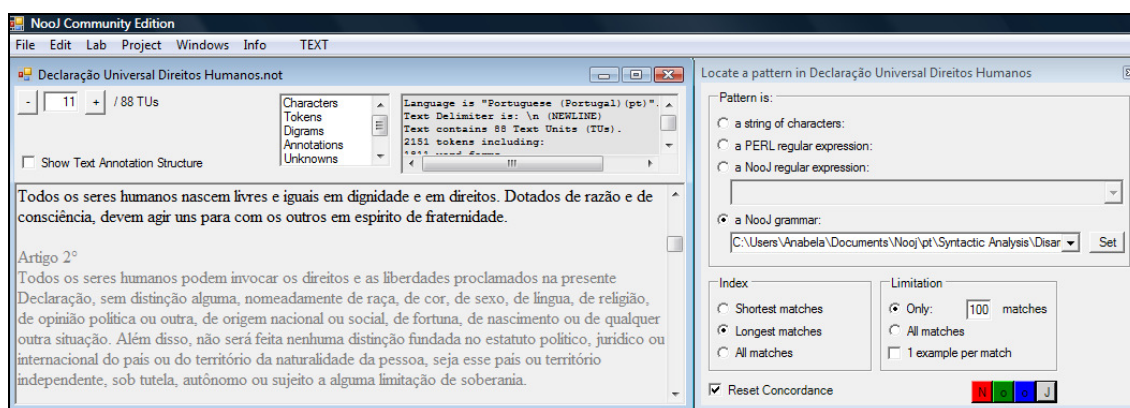
[Figura 9](#): Painel de controlo das preferências

3.2.1 Aplicação de gramáticas

As gramáticas podem ser aplicadas aos textos para a identificação e análise de vários tipos de fenómenos linguísticos: reconhecimento e anotação de padrões morfológicos, lexicais e sintático-semânticos; identificação e extracção de unidades semânticas, tais

como datas, entidades mencionadas e unidades terminológicas; reconhecimento e anotação de palavras e unidades multipalavra; identificação de constituintes sintácticos, tais como sintagmas nominais e outros constituintes; extracção de relações semânticas e desambiguação. Também podem ser especificadas gramáticas para o reconhecimento, parafraseamento e tradução de fenómenos linguísticos como construções com verbos suporte, do tipo *'tomar uma decisão'*. Nas [Secções 4.2 e 4.3](#) serão ilustrados estes elementos.

A título de exemplificação, para aplicar uma gramática de desambiguação ao texto apresentado na [Figura 8](#), clique em **TEXT > Locate** (tem que ter o cursor no texto). No painel de controlo, seleccione a gramática de desambiguação, clicando em **a NooJ grammar** e seguidamente em **Set**, tal como ilustra a [Figura 10](#). Neste mesmo painel, pode também pesquisar padrões. A pesquisa pode ser efectuada através de expressões regulares de Perl ou NooJ (consulte o manual do NooJ).



[Figura 10](#): Painel de controlo para pesquisar padrões ou aplicar gramáticas

Após seleccionar uma gramática específica, o sistema abre a janela onde estão os ficheiros das gramáticas na pasta **Syntactic analysis/**. Seleccione a gramática de exemplo que vem no **Pacote de Recursos do Port4NooJ**, *PT-Disamb-V-Pron* e clique num dos botões da palavra NooJ, o que tiver a cor da sua preferência. Surge-lhe no ecrã uma janela que contém uma concordância em que a sequência reconhecida e desambiguada deverá aparecer. Se aparecerem sequências incorrectas, isso significa que a gramática deve ser refinada. Pode seleccionar as sequências indesejadas, clicar no botão direito do seu rato e eliminar essas sequências, fazendo **Filter out selected lines**. Pode também exportar a concordância (**Export Concordance**). [A Figura 11](#)

apresenta uma concordância resultante da aplicação da gramática de resolução de ambiguidades. Os padrões encontrados apresentam estruturas já cobertas pela gramática. O último padrão foi mal reconhecido e pode/deve ser eliminado da concordância.

| Text | Before | Seq. | After |
|---|---|---|-------|
| advento de um mundo em uma liberdade mais ampla; Considerando acusada de um acto delituoso de abandonar o país em dos poderes públicos: e deve uma remuneração equitativa e satisfatória, progresso científico e nos benefícios | que os seres que os Estados presume-se que se encontra exprimir-se que lhe permita que deste resultam | humanos sejam livres de falar membros se comprometeram a promover inocente até que a sua , incluindo o seu, e o através de eleições honestas a e à sua família uma . 2. Todos têm direito à protecção | |

