



UNIVERSIDADE DE ÉVORA

Mestrado em Engenharia Informática

Agente inteligente para análise  
sintáctica/semântica e marcação de nomes  
próprios

*João Tomé da Silva Laranjinho*

orientador: *Prof. Doutora Irene Pimenta Rodrigues*

3 de Maio de 2010



UNIVERSIDADE DE ÉVORA

Mestrado em Engenharia Informática

Agente inteligente para análise  
sintáctica/semântica e marcação de nomes  
próprios

*João Tomé da Silva Laranjinho*

orientador: *Prof. Doutora Irene Pimenta Rodrigues*



185 666

3 de Maio de 2010

# Agradecimentos

Quero começar por agradecer à professora Irene Rodrigues por toda a ajuda que me deu no desenvolvimento desta dissertação.

Quero agradecer também aos restantes professores do Departamento de Informática, e claro, sem esquecer os professores de outros Departamentos que me deram aulas.

Finalmente, quero agradecer aos meus pais (Tomé e Miquelina), aos meus amigos e colegas de curso, e ao meu grande amigo Gonçalo (o Sr. da Matemática).



# Sumário

Nesta dissertação apresenta-se a implementação e avaliação de um sistema de análise sintáctico/semântico e um marcador automático de nomes próprios para o processamento de frases interrogativas em sistemas de pergunta-resposta.

O sistema de análise sintáctico/semântico é composto por duas fases. A primeira fase tem como objectivo verificar se as frases de um texto se encontram sintacticamente correctas e inferir para cada frase, as estruturas sintácticas correspondentes. A segunda fase tem como objectivo encontrar para cada estrutura sintáctica, as estruturas semânticas que a representam.

O marcador automático de nomes próprios tem como objectivo identificar e classificar os nomes próprios presentes num texto nas categorias data, hora e numa classe mais geral de nome, nome próprio. Este sistema recorre a uma função heurística de forma a pontuar os nomes próprios marcados, assim como, pontuar às interpretações de uma frase que são representáveis por estruturas sintácticas.

Ainda em relação aos dois sistemas, sempre que estes detectam palavras ou nomes próprios que não se encontram no dicionário local, fazem a sua pesquisa em dicionários *online* através da Internet. A pesquisa das palavras é feita no dicionário da Priberam, enquanto que a pesquisa de nomes próprios é feita na Wikipédia.



# Abstract

## **An intelligent agent for syntactic/semantic analysis and named entities recognition**

This dissertation presents the implementation and evaluation of a syntactic/semantic analyser and of a proper names recogniser for processing interrogative sentences in question-answer systems.

The syntactic/semantic analyser consists of two phases. The first phase aims to verify whether the sentences of a text are syntactically correct and infer for each sentence, all possible syntactic structures. The second phase aims at finding for each syntactic structure, the semantic structures that represent it.

The proper names recogniser aims to identify and classify the proper names present in texts in the categories date, time and a more general class name, proper name. We use an heuristic function in order to score the marked proper names that takes into account the fact that the sentence with that proper name has a syntactic structure.

In these systems when a word or a proper name is not in the local dictionary, an automatic search on online dictionaries is done. The search for words is done in the dictionary of Priberam, and the search for proper names is done on Wikipedia.





# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Trabalho Relacionado</b>	<b>5</b>
2.1	Análise Sintáctica/Semântica . . . . .	6
2.1.1	Dicionário . . . . .	6
2.1.2	Gramáticas . . . . .	6
2.1.3	Parser . . . . .	7
2.1.4	Teoria de Representação do Discurso . . . . .	8
2.2	Reconhecimento de Entidades Mencionadas . . . . .	8
2.2.1	Objectivo . . . . .	8
2.2.2	Técnicas Usadas . . . . .	8
2.2.3	Métricas . . . . .	9
2.3	Analísadores Sintácticos . . . . .	9
2.3.1	Projecto VISL . . . . .	10
2.3.2	PEN . . . . .	10
2.4	Sistemas de REM . . . . .	11
2.4.1	Cortex . . . . .	11
2.4.2	REMBRANDT . . . . .	11
2.4.3	SIEMÊS . . . . .	12
2.4.4	Balie . . . . .	13
2.4.5	Avaliação de Sistemas HAREM . . . . .	14
<b>3</b>	<b>Dicionário e Análise Sint/Sem</b>	<b>17</b>
3.1	Dicionários . . . . .	18
3.1.1	Arquitectura do Dicionário Local . . . . .	18
3.1.2	Consulta Priberam . . . . .	18
3.1.3	Consulta Wikipédia . . . . .	23
3.2	Análise Sintáctica . . . . .	25
3.2.1	Analísador Sintáctico . . . . .	25
3.2.2	Top-down . . . . .	25
3.2.3	Chart Parser . . . . .	27
3.2.4	Reescrever Estruturas Sintácticas . . . . .	31
3.3	Análise Semântica . . . . .	31
3.3.1	Analísador Semântico . . . . .	31

<b>4</b>	<b>Analizador Sintático/Semântico</b>	<b>39</b>
4.1	Arquitectura do Analizador Sintático/Semântico . . . . .	40
4.1.1	Pré-processamento . . . . .	40
4.1.2	Análise Lexical . . . . .	42
4.1.3	Análise Sintáctica . . . . .	43
4.1.4	Análise Semântica . . . . .	44
4.2	Avaliação de Resultados . . . . .	44
4.2.1	Avaliação de DRS's . . . . .	45
4.2.2	Número de ES/DRS . . . . .	49
<b>5</b>	<b>Marcador de Nomes Próprios</b>	<b>53</b>
5.1	Arquitectura do Marcador de Nomes Próprios . . . . .	54
5.1.1	Pré-processamento . . . . .	56
5.1.2	Análise Lexical . . . . .	60
5.1.3	Pontuar Interpretações . . . . .	61
5.1.4	Saída . . . . .	62
5.1.5	Avaliação . . . . .	64
5.2	Avaliação de Resultados . . . . .	64
5.2.1	Pontuações Atribuídas . . . . .	65
5.2.2	CLEF . . . . .	67
5.2.3	Público . . . . .	69
5.2.4	HAREM . . . . .	71
<b>6</b>	<b>Conclusões e Trabalho Futuro</b>	<b>73</b>
6.1	Conclusões . . . . .	74
6.2	Trabalho Futuro . . . . .	76
<b>7</b>	<b>Anexos</b>	<b>81</b>

# Lista de Figuras

3.1	Dicionário Local	18
3.2	Consulta Priberam	19
3.3	Arquitetura da Consulta Priberam	19
3.4	Pesquisar Palavras	20
3.5	Processar Pesquisas	21
3.6	Entrada de Palavras	23
3.7	Consulta Wikipédia	23
3.8	Arquitetura da Consulta Wikipédia	24
3.9	Pesquisar Nomes Próprios	24
3.10	Entrada de Nomes Próprios	24
3.11	Analisador Sintático	25
3.12	Analisador Semântico	32
3.13	Regras para reescrever uma ES numa DRS	33
3.14	Reescrever ES em DRS	34
4.1	Analisador Sintático/Semântico	40
4.2	Arquitetura do Analisador Sintático/Semântico	41
4.3	Remover Expressões	42
4.4	Movimentar Pronomes Clíticos	42
4.5	Remover Verbos	43
4.6	Acrescentar Verbos	43
4.7	Encontrar Estruturas Sintáticas	44
4.8	Encontrar Drs's	45
5.1	Marcador de nomes próprios	54
5.2	Arquitetura do marcador de nomes próprios	55
5.3	Separar em frases	56
5.4	Separar em tokens	57
5.5	Separação de frase em tokens	57
5.6	Marcar candidatos a nomes próprios	58
5.7	Marcação de candidatos a nomes próprios numa frase	59
5.8	Desdobrar Contrações	59
5.9	Encontrar Palavras Desconhecidas	60
5.10	Encontrar nomes próprios	60
5.11	Pontuar nomes próprios	62
5.12	Pontuar estruturas sintáticas	63
5.13	Escrever saída	63
5.14	Avaliar saída	64

5.15	Testes com o corpus do CLEF . . . . .	68
5.16	Testes com o corpus do CLEF . . . . .	68
5.17	Testes com o corpus do Público . . . . .	70
5.18	Testes com o corpus do Público . . . . .	70
5.19	Testes com o corpus do HAREM . . . . .	72
5.20	Testes com o corpus do HAREM . . . . .	72

# Lista de Tabelas

2.1	Resultados do Primeiro HAREM . . . . .	14
2.2	Resultados no Mini-HAREM . . . . .	15
2.3	Resultados do Segundo HAREM . . . . .	15
3.1	Informação sobre palavra “é” . . . . .	21
3.2	Regras de processamento das tabelas obtidas do Priberam . . . . .	22
3.3	Estruturas resultantes de processar a tabela do Priberam . . . . .	22
3.4	Regras para o Top-Down . . . . .	26
3.5	Estruturas sintáticas do Top-down . . . . .	27
3.6	Regras para o Chart Parser . . . . .	28
3.7	Regras para o Chart Parser . . . . .	29
3.8	Estruturas sintáticas do Chart Parser . . . . .	30
3.9	Reescrever Estruturas Sintáticas do Chart Parser . . . . .	31
3.10	ES’s de uma frase . . . . .	35
3.11	DRS’s de uma frase . . . . .	36
3.12	ES’s de uma frase . . . . .	37
3.13	DRS’s de uma frase . . . . .	37
3.14	ES’s, DRS’s e tempo de execução de uma frase . . . . .	38
4.1	Perguntas do CLEF avaliadas com Top-down e Chart Parser . . . . .	46
4.2	Perguntas do CLEF avaliadas com Top-down . . . . .	47
4.3	Perguntas do CLEF avaliadas com Top-down e Chart Parser . . . . .	48
4.4	Médias com Top-down e Chart Parser (100 frases CLEF) . . . . .	49
4.5	ES e DRS com Top-down e Chart Parser (10 frases CLEF) . . . . .	50
4.6	Média ES e DRS com Top-down e Chart Parser (700 frases CLEF) . . . . .	50
4.7	Nº de frases com análises sintática e semântica (700 frases CLEF) . . . . .	50
4.8	Média ES e DRS com Top-down e Chart Parser (300 frases Público) . . . . .	50
4.9	Nº de frases com análises sintática e semântica (300 frases Público) . . . . .	51
4.10	Média ES e DRS com Top-down e Chart Parser (200 frases HAREM) . . . . .	51
4.11	Nº de frases com análises sintática e semântica (300 frases Público) . . . . .	51
4.12	Média ES e DRS com Top-down e Chart Parser . . . . .	52
4.13	Porcentagem de frases com análise . . . . .	52
5.1	Heurística . . . . .	66
5.2	Pontuações atribuídas nos 8 testes . . . . .	66
5.3	Testes com corpus do CLEF . . . . .	67
5.4	Testes com corpus do Público . . . . .	70
5.5	Testes com corpus do HAREM . . . . .	71

# **Capítulo 1**

## **Introdução**

Neste capítulo faz-se uma breve introdução sobre a área desta dissertação e finalmente é apresentada a estrutura da dissertação.

A possibilidade de automatizar a compreensão da informação escrita em Língua Natural é uma tarefa importante na área da Pesquisa e Recuperação de Informação.

Um exemplo desta automatização, são os sistemas de pergunta-resposta. Estes sistemas, respondem a perguntas com base na informação de que dispõem em bases de documentos locais ou na Web. As respostas às perguntas podem ser tanto mais exactas, quanto melhor tiver sido a “compreensão” da informação pelo sistema. Este “compreender” da informação pelo sistema, pode passar pela análise sintáctica e semântica das perguntas e das frases dos textos dos documentos.

Perceber num texto, quais são as Entidades Mencionadas, e quais as suas classificações, também pode contribuir para o entendimento do mesmo.

Finalmente, a utilização da Internet, permite que tenhamos acesso rápido a enormes quantidades de dados, em qualquer parte e de qualquer parte do mundo, de forma gratuita em muitos casos. Por exemplo, a Internet pode ser utilizada no acesso a dados que se encontram em dicionários ou enciclopédias *online*.

Nesta dissertação apresenta-se um sistema de análise sintáctico/semântico e um marcador automático de nomes próprios que foram desenvolvidos com objectivo de processar frases interrogativas em sistemas de pergunta-resposta. Nalguns [QRPV06] sistemas de pergunta-resposta a representação semântica da pergunta é um requisito necessário para poder encontrar uma resposta adequada à pergunta do utilizador.

Como se pode ver nos capítulos 4 e 5, onde se apresentam resultados da avaliação do analisador sintáctico/semântico para 3 domínios diferentes de frases, estes sistemas que foram inicialmente desenvolvidos para analisar perguntas, podem ser usados em outras aplicações de Processamento de Língua Natural necessitando, no entanto, de algumas adaptações, por exemplo aumentar as regras da gramática de forma a ter uma melhor cobertura do Português.

Para a análise sintáctica das frases desenvolveu-se, inicialmente, um conjunto de regras para um analisador Top-down. Este analisador mostrou bom desempenho nos testes efectuados em perguntas. Com o objectivo de poder obter análises parciais das frases este conjunto de regras foi adaptado para um analisador Bottom-up.

A análise sintáctica é representada por uma estrutura sintáctica que é reescrita pelo analisador semântico numa Estrutura de Representação do Discurso (DRS). Esta ferramenta foi avaliada mostrando que tem um desempenho adequado à expectativa inicial.

O que não se esperava, e não tem resposta neste trabalho, é o número de análises sintácticas ou semânticas que uma frase pode ter, pode ir de uma DRS, no caso de algumas perguntas no corpus usado na avaliação, a 400 DRSs no caso de algumas frases retiradas de notícias do Público. Para utilizar estas ferramentas (analisador sintáctico/semântico) noutras domínios será necessário lidar com esta questão.

O desenvolvimento do marcador de nomes próprios foi necessário para melhorar o desempenho do analisador sintáctico. Este sistema procurou tirar partido do analisador sintáctico na decisão de marcar nomes próprios. O analisador sintáctico e o marcador de nomes próprios recorrem a recursos linguísticos e enciclopédias disponíveis na WEB (por exemplo, o Dicionário *online* da Priberam e a Wikipédia).

No próximo capítulo são apresentados alguns conceitos relacionados com a área em que esta dissertação se insere: análise sintáctica, análise semântica e o Reconhecimento de Entidades Mencionadas. Neste capítulo também são descritos alguns trabalhos relacionados com o tema do trabalho desenvolvido, analisadores sintácticos e sistemas de Reconhecimento de Entidades Mencionadas incluindo uma avaliação conjunta de sistemas de Reconhecimento de Entidades Mencionadas.

No capítulo 3 descrevem-se os dicionários utilizados e a análise sintáctica e semântica.

Nos capítulos 4 e 5 apresenta-se e avalia-se o analisador sintáctico/semântico e o marcador de nomes próprios, respectivamente.

No capítulo 6 são dadas a conhecer as conclusões e fazem-se algumas sugestões para o melhoramento do analisador sintáctico/semântico e do marcador automático de nomes próprios.

Finalmente, no capítulo 7 encontra-se os 3 corpus (7 textos de 100 frases cada do CLEF, 3 textos de 100 frases do jornal, Público e 200 frases do HAREM) utilizados na avaliação do analisador sintáctico/semântico e do marcador de nomes próprios.





## Capítulo 2

# Trabalho Relacionado

Neste capítulo estão resumidos alguns conceitos relacionados com área desta dissertação.

Na secção 2.1 são tratadas as análises sintáctica e semântica, onde se explica a importância de dicionários locais (2.1.1), de gramáticas (2.1.2), de *parser's* (2.1.3) e da Teoria de Representação do Discurso (2.1.4) no processamento de Língua Natural.

A secção 2.2 é dedicada ao Reconhecimento de Entidades Mencionadas. Nesta secção é referido o objectivo do Reconhecimento de Entidades Mencionadas (2.2.1), são apresentadas técnicas utilizadas no Reconhecimento de Entidades Mencionadas (2.2.2) e descritas três métricas usadas na medição do desempenho de sistemas de Recuperação de Informação (2.2.3).

Na secção 2.3 são apresentados dois analisadores sintácticos, o VISL (2.3.1) e o PEN (2.3.2).

Na secção 2.4 são apresentados os seguintes sistemas de Reconhecimento de Entidades Mencionadas: Cortex (2.4.1), REMBRANDT (2.4.2), SIEMÊS (2.4.3) e Balie (2.4.4). Finalmente, é apresentada a avaliação conjunta de sistemas de Reconhecimento de Entidades Mencionadas, HAREM.

## 2.1 Análise Sintáctica/Semântica

A compreensão automática de uma língua pode beneficiar com a análise sintáctica e semântica. A análise sintáctica estuda a constituição e a classificação das palavras ou grupos de palavras nas frases e verifica se a estrutura de uma frase é válida.

Na análise semântica atribuí-se significado às expressões da língua. Esta atribuição pode ser feita através de teorias para a representação de frases ou conjuntos de frases.

Nesta secção são descritos alguns conceitos relacionadas com a análise sintáctica e semântica no Processamento de Língua Natural.

### 2.1.1 Dicionário

A existência de um dicionário local é fundamental para a maioria dos sistemas [VdL01] de processamento de Língua Natural. Um dicionário local num sistema pode contribuir para uma maior rapidez no acesso à informação.

Um dicionário é composto por um conjunto de entradas de palavras com informação lexical. A informação é constituída pela classe gramatical (substantivo, adjectivo, verbo, etc) da palavra e por etiquetas como: o género, número, pessoa, tempo, modo, etc.

Na representação das entradas de um dicionário pode usar-se dois tipos de estruturas: de formas ou de bases [VdL01].

Nas estruturas de formas, as palavras são colocadas por extenso, ou seja, não estão divididas em morfemas (unidades menores). Neste tipo de estruturas, a pesquisa de uma palavra é directa.

As palavras nas estruturas de bases encontram-se estruturadas em morfemas. A pesquisa de uma palavras neste tipo de estrutura requer um algoritmo para reconstruir palavras através dos seus morfemas. A utilização de estruturas de bases faz mais sentido em dicionários grandes.

