

Novas Ferramentas e Recursos Linguísticos para a Tradução Automática

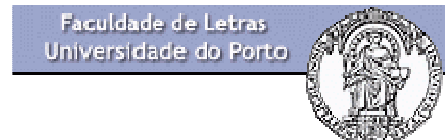
Por ocasião d'O Fim do Início de uma Nova Era
no Processamento da Língua Portuguesa

Anabela Barreiro

barreiro_anabela@hotmail.com

FLUP & CLUP-Linguatca

New York University



Projecto de Doutoramento

Objectivos:

1. Melhoramento da tradução automática através de um conhecimento linguístico sobre paráfrases
2. Obtenção de paráfrases monolingues para a edição de texto e escrita de linguagem técnica e obtenção de paráfrases bilingues para a tradução automática
3. Análise e reconhecimento de expressões multipalavra e construção de dicionários monolingues e bilingues para estas expressões

Recursos e Ferramentas Utilizados

- Dicionário inglês-português do sistema **OpenLogos**
- Ambiente de desenvolvimento linguístico **NooJ**
- Corpora anotados da **Lingueca**

COMPARA: inventariação de fenómenos linguísticos contrastivos e de criação de algumas regras parafrásticas bilingues

CETEMPúblico e outros: extracção de exemplos

- Outros recursos da **Lingueca**

METRA: avaliação da tradução automática

Corpógrafo: desenvolvimento e integração de novos recursos

Recursos e Ferramentas Criados

➤ **Port4NooJ**

Sistema baseado em ontologias lexicais construído com base em dicionários e gramáticas locais, com conteúdo sintático e semântico. O dicionário geral foi integrado no Corpógrafo e serve de alicerce para um sistema de tradução automática de português para inglês, em fase de desenvolvimento

➤ **DicTuM**

Dicionário de **T**ermos e **U**nidades ou Expressões **M**ultipalavra

➤ **ReWriter e ParaMT**

Ferramentas automáticas que permitem reconhecer expressões multipalavra e gerar paráfrases

Associações Lexicais, Sintáticas e Semânticas

Tratamento da Derivação

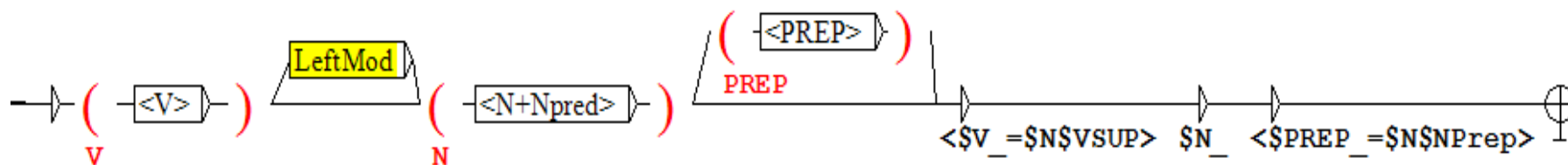
(implicações ao nível morfo-lexical, sintático e semântico)

- Identificação das **nominalizações** e **adjectivos predicativos** nas entradas verbais
- Identificação de **verbos suporte** (elementares e não elementares) que ocorrem com os nomes predicativos e de **verbos copulativos** que ocorrem com os adjectivos predicativos > variantes estilísticas
- Identificação dos paradigmas derivacionais das **adverbializações**
- Identificação de **nomes predicativos autónomos**
- Identificação de **adjectivos predicativos**

ReWriter: Características

- Gerador de Paráfrases Monolíngues (= Parafraseador Monolíngue)
- Autónomo e multifuncional
- Estado actual
 - reconhecimento e extracção de construções com verbos suporte elementares (*fazer uma operação*);
 - parafraseamento por um verbo lexical correspondente (*operar*)
 - parafraseamento por uma construção com verbo suporte não elementar equivalente (*realizar/effectuar uma operação*)
- Aplicações:
 - edição de textos
 - escrita de linguagem técnica
 - pesquisa e extracção de informação

ReWriter/ReEscreve: um Parafraseador Monolingue



gosto de ver o comboio a	fazer corridas /correr	à velocidade máxima ao long
o de cheque especial para	fazer doações /doar	às entidades que escolher. A
cores e, quando é preciso ir	fazer filmagens/filmar	fora do estúdio, às vezes fic
se queria trocar de pares e	fazer um jogo /jogar	ao melhor de três sets , mas
o dra deu-me um papel para	fazer uma lista de/listar	todas as coisas boas que ex
res foram à caracterização	fazer uns retoques/retocar	, outros estão a descansar n

Reconhecimento e parafraseamento monolingue
de construções com verbos suporte elementares
(construção com verbo suporte / verbo lexical equivalente)