Query 7/7

Figura 11: Concordância resultante de aplicação de gramática de desambiguação

3.2.2 Aplicação de dicionários

A aplicação de dicionários pode ser especificada no painel de controlo **Preferences**. Em **Lexical Analysis** pode incluir todos os seus dicionários, dando-lhe uma ordem de prioridade. A ordem de prioridade deve ser definida de acordo com as características do seu texto. Pode também optar por não seleccionar determinado(s) dicionário(s) para a análise de textos específicos, se assim achar conveniente. Quando altera os seus recursos, deve primeiro compilar o dicionário ou dicionários e depois fazer nova análise linguística do texto. Para compilar dicionários, clique em **Lab > Dictionary**. Surge o painel de controlo do **Dictionary Lab**, ilustrado na [Figura 3](#) na [Secção 3.1.1](#). Nesse painel de controlo, deve seleccionar o dicionário a ser compilado. Para isso, clique em **Set** e escolha o dicionário a partir da janela da pasta **Lexical Analysis**, onde se encontram os dicionários. O caminho para o dicionário surge na linha a seguir a “Enter a NooJ Dictionary file”. Seguidamente, clique em **Compile**. Enquanto a compilação está a ser realizada, pode ver o processo a decorrer no fundo da página. Se o dicionário não contiver erros, a compilação decorrerá com sucesso. Quando o processo de compilação terminar pode fechar a janela do painel de controlo. O dicionário está compilado! Se o dicionário contiver erros, o processo de compilação é interrompido. Surge uma nova janela com os erros listados a vermelho. Corrija os erros que surgem nessa janela no dicionário. Antes da nova compilação, pode/deve alinhar o

dicionário alfabeticamente e verificar o formato. Para alinhar o dicionário, (com o dicionário aberto) deve clicar em **DICTIONARY > Sort**. Quando o alinhamento alfabético tiver terminado aparece uma pequena janela com a informação “Dictionary sorted”. Clique em OK. Para verificar o formato, deve clicar em **DICTIONARY > Check format**. Se não existirem problemas de formato, imediatamente aparece a janela com a mensagem “No problem found”. Clique em OK.

Quando o seu dicionário estiver compilado e pretender aplicar as alterações ao texto, com o texto aberto, deve clicar em **TEXT > Linguistic Analysis**, para que os novos recursos passem a ter efeito nos textos. O processo de análise linguística pode ser acompanhado no final da página (a letras vermelhas). A seguir pode localizar padrões no seu texto já pré-processado e aplicar gramáticas.

3.3 Experimente você!

Apesar do Port4NooJ já conter alguns recursos que pode utilizar livremente, você é incentivado a construir recursos que respondam às suas necessidades específicas. Pode criar os seus próprios corpora, dicionários e gramáticas.

3.3.1 Use os seus próprios textos

Pode utilizar os textos que desejar, no formato que desejar. Consulte o manual do NooJ para ter conhecimento acerca de formatos e outra informação relacionada com os textos. Recomenda-se que coloque todos os seus ficheiros de corpora na directoria **_Projects**.

3.3.2 Construa os seus próprios dicionários

Pode criar novos dicionários através do comando **File > New > Dictionary**. Os dicionários NooJ estão associados a uma única língua de partida (input language), mesmo os que contêm mais do que uma língua de chegada (output language). Tem que especificar qual a sua língua de partida quando criar um novo dicionário. Para os dicionários portugueses tem que seleccionar a língua **pt**, conforme ilustra a [Figura 12](#).