### 2.1.2 Gramáticas

Uma gramática é constituída por um conjunto de regras que definem as expressões válidas de uma língua, deve ser expressiva e reconhecer facilmente uma expressão. As expressões são compostas de relações entre palavras ou conjuntos de palavras [VdL01].

Noam Chomsky [Cho56] classificou as gramáticas em quatro tipos:

- Tipo 3 (regulares) - são as mais limitadas e mais fáceis de reconhecer, no entanto, são insuficientes para exprimir as regras de Língua Natural;
- Tipo 2 (livres de contexto) - mais poderosas que as regulares, no entanto, apresentam problemas para exprimir as dependências (caso da concordância);
- Tipo 1 (sensíveis ao contexto) - permitem exprimir dependências, mas apresentam problemas de complexidade no reconhecimento;
- Tipo 0 (sem restrições) - não têm limitações impostas.

Uma gramática pode ser escrita através de vários formalismos [VdL01]. Entre os formalismos pode destacar-se: Redes de Transição [Woo70] [JM92] [FG92], Gramáticas de Constituintes Imediatos [Gaz82], Gramáticas de Constituintes Imediatos Generalizados [GGS85], Gramáticas de Unificação Funcional [KAY79], PATR-II [Shi84] e HPSG (head-driven phrase structure grammar) [PS94].

Alguns estudos indicam que as melhores representações de língua natural se situam entre as gramáticas livres de contexto e as sensíveis ao contexto.

### 2.1.3 Parser

Um *parser* é um programa que se baseia num dicionário e nas regras de uma gramática, que permite classificar expressões de uma língua. Um *parser* além de verificar se uma expressão pertence à língua pode inferir a sua estrutura sintáctica [Mar04].

Para se poder visualizar a representação da estrutura sintáctica de uma expressão pode recorrer-se a uma árvore [Mar04]. Uma árvore além de expor de forma clara as dependências entre as sub-partes da expressão é muito simples de representar.

Deve ter-se em atenção que a utilização de uma gramática incorrecta num *parser* pode provocar a inferência de estruturas sintácticas incorrectas.

Um *parser* é essencial em algumas aplicações [Mar04]. Por exemplo, é indispensável em compiladores de linguagens de programação.

As técnicas de *parser* que vão ser apresentadas são: Top-down, Bottom-up, Left-corner e Chart Parser.

#### Top-down

Um *parser* na análise descendente ou Top-down constrói a estrutura sintáctica começando com as regras mais gerais e terminando com as regras mais específicas [Mar04].

A construção das estruturas sintácticas é feita da esquerda para a direita neste tipo de *parser*.

No *parser* Top-down deve ter-se cuidado com a existência de regras recursivas à esquerda. A recursividade à esquerda nas regras pode impedir o *parser* de terminar a execução.

#### Bottom-up

Um *parser* na análise ascendente ou Bottom-up constrói a estrutura sintáctica começando com as regras mais específicas e terminando com as regras mais gerais [Mar04].

As palavras são combinadas em constituintes e, por sua vez, os constituintes em outros constituintes mais gerais no *parser* Bottom-up. Neste tipo de *parser* não existem problemas de ciclo devido a regras recursivas à esquerda, porque a análise começa sempre com a combinação de palavras.

Um dos problemas do *parser* Bottom-up está relacionado com a presença de constituintes vazios.

#### Left-corner

O Left-corner é um *parser* que combina estratégias Top-down e Bottom-up [VdL01]. Quando uma palavra é encontrada, ele verifica o tipo de constituinte que se inicia com a palavra e faz o resto da análise de forma Top-down para aquele constituinte. Neste *parser* há problemas se se escreverem regras recursivas à esquerda.

#### Chart Parser

O Chart Parser é um *parser* do tipo Bottom-up [VdL01]. Este *parser* consegue lembrar-se das estruturas já analisadas. Quando inicia uma pesquisa, primeiro verifica se a informação já existe nos seus registos.

### 2.1.4 Teoria de Representação do Discurso

A Teoria de Representação do Discurso (DRT) é uma teoria para a representação semântica do discurso [KR93].

A DRT representa o significado das frases no seu contexto (do discurso). Na DRT, o discurso (sequências de frases) é representado por Estruturas de Representação do Discurso (DRS). Uma DRS é uma estrutura semântica que representa o significado de uma ou mais frases do discurso.

As DRS resultam da reescrita das estruturas obtidas na análise sintáctica, ou seja das estruturas sintácticas.

Uma DRS é composta por um conjunto de referentes (variáveis) e um conjunto de condições (predicados).

Os referentes identificam as entidades presentes no discurso. De uma forma geral, as condições expressam as relações entre as entidades do discurso.

## 2.2 Reconhecimento de Entidades Mencionadas

A compreensão de um texto passa pelo Reconhecimento das Entidades Mencionadas (REM) no texto [SC07]. Quanto mais exacta for a tarefa de REM num texto melhor poderão ser feitas outras tarefas de processamento automático de documentos como, por exemplo, a análise sintáctica.

Nesta secção é descrito o objectivo do REM, as técnicas usadas e algumas métricas que podem ser utilizadas para medir o seu desempenho.

### 2.2.1 Objectivo

O REM num texto procura identificar e classificar as EM em categorias predefinidas. Entre as categorias de EM podem destacar-se: pessoas, organizações, locais, acontecimentos, datas, valores numéricos, etc.

As palavras da classe substantivo, assim como quantidades, referências e alguns advérbios podem ser considerados EM.

Os sistemas de REM que apresentaram melhores resultados foram desenvolvidos para domínios específicos.

No REM, é necessário identificar a classe gramatical das palavras de forma a deduzir os limites da entidade e a sua classificação no conjunto de categorias considerado.

Alguns sistemas de REM são compostos por duas etapas principais. Na primeira etapa faz-se a identificação de todos os possíveis candidatos a EM. Na segunda etapa pode ser feita a revisão de alguns passos da etapa de identificação e é feita a classificação dos candidatos no conjunto de categorias considerado. Em caso de ambiguidade na classificação pode adicionar-se um novo passo para desambiguar.

Os sistemas de REM podem contribuir com os seus resultados para melhorar o desempenho de vários sistemas de Processamento Língua Natural [SC07] como por exemplo: Construção de Resumos, Mineração de Textos, Web Semântica, sistemas Perguntas-Respostas e Tradução Automática.

### 2.2.2 Técnicas Usadas

As técnicas usadas nos sistemas de REM dividem-se em: modelos baseadas em gramáticas e modelos estatísticos.

Os modelos baseados em gramáticas têm apresentado melhores resultados, no entanto, requerem muito trabalho feito por linguistas na sua implementação.

Os sistemas baseados em modelos estatísticos ou em aprendizagem requerem grandes quantidades de dados anotados manualmente para a fase de treino.

Alguns dos recursos utilizados no REM [AA08] são:

- **Corpus** - conjuntos de textos anotados. Normalmente em conjunto com corpus, são utilizadas estratégias de aprendizagem. Um dos problemas do corpus está relacionado com a dificuldade da sua anotação;
- **Almanaques** - dicionários de EM;
- **Metapalavras** - representam as palavras próximas das EM. Estas são palavras que dão alguma informação sobre as entidades, normalmente são utilizadas na fase de desambiguação. Exemplos: rio, rua, etc;
- **Abreviaturas** - palavras que fazem parte das entidades e dão informação sobre a classe da entidade. Normalmente são utilizadas na classificação. Exemplo: Dr., Sr., etc;
- **Regras de similaridade** - um conjunto de regras que definem semelhanças entre as entidades a serem classificadas e as entidades que existem em Almanagues. Normalmente estas regras são implementadas em expressões regulares.

### 2.2.3 Métricas

O desempenho de um sistema de REM pode ser medido através de diversas métricas [SC07] que representam o desempenho em valores numéricos.

As três métricas que normalmente são utilizadas para avaliar o desempenho de sistema de Recuperação de Informação são as seguintes: Abrangência (Recall), Precisão (Precision) e Medida-F (F-Measure).

- A Abrangência mede a relação entre o número de resultados correctos e o número de resultados existentes. A fórmula da Abrangência é a seguinte:

$$\text{Abrangência} = \frac{\text{Resultados Correctos} \cap \text{Resultados Existentes}}{\text{Resultados Existentes}}$$

- A Precisão mede a relação entre o número de resultados correctos e o número de resultados obtidos. A fórmula da Precisão é a seguinte:

$$\text{Precisão} = \frac{\text{Resultados Correctos} \cap \text{Resultados Obtidos}}{\text{Resultados Obtidos}}$$

- A Medida-F é uma média harmónica de Precisão ( $P$ ) e Abrangência ( $A$ ). A fórmula da Medida-F é a seguinte:

$$\text{Medida-F} = 2 * \frac{P * A}{P + A}$$

## 2.3 Analisadores Sintácticos

O objectivo de um analisador sintáctico é o de verificar se uma frase tem análise sintáctica e encontrar a(s) estrutura(s) sintáctica(s) associada(s) à frase.

Nesta secção apresentam-se dois analisadores sintácticos, o VISL [Bic00] e o PEN [OG08].

### 2.3.1 Projecto VISL

O projecto VISL<sup>1</sup> (Visual Interactive Syntax Learning) é um projecto de investigação do Institute of Language and Communication, University of Southern Denmark. Este projecto tem como objectivo desenvolver ferramentas gramaticais que recorrem à Internet [Bic00].

Algumas palavras de textos em Língua Natural são ambíguas quando vistas isoladas relativamente ao tipo de classe gramatical, inflexão, função sintáctica e conteúdo semântico, sendo o contexto o que determina como uma palavra deve ser classificada.

O formalismo utilizado na desambiguação das possíveis interpretações no projecto VISL é o das Gramáticas de Restrições (Constrain Grammar, CG).

Este formalismo tenta desambiguar as interpretações possíveis para cada palavra escolhendo-as ou rejeitando-as em determinado contexto.

Uma CG necessita dos resultados de um analisador morfológico antes do processo de desambiguação.

No analisador morfológico, utilizado no VISL, são tratadas questões de flexão, derivação e controle de lemas. Além disso, é retirada informação de regência e semântica, para a sua contextualização.

Algumas palavras no input da CG apresentam mais do que uma análise morfológica. Para resolver estas ambiguidades são utilizadas regras que seleccionam as análises morfológicas correctas.

Na desambiguação morfológica são acrescentadas funções sintácticas a cada palavra com a utilização de condições contextuais. Para encontrar as funções sintácticas correctas são utilizadas regras que descartam hipóteses. As hipóteses que subsistirem a estas regras são consideradas correctas.

Os resultados obtidos têm estruturas sintácticas associadas, que podem ser utilizados por aplicações de processamento automático de Língua Natural.

### 2.3.2 PEN

O PEN [OG08] é um analisador sintáctico que é mantido e disponibilizado pelo Pólo de Coimbra da Linguateca. Este analisador foi implementado através do algoritmo de Earley.

A entrada do analisador contém dois ficheiros de texto. Um dos ficheiros deve conter o texto a analisar e outro a descrição da gramática.

Cada regra do PEN deve ocupar uma linha e deve conter no início uma sequência de caracteres maiúsculos que identificam a regra.

Os símbolos terminais não deve começar com maiúscula para que não sejam “confundidos” com nomes de regras pelo analisador. As primeiras regras deste analisador devem chamar-se RAIZ.

Para uma frase, o analisador PEN retorna todas as derivações possíveis. Também é possível que, para algumas frases não existam derivações, no caso em que, não existem regras reconheçam essas frases.

No PEN é possível atribuir um peso a cada regra da gramática. A utilização de pesos nas regras pode ser um meio a utilizar na escolha da melhor derivação.

---

<sup>1</sup><http://visl.sdu.dk/>

## 2.4 Sistemas de REM

Os sistemas de REM são utilizadas na identificação e classificação [SC07] de EM. Estes sistemas podem utilizar-se em outros sistemas que necessitam das EM destacadas das restantes palavras.

Nesta secção estão apresentados quatro sistemas de REM, o Cortex, o REMBRANDT, o SIEMÊS e o Balie. No final da secção é apresentada uma avaliação conjunta de sistemas REM, o HAREM, e os resultados de alguns dos sistemas referidos em edições HAREM.

### 2.4.1 Cortex

O sistema Cortex [SC07] foi desenvolvido por Christian Nunes Aranha durante a sua tese de doutoramento na Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

O Cortex é um sistema para o processamento de Língua Natural tendo como propósito reproduzir alguns comportamentos humanos.

O processamento no Cortex é feito em várias etapas e, em cada uma dessas etapas é possível corrigir alguns passos das anteriores.

O sistema Cortex é composto pelas seguintes etapas:

- Separar em palavras - as frases são decompostas em palavras;
- Agrupar palavras - é feito o reconhecimento com um autómato das EM que são constituídas por várias palavras. O resultado do reconhecimento é acrescentado ao léxico e finalmente à base de dados;
- Classificação - é feita a classificação das EM encontradas na etapa anterior. Na classificação é consultada informação lexical que se encontra na base de dados.

As informações do Cortex são obtidas de quatro fontes diferentes. As fontes são as seguintes:

- Almanaque - contém EM que provêm de uma enciclopédia;
- Metapalavras - composto de palavras que surgem junto a EM;
- Abreviaturas - composto de palavras que compõem as EM;
- Léxico - reúne a informação adquirida no processamento de textos.

Estas quatro fontes são utilizadas nas regras de identificação e classificação de EM.

### 2.4.2 REMBRANDT

O REMBRANDT [Car] (Reconhecimento de Entidades Mencionadas Baseado em Relações e ANálise Detalhada do Texto) foi desenvolvido por Nuno Cardoso, durante a sua tese de doutoramento na Faculdade de Ciências, Universidade de Lisboa.

O REMBRANDT é um sistema de REM e de detecção de relações entre EM. Este sistema recorre à Wikipédia e a um conjunto de regras gramaticais que procuram evidências internas e externas das EM. A pesquisa da informação na Wikipédia faz-se com interface SASKIA.



Para processar grandes quantidades de texto, o REMBRANDT permite distribuir as tarefas de REM por vários computadores.

No REMBRANDT, cada EM de um texto pode sofrer várias anotações. Estas anotações são guardadas junto de cada EM.

O funcionamento do REMBRANDT pode ser dividido em três etapas:

- Reconhecimento de expressões numéricas e geração de EM candidatas - o texto é separado em frases e em *tokens*. Em seguida é utilizado um conjunto de regras que identificam expressões numéricas. Depois são aplicadas novas regras para reconhecer expressões temporais e valores com o auxílio das expressões numéricas identificadas. Finalmente, são geradas as EM candidatas, que podem conter pelo menos uma letra em maiúscula ou um algarismo, além de elementos de ligação de nomes próprios (de, da, do, das, dos, etc);
- Classificação semântica de EM - cada EM é classificada pela SASKIA e por um conjunto de regras gramaticais. De seguida, é aplicado um segundo conjunto de regras gramaticais que se serve das classificações anterior de forma a detectar EM mais elaboradas. Nesta etapa, as EM que possuem elementos de ligação de nomes próprios são classificadas ou divididas em EM mais pequenas e classificadas novamente;
- Repescagem de EM sem classificação - é feita a detecção de relações entre EM com o auxílio de um conjunto de regras gramaticais. Estas relações são utilizadas de forma a classificar EM que permaneceram sem classificação e que se encontram relacionadas com EM correctamente classificadas. Finalmente, as EM que não foram classificadas são descartadas.

### 2.4.3 SIEMÊS

O SIEMÊS [SC07] (Sistema de Identificação de Entidades Mencionadas com Estratégia Siamesa) foi desenvolvido por uma equipa de três elementos (Luís Sarmiento, Luís Cabral e Ana Sofia Pinto) do Pólo do Porto da Linguatca.

O SIEMÊS é um sistema de identificação e classificação de EM que combina consultas a almanaques com regras de análise de contexto.

A informação dos almanaques é utilizada para gerar um conjunto de hipóteses de classificação para cada candidato a EM, enquanto que as regras de análise de contexto são utilizadas para desambiguar a respectiva classe do candidato em questão.

Na primeira participação no HAREM, o sistema SIEMÊS utilizou o almanaque REPENTINO.

No REPENTINO, as EM encontram-se distribuídas em classes e subclasses. Estas EM foram obtidas de grandes corpora com métodos semi-automáticos e de páginas web.

A segunda versão do SIEMÊS, SIEMÊS v2, resultou da re-implementação da primeira, conservando a estratégia assente em regras de semelhança sobre os almanaques.

A arquitectura do SIEMÊS v2 encontra-se dividida por módulos, sendo possível avaliar o sistema por componentes.

Na versão v2 é possível criar regras externas ao sistema que são interpretadas por um motor para tratar de contextos.

O funcionamento da segunda versão pode ser divididas nas seguintes etapas:

- Identificação de candidatos - esta etapa utiliza um conjunto de regras para identificar candidatos alfabéticos e um outro para identificar e classificar semanticamente entidades numéricas;
- Classificação de entidades alfabéticas - esta etapa utiliza uma cadeia de classificação com cinco blocos que permitem gerar hipótese para classificação dos candidatos.

Os cinco blocos utilizados na cadeia de classificação são os seguintes:

1. Bloco de regras “simples” sobre o contexto - as regras deste bloco pesquisam informações de contexto. Estas regras são exteriores ao SIEMÊS;
2. Bloco de pesquisa directa no REPENTINO - neste bloco por cada candidato a EM encontra-se o número de entradas com representação lexical idêntica no REPENTINO. Esta informação é guardada para mais tarde ser utilizada na classificação;
3. Bloco de emparelhamento de prefixo sobre o REPENTINO - neste bloco por cada candidato a EM pesquisam-se as EM do REPENTINO que contêm prefixos iguais para classificar o candidato. Se não for possível classificar, são removidas palavras ao candidato e procura-se fazer nova classificação;
4. Bloco de semelhança sobre o REPENTINO - neste bloco são utilizadas duas funções heurísticas que encontram semelhanças entre os candidatos a EM e as EM do REPENTINO;
5. Bloco posterior de recurso - as regras deste bloco pesquisam informações de contexto.