ReWriter/ReEscreve: Exemplos

nça, o cirurgião Faivre, ao	fazer uma amputação	amputar	
nça, o cirurgião Faivre, ao	fazer uma amputação	efectuar	uma amputação
nça, o cirurgião Faivre, ao	fazer uma amputação	realizar	uma amputação
omista britânico, conseguiu	fazer uma transfusão de sangue	realizar	uma transfusão de sangue
omista britânico, conseguiu	fazer uma transfusão de sangue	efectuar	uma transfusão de sangue
os pacientes que precisam	fazer uma transfusão de sangue	sujeitar-se a	uma transfusão de sangue
os pacientes que precisam	fazer uma transfusão de sangue	submeter-se a	uma transfusão de sangue
os pacientes que precisam	fazer uma transfusão de sangue	realizar	uma transfusão de sangue
os pacientes que precisam	fazer uma transfusão de sangue	efectuar	uma transfusão de sangue

CVS Elementar > Verbo Lexical

CVS Elementar > CVS não elementar

realizar/efectuar

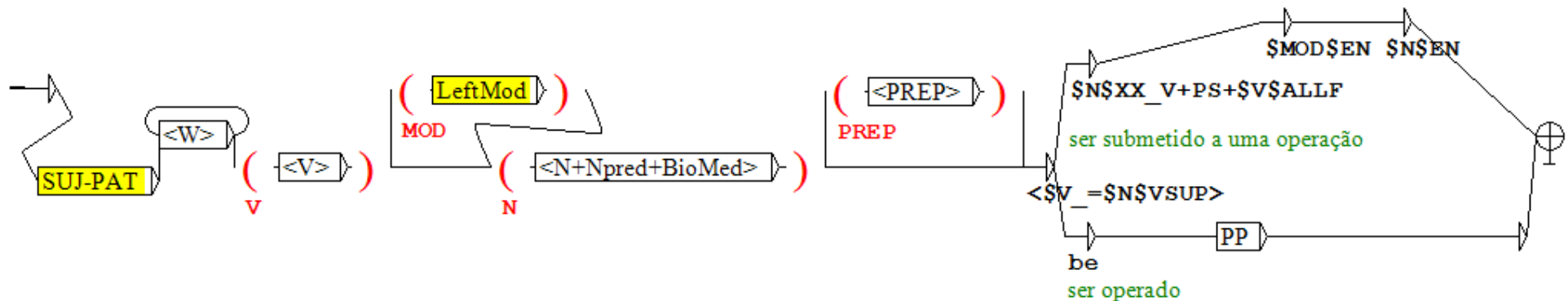
CVS Elementar > *sujeitar-se a*

submeter-se a

APENAS se o SUJEITO fôr um paciente

Reconhecimento e parafraseamento de construções com verbos suporte elementares que co-ocorrem com nomes predicativos da área biomédica

(construção com verbo suporte / verbo lexical equivalente ou variante estilística)



ReWriter/ReEscreve: Exemplo de Aplicação

1. Toda a pessoa tem o direito de tomar parte na direcção dos negócios, públicos do seu país, quer directamente, quer por intermédio de representantes livremente escolhidos.

2. Toda a pessoa tem direito de acesso, em condições iguais, às funções públicas do seu país.

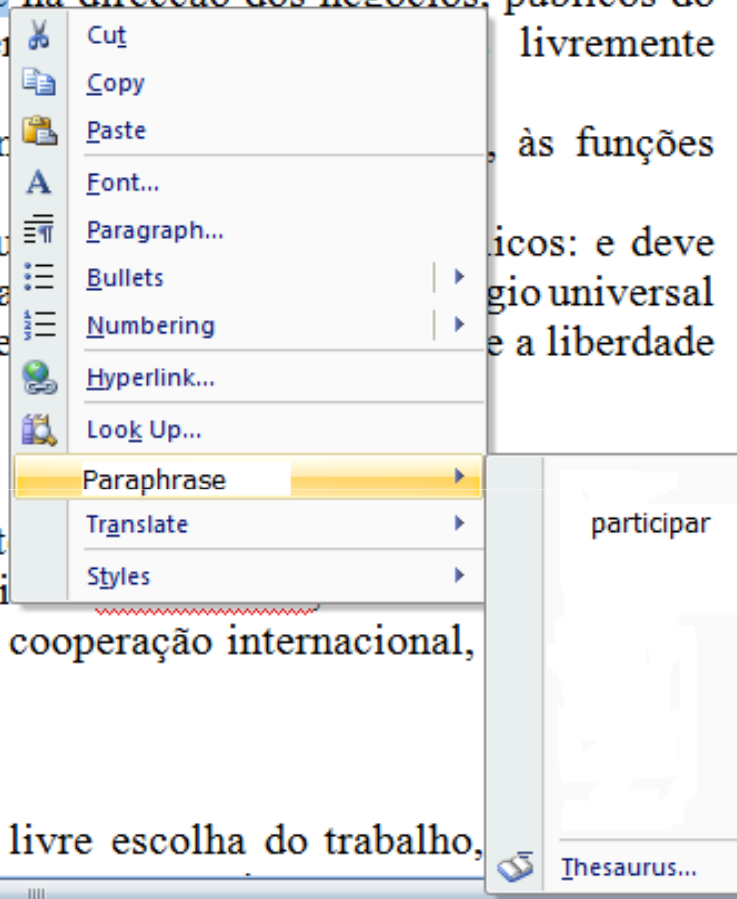
3. A vontade do povo é o fundamento da autoridade dos poderes públicos: e deve exprimir-se através de eleições honestas a realizar-se em prazo regular, igual e igual, com voto secreto ou segundo processo eleitoral, e a liberdade de voto.