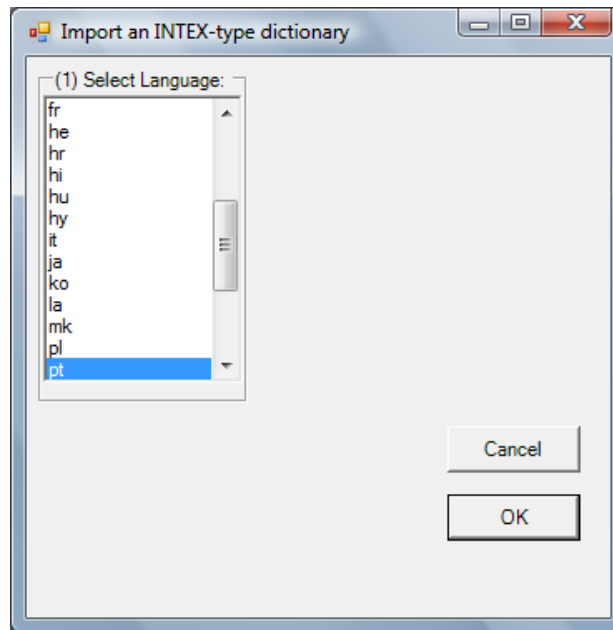


Figura 12: Painel de controlo para selecção da língua do dicionário

3.3.3 Construa as suas próprias gramáticas

Pode criar novas gramáticas através do comando **File > New > Grammar**. Surgir-lhe-á o painel de controlo para a criação de uma nova gramática, conforme ilustra a Figura 13.

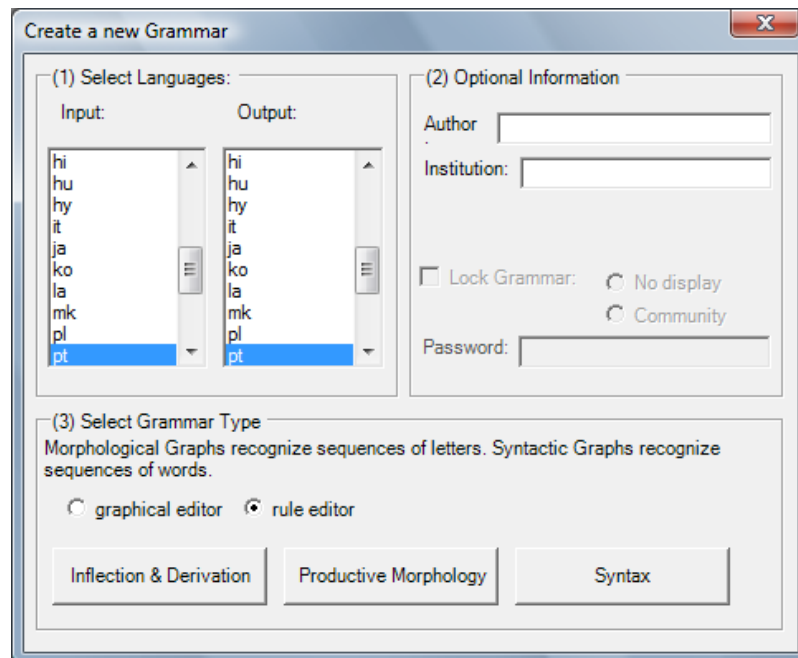


Figura 13: Painel de controlo para criação de uma nova gramática

Pode seleccionar o modo de construção da gramática: gráfico ou textual (consulte o manual NooJ). Os comandos ao fundo do painel servem para especificar o tipo de

gramática. Como o Port4NooJ já integra flexão e derivação, as gramáticas que criar para o português, quer através de gráficos, quer textualmente, serão essencialmente gramáticas sintáticas.

4 ALGUNS EXEMPLOS DE APLICAÇÕES

4.1 Etiquetagem de fenómenos linguísticos

Os recursos linguísticos do Port4NooJ servem, entre várias outras aplicações, para a anotação de unidades linguísticas, tais como entidades mencionadas do tipo das que a [Figura 14](#) apresenta.