Depois da utilização da cadeia de classificação é escolhida a classificação mais correcta com o auxílio de informação de contexto.

#### 2.4.4 Balie

O sistema Balie [Nad05] foi desenvolvido por David Nadeau na School of Information technology & Engineering University of Ottawa.

O Balie é um sistema para a extracção de informação em textos escritos em diferentes línguas.

O sistema Balie utiliza técnicas de aprendizagem máquina. Na aprendizagem é utilizada a ferramenta Weka e um conjunto de corpora anotados.

O sistema Balie oferece os seguintes serviços:

- Identificação da língua - a identificação da língua é feita através das análises de frequência de palavras e de frequência de sequências de caracteres;
- Tokenização - na tokenização, o texto é decomposto em palavras, números e sinais de pontuação;
- Detecção dos limites das frases - a detecção dos limites das frases faz-se com o auxílio de sinais de pontuação e regras. No entanto é tido em consideração que nem todos os sinais de pontuação indicam o fim de uma frase, exemplo disso, são as abreviaturas;

- Classificação dos *tokens* - na classificação, a cada *token* é atribuído uma classe. Na classificação é utilizada a ferramenta qTag.

O resultado do processamento de um texto pelo sistema Balie é uma representação em XML. A representação em XML é útil no intercâmbio, assim como, na facilidade da sua visualização.

### 2.4.5 Avaliação de Sistemas HAREM

O HAREM [SC07] é um evento realizado pela Linguateca no qual se faz a avaliação de sistemas de REM para textos escritos em português. No HAREM são focadas as tarefas de identificação e classificação de EM.

Os sistemas no HAREM são avaliados com o auxílio de um conjunto de recursos comuns e métricas consensuais.

A possibilidade de avaliar conjuntamente vários sistemas contribui para o desenvolvimento de melhores sistemas. Além disso, permite definir, em conjunto, uma área para avaliar e comparar diferentes tecnologias. Assim como, tornar público um conjunto de recursos com o intuito de avaliar e treinar sistemas.

A metodologia utilizada no HAREM inclui a definição de directivas de etiquetagem, a especificação de tarefas de avaliação e a colecção de texto.

No âmbito do HAREM foram realizados três eventos: Primeiro HAREM, Mini-HAREM (dentro do Primeiro HAREM) e Segundo HAREM.

#### Resultados do Primeiro HAREM

Nas tabela 2.1 encontram-se expostos os resultados obtidos pelos sistemas SIEMÊS e Cortex no Primeiro HAREM na tarefa de identificação, classificação morfológica e classificação semântica.

Tarefa	Sistema	Precisão (%)	Abrangência (%)	Medida F (%)
identificação	SIEMÊS	0,7700	0,8396	0,8033
identificação	Cortex	0,6262	0,4344	0,5129
classificação morfológica	SIEMÊS	0,7700	0,8396	0,8033
classificação semântica	SIEMÊS	0,5704	0,4929	0,5288
classificação semântica	Cortex	0,4775	0,3295	0,3899

Tabela 2.1: Resultados do Primeiro HAREM

#### Resultados do Mini-HAREM

Na tabela 2.2 encontram-se expostos os resultados obtidos pelos sistemas Cortex e SIEMÊS no Mini-HAREM nas tarefas de identificação e classificação morfológica.

#### Resultados do Segundo HAREM

Na tabela 2.3 encontram-se expostos os resultados do sistema REMBRANDT no Segundo HAREM nas tarefas de identificação e classificação semântica.

Tarefa	Sistema	Precisão (%)	Abrangência (%)	Medida F (%)
identificação	Cortex	0,8229	0,7702	0,7850
identificação	SIEMÊS	0,7605	0,5474	0,6132
classificação morfológica	SIEMÊS	0,6318	0,4868	0,5462
classificação morfológica	Cortex	0,5471	0,3815	0,4305

Tabela 2.2: Resultados no Mini-HAREM

Tarefa	Sistema	Precisão (%)	Abrangência (%)	Medida F (%)
identificação	REMBRANDT	0,7526	0,5901	0,6615
classificação	REMBRANDT	0,6393	0,4019	0,4935

Tabela 2.3: Resultados do Segundo HAREM



## **Capítulo 3**

# **Dicionário e Análise Sint/Sem**

Este capítulo encontra-se dividido em 3 secções. Na secção 3.1 apresentam-se os dicionários, na secção 3.2 a análise sintáctica e na secção 3.2 a análise semântica.

### 3.1 Dicionários

Nas aplicações que necessitam de informação sobre palavras e nomes próprios, a possibilidade de aceder a dicionários *online* pode ser uma mais valia, na medida em que se pode aceder a informações de vários locais.

Nessas aplicações, a existência de um dicionário local pode reduzir o tempo de execução das mesmas.

Nas seguintes subsecções são descritos: a arquitectura do dicionário local (3.1.1), uma aplicação para pesquisa de palavras no Priberam (3.1.2) e uma aplicação para pesquisa de nomes próprios na Wikipédia (3.1.3).

#### 3.1.1 Arquitectura do Dicionário Local

O dicionário local desenvolvido é composto por um conjunto de entradas com informação sobre palavras e nomes próprios. Na Figura 3.1 pode ver-se uma representação do dicionário local com os vários tipos de entradas.

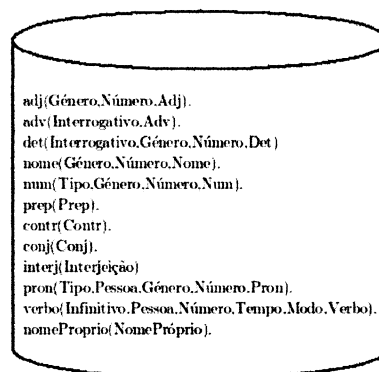


Figura 3.1: Dicionário Local

No dicionário local desenvolvido, a informação sobre as palavras é relativa às classes gramaticais, além de valores para variáveis morfo-sintáctico-semânticas. Enquanto que a informação sobre os nomes próprios é relativa a sua existência.

As entradas no dicionário local consideradas foram as seguintes: adjectivo (adj), advérbio (adv), determinante (det), substantivo comum (nome), número (num), preposição (prep), contracção (contr), conjunção (conj), interjeição (interj), pronome (pron), verbo (verbo) e nome próprio (nomeProprio).

#### 3.1.2 Consulta Priberam

O Priberam<sup>1</sup> é um dicionário *online*, no qual podem ser feitas consultas sobre palavras através da Internet. Em cada consulta obtém-se informações morfo-sintáctico-semânticas acerca da palavra pesquisada.

De forma a consultar o Priberam foi desenvolvida uma aplicação que estabelece ligação ao mesmo.

<sup>1</sup><http://priberam.pt/dlpo/dlpo.aspx>

### Arquitectura da Consulta Priberam

A aplicação recebe um ficheiro de texto com as palavras a pesquisar e retorna um ficheiro de entradas para o dicionário local. Na Figura 3.2 apresenta-se a aplicação para pesquisa de palavras no Priberam.

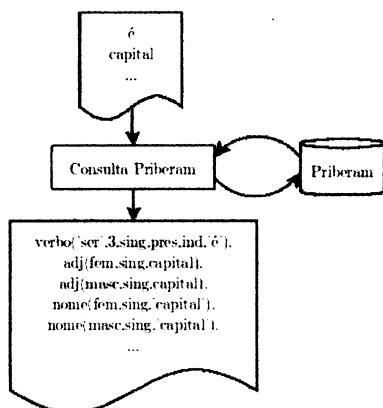


Figura 3.2: Consulta Priberam

Na pesquisa de cada palavra estabelece-se uma ligação ao dicionário da Priberam que obtém informação acerca da palavra. A informação posteriormente é processada por um *parser* e finalmente são geradas entradas para o dicionário local.

A aplicação contém três módulos. Na Figura 3.3 pode ver-se a arquitectura da consulta no dicionário da Priberam.

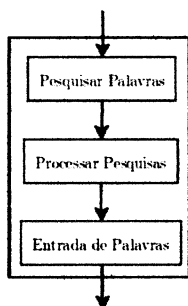


Figura 3.3: Arquitectura da Consulta Priberam

Os módulos são: pesquisar palavras, processar pesquisas e entrada de palavras.

**Pesquisar Palavras** A aplicação recebe um ficheiro de texto com as palavras a pesquisar no dicionário da Priberam. Na Figura 3.4 pode ver-se a pesquisa de palavras no dicionário da Priberam.

A informação da palavra no dicionário da Priberam consiste numa tabela. Cada tabela é composta por duas partes (dois tr's). Na primeira parte encontra-se informação directa sobre a palavra em pesquisa. Na segunda parte além de existir informação



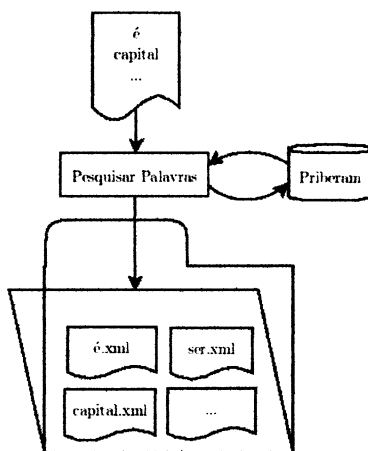


Figura 3.4: Pesquisar Palavras

directa, existe também informação indirecta. A informação indirecta é complementada com *links* para outras palavras que se encontram relacionadas com a palavra em pesquisa. Por exemplo, na pesquisa da palavra “é” obtém-se a informação da Tabela 3.1.

Relativamente à pesquisa da palavra “é”, a primeira parte da tabela refere que esta palavra é um substantivo comum de género masculino. A segunda parte menciona que a palavra “é” é um verbo que se encontra na 3ª pessoa do singular do presente do indicativo. Ainda na segunda parte da tabela é feita referência ao verbo “ser” que é o infinitivo da conjugação “é”.

**Processar Pesquisas** Para processar as pesquisas feitas ao Priberam foi criado um *parser*. Na Figura 3.5 pode ver-se o resultado do processamento pelo *parser* de algumas pesquisas.

O *parser* por cada pesquisa cria uma estrutura com o nome da palavra pesquisada e uma lista com as suas características.

Para que um ficheiro obtido numa pesquisa ao Priberam possa ser processado pelo *parser* deve obedecer às regras da Tabela 3.2.

As estruturas que resultam do *parser* no processamento das tabelas obtidas nas pesquisas ao Priberam podem ser vistas na Tabela 3.3.

**Entrada de Palavras** As estruturas obtidas no processamento das pesquisas ao Priberam devem ser convertidas em entradas para o dicionário local. Na Figura 3.6 pode ver-se a conversão de estruturas resultantes do processamento de pesquisas ao Priberam em entradas para o dicionário local.

Na conversão das estruturas em entradas, as estruturas podem dar origem a várias entradas. Mais detalhes sobre o tipo de entradas do dicionário local podem ser encontradas na sub-secção 3.1.1.

```

<xml search="é">
  <table style="background-color:\#eee; width:100%;"
    cellpadding="4"
    cellspacing="0"
    border="0"
    bordercolor="\#cccccc">
    <tr>
      <td>
        <div>
          <b>é</b> | <em>s. m.</em>
        </div>
      </td>
    </tr>
    <tr>
      <td>
        <div>3ª pess. sing. pres. ind. de
          <a href="&#xA; default.aspx?pal=ser">ser</a>
        </div>
      </td>
    </tr>
  </table>
</xml>

```

Tabela 3.1: Informação sobre palavra “é”

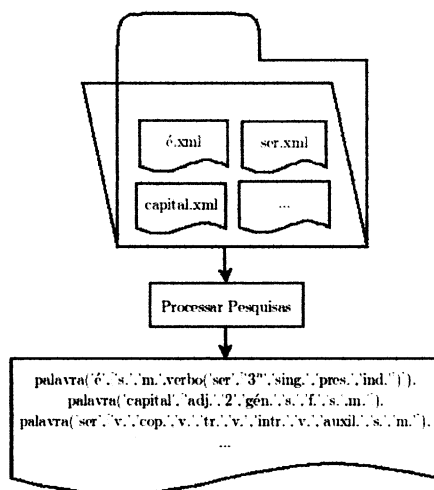


Figura 3.5: Processar Pesquisas

programa -->	<xml search=\" string \"> <table style=\"background-color:\#eee; width:100%;\ cellpadding=\"4\ cellspacing=\"0\ border=\"0\ bordercolor=\"\#cccccc\ </xml>
tabela -->	<tr> tr </tr><tr> tr </tr>
tr -->	<td />
tr -->	<td> td </td>
td -->	divs
divs -->	<div> div </div>
divs -->	<div> div </div> divs
div -->	<b> b </b>   ems
div -->	string <a href=\"&#xA; default.aspx?pal= string \"> string </a>
div -->	spans ems
div -->	<b> b </b>
b -->	string
ems -->	em
ems -->	em ems
em -->	<em> string </em>
spans -->	span
spans -->	span spans
span -->	<span class=\" string \"> <b> string </b></span>

Tabela 3.2: Regras de processamento das tabelas obtidas do Priberam

programa -->	palavra( string , [ tabela ] ).
tabela -->	tr
tabela -->	tr , tr
tr -->	td
td -->	divs
divs -->	div
divs -->	div , divs
div -->	ems
div -->	verbo( string , [ sintrig ] )
div -->	div_a( string , [ sintrig ] )
ems -->	em
ems -->	em , ems
em -->	string

Tabela 3.3: Estruturas resultantes de processar a tabela do Priberam

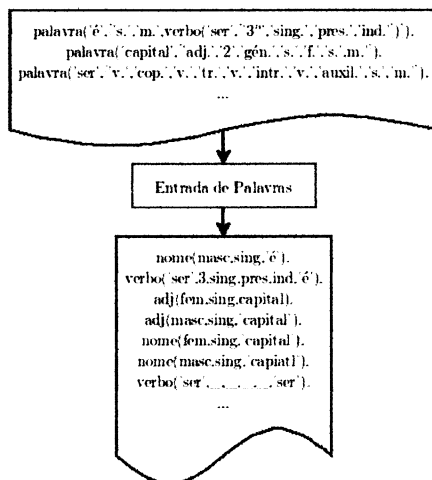


Figura 3.6: Entrada de Palavras

### 3.1.3 Consulta Wikipédia

A Wikipédia<sup>2</sup> é uma enciclopédia, na qual podem ser feitas consultas via Internet sobre nomes próprios e outras palavras. Esta enciclopédia pode ser utilizada de forma a verificar se determinado nome próprio existe.

De forma a verificar a veracidade de nomes próprios, foi desenvolvida uma aplicação que estabelece ligação à Wikipédia.

#### Arquitectura da Consulta Wikipédia

A aplicação recebe um ficheiro de texto com os nomes próprios a consultar e retorna um ficheiro com entradas para o dicionário local. Na Figura 3.7 pode ver-se a consulta de nomes próprios na Wikipédia.

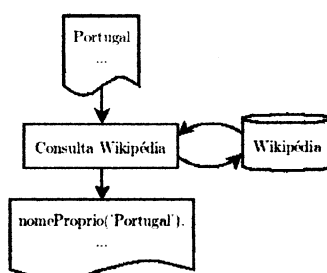


Figura 3.7: Consulta Wikipédia

Na consulta de cada nome próprio estabelece-se uma ligação à Wikipédia de forma a verificar se o nome próprio tem entrada. Por cada nome próprio que tenha entrada na Wikipédia gera-se uma entrada para o dicionário local.

<sup>2</sup>[http://pt.wikipedia.org/wiki/Página\\_principal](http://pt.wikipedia.org/wiki/Página_principal)

A aplicação é composta por dois módulos. Na Figura 3.8 pode ver-se a arquitectura da consulta na Wikipédia.

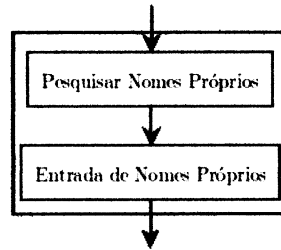


Figura 3.8: Arquitectura da Consulta Wikipédia

Os módulos são: pesquisar nomes próprios e entrada de nomes próprios.

**Pesquisar Nomes Próprios** Na pesquisa dos nomes próprios é estabelecida uma ligação à Wikipédia, que verifica se existem entradas para os nomes próprios. Na Figura 3.9 pode ver-se a pesquisa de nomes próprios.

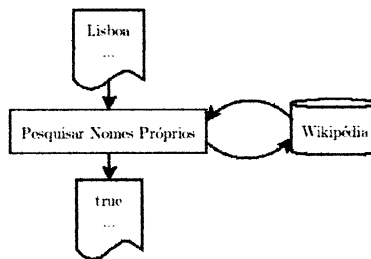


Figura 3.9: Pesquisar Nomes Próprios

Por cada nome próprio procurado na Wikipédia é devolvido um valor booleano, “true” ou “false”, consoante tenha ou não entrada na Wikipédia.

**Entrada de Nomes Próprios** As entradas dos nomes próprios para o dicionário local são geradas após a pesquisa dos nomes próprios na Wikipédia. Na Figura 3.10 pode ver-se a geração de entradas de nomes próprios para o dicionário local.

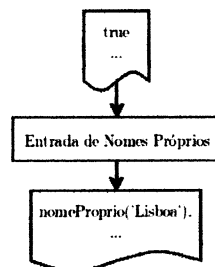


Figura 3.10: Entrada de Nomes Próprios

Por cada nome próprio que tenha entradas na Wikipédia é gerada uma entrada para o dicionário local.

## 3.2 Análise Sintáctica

A análise sintáctica é feita com um analisador sintáctico ou *parser* (subsecção 3.2.1) Top-Down ou Bottom-Up.

Os *parser's* Top-Down e Bottom-Up utilizados nesta dissertação encontram-se descritos nas sub-secções 3.2.2 e 3.2.3, respectivamente.

### 3.2.1 Analisador Sintáctico

O analisador sintáctico verifica se as frases de um texto são representáveis por estruturas sintácticas e infere essas estruturas. Uma estrutura sintáctica expressa as relações entre os constituintes de uma frase.