Artigo 22º

Toda a pessoa, como membro da sociedade, tem direito de participar legitimamente exigir a satisfação dos direitos e interesses indispensáveis, graças ao esforço nacional e à cooperação internacional, com a organização e os recursos de cada país.

Artigo 23º

1. Toda a pessoa tem direito ao trabalho, à livre escolha do trabalho,



Uso interactivo do ReWriter
Edição de texto com paráfrases para aplicações em
processamento de texto

ParaMT: Características

- Gerador de Paráfrases Bilingues/Multilingues (= Parafraseador Bilingue/Multilingue)
- Estado actual
 - reconhecimento e extracção de construções com verbos suporte elementares em português (*fazer uma operação*)
 - parafraseamento por um verbo lexical correspondente em inglês (*operate*)
 - parafraseamento por uma construção com verbo suporte não elementar equivalente em inglês (*perform an operation*)
 - desambiguação e tradução de *fazer uma operação* por *to have an operation* ou *to be operated* de acordo com a estrutura argumental do predicado
- Aplicações:
 - Tradução automática
 - Dicionários bilingues de expressões multipalavra

ParaMT: um Parafraseador Bilingue/Multilingue

a fazer um estágio para	dar aulas de/teach	religião, mas não se import
m -- os filhos -- juntos e	fizeram a mudança para/change	Johannesburg, e ensinaram
. Necessitava apenas de	ter a certeza de/know	que não escapara à sua
ente hipotética. -- Deves	ter alguma ideia/know	. Dorothy andava a fazer um
. não podemos deixar de	ter cautela/beware	. Pobre Caro, pensou Lync
ra dos chinelos, antes de	ter chance de/can	mudar de idéia. Como pos
ope a Jean, esta pareceu	ter dificuldade em/avoid	olhá-lo nos olhos. Deixou
ao Kiss dela. Apesar de	ter falta de/lack	amor-póprio, isso não sign
igos e imprensa estava a	ter lugar /occur	numa longa galeria com car
guiu ter filhos. -- Tens de	ter mão /control	nessa confusão toda. Sam
spondi, minha mãe deve	ter medo de/fear	cobras. Eu disse no Gabin
da loja antes de ele	ter tempo de/could	chamar a brigada de narcó
a triste aventura havia de	ter um fim/finish	.
Ela ouvira a tia Velma	ter uma discussão com/argue	Jack acerca de mostarda r
de olhos fechados para	ter uma ideia de/know	como seria ser cego e
ter paciência.» «Voltei a	ter uma imensa vontade de/want	viver. A conversa parecia :

Reconhecimento e parafraseamento bilingue (tradução)
de construções com verbos suporte elementares

(construção com verbo suporte em português / verbo lexical equivalente em inglês)

Primeiros Resultados

500 frases

100 para cada verbo suporte elementar

	Reconhecimento Precisão	Reconhecimento Abrangência	Parafraseamento Precisão
Pôr	73/73 - 100%	73/100 - 73%	72/73 - 98.6%
Tomar	75/75 - 100%	75/100 - 75%	68/73 - 93.1%
Ter	65/65 - 100%	65/100 - 65%	59/65 - 90.7%
Dar	57/60 - 95%	57/100 - 57%	46/51 - 90.1%
Fazer	43/45 - 95.5%	43/100 - 43%	40/45 - 88.8%
Average	62.6/63.6 - 98.4%	62.6/100 - 62.6%	57/61 - 93.4%

Avaliação do reconhecimento e parafraseamento de construções com verbos suporte elementares

Parafraseador automático vs anotação manual

Como Aceder aos Recursos

Port4NooJ

Disponível publicamente em:

<http://www.nooj4nlp.net> (Resources/Portuguese)

<http://www.linguateca.pt/Repositorio/Port4Nooj/>

ReWriter

Interface em fase de desenvolvimento para disponibilização como um serviço na Linguateca

ParaMT

A ser desenvolvido para integração num sistema de tradução automática

Agradecimentos

Obrigada pela atenção!

Este trabalho foi parcialmente financiado pela *Fundação para a Ciência e a Tecnologia*, através da bolsa SFRH/BD/14076/2003, co-financiado pelo POSI e pela *Fundação para a Computação Científica Nacional*.