| | | |
|---------------------|---|---|
| certo ponto, o | Dom Quixote/ PERSON+Title+Name | da sociedade velha. |
| o. O que eu fiz na | Dona Branca/ PERSON+Title+Name | é pouco e mal esboçado à pressa. O grande mago lusitano |
| A | Dona Branca/ PERSON+Title+Name | três, Frei Soeiro, Frei Lopo e S. Frei Gil - faz quatro |
| esús, logo o das | Donas/ PERSON+Title | , depois o de S. Domingos, célebre pelo jazigo do no |
| - S. Frei Gil e o | Doutor Fausto/ PERSON+Title+Name | . - De como o A. foi ao túmulo do santo bruxo e o achou va |
| antes e depois do | Doutor Fausto/ PERSON+Title+Name | . Mas sem Homero ou Goethe é que se não chega à rep |
| amei já o nosso | Doutor Fausto/ PERSON+Title+Name | : e é com efeito. Não lhe falta senão o seu Goethe. |
| vê que o nosso | doutor/ PERSON+Title | de bivaque, o soldado que lhe chamou maluco ao pensad |
| impressão era a | Duquesa de Abrantes/ PERSON+Title+Place | . Mas em meia hora de conversação, de trato, descobriam- |
| que lhe deram. - A | Duquesa de Abrantes/ PERSON+Title+Place | . - Chega-se enfim ao Vale de Santarém. |
| adíssimo bobo de | el-rei de Dinamarca/ PERSON+Title+Place | , o que alguns anos depois ressuscitou em Sterne com tã |
| e e o Condestável, | el-rei D. Fernando/ PERSON+Title+Name | e a Rainha D. Leonor, Camões desterrado aqui, Frei Lu |
| meiro-ministro de | el-rei D. José/ PERSON+Title+Name | ? Por onde está Ixião e Tântalo, por onde demora Sísifo |
| seu augusto amo | el-rei de Dinamarca/ PERSON+Title+Place | . Por pouco mais que se generalize o princípio, |
| s promessas - como | el-rei de Prússia/ PERSON+Title+Place | prometeu uma constituição; e não faltou ainda, porque, |
| o.- De como S. M. | El-Rei de Dinamarca/ PERSON+Title+Place | tinha menos juízo do que Yorick, seu bobo. - Doutrina deste |
| a, em que domina | el-rei Sancho/ PERSON+Title+Name | Depois há |
| ssível. Desenganado | el-rei/ PERSON+Title | de que um poder sobre-humano não permitia que els |
| ho, beijou a mão à | el-rei/ PERSON+Title | , e daí tomou um dia o caminho de Santarém, chegou àquel |
| nosso Carlos; e o | enfermeiro/ PERSON+Title | que o velava, uma bela mulher de estatura não acima de |
| ernada elegância do | enfermeiro/ PERSON+Title | que o velava. O quarto era com efei |
| - O hospital. - O | enfermeiro/ PERSON+Title | . - Georgina. |
| parte da nossa | Espanha/+Place | é, geologicamente falando, já tão África, tão pouco |
| rase didática das | Espanhas/+Place | chamou romance em endechas. Eu, adotando para ele, |

Figura 14: Reconhecimento e anotação de entidades mencionadas do tipo PESSOA

A concordância apresentada na figura resulta da aplicação de uma gramática de reconhecimento de entidades mencionadas. As entidades mencionadas são reconhecidas e anotadas de acordo com o seu tipo. Por exemplo, *Dom Quixote* e *Doutor Fausto* são anotados como PERSON+Title+Name porque tanto *Dom* como *Doutor* são títulos de pessoas, e *Quixote* e *Fausto* são nomes; *Duquesa de Abrantes* e *El-Rei de Dinamarca* são anotados como PERSON+Title+Place porque se referem a pessoas que desempenham um cargo oficial no local que também contribui para o nome do seu título.

4.2 Paráfrases

O parafraseamento de expressões é outra das aplicações dos recursos linguísticos do Port4NooJ. A concordância apresentada na [Figura 15](#) mostra o parafraseamento de construções com verbos suporte em verbos lexicais.