Na Figura 3.11 pode ver-se a estrutura sintáctica gerada pelo analisador sintáctico para a frase, “Lisboa é a capital de Portugal”.

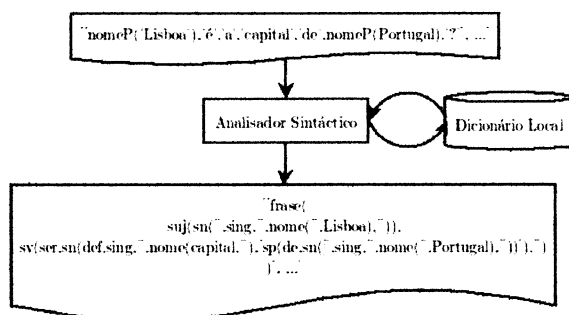


Figura 3.11: Analisador Sintáctico

O analisador sintáctico para cada interpretação de uma frase gera o conjunto de estruturas que a representam. Para cada palavra da frase a analisar são verificadas as suas características no dicionário local, de forma a construir a estrutura sintáctica correspondente.

### 3.2.2 Top-down

Para a análise sintáctica foi desenvolvido um *parser* Top-down com recurso às DCG's de Prolog.

#### Regras Gramaticais

As regras gramaticais do *parser* Top-down encontram-se na tabela 3.4.

Cada regra no Top-down é composta por duas parte, o lado esquerdo da regra contém um *token* e o lado direito, a sequência de *tokens* que formam o *token* do lado esquerdo.



F	-->	SN, SV.
F	-->	SV.
F	-->	SP.INTERR, SV.
F	-->	SP.INTERR, SV, SN.
F	-->	SP.INTERR, SN, SV.
F	-->	adv.interr, SV.
F	-->	adv.interr, SV, SN.
F	-->	adv.interr, SN, SV.
SV	-->	VERBO.T, COMP.
SV	-->	VERBO.T, COMP, SN, COMP.
VERBO.T	-->	verbo, verbo, verbo.
VERBO.T	-->	verbo, verbo.
VERBO.T	-->	verbo.
COMP	-->	[].
COMP	-->	SP, COMP.
COMP	-->	adv, COMP.
SP	-->	prep, SN.
SN	-->	pron.
SN	-->	det, pron.
SN	-->	COMP.SN, NOME.T, COMP.SN.
SN	-->	det, COMP.SN, NOME.T, COMP.SN.
SN	-->	det, numero, COMP.SN, NOME.T, COMP.SN.
SN	-->	numero.
SN	-->	det, numero.
SN	-->	det, numero, COMP.SN.
NOME.T	-->	nomeComum, NOME.P.
NOME.T	-->	NOME.P.
NOME.P	-->	nomeProprio.
NOME.P	-->	[].
COMP.SN	-->	[].
COMP.SN	-->	prop_rel, F, COMP.SN.
COMP.SN	-->	adj, COMP.SN.
COMP.SN	-->	SP, COMP.SN.
COMP.SN	-->	adv, COMP.SN.

Tabela 3.4: Regras para o Top-Down

### Estruturas Sintáticas

As estruturas sintáticas que resultam da análise com o Top-down podem ser vistas na tabela 3.5.

F	-->	frase(suj(SN),SV).
F	-->	frase(suj([]),SV).
SV	-->	sv(VERBO.T,[],COMP)
SV	-->	sv(VERBO.T,SN,COMP)
VERBO.T	-->	verbo
COMP	-->	[]
COMP	-->	[SP COMP]
COMP	-->	[adv COMP]
SP	-->	sp(pre,SN)
SN	-->	sn([],num,[],pron(pron),COMP_SN)
SN	-->	sn(tipo_pron,num,[],pron(pron),COMP_SN)
SN	-->	sn([],num,[],NOME.T,COMP_SN)
SN	-->	sn(tipo_det,num,[],NOME.T,COMP_SN)
SN	-->	sn([],num,numero,NOME.T,COMP_SN)
SN	-->	sn(tipo_det,num,numero,NOME.T,COMP_SN)
SN	-->	sn([],num,numero,[],COMP_SN)
SN	-->	sn(det_tipo,num,numero,[],COMP_SN)
NOME.T	-->	nome(nomeComum,NOME.P)
NOME.T	-->	nome([],NOME.P)
NOME.P	-->	nomeProprio
NOME.P	-->	[]
COMP_SN	-->	[]
COMP_SN	-->	[frase_rel(suj(pron_rel),SV) COMP_SN]
COMP_SN	-->	[adj COMP_SN]
COMP_SN	-->	[SP COMP_SN]
COMP_SN	-->	[adv COMP_SN]

Tabela 3.5: Estruturas sintáticas do Top-down

### 3.2.3 Chart Parser

Para a análise de frases foram desenvolvidas regras para um *parser* Bottom-up, Chart Parser<sup>3</sup>.

#### Regras Gramaticais

As regras gramaticais do *parser* Chart Parser encontram-se na tabela 3.6.

#### Estruturas Sintáticas

As estruturas sintáticas que resultam da análise com o Chart Parser podem ser vistas na tabela 3.8.

<sup>3</sup><http://www.brawer.ch/prolog/earley/>



rule(F, [SN, SV]). rule(F, [SV]). rule(F, [SP_INTERR, SV]). rule(F, [SP_INTERR, SN, SV]). rule(F, [SP_INTERR, SV, SN]). rule(F, [adv_interr, SV]). rule(F, [adv_interr, SN, SV]). rule(F, [adv_interr, SV, SN]).
rule(SV, [VERBO_T]). rule(SV, [VERBO_T, COMP]). rule(SV, [VERBO_T, SN]). rule(SV, [VERBO_T, COMP, SN]). rule(SV, [VERBO_T, COMP, SN, COMP]).
rule(VERBO_T, [verbo, verbo, verbo]). rule(VERBO_T, [verbo, verbo]). rule(VERBO_T, [verbo]).
rule(COMP, [SP]). rule(COMP, [SP, COMP]). rule(COMP, [adv]). rule(COMP, [adv, COMP]).
rule(SP, [prep, SN]).
rule(SN, [pron]). rule(SN, [det, pron]). rule(SN, [NOME_T]). rule(SN, [COMP_SN, NOME_T]).

Tabela 3.6: Regras para o Chart Parser

<p>rule(SN, [NOME_T, COMP_SN]).  rule(SN, [COMP_SN, NOME_T, COMP_SN]).  rule(SN, [det, NOME_T]).  rule(SN, [det, COMP_SN, NOME_T]).  rule(SN, [det, NOME_T, COMP_SN]).  rule(SN, [det, COMP_SN, NOME_T, COMP_SN]).  rule(SN, [num, NOME_T]).  rule(SN, [num, COMP_SN, NOME_T]).  rule(SN, [num, NOME_T, COMP_SN]).  rule(SN, [num, COMP_SN, NOME_T, COMP_SN]).  rule(SN, [det, num, NOME_T]).  rule(SN, [det, num, COMP_SN, NOME_T]).  rule(SN, [det, num, NOME_T, COMP_SN]).  rule(SN, [det, num, COMP_SN, NOME_T, COMP_SN]).  rule(SN, [num]).  rule(SN, [num, COMP_SN]).  rule(SN, [det, num]).  rule(SN, [det, num, COMP_SN]).</p>
<p>rule(NOME_T, [nomeComum]).  rule(NOME_T, [nomeProprio]).  rule(NOME_T, [nomeComum, nomeProprio]).</p>
<p>rule(COMP_SN, [pron_rel, F]).  rule(COMP_SN, [pron_rel, F, COMP_SN]).  rule(COMP_SN, [adj]).  rule(COMP_SN, [adj, COMP_SN]).  rule(COMP_SN, [SP]).  rule(COMP_SN, [SP, COMP_SN]).  rule(COMP_SN, [adv]).  rule(COMP_SN, [adv, COMP_SN]).</p>

Tabela 3.7: Regras para o Chart Parser

F	-->	frase(suj(SN),SV).
F	-->	frase(suj([]),SV).
SV	-->	sv(VERBO.T,[],COMP,[])
SV	-->	sv(VERBO.T,SN,COMP,COMP)
VERBO.T	-->	verbo
COMP	-->	[]
COMP	-->	[SP COMP]
COMP	-->	[adv COMP]
SP	-->	sp(prepos,SN)
SN	-->	sn([],num,[],pron(pron),[],[])
SN	-->	sn(tipo-pron,num,[],pron(pron),[],[])
SN	-->	sn([],num,[],NOME.T,COMP,COMP)
SN	-->	sn(tipo-det,num,[],NOME.T,COMP,COMP)
SN	-->	sn([],num,numero,NOME.T,COMP,COMP)
SN	-->	sn(tipo-det,num,numero,NOME.T,COMP,COMP)
SN	-->	sn([],num,numero,[],COMP,[])
SN	-->	sn(det-tipo,num,numero,[],COMP,[])
NOME.T	-->	nome(nomeComum,[])
NOME.T	-->	nome([],nomeProprio)
NOME.T	-->	nome(nomeComum,nomeProprio)
COMP.SN	-->	[]
COMP.SN	-->	[frase_rel(suj(pron_rel),SV) COMP]
COMP.SN	-->	[adj COMP]
COMP.SN	-->	[SP COMP]
COMP.SN	-->	[adv COMP]

Tabela 3.8: Estruturas sintáticas do Chart Parser

### 3.2.4 Reescrever Estruturas Sintáticas

Algumas das estruturas obtidas com o *parser* Top-down são diferentes das obtidas com o Chart Parser.

Pode ver-se nas tabelas 3.5 e 3.8, que as diferenças nas estruturas residem ao nível dos complementos dos sintagmas verbais e nominais.

No caso Top-down existe um único conjunto de complementos nos sintagmas verbais e nominais, enquanto que no Chart Parser existem dois conjuntos de complementos.

A obtenção de estruturas sintáticas iguais em ambos os *parser's* é um passo importante, porque na fase de análise semântica (secção 3.3) pode utilizar-se a mesma aplicação para as estruturas sintáticas geradas pelos dois *parser's*.

De forma a obter estruturas sintáticas iguais, as estruturas do Chart Parser são reescritas. As modificações das estruturas podem ser vista na tabela 3.9.

<code>sv(VERBO_T, [], COMP, []) =&gt; sv(VERBO_T, [], COMP)</code>
<code>sv(VERBO_T, SN, COMP, COMP) =&gt; sv(VERBO_T, SN, COMP)</code>
<code>sn([], num, [], pron(pron), [], []) =&gt;</code> <code>sn([], num, [], pron(pron), [])</code>
<code>sn(tipo_pron, num, [], pron(pron), [], []) =&gt;</code> <code>sn(tipo_pron, num, [], pron(pron), [])</code>
<code>sn([], num, [], NOME_T, COMP, COMP) =&gt;</code> <code>sn([], num, [], NOME_T, COMP)</code>
<code>sn(tipo_det, num, [], NOME_T, COMP, COMP) =&gt;</code> <code>sn(tipo_det, num, [], NOME_T, COMP)</code>
<code>sn([], num, numero, NOME_T, COMP, COMP) =&gt;</code> <code>sn([], num, numero, NOME_T, COMP)</code>
<code>sn(tipo_det, num, numero, NOME_T, COMP, COMP) =&gt;</code> <code>sn(tipo_det, num, numero, NOME_T, COMP)</code>
<code>sn([], num, numero, [], COMP, []) =&gt;</code> <code>sn([], num, numero, [], COMP)</code>
<code>sn(tipo_det, num, numero, [], COMP, []) =&gt;</code> <code>sn(tipo_det, num, numero, [], COMP)</code>

Tabela 3.9: Reescrever Estruturas Sintáticas do Chart Parser

Na reescrita das estruturas do Chart Parser, os dois conjuntos de complementos dos sintagmas verbais são unidos num conjunto de complementos. De forma igual são unidos os dois conjuntos complementos de cada sintagma nominal.

## 3.3 Análise Semântica

A análise semântica é feita com um analisador semântico (subsecção 3.3.1). Este transforma estruturas sintáticas em estruturas semânticas ou DRS's.

### 3.3.1 Analisador Semântico

Um analisador semântico reescreve estruturas sintáticas em estruturas semânticas. Uma estrutura semântica ou DRS é um modelo de representação do discurso. Um

discurso é representado por um conjunto de referentes (variáveis) e um conjunto de condições (predicados).

O modelo adoptado para a representação do discurso tem um conjunto de referentes (as entidades do discurso) e dois conjuntos de condições (as condições da pressuposição e as condições que resultam da representação do predicado principal da frase).

O conjunto de referentes tem informação sobre os sintagmas nominais e advérbios que os introduziram no discurso. Um referente de um sintagma nominal contém o tipo do sintagma nominal (definido ou indefinido), o número do sintagma nominal (singular ou plural), um número caso o sintagma nominal seja composto por mais que um elemento, a variável que representa a entidade introduzida pelo sintagma nominal e uma lista com as variáveis que representam os advérbios que estão na lista de complementos do sintagma nominal. Um referente de advérbio contém o nome do advérbio e a variável que representa a entidade introduzida no discurso pelo advérbio.

O primeiro conjunto de condições contém predicados que resultam da representação de nomes comuns, nomes próprios, adjectivos, etc.

O segundo grupo de condições resulta da reescrita do verbo principal da frase.

Na Figura 3.12 pode ver-se a transformação de uma estrutura sintáctica da frase, “Lisboa é a capital de Portugal”, numa DRS.

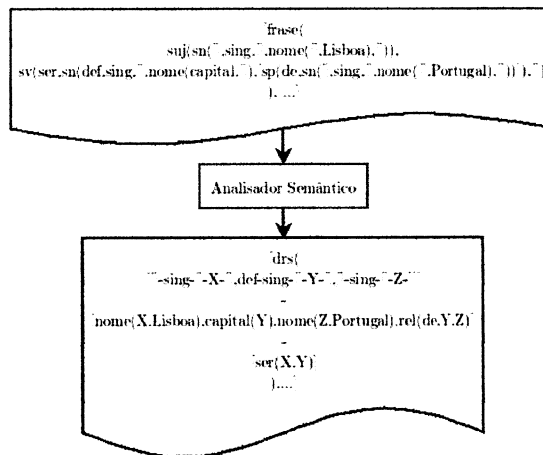


Figura 3.12: Analisador Semântico

O analisador semântico para cada estrutura sintáctica de uma frase encontra o conjunto de estruturas semânticas que a representam.

No exemplo da Figura 3.12, o conjunto de referentes contém as variáveis X, Y e Z, obtidas na reescritas dos substantivos “Lisboa”, “capital” e “Portugal”, respectivamente.

O primeiro conjunto de condições contém as condições “nome(X,Lisboa)”, “capital(Y)”, “nome(Z,Portugal)” que designam os substantivos, “Lisboa”, “capital” e “Portugal”, respectivamente, e a condição, “rel(de,Y,Z)”, representa o sintagma preposicional “capital de Portugal”.

O segundo conjunto de condições contém uma condição que resulta da representação do verbo “ser”.

Na Figura 3.13 podem ver-se as regras para a reescrita de uma estrutura sintáctica numa DRS.

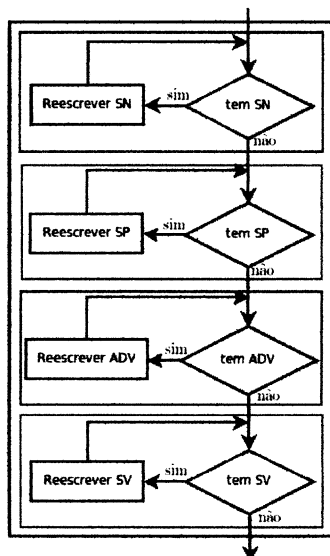


Figura 3.13: Regras para reescrever uma ES numa DRS

As regras para reescrever uma estrutura sintáctica numa DRS são as seguintes:

- Reescrever SN - esta é a primeira regra a aplicar na reescrita de uma estrutura sintáctica. Cada sintagma nominal da estrutura sintáctica é reescrito por variável;
- Reescrever SP - esta regra é aplicada após a reescrita dos sintagmas nominais. Cada sintagma preposicional da estrutura sintáctica resulta numa condição *rel* que contém três argumentos, a preposição e as duas variáveis que resultam da reescrita dos sintagmas nominais e do sintagma preposicional;
- Reescrever ADV - esta regra é aplicada em simultâneo com a regra para reescrever os sintagmas preposicionais. Cada advérbio é reescrito por uma variável (um referente);
- Reescrever SV - última regra a aplicar na reescrita de uma estrutura sintáctica. Cada sintagma verbal da estrutura sintáctica resulta numa condição com o nome do verbo no infinitivo e os argumentos são as preposições e as variáveis que resulta da reescrita dos sintagmas nominais e dos advérbios do sintagma verbal.

Na Figura 3.14 pode ver ao pormenor a reescrita de uma estrutura sintáctica da frase “Lisboa é a capital de Portugal” numa DRS.

A reescrita da estrutura sintáctica inicia-se com a estrutura sintáctica encontrada com o analisador sintáctico e uma DRS sem informação, ou seja, uma estrutura com três estruturas vazias.

Na reescrita, à medida que a estrutura sintáctica é reescrita, vai-se construindo a DRS. A reescrita termina quando a estrutura sintáctica estiver totalmente reescrita numa estrutura vazia, “[ ]”.

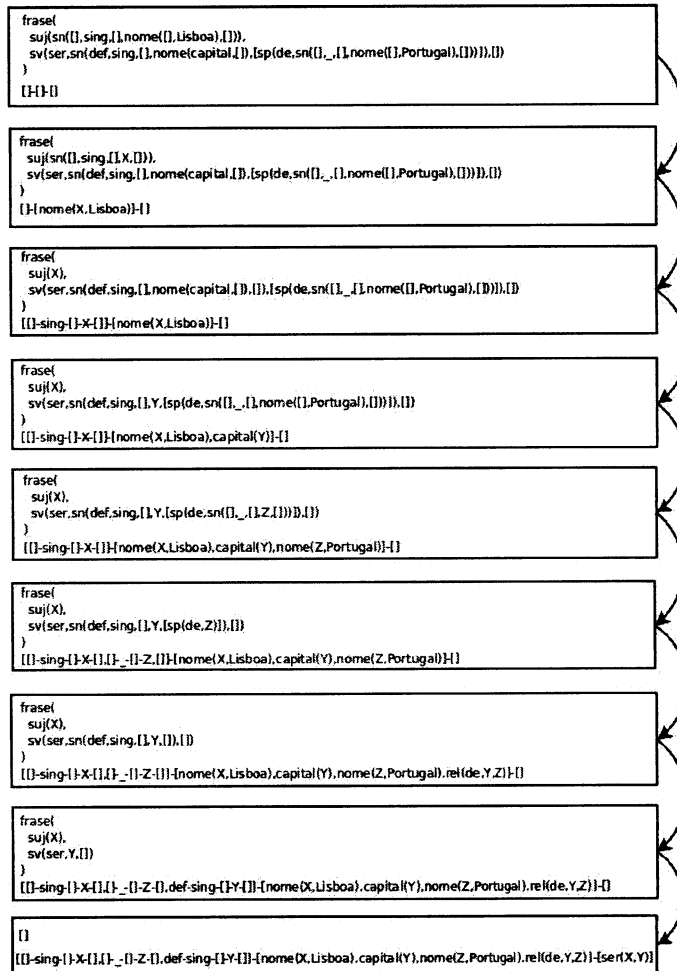


Figura 3.14: Reescrever ES em DRS

### Escolha da Melhor DRS

Algumas frases podem ser representadas por mais do que uma estrutura sintáctica e mais do que uma DRS.

Para escolher a melhor representação de uma frase decidiu-se utilizar apenas as DRS's.

Na escolha de entre as DRS's a que melhor representa uma frase, devem escolher-se aquelas que têm menor número de referentes e de entre estas as que têm mais referentes definidos e menor número de argumentos nos predicados que correspondem à reescrita dos verbos.

Nas Tabelas 3.10 e 3.11 podem ver-se as estruturas sintácticas e as DRS's, geradas pelo analisador sintáctico/semântico para a pergunta nº 10 do CLEF, "Quando foi atribuído o Prémio Nobel da Paz a Yasser Arafat?".

<code>frase(suj([], sv(atribuir, sn(def, sing, [], nome([], 'Prémio Nobel'), []), [quando, sp(de, sn(def, sing, [], nome([], 'Paz'), []), sp(a, sn([], \_, [], nome([], 'Yasser Arafat'), [])))]))</code>
<code>frase(suj([], sv(atribuir, sn(def, sing, [], nome([], 'Prémio Nobel'), []), [quando, sp(de, sn(def, sing, [], nome([], 'Paz'), [sp(a, sn([], \_, [], nome([], 'Yasser Arafat'), [])))]))</code>
<code>frase(suj([], sv(atribuir, sn(def, sing, [], nome([], 'Prémio Nobel'), [sp(de, sn(def, sing, [], nome([], 'Paz'), []))], [quando, sp(a, sn([], \_, [], nome([], 'Yasser Arafat'), [])))]))</code>
<code>frase(suj([], sv(atribuir, sn(def, sing, [], nome([], 'Prémio Nobel'), [sp(de, sn(def, sing, [], nome([], 'Paz'), []), sp(a, sn([], \_, [], nome([], 'Yasser Arafat'), []))], [quando]))</code>
<code>frase(suj([], sv(atribuir, sn(def, sing, [], nome([], 'Prémio Nobel'), [sp(de, sn(def, sing, [], nome([], 'Paz'), [sp(a, sn([], \_, [], nome([], 'Yasser Arafat'), [])))]))], [quando]))</code>
<code>frase(suj([], sv(atribuir, sn(def, sing, [], nome([], 'Prémio Nobel da Paz'), [], [quando, sp(a, sn([], \_, [], nome([], 'Yasser Arafat'), []))]))</code>
<code>frase(suj([], sv(atribuir, sn(def, sing, [], nome([], 'Prémio Nobel da Paz'), [sp(a, sn([], \_, [], nome([], 'Yasser Arafat'), []))], [quando]))</code>

Tabela 3.10: ES's de uma frase

Na escolha da melhor DRS, as DRS's (5) e (6) devem ser escolhidas, visto terem menor número de referentes. De entre as DRS's (5) e (6), devia-se escolher-se a DRS (6), porque é aquela que contém menor número de argumentos no predicado que representa o verbo da frase. No entanto, o verbo "atribuir" necessita de complemento indirecto, então deve escolher-se a DRS (5).

Nas Tabelas 3.12 e 3.13 podem ver-se as estruturas sintácticas e as DRS's, geradas pelo analisador sintáctico/semântico para a pergunta nº 14 do CLEF, "Como se chamavam os pilotos suicidas japoneses?".

Na escolha da melhor DRS, a DRS (1) deve ser escolhidas, visto ter menor número de referentes. Ainda, a DRS (1) é aquela que contém menor número de argumentos no predicado que representa o verbo da frase.

A utilização do analisador sintáctico/semântico em algumas frases produz um número elevado de DRS's. Por exemplo, as frases que têm muitos sintagmas preposicionais e verbos com múltiplas interpretações dão origem a muitas estruturas sintácticas e consequentemente de DRS's.



(1) drs([quando-1, []-_-[]-2-[], def-sing-[]-3-[], def-sing-[]-4-[]], [nome(2, 'Yasser Arafat'), nome(3, 'Prémio Nobel'), nome(4, 'Paz')], rel(a,4,2)], [atribuir(3,1,de:4)]).
(2) drs([quando-1, []-_-[]-2-[], def-sing-[]-3-[], def-sing-[]-4-[]], [nome(2, 'Yasser Arafat'), nome(3, 'Prémio Nobel'), nome(4, 'Paz')], rel(de,3,4)], [atribuir(3,1,a:2)]).
(3) drs([quando-1, []-_-[]-2-[], def-sing-[]-3-[], def-sing-[]-4-[]], que-sing-[]-4-[]], [nome(2, 'Yasser Arafat'), nome(3, 'Prémio Nobel'), nome(4, 'Paz')], rel(a,3,2), rel(de,3,4)], [atribuir(3,1)]).
(4) drs([quando-1, []-_-[]-2-[], def-sing-[]-3-[], def-sing-[]-4-[]], [nome(2, 'Yasser Arafat'), nome(3, 'Prémio Nobel'), nome(4, 'Paz')], rel(a,4,2), rel(de,3,4)], atribuir(3,1)).
(5) drs([quando-1, []-_-[]-2-[], def-sing-[]-3-[]], [nome(2, 'Yasser Arafat'), nome(3, 'Prémio Nobel da Paz')], [atribuir(3,1,a:2)]).
(6) drs([quando-1, []-_-[]-2-[], def-sing-[]-3-[]], [nome(2, 'Yasser Arafat'), nome(3, 'Prémio Nobel da Paz')], rel(a,3,2)], [atribuir(3,1)]).

Tabela 3.11: DRS's de uma frase

Na Tabela 3.14 encontram-se os números de estruturas sintáticas e DRS geradas pelo analisador sintático/semântico com os *parser's* Top-down e Chart Parser para a frase, “O grupo parlamentar do PS vai dar início à discussão interna sobre a concretização de um referendo à regionalização ainda na presente legislatura.”, retirada do jornal Público. Ainda na mesma tabela são apresentados os tempos de execução da frase pelo analisador sintático/semântico com cada um dos *parser's*.

O número de estruturas sintáticas encontradas com o Chart Parser relativamente às encontradas com o Top-Down é de 25% ( $594/2376 = 0,25$ ), enquanto que os números de DRS's são muito próximos, cerca de 94% ( $340/360 = 0,94(4)$ ).

Quanto aos tempos de execução com o Chart Parser relativamente ao Top-down é de cerca de 10% ( $116.579/1.161.641 = 0,10$ ).

Apesar de no Chart Parser, o número de estruturas sintáticas encontradas ser de 25% das encontradas com o Top-down, o tempo de execução com o Chart Parser foi muito inferior relativamente ao Top-down, cerca de 10%. Isto deve-se à optimização do Chart Parser relativamente ao Top-down, que quando inicia uma pesquisa primeiro procura informação nos seus registos sobre estruturas anteriormente analisadas.

frase(suj([], sv(chamar, sn(def, plu, [], nome(pilotos, []), [suicidas, japoneses]), [como])).
frase(suj([], sv(chamar, sn(def, plu, [], nome(pilotos, suicidas, []), [japoneses]), [como])).
frase(suj([], sv(chamar, sn(def, plu, [], nome(suicidas, []), [pilotos, japoneses]), [como])).
frase(suj([], sv(chamar, sn(def, plu, [], nome(suicidas, japoneses, []), [pilotos]), [como])).
frase(suj([], sv(chamar, sn(def, plu, [], nome(japoneses, []), [pilotos, suicidas]), [como])).
frase(suj(sn(def, plu, [], nome(pilotos, []), [suicidas, japoneses]), sv(chamar, [], [como])).
frase(suj(sn(def, plu, [], nome(pilotos, suicidas, []), [japoneses]), sv(chamar, [], [como])).
frase(suj(sn(def, plu, [], nome(suicidas, []), [pilotos, japoneses]), sv(chamar, [], [como])).
frase(suj(sn(def, plu, [], nome(suicidas, japoneses, []), [pilotos]), sv(chamar, [], [como])).
frase(suj(sn(def, plu, [], nome(japoneses, []), [pilotos, suicidas]), sv(chamar, [], [como])).
frase(suj(sn(def, plu, [], nome(pilotos, []), []), sv(chamar, sn([], plu, [], nome(suicidas, []), [japoneses]), [como])).
frase(suj(sn(def, plu, [], nome(pilotos, []), []), sv(chamar, sn([], plu, [], nome(suicidas, japoneses, []), []), [como])).
frase(suj(sn(def, plu, [], nome(pilotos, []), []), sv(chamar, sn([], plu, [], nome(japoneses, []), [suicidas]), [como])).
frase(suj(sn(def, plu, [], nome(pilotos, []), [suicidas]), sv(chamar, sn([], plu, [], nome(japoneses, []), []), [como])).
frase(suj(sn(def, plu, [], nome(pilotos, suicidas, []), []), sv(chamar, sn([], plu, [], nome(japoneses, []), []), [como])).
frase(suj(sn(def, plu, [], nome(suicidas, []), [pilotos]), sv(chamar, sn([], plu, [], nome(japoneses, []), []), [como])).

Tabela 3.12: ES's de uma frase

(1) drs([como-1, def-plu-[]-2-[]], [japoneses(2), pilotos(2), suicidas(2)], [chamar(2,1)]).
(2) drs([como-1, []-plu-[]-2-[], def-plu-[]-3-[]], [japoneses(2), pilotos(3), suicidas(2)], [chamar(3,2,1)]).
(3) drs([como-1, []-plu-[]-2-[], def-plu-[]-3-[]], [japoneses(2), pilotos(3), suicidas(3)], [chamar(3,2,1)]).

Tabela 3.13: DRS's de uma frase

	Top Down	Chart-Parser
ES's	2376	594
DRS's	360	340
Tempo de execução	1.161.641 ms	116.579 ms

Tabela 3.14: ES's, DRS's e tempo de execução de uma frase

## **Capítulo 4**

# **Analizador Sintáctico/Semântico**

Neste capítulo está descrito o sistema proposto para análise sintáctica e semântica de frases em textos.

O capítulo está dividido em 2 secções: na secção 4.1 é apresentado o sistema e na secção 4.2 é feita a avaliação do sistema.

## 4.1 Arquitectura do Analisador Sintáctico/Semântico

Dado um ficheiro de texto, o sistema faz as análises sintáctica e semântica correspondentes. A Figura 4.1 apresenta o analisador sintáctico e semântico.

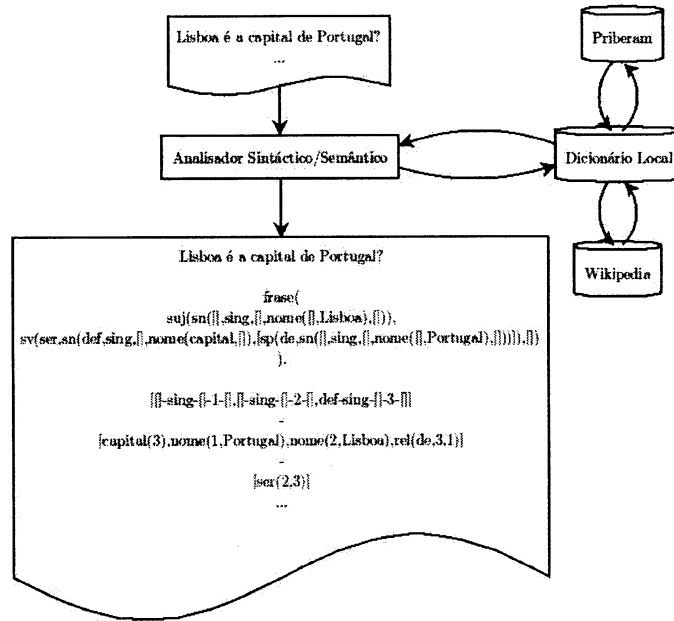


Figura 4.1: Analisador Sintáctico/Semântico

O analisador recebe o ficheiro texto e para cada frase encontra as estruturas sintácticas e semânticas. Durante a análise, sempre que são encontradas palavras ou nomes próprios que não existem no dicionário local são feitas consultas via Internet.

As consultas das palavras são feitas no dicionário Priberam, enquanto que as consultas dos nomes próprios são feitas na Wikipédia.

O analisador é composto por 4 módulos. Na Figura 4.2 pode ver-se a arquitectura do analisador sintáctico/semântico.

Os módulos são os seguintes: pré-processamento, análise lexical, análise sintáctica e análise semântica.

No pré-processamento, o texto é preparado para a análise. Na análise lexical, os *tokens* do texto são identificados. Na análise sintáctica são encontradas as estruturas sintácticas que representam cada uma das frases. Na análise semântica são encontradas as estruturas semânticas que representam cada uma das frases. Nas seguintes subsecções estão descritos cada um dos módulos.

### 4.1.1 Pré-processamento

No pré-processamento, o texto é preparado para a análise. Neste módulo, o texto é dividido em frases e as frases em *tokens* (palavras, números e sinais de pontuação). Além disso, os candidatos a nomes próprios são marcados e as contracções desdobradas, assim como, são removidas algumas expressões e movimentados os pronomes clíticos.

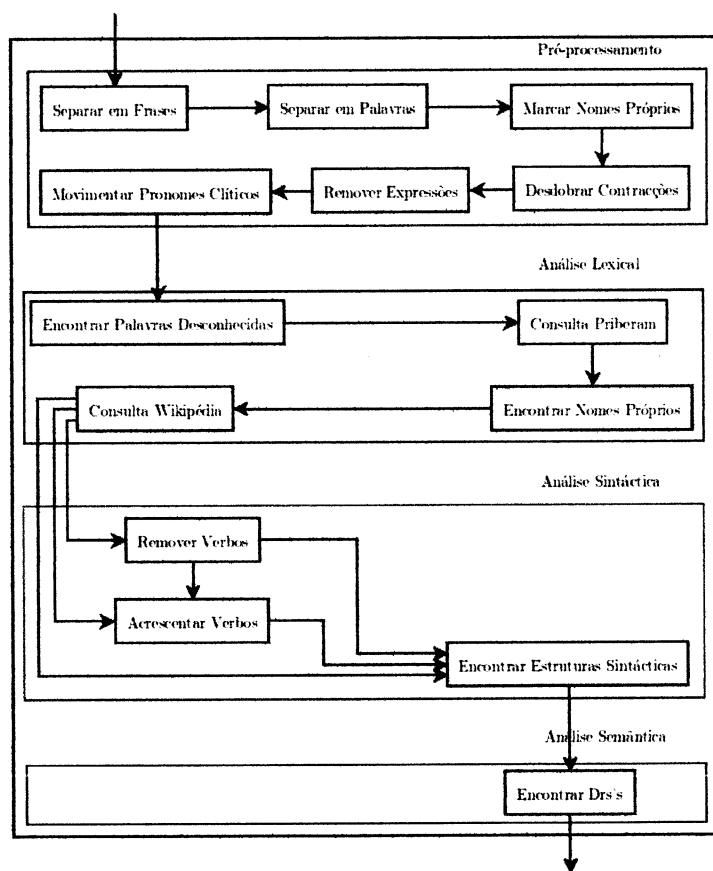


Figura 4.2: Arquitectura do Analisador Sintático/Semântico

Alguns dos passos deste módulo estão descritos nesta subsecção, os restantes estão descritos no capítulo 5, marcador de nomes próprios, na subsecção 5.1.1.

### Remover Expressões

Nem sempre é simples fazer a análise de uma frase. Em determinadas circunstâncias é necessário remover expressões.

Por exemplo, na frase “Quando é que John Lennon foi assassinado?”, a expressão “é que” deve ser removida. Na Figura 4.3 pode ver-se a remoção desta expressão.

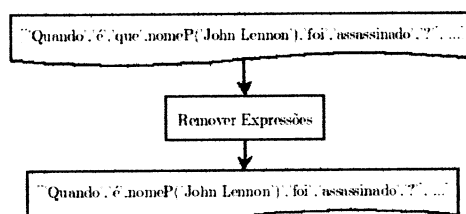


Figura 4.3: Remover Expressões

O sistema para cada interpretação de uma frase procura pela sequência de palavras “é” e “que”, se forem encontradas, são removidas da interpretação.

### Movimentar Pronome Clíticos

Nas frases a presença de pronomes clíticos antes de verbos dificulta a análise. A permutação dos pronomes clíticos com os verbos torna a análise das frases possível. Na Figura 4.4 pode ver-se a permutação do pronome clítico “se” com a conjugação “entregam” do verbo “entregar”.