| Text | Before | Seq. | After |
|---------------------------------------|-------------------------------------|------|---------------------------------------|
| militar, para que Edward pudesse | dar início à/iniciar | | carreira de clínico geral. Quando |
| militar, para que Edward pudesse | dar início à/começar | | carreira de clínico geral. Quando |
| elegante, se me é permitido | dar uma opinião/ter uma opinião | | . Bernard aproveitou para dar uma |
| elegante, se me é permitido | dar uma opinião/dar uma opinião | | . Bernard aproveitou para dar uma |
| elegante, se me é permitido | dar uma opinião/opinar | | . Bernard aproveitou para dar uma |
| Mr. Walsh. -- Quanto se deve | dar de gorjeta/gratificar | | ? -- O que quis dizer foi |
| sinto que, se puder, devo | dar uma ajuda ao/ajudar | | Council. Um deles abanou a |
| servia isso mandaram-no ir | dar uma volta/passear | | . Uma mulher com a saia |
| sugestão de Persse, começa a | dar exemplos/exemplificar | | : O Espelho da Sinceridade (Péricles |
| parecesse irracional impedia-a de | dar voz ao/ouvir | | medo. Ainda não eram cinco |
| parecesse irracional impedia-a de | dar voz ao/escutar | | medo. Ainda não eram cinco |
| e sentia-se capaz de | dar um murro em/esmurrar | | quem tentasse detê-lo. Mas |
| seu carrossel. Gostávamos de lhe | dar uma palavrinha/falar | | . -- Então vamos dar um aperto |
| dar uma palavrinha. -- Então vam... | dar um aperto de/apertar | | mão. «Deixar que o seu |
| múltiplos relatórios, «poderia não... | dar lugar a uma/originar | | colisão qualquer, mas quem sabe |
| múltiplos relatórios, «poderia não... | dar lugar a uma/causar | | colisão qualquer, mas quem sabe |
| putrefacto? Às vezes ouvia-a | dar instruções a/instruir | | Julie acerca das compras ou |
| o problema do altruísmo, podia | dar sempre origem a/originar | | confusões. O Oliver gosta de |
| ligado a um tractor. -- Posso | dar uma olhadela/olhar | | ? -- perguntou Zoe. E diz ao |
| pai que gostava de lhe | dar uma palavrinha/falar | | , está bem? Harriet levantou um |
| quando eu não estiver a | dar aulas/leccionar | | , e nos fins-de-semana |
| quando eu não estiver a | dar aulas/ensinar | | , e nos fins-de-semana |
| acenou com o charuto para | dar ênfase às/enfatizar | | suas palavras. «Gostaria primeiro de |
| Safie, Agatha e Félix foram | dar um passeio pelo/passear | | campo e deixaram o velho |
| Safie, Agatha e Félix foram | dar um passeio pelo/andar | | campo e deixaram o velho |
| e fui incapaz de lhe | dar resposta/responder | | . Apenas me preocupava com um |
| declarou-lhe. Não precisa de | dar explicações/explicar | | . Reza fita-me como se |
| brancos. -- Se me permite, posso | dar uma sugestão/fazer uma sugestão | | . |
| brancos. -- Se me permite, posso | dar uma sugestão/dar uma sugestão | | . |
| brancos. -- Se me permite, posso | dar uma sugestão/sugerir | | . |

[Figura 15](#): Parafraseamento de construções com verbos suporte em verbos lexicais

4.3 Tradução automática

A tradução automática é outra aplicação dos recursos linguísticos do Port4NooJ. A concordância apresentada na [Figura 16](#) mostra a tradução de português para o inglês de construções com verbos suporte.

| Text | Before | Seq. | After |
|------|---------------------------------------|---------------------------|---------------------------------------|
| | a fazer um estágio para | dar aulas de/tutor | religião, mas não se importava |
| | a fazer um estágio para | dar aulas de/lecture | religião, mas não se importava |
| | a fazer um estágio para | dar aulas de/teach | religião, mas não se importava |
| | militar, para que Edward pudesse | dar inicio /start | à carreira de clínico geral. Quando |
| | militar, para que Edward pudesse | dar inicio /initiate | à carreira de clínico geral. Quando |
| | militar, para que Edward pudesse | dar inicio /commence | à carreira de clínico geral. Quando |
| | militar, para que Edward pudesse | dar inicio /begin | à carreira de clínico geral. Quando |
| | «Casamos no Natal.» -- Estou a | dar aulas para/tutor | me livrar da tropa -- explicou |
| | «Casamos no Natal.» -- Estou a | dar aulas para/lecture | me livrar da tropa -- explicou |
| | uma opinião. Bernard aproveitou para | dar uma volta /walk | pela cozinha. Ainda há-de dar |
| | Mas quem é que vai | dar um subsídio/subsidize | capaz para estudar os Oof |
| | Mas quem é que vai | dar um subsídio/subsidise | capaz para estudar os Oof |
| | a dizer-te que pode | dar a impressão de/seem | que não confio em ti |
| | sinto que, se puder, devo | dar uma ajuda /help | ao Council. Um deles abanou a |
| | sinto que, se puder, devo | dar uma ajuda /aid | ao Council. Um deles abanou a |
| | sinto que, se puder, devo | dar uma ajuda /assist | ao Council. Um deles abanou a |
| | sugestão de Persse, começa a | dar exemplos/exemplify | : O Espelho da Sinceridade (Péricles |
| | parecesse irracional impedia-a de | dar voz /hear | ao medo. Ainda não eram cinco |
| | parecesse irracional impedia-a de | dar voz /listen | ao medo. Ainda não eram cinco |
| | e sentia-se capaz de | dar um murro em/punch | quem tentasse detê-lo. Mas |
| | seu carrossel. Gostávamos de lhe | dar uma palavrinha/speak | . -- Então vamos dar um aperto |
| | dar uma palavrinha. -- Então vamos | dar um aperto de/shake | mão. «Deixar que o seu |
| | múltiplos relatórios, «poderia não só | dar lugar a/originate | uma colisão qualquer, mas quem |