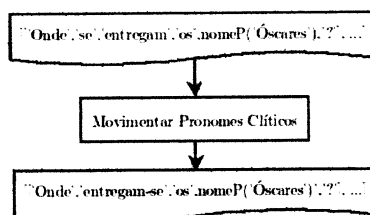


Figura 4.4: Movimentar Pronomes Clíticos

O sistema para cada interpretação de uma frase procura pela sequência pronome clítico seguido de verbo. Se a sequência ocorrer, o pronome permuta com o verbo, procurando-se por nova sequência.

Os pronomes clíticos considerados foram os seguintes: “me”, “te”, “se”, “lhe”, “nos”, “vos” e “lhes”.

### 4.1.2 Análise Lexical

Na análise lexical, as palavras do texto em análise que não se encontram do dicionário local são reunidas num ficheiro para consulta, assim como, os nomes próprios. Os

passos deste módulo estão descritos no capítulo 5, marcador de nomes próprios, na subsecção 5.1.2.

### 4.1.3 Análise Sintáctica

Na análise sintáctica são encontradas as estruturas sintácticas para cada frase do texto em análise. Neste módulo fala-se sobre a importância de acrescentar ou remover verbos em frases, nas quais não é possível obter análise sintáctica.

#### Remover Verbos

Algumas frases não são possíveis de analisar sem a remoção do verbo que surge no início. A remoção desse verbo pode tornar possível a análise. Na Figura 4.5 pode ver-se a remoção da conjugação “diga” do verbo “dizer” do início da frase.

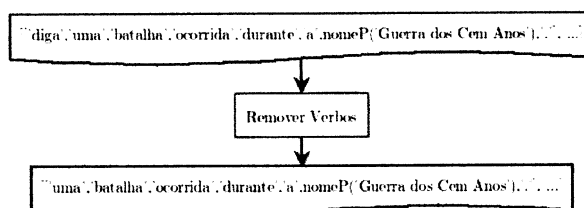


Figura 4.5: Remover Verbos

O sistema recebe o texto a analisar e, para cada interpretação de uma frase verifica se a primeira palavra é um dos verbos considerados para a remoção.

Os verbos considerados para a remoção são os seguintes: “dizer”, “indicar”, “encontrar”, “nomear”, “mencionar” e “referir”.

Se a primeira palavra de uma interpretação de uma frase for um destes verbos, então é removido.

#### Acrescentar Verbos

Algumas frases não são possíveis de analisar sem lhes serem acrescentadas uma conjugação do verbo “ser”. Por exemplo, as frases que têm no início um dos pronomes “qual” ou “quais” e não possuem quaisquer verbos. Na Figura 4.6 pode ver-se a adição numa interpretação de uma frase da conjugação “é” do verbo “ser”.

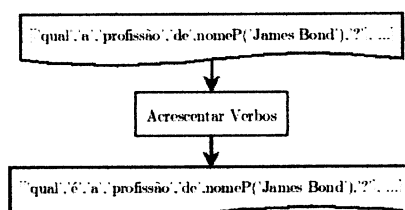


Figura 4.6: Acrescentar Verbos

O sistema recebe o texto a analisar e, para cada interpretação de uma frase que não tem quaisquer verbos e possui o pronome “qual” no início, acrescenta a conjugação



“é” do verbo “ser” após o pronome. Para o caso do pronome “quais” é acrescentada a conjugação “são”.

### Encontrar Estruturas Sintáticas

As frases sintacticamente correctas são representadas por uma ou várias estruturas sintáticas. Na Figura 4.7 pode ver-se a estrutura sintáctica que corresponde a uma interpretação da frase, “Lisboa é a capital de Portugal”.

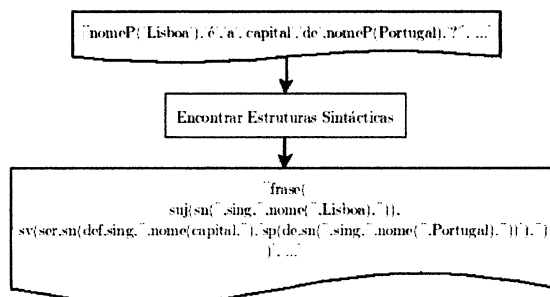


Figura 4.7: Encontrar Estruturas Sintáticas

O sistema recebe o texto a analisar, e para cada interpretação de uma frase encontra o conjunto de estruturas sintáticas que a representam.

### 4.1.4 Análise Semântica

Na análise semântica são encontradas as estruturas semânticas para cada frase do texto em análise. Neste módulo fala-se obtenção das estruturas semânticas (DRS) num texto.

#### Encontrar DRS's

Encontrar as DRS's de uma frase, corresponde a reescrever as estruturas sintáticas em DRS's.

Uma DRS é composta por um conjunto de referentes ou variáveis e, dois conjuntos de condições.

Na Figura 4.8 pode ver-se obtenção da DRS que corresponde a uma estruturas sintáctica da frase, “Lisboa é a capital de Portugal”.

Na geração das DRS, o sistema recebe as estruturas sintáticas de cada frase de um texto e, retorna as DRS que representam cada uma das estruturas sintáticas.

## 4.2 Avaliação de Resultados

A avaliação de resultados encontra-se dividida em duas subsecções. Na subsecção 4.2.1 é feita a avaliação das DRS's encontradas com o analisador sintáctico/semântico em 100 perguntas do CLEF (Texto 1 do CLEF - capítulo 7), utilizando os *parser's* Top-down e Chart Parser. Na subsecção 4.2.2 é feita a avaliação do número de estruturas sintáticas e DRS's encontradas em 3 corpora (700 perguntas do CLEF, 300 notícias do jornal, Público, e 200 frases do HAREM - capítulo 7) com o analisador

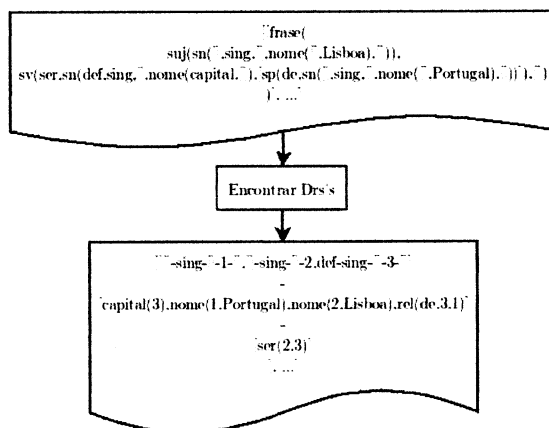


Figura 4.8: Encontrar Drs's

sintático/semântico utilizando os *parser's* Top-down e Chart Parser, sem recurso e com recurso ao marcador de nomes próprios.

#### 4.2.1 Avaliação de DRS's

Na avaliação das DRS's geradas pelo analisador sintático/semântico com os *parser's* Top-down e Chart Parser foram utilizadas 100 perguntas do CLEF (Texto 1 do CLEF - capítulo 7).

As variáveis a estudar na avaliação da DRS's são as seguintes: o número de DRS's encontradas (Enc), o número de DRS's correctas (Corr) das encontradas e o número de DRS's incorrectas (Inc) das encontradas de cada uma das perguntas.

Os resultados obtidos com o Top-down e Chart Parser na avaliação das DRS's podem ser vistos nas tabelas 4.1, 4.2 e 4.3.

As perguntas para as quais não se conseguiu obter qualquer DRS ou DRS correcta com o Top-down foram as seguintes:

- pergunta 65 - "Quando é que o príncipe Carlos e Diana se casaram?"
- pergunta 84 - "Em que ano teve lugar a guerra do Yom Kippur?"
- pergunta 86 - "Qual o vulcão activo mais alto da Europa?"

Na pergunta 65, o analisador não conseguiu encontrar quaisquer DRSs, porque não foram criadas regras na gramática para a coordenação de frases. O analisador reconhece o sintagma nominal "o príncipe Carlos e Diana" como estando no singular e tenta fazer a concordância sem sucesso com a conjugação "casarem" do verbo "casar" que se encontra no plural.

Na pergunta 84, o analisador encontrou duas DRS's que se encontram incorrectas. A expressão "teve lugar" nesta frase surge com um sentido que o sistema desconhece. O sentido correcto para esta expressão será "ocorreu".

Na pergunta 86, o analisador encontrou uma DRS que se encontra incorrecta. O analisador encontrou como verbo principal da frase a palavra "activo". Esta palavra nesta frase deve ser vista como adjectivo da palavra "vulcão". A resolução do problema passa por acrescentar a conjugação "é" do verbo "ser" após o pronome "Qual".

Nº da Frase	Top-down			Chart Parser		
	Enc	Corr	Inc	Enc	Corr	Inc
1	6	4	2	6	2	4
2	3	2	1	3	2	1
3	10	2	8	10	4	6
4	6	4	2	6	4	2
5	6	6	0	6	6	0
6	2	2	0	2	2	0
7	22	3	19	20	2	18
8	3	2	1	3	2	1
9	3	2	1	3	2	1
10	6	5	1	6	5	1
11	3	2	1	3	2	1
12	7	6	1	7	6	1
13	6	3	3	6	2	4
14	3	1	2	3	1	2
15	15	3	12	15	2	13
16	28	3	25	28	3	25
17	5	2	3	5	2	3
18	5	2	3	5	2	3
19	14	2	12	4	2	2
20	2	2	0	2	2	0
21	6	4	2	6	6	0
22	6	2	4	6	2	4
23	6	4	2	6	2	4
24	5	4	1	5	2	3
25	3	2	1	3	2	1
26	3	2	1	3	2	1
27	124	8	116	124	8	116
28	1	1	0	1	1	0
29	1	1	0	1	1	0
30	2	1	1	3	2	1
31	7	3	4	7	3	4
32	3	2	1	3	2	1
33	18	1	17	18	1	17
34	6	2	4	6	2	4

Tabela 4.1: Perguntas do CLEF avaliadas com Top-down e Chart Parser

Nº da Frase	Top-down			Chart Parser		
	Enc	Corr	Inc	Enc	Corr	Inc
35	4	2	2	4	2	2
36	10	7	3	10	7	3
37	1	1	0	1	1	0
38	1	1	0	2	1	1
39	2	2	0	2	2	0
40	2	2	0	1	1	0
41	18	5	13	18	5	13
42	4		3	4	1	3
43	4	2	2	6	3	3
44	12	7	5	12	7	5
45	6	3	3	6	2	4
46	1	1	0	0	0	0
47	6	2	4	6	1	5
48	12	4	8	12	4	8
49	3	2	1	3	2	1
50	8	3	5	8	2	6
51	2	2	0	2	2	0
52	6	6	0	6	6	0
53	2	1	1	2	2	0
54	2	1	1	2	2	0
55	1	1	0	2	1	1
56	3	2	1	3	2	1
57	3	3	0	3	2	1
58	2	2	0	2	2	0
59	6	1	5	6	2	4
60	2	1	1	2	1	1
61	1	1	0	1	1	0
62	4	2	2	4	2	2
63	2	2	0	2	2	0
64	25	2	23	25	2	23
65	0	0	0	0	0	0
66	2	2	0	2	2	0
67	9	7	2	9	5	4
68	1	1	0	1	1	0

Tabela 4.2: Perguntas do CLEF avaliadas com Top-down

Nº da Frase	Top-down			Chart Parser		
	Enc	Corr	Inc	Enc	Corr	Inc
69	4	2	2	4	3	1
70	3	2	1	3	2	1
71	4	4	0	4	4	0
72	4	1	3	4	2	2
73	5	2	3	5	2	3
74	3	2	1	3	2	1
75	3	3	0	3	2	1
76	6	3	3	6	3	3
77	14	2	12	14	2	12
78	6	2	4	9	2	7
79	3	2	1	3	2	1
80	3	2	1	3	2	1
81	1	1	0	2	2	0
82	4	2	2	4	2	2
83	4	4	0	4	3	1
84	2	0	2	2	0	2
85	22	3	19	0	0	0
86	1	0	1	1	0	1
87	12	1	11	12	1	11
88	2	2	0	2	2	0
89	3	2	1	3	2	1
90	3	3	0	3	3	0
91	2	2	0	2	2	0
92	4	4	0	4	4	0
93	10	10	0	10	10	0
94	13	1	12	13	1	12
95	3	3	0	3	3	0
96	3	2	1	3	2	1
97	2	1	1	2	1	1
98	4	1	3	4	1	3
99	2	2	0	1	1	0
100	20	11	9	20	11	9

Tabela 4.3: Perguntas do CLEF avaliadas com Top-down e Chart Parser

As perguntas para as quais não se conseguiu obter qualquer DRS ou DRS correcta com o Chart Parser foram as seguintes:

- pergunta 46 - “Quantos anos de prisão cumpriu Nelson Mandela?”
- pergunta 65 - “Quando é que o príncipe Carlos e Diana se casaram?”
- pergunta 84 - “Em que ano teve lugar a guerra do Yom Kippur?”
- pergunta 85 - “Quantas pessoas morrem devido a minas anti-pessoais anualmente?”
- pergunta 86 - “Qual o vulcão activo mais alto da Europa?”

As perguntas 65, 84 e 86 falham com o Chart Parser pelas mesmas razões que falham com o Top-down, enquanto que as perguntas 46 e 85 falham devido a problemas na implementação das regras gramaticais do Chart Parser.

Na tabela 4.4 encontram-se as médias de DRS’s encontradas, DRS’s correctas das encontradas e DRS’s incorrectas das encontradas de 100 perguntas do CLEF para os *parser’s* Top-down e Chart Parser.

Nº de Frases	Top-down			Chart Parser		
	Enc	Corr	Inc	Enc	Corr	Inc
100	6,88	2,59	4,29	6,60	2,48	4,12

Tabela 4.4: Médias com Top-down e Chart Parser (100 frases CLEF)

Relativamente aos resultados obtidos nas 100 frases do CLEF com o analisador sintáctico/semântico utilizando os *parser’s* Top-down e Chart Parser, o analisador com o Chart Parser apresentou ligeiramente resultados inferiores com uma percentagem de 62,42% (4,12/6,60) de DRS’s incorrectas enquanto que com o Top-down apresentou uma percentagem de 62,35% (4,29/6,88).

#### 4.2.2 Número de ES/DRS

De forma a avaliar os números de estruturas sintácticas e DRS’s geradas pelo analisador sintáctico/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando os *parser’s* Top-down e Chart Parser, foram utilizados 3 corpora (capítulo 7): 700 perguntas (7 textos de 100 frases cada) do CLEF, 300 notícias (3 textos de 100 frases cada) do jornal, Público e 200 frases (2 textos de 100 frases cada) do HAREM.

Na Tabela 4.5 encontram-se os números de estruturas sintácticas e DRS’s de 10 perguntas do CLEF (10 primeiras perguntas do Texto 1 do CLEF), obtidas com o analisador sintáctico/semântico sem recurso e com recurso ao marcador, utilizando os *parser’s* Top-down e Chart Parser.

Nas tabelas 4.6, 4.8, 4.10 encontram-se as médias de estruturas sintácticas e DRS’s de cada um dos textos dos 3 corpora, obtidas com o analisador sintáctico/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando os *parser’s* Top-down e Chart Parser.

Nas tabelas 4.7, 4.9, 4.11 encontram-se os números de frases que obtiveram análises sintácticas e semânticas em cada um dos 3 corpora com o analisador sintáctico/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando os *parser’s* Top-down e Chart Parser.

Nº da Frase	Analisador				Analisador c/ Marcador			
	Top-down		Chart Parser		Top-down		Chart Parser	
	ES	DRS	ES	DRS	ES	DRS	ES	DRS
1	10	6	9	6	2	2	9	6
2	6	3	6	3	6	3	6	3
3	18	10	17	10	18	10	17	10
4	20	6	20	6	20	6	20	6
5	24	6	14	6	24	6	14	6
6	8	2	8	2	8	2	4	2
7	40	22	36	20	24	16	36	20
8	3	3	3	3	3	3	3	3
9	3	3	3	3	3	3	3	3
10	7	6	7	6	2	2	3	3
MÉDIA	13,90	6,97	11,90	6,50	11,80	5,70	11,50	6,20

Tabela 4.5: ES e DRS com Top-down e Chart Parser (10 frases CLEF)

Texto	Nº Frases	Analisador				Analisador c/ Marcador			
		Top-down		Chart parser		Top-down		Chart parser	
		ES	DRS	ES	DRS	ES	DRS	ES	DRS
Texto1	100	23,63	6,88	14,38	6,68	22,05	6,23	11,84	5,55
Texto2	100	54,14	17,95	39,72	17,43	25,61	8,84	18,24	8,56
Texto3	100	18,41	7,53	13,91	7,28	14,58	6,29	12,36	6,48
Texto4	100	12,55	5,69	10,33	5,77	9,27	4,35	7,76	4,39
Texto5	100	15,35	7,97	11,61	7,88	10,48	5,36	7,74	5,38
Texto6	100	43,96	14,12	30,21	14,06	16,94	4,96	11,81	5,06
Texto7	100	22,73	10,63	20,18	10,61	18,52	8,22	16,43	8,22
Média	700	26,68	10,11	20,05	9,96	16,78	6,32	12,31	6,23

Tabela 4.6: Média ES e DRS com Top-down e Chart Parser (700 frases CLEF)

	Nº de Frase	Analisador		Analisador c/ Marcador	
		Top-down	Chart Parser	Top-down	Chart Parser
C/ Análise	700	663	662	663	662

Tabela 4.7: Nº de frases com análises sintática e semântica (700 frases CLEF)

Texto	Nº Frases	Analisador				Analisador c/ Marcador			
		Top-down		Chart Parser		Top-down		Chart Parser	
		ES	DRS	ES	DRS	ES	DRS	ES	DRS
Texto1	100	20,25	13,62	19,27	13,62	17,67	11,83	16,83	11,83
Texto2	100	26,38	12,07	23,28	11,93	22,32	9,16	8,28	5,35
Texto3	100	37,43	8,44	15,26	8,44	35,07	7,56	10,06	5,80
Média	300	28,02	11,38	19,27	11,33	25,02	9,52	11,73	7,66

Tabela 4.8: Média ES e DRS com Top-down e Chart Parser (300 frases Público)

	Nº Frase	Analisador		Analisador c/ Marcador	
		Top-down	Chart Parser	Top-down	Chart Parser
C/ Análise	300	34	34	34	34

Tabela 4.9: N° de frases com análises sintáctica e semântica (300 frases Público)

	Nº Frases	Analisador				Analisador c/ Marcador			
		Top-down		Chart Parser		Top-down		Chart Parser	
Texto		ES	DRS	ES	DRS	ES	DRS	ES	DRS
Texto1	100	1,86	0,75	0,80	0,44	1,84	0,74	0,79	0,44
Texto2	100	1,04	0,57	0,71	0,57	1,04	0,57	0,71	0,57
Média	200	1,45	0,66	0,75	0,51	1,44	0,66	0,75	0,51

Tabela 4.10: Média ES e DRS com Top-down e Chart Parser (200 frases HAREM)

Na tabela 4.12 encontram-se as médias de estruturas sintáticas e DRS's de cada um dos 3 corpora, obtidas com o analisador sintático/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando os *parser's* Top-down e o Chart Parser.

A tabela 4.12 mostra que o corpus que obteve em média mais estruturas sintáticas com o analisador sintático/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando o *parser* Top-down, foi o corpus do Público (28,02 s/ marcador e 25,02 c/ marcador). Relativamente ao *parser* Chart Parser, o corpus que apresentou mais estruturas sintáticas sem recurso e com recurso ao marcador de nomes próprios foi o corpus do CLEF (20,05 s/marcador e 12,31 c/ marcador).

Quanto ao número de DRS's, o corpus que apresentou em média mais DRS's com o analisador sintático/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando o *parser* Top-down, foi o corpus do Público (11,38 s/ marcador e 9,52 c/ marcador). Relativamente ao *parser* Chart Parser, foi o corpus do Público que apresentou novamente mais DRS's (11,33 s/ marcador e 7,66 c/ marcador).

Na tabela 4.13 encontram-se as percentagens de frases com análises sintáticas e semânticas de cada um dos 3 corpora, analisadas com o analisador sintático/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando os *parser's* Top-down e Chart Parser.

A tabela 4.13 mostra que o corpus que apresentou maior percentagem de frases analisadas com o analisador sintático/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando o *parser* Top-down, foi o corpus do CLEF (94,71% c/ e s/ marcador). Relativamente ao *parser* Chart Parser, foi novamente o corpus do CLEF que apresentou novamente maior percentagem de frases analisadas (93,14% c/ e s/ marcador)

Os melhores resultados com o corpus do CLEF devem-se ao analisador sintácti-

	Nº Frase	Analisador		Analisador c/ Marcador	
		Top-down	Chart Parser	Top-down	Chart Parser
C/ Análise	200	34	29	34	29

Tabela 4.11: N° de frases com análises sintáctica e semântica (300 frases Público)



Corpus	Nº Frases	Analisador				Analisador c/ Marcador			
		Top-down		Chart Parser		Top-down		Chart Parser	
		ES	DRS	ES	DRS	ES	DRS	ES	DRS
CLEF	700	26,68	10,11	20,05	9,96	16,78	6,32	12,31	6,23
Público	300	28,02	11,38	19,27	11,33	25,02	9,52	11,73	7,66
HAREM	200	1,45	0,66	0,75	0,51	1,44	0,66	0,75	0,51

Tabela 4.12: Média ES e DRS com Top-down e Chart Parser

Corpus	Nº Frases	Analisador		Analisador c/ Marcador	
		Top-down	Chart Parser	Top-down	Chart Parser
CLEF	700	0,9471	0,9314	0,9471	0,9314
Público	300	0,1130	0,1130	0,1130	0,1130
HAREM	200	0,1700	0,1450	0,1700	0,1450

Tabela 4.13: Percentagem de frases com análise

co/semântico ter sido desenvolvido para a análise de frases interrogativas, e visto este corpus conter apenas frases interrogativas. Os resultados inferiores com os corpora do jornal Público e HAREM devem-se à maioria das frases nestes corpora não serem interrogativas. Além disso, grande parte das frases nestes dois últimos corpora necessitam de coordenação de frases, que a gramática não cobre.

Quanto ao *parser* que apresentou maior percentagem de frases analisadas foi o Top-down.

Em relação ao número de estruturas sintáticas e DRSs, apesar de em média um corpus apresentar valores altos das mesmas, não significa que a maior parte das frases tenham sido analisadas. Exemplo disso, é o corpus do Público que apresentou valores altos em média de estruturas sintáticas (aproximadamente 28 ES's por frase) e DRSs (aproximadamente 11 DRS's por frase) com o analisador sintático/semântico, utilizando o *parser* Top-down sem recurso ao marcador de nomes próprios, e que no entanto analisou apenas cerca de 11% das frases.

## **Capítulo 5**

# **Marcador de Nomes Próprios**

Neste capítulo está descrito o sistema proposto para a marcação de nomes próprios em textos.

O capítulo está dividido em 2 secções: na secção 5.1 é apresentado o sistema e na secção 5.2 é feita a avaliação do sistema.

## 5.1 Arquitectura do Marcador de Nomes Próprios

Dado um ficheiro de texto, o sistema faz a marcação dos nomes próprios e avalia o resultado. Na Figura 5.1 é apresentado o sistema.

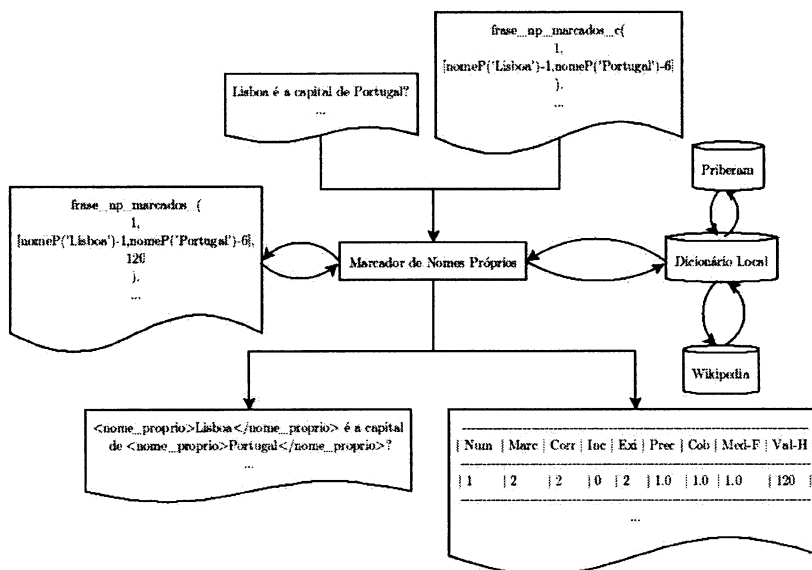


Figura 5.1: Marcador de nomes próprios

O sistema recebe dois ficheiros, um com o texto para marcação dos nomes próprios e um outro criado manualmente com todos os nomes próprios do primeiro ficheiro para usar na avaliação do sistema.

Na marcação dos nomes próprios recorre-se a regras internas ao sistema e ao dicionário local. Na ausência de palavras ou nomes próprios no dicionário local são feitas consultas via Internet. As consultas das palavras são feitas no dicionário Priberam, enquanto que as consultas dos nomes próprios são feitas na Wikipédia.

No resultado da marcação dos nomes próprios são obtidos dois ficheiros, um com o texto em que os nomes próprios surgem marcados e um outro com todos os nomes próprios que o sistema marcou.

Finalmente na avaliação são comparados o ficheiro criado manualmente que contém a lista de nomes próprios e o ficheiro gerado pelo sistema que contém a lista de nomes próprios.

O sistema é composto por 5 módulos. Na Figura 5.2 é apresentada a arquitectura do sistema.

No pré-processamento, o texto é decomposto em frase e as frases decompostas em *tokens* (palavras, números, sinais de pontuação, etc). Posteriormente é feita a marcação dos candidatos a nomes próprios e as contracções são desdobradas.

Na análise lexical são reunidas para consulta, as palavras e os candidatos a nomes próprios que não se encontram no dicionário local. As palavras são consultadas no dicionário Priberam e os candidatos a nomes próprios na Wikipédia.

Na Hipótese de Nomes Próprios, são atribuídas pontuações aos candidatos a nomes próprios e às interpretações das frases que apresentam estruturas sintácticas.

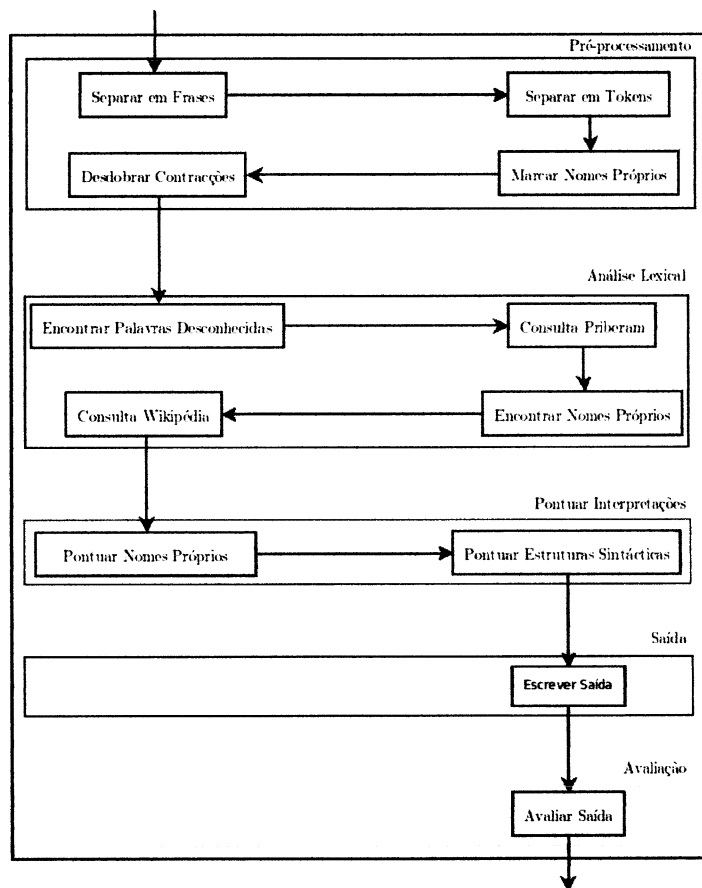


Figura 5.2: Arquitectura do marcador de nomes próprios

Na Saída são obtidos dois ficheiros, um com o texto em que os nomes próprios surgem marcados e um outro com a lista de nomes próprios marcados pelo sistema.

Na avaliação são avaliadas as marcações dos nomes próprios efectuadas pelo sistema relativamente a marcações manuais.

Nas seguintes sub-seções são descritos os módulos que constituem o sistema para a marcação de nomes próprios.

### 5.1.1 Pré-processamento

#### Separar em Frases

Para que o marcador analise um texto, o texto deve encontrar-se separado em frases. Na Figura 5.3 pode ver-se a separação de um texto em frases.

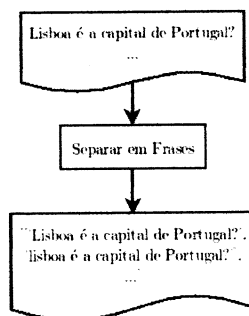


Figura 5.3: Separar em frases

Na separação das frases é utilizada uma estrutura que guarda, em cada posição, um conjunto de interpretações de uma frase. Uma interpretação de uma frase é composta por uma sequência de caracteres delimitados por plicas (“”). As frases que começam com maiúscula são desdobradas em duas interpretações (interpretação igual e interpretação que começa com minúscula) e as frases que começam com minúscula são desdobradas numa interpretação (interpretação igual).

#### Separar em Tokens

Para que uma frase possa ser analisada é necessário que esteja separada em *tokens*. Um *token* é constituído por uma sequência de caracteres sem espaços em branco. Os *tokens* podem ser palavras, números, sinais de pontuação, sequências de letras alternadas com números, etc. Na Figura 5.4 pode ver-se a separação de cada interpretação de uma frase em *tokens*.

O sistema recebe um conjunto de interpretações de frases, e por cada interpretação faz a sua separação em *tokens*, passando a interpretação a ser representada por uma estrutura de *tokens*.

Na figura 5.5 pode ver-se o processo de separação de uma interpretação de uma frase em *tokens*.

No processo de separação de uma interpretação de uma frase em *tokens* são utilizadas duas estruturas, numa guardam-se os caracteres do *token* em leitura e na outra os vários *tokens* da interpretação.

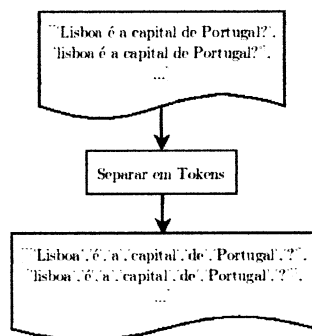


Figura 5.4: Separar em tokens

Interpretação	Token	Lista de Tokens
'Lisboa é a capital de Portugal?'	..	..
'lisboa é a capital de Portugal?'	↓ 'L'	..
'sboa é a capital de Portugal?'	↓ 'li'	..
'bon é a capital de Portugal?'	↓ 'lis'	..
'oa é a capital de Portugal?'	↓ 'lisb'	..
'a é a capital de Portugal?'	↓ 'lisbo'	..
' é a capital de Portugal?'	↓ 'lisboa'	..
' é a capital de Portugal?'	↓ ..	'Lisboa'
...	...	..
..	↓ ..	'Lisboa', 'é', 'a', 'capital', 'de', 'Portugal', '?'

Figura 5.5: Separação de frase em tokens

### Marcar Nomes próprios

Nas marcações dos candidatos a nomes próprios são reunidos os constituintes de possíveis nomes próprios. Os nomes próprios são constituídos por uma ou várias palavras que começam com maiúscula, podendo possuir números e elementos de ligação de nomes próprios. Os elementos de ligação de nomes próprios considerados são os seguintes: “de”, “da”, “do”, “das”, “dos” e “e”. Na figura 5.6 pode ver-se a marcação dos candidatos a nomes próprios em interpretações de frases.

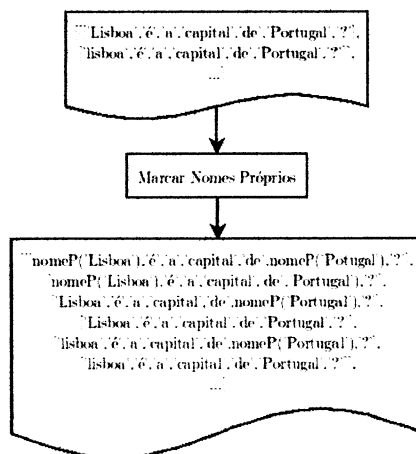


Figura 5.6: Marcar candidatos a nomes próprios

O sistema recebe um conjunto de interpretações de frases na marcação dos candidatos a nomes próprios. Em cada interpretação de uma frase são marcados os possíveis candidatos a nomes próprios, podendo uma interpretação ser desdobrada em várias outras interpretações.

Ainda relativamente aos nomes próprios, estes são divididos em datas, horas e numa categoria mais geral, o nome próprio.

Na Figura 5.7 pode ver-se o processo de marcar os candidatos a nomes próprios de uma interpretação de uma frase.

No processo de marcação dos candidatos a nomes próprios, são utilizadas duas estruturas auxiliares, numa guardam-se os *tokens* do candidato a nome próprio em leitura e na outra os candidatos a nomes próprios marcados e os restantes *tokens*.

O sistema para uma interpretação que tem  $N$  candidatos a nomes próprios produz  $2^N$  interpretações. As diferenças entre às interpretações são ao nível do número de candidatos a nomes próprios marcados que variam entre 0 e  $N$ .

### Desdobrar Contrações

No desdobramento das contrações, estas são desdobradas nos seus constituintes. Uma contração resulta da junção de duas palavras. Na Figura 5.8 pode ver-se o desdobramento das contrações.

O sistema recebe um conjunto interpretações de frases, e em cada contração que encontra numa interpretação faz o seu desdobramento. Na Figura 5.8 a contração “da” é desdobrada nos constituintes “de” e “a”.

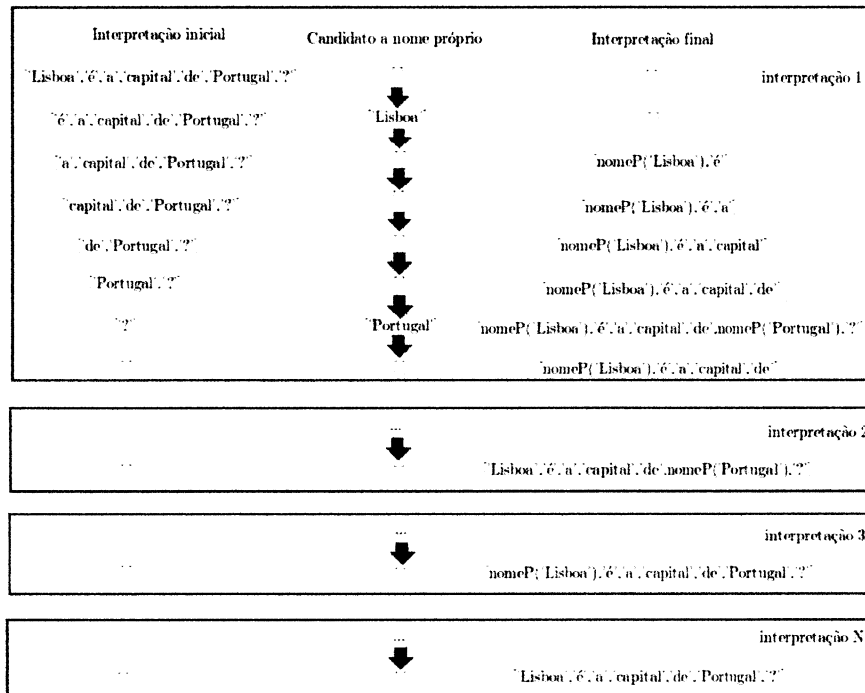


Figura 5.7: Marcação de candidatos a nomes próprios numa frase

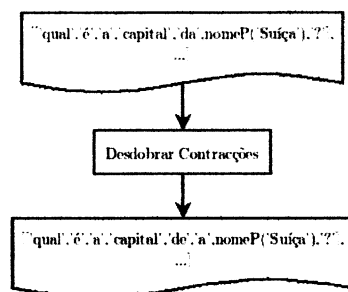


Figura 5.8: Desdobrar Contrações



### 5.1.2 Análise Lexical

#### Encontrar Palavras Desconhecidas

Na análise de um texto pelo marcador de nomes próprios, a consulta das características gramaticais de cada palavra é uma tarefa indispensável. Na Figura 5.9 pode ver-se a inserção num ficheiro das palavras que não se encontram no dicionário local.