Figura 16: Tradução de construções com verbos suporte para inglês

5 RECURSOS PARTILHÁVEIS

A criação de recursos partilháveis facilita o trabalho do investigador tornando-o também mais produtivo. A possibilidade de usufruir de informação/dados e recursos que não tenha ao seu dispor e o intercâmbio fortalecem o trabalho da comunidade científica. A seguir são apresentadas algumas ideias de como poderá colaborar no desenvolvimento e melhoria dos recursos partilháveis.

5.1 Aumento da cobertura

A cobertura do Port4NooJ é ainda limitada. Qualquer pessoa pode contribuir para o aperfeiçoamento e aumento do dicionário de cobertura larga ou para a criação de dicionários mais específicos (como, por exemplo, dicionários de termos de uma determinada área técnica ou científica). As necessidades mais prementes neste momento são o aumento das formas verbais e a inclusão de compostos.

O NooJ indica em cada texto quais as palavras desconhecidas e lista-as em formato de dicionário, como se pode observar através da Figura 17. Experimente abrir o texto

da *Declaração Universal dos Direitos Humanos* e clicar em **Unknowns**. Verificará que algumas dessas palavras são erros ou variantes do português do Brasil (que também deverão ser dicionarizadas e anotadas como tal), mas muitas outras são candidatos válidos a entradas do dicionário.

```
# NooJ V2
# Dictionary
#
# Input Language is: pt
#
# Alphabetical order is not required.
#
# Use inflectional & derivational paradigms' and proper
# Special Command: #use properties.def
# Special Command: #use paradigms.nof
#
# Special Features: +NW (non-word) +FXC (frozen express
#                   +FLX= (inflectional paradigm) +DRV=
#
# Special Characters: '\ ' "' '+' ',' '# ' ' '
#
agir,UNKNOWN
amistosas,UNKNOWN
Assembléia,UNKNOWN
autônomo,UNKNOWN
contrariamente,UNKNOWN
cruéis,UNKNOWN
dar,UNKNOWN
delituoso,UNKNOWN
dentro,UNKNOWN
econômicos,UNKNOWN
escolher,UNKNOWN
fazer,UNKNOWN
filiar,UNKNOWN
fique,UNKNOWN
fruir,UNKNOWN
Homen,UNKNOWN
idéias,UNKNOWN
inerente,UNKNOWN
```

Figura 17: Listas de palavras desconhecidas – potenciais candidatos a entradas lexicais

5.2 Revisão dos actuais recursos

O Port4NooJ foi desenvolvido com base em recursos adaptados do inglês. Existem erros de herança. Alguns dos erros mais comuns são: entradas com traduções incorrectas, sobregeração de formas, atribuição de um paradigma flexional incorrecto, erros ortográficos ou tipográficos, flexão incorrecta de compostos e unidades multipalavra, especificação de morfemas incorrectos para as formas superlativas, diminutivas, etc., atribuição de paradigma flexional a palavras invariáveis, atribuição de paradigma flexional incorrecto às palavras derivadas, entre outros.

A detecção e relato de erros proporciona a sua correcção e, conseqüentemente, o aperfeiçoamento dos recursos utilizados no pré-processamento dos textos. Os recursos partilháveis serão alvo de uma avaliação continuada. Se encontrar erros ou

inconsistências nos recursos ou tenha sugestões ou comentários a fazer, por favor contacte-nos.

5.3 Criação de novas gramáticas

A criação de gramáticas monolingues e para a tradução automática é um desafio para os que queiram trabalhar em investigação, para mestrandos ou doutorandos. Existe muito trabalho interessante a fazer nesta área.

6 CONCLUSÃO

Com este breve tutorial introdutório, tentámos responder às questões mais elementares para que os novos utilizadores do NooJ possam dar início à sua actividade linguística de uma forma mais suave. Tentámos mostrar os recursos paralelamente à demonstração de algumas das funcionalidades mais populares. Um tutorial mais avançado a desenvolver futuramente responderá a questões aqui não abordadas.