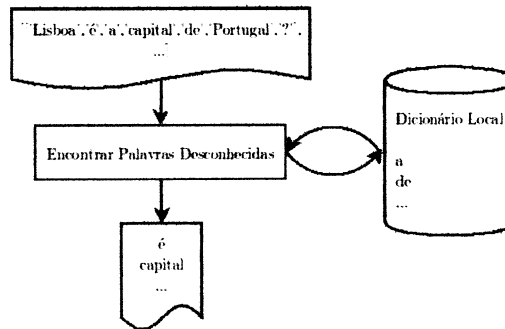


Figura 5.9: Encontrar Palavras Desconhecidas

O sistema para cada frase reúne as palavras que não se encontram no dicionário local num ficheiro para consulta no dicionário da Priberam.

A consulta das palavras no dicionário da Priberam é feita com a aplicação apresentada na subsecção 3.1.2.

#### Encontrar Nomes Próprios

Na análise de um texto deve dispor-se de informação sobre os candidatos a nomes próprios marcados, em particular se têm entrada na Wikipédia. Na Figura 5.10 pode ver-se a inserção num ficheiro dos candidatos a nomes próprios que não se encontram no dicionário local.

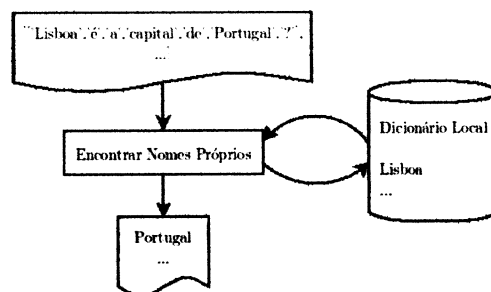


Figura 5.10: Encontrar nomes próprios

O sistema para cada frase reúne os candidato a nomes próprios que não se encontram no dicionário local para consulta num ficheiro.

A consulta dos candidatos a nomes próprios é feita com apresentada na subsecção 3.1.3.

### 5.1.3 Pontuar Interpretações

A marcação dos candidatos a nomes próprios numa frase pode produzir várias interpretações. A melhor interpretação de uma frase é aquela que contém o maior número de nomes próprios marcados correctamente e que é possível de ser representada por uma estrutura sintáctica. Para encontrar a melhor interpretação de uma frase pode recorrer-se a uma função heurística que pontua cada interpretação.

#### Pontuar Nomes Próprios

As palavras que começam com maiúscula e valores numéricos podem ser marcados como constituintes de nomes próprios, no entanto, algumas dessas marcações são incorrectas. A atribuição de pontuação aos candidatos a nomes próprios numa interpretação de uma frase é um passo importante na escolha da melhor interpretação.

Nos candidatos a nomes próprios existem características comuns. Essas características podem ser utilizadas para agrupar os candidatos com características iguais e atribuir-lhes iguais pontuações.

Os tipos de candidatos a nomes próprios considerados são os seguintes:

- NP WIKI - candidato a nome próprio que tem entrada na Wikipédia;
- NP SIMPLES - candidato a nome próprio que não tem entrada na Wikipédia e, é composto por apenas uma palavra das seguintes classes gramaticais: substantivo comum, adjectivo ou nome próprio;
- NP COMPOSTO - candidato a nome próprio que não tem entrada na Wikipédia e, é composto por mais que uma palavra, em que a primeira palavra pode ser das seguintes classes gramaticais: substantivos comum, adjectivos, verbo ou nome próprio;
- NP NUM - candidato a nome próprio que não tem entrada na Wikipédia e, composto por um valor numérico;
- NP ASPAS - candidato a nome próprio delimitado por aspas (“”);
- NP DATA - candidato a nome próprio marcado como data;
- NP HORA - candidato a nome próprio marcado como hora;
- NAO NP - candidato a nome próprio que não reúne nenhuma das características anteriores.

Na Figura 5.11 pode ver-se a atribuição de pontuação aos candidatos a nomes próprios.

O sistema recebe um conjunto de interpretações de frases pontuando os candidatos a nomes próprios de cada interpretação com uma função heurística. Estas pontuações são acrescentadas às interpretações correspondentes.

#### Pontuar Estruturas Sintácticas

As interpretações que são representáveis por estruturas sintácticas possuem um grau de importância elevado. A atribuição de pontuações a essas interpretações, é um passo importante na escolha da melhor interpretação.

As interpretações, que são representáveis por estruturas sintácticas, podem ser divididas em:

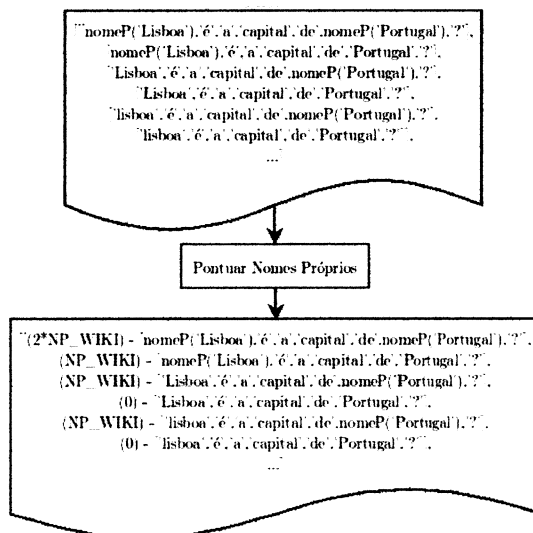


Figura 5.11: Pontuar nomes próprios

- TOTAL REP - interpretação totalmente representável por uma estrutura sintática;
- PARCIAL REP - interpretação parcialmente representável por uma estrutura sintática;

Na Figura 5.12 pode ver-se a atribuição de pontuação às interpretações que são representáveis por estruturas sintáticas.

O sistema recebe um conjunto de interpretações de frases pontuando as interpretações que são representáveis por estruturas sintáticas. Estas pontuações são somadas às pontuações atribuídas aos candidatos a nomes próprios.

### 5.1.4 Saída

#### Escrever Saída

A saída de um ficheiro processado pelo marcador de nomes próprios é constituída pelas interpretações com maiores pontuações atribuídas pela função heurística aos candidatos a nomes próprios e as interpretações representáveis com estruturas sintáticas.

Na Figura 5.13 pode ver-se a saída de um ficheiro processado pelo marcador de nomes próprios.

O sistema recebe um conjunto de interpretações de frases com as respectivas pontuações atribuídas aos candidatos a nomes próprios e às interpretações que contêm estruturas sintáticas, escolhendo a interpretação de cada frase que contém a maior pontuação para utilizar na escrita da saída do texto. Na saída, os nomes próprios são destacados das restantes palavras com *tags*.

Os tipos de nomes próprios considerados e as respectivas etiquetas são os seguintes:

- Nome próprio data é delimitado pelas etiquetas <data> e </data>;
- Nome próprio hora é delimitados pelas etiquetas <hora> e </hora>;

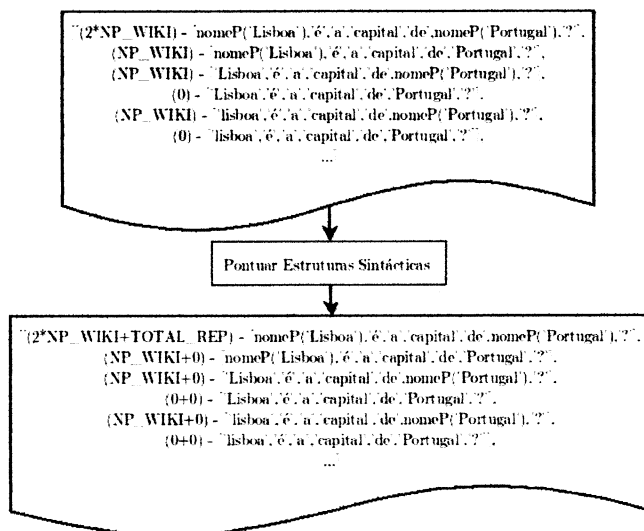


Figura 5.12: Pontuar estruturas sintáticas

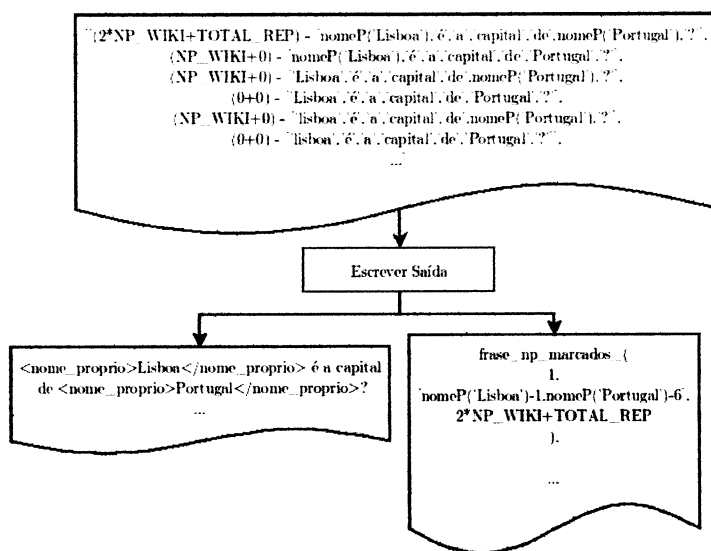


Figura 5.13: Escrever saída

- Restantes nomes próprios são delimitados pelas etiquetas <nome\_proprio> e </nome\_proprio>.

### 5.1.5 Avaliação

Na avaliação do marcador de nomes próprios foram utilizadas as métricas que se encontram na secção 2.2.3. Estas métricas foram adaptadas ao problema de marcação de nomes próprios.

#### Avaliar Saída

Na avaliação do marcador de nomes próprios são comparados os ficheiros com a lista de nomes próprios marcados manualmente e a listas de nomes próprios marcados pelo marcador.

Na Figura 5.14 pode ver-se o resultado da avaliação do marcador na frase “Lisboa é a capital de Portugal?”.

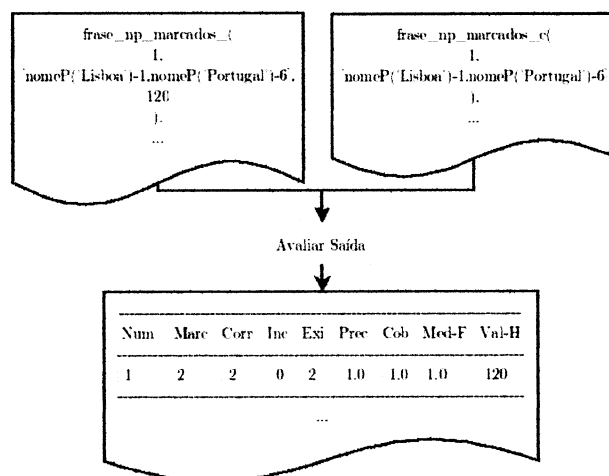


Figura 5.14: Avaliar saída

Na avaliação, cada frase marcada manualmente é comparada com a sua equivalente marcada pelo marcador. No resultado da comparação obtém-se uma tabela, em que cada entrada corresponde a uma frase com as seguintes variáveis: número da frase, número de nomes próprios marcados, número de nomes próprios correctos dos marcados, número de nomes próprios incorrectos dos marcados, número de nomes próprios existentes, a precisão, a cobertura, a medida-F e o valor da heurística.

## 5.2 Avaliação de Resultados

Na avaliação do sistema de marcação de nomes próprios foram utilizadas 3 corpora: 700 perguntas do CLEF, 300 notícias do jornal, o Público e 200 frases do HAREM.

As variáveis a estudar na avaliação do sistema de marcação de nomes próprios são as seguintes: os números por texto de nomes próprios marcados (Marc), nomes próprios correctos (Corr) dos marcados, nomes próprios incorrectos (Inc) dos marcados

e nomes próprios existentes (Exi), e as médias por texto de precisão (Prec), cobertura (Cob), medida-F (Med-F) e valor de heurística (Val-H).

Esta secção encontra-se dividida em quatro subsecções. Na subsecção 5.2.1 encontram-se as pontuações atribuídas em 8 testes aos candidatos a nomes próprios e às interpretações de frases que contêm estruturas sintácticas. Nas subsecções 5.2.2, 5.2.3 e 5.2.4 encontram-se os resultados obtidos com o marcador de nomes próprios ao processar cada um dos corpora (CLEF, Público e Harem).

### 5.2.1 Pontuações Atribuídas

Nos testes realizados com o sistema de marcação de nomes próprios foram feitas variações nas pontuações atribuídas aos candidatos a nomes próprios e às interpretações de frases representáveis por estruturas sintácticas. Mais informações sobre os candidatos a nomes próprios e as interpretações de frases representáveis por estruturas sintácticas podem ser encontradas na subsecção 5.1.3.

Em baixo podem ver-se as pontuações atribuídas pela função heurística no teste 1. Estas pontuações correspondem às que dão o melhor desempenho ao sistema nos diferentes corpora. Estas pontuações foram encontradas por tentativa e erro e não se garante que sejam as melhores.

- NP WIKI =  $10 * (1 + Es)$  (Es -é número de espaços em branco numa string).

Neste teste valoriza-se o facto de os nomes próprios terem entrada na Wikipédia e também se prefere os nomes próprios mais compridos (com mais palavras) se existirem na Wikipédia.

- NP SIMPLES = 4 e NP COMPOSTO =  $9 * (1 + Es)$  (Es -é número de espaços em branco numa string)

Valoriza-se o comprimento dos nomes próprios mesmo quando não se tem entrada na Wikipédia.

- NP NUM = -9

Desvalorizam-se números isolados que sejam marcados como nomes próprios.

- NP ASPAS = 15, NP DATA = 15 e NP HORA = 15

Valorizam-se expressões entre aspas que sejam marcadas como nomes próprios e outras expressões sejam marcadas como data e horas.

- NAO NP =  $-10 * (1 + Es)$  (Es -é número de espaços em branco numa string)

Desvalorizam-se expressões que não satisfazem os candidatos a nomes próprios anteriores.

- TOTAL REP = 100 e PARCIAL REP = 60

Com estas pontuações valorizam-se bastante as interpretações de frases que têm análise sintáctica (TOTAL REP=100 e PARCIAL REP=60)

Na tabela 5.1 pode ver-se heurística utilizada no Teste N.

Os parâmetros NP1, NP2, NP3, NP4, NP5, NP6, NP7 e NP8 são parâmetros para afinação das pontuações atribuídas aos candidatos a nomes próprios. Os parâmetros I1 e I2 são parâmetros para afinação das pontuações atribuídas às interpretações de

	Heurística
NP WIKI	$NP1*(1+Es)$
NP SIMPLES	NP2
NP COMPOSTO	$NP3*(1+Es)$
NP NUM	NP4
NP ASPAS	NP5
NP DATA	NP6
NP HORA	NP7
NAO NP	$NP8*(1+Es)$
TOTAL REP	I1
PARCIAL REP	I2

Tabela 5.1: Heurística

frases representáveis por estruturas sintáticas. O parâmetro Es é um parâmetro para contagem de palavras num candidato a nome próprio.

Na tabela 5.2 podem ver-se as pontuações utilizadas nos parâmetros para afinação das pontuações atribuídas aos candidatos a nomes próprios, às interpretações de frases representáveis por estruturas sintáticas e à contagem de palavras num candidato a nome próprio.

	Teste1	Teste2	Teste3	Teste4	Teste5	Teste6	Teste7	Teste8
I1	100	0	100	100	100	100	100	100
I2	60	0	60	60	60	60	60	60
NP1	10	10	0	10	10	10	10	10
NP2	4	4	4	5	4	4	4	4
NP3	9	9	9	5	9	9	9	9
NP4	-9	-9	-9	-9	0	-9	-9	-9
NP5	15	15	15	15	15	0	15	15
NP7	15	15	15	15	15	15	0	15
NP8	15	15	15	15	15	15	15	0
Es	N	N	N	0	N	N	N	N

Tabela 5.2: Pontuações atribuídas nos 8 testes

Com estes testes pretende ver-se o impacto:

- Teste 2 – da análise sintáctica. No Teste 2 não se tem em conta a existência de estrutura sintáctica para a frase com os nomes próprios.
- Teste 3 – da Wikipédia. No Teste 3 não se tem em conta a informação sobre a existência de entrada para o nome próprio na Wikipédia.
- Teste 4 – do comprimento do nome próprio. No Teste 4 não se valoriza o número de palavras do nome próprio.
- Teste 5 – dos números isolados. No Teste 5 não se valorizam números isolados que sejam marcados como nomes próprios.
- Teste 6 – das expressões entre aspas. No Teste 6 não se valorizam expressões entre aspas que sejam marcadas como nomes próprios.

- Teste 7 – das datas. No Teste 7 não se valorizam expressões que sejam marcadas como datas.
- Teste 8 – das horas. No Teste 8 não se valorizam expressões que sejam marcadas como horas.

### 5.2.2 CLEF

O CLEF (Cross-Language Evaluation Forum) é uma série de avaliações conjuntas que promove a pesquisa e desenvolvimento na área de recolha de informação entre várias línguas.

A participação do português tem sido financiada pela Linguateca, a nível de recursos humanos, e pelo diário Público (Portugal) e Folhas de São Paulo (Brasil), a nível de fornecimento de recursos.

A Linguateca disponibiliza a colecção CHAVE que contém textos, tópicos e perguntas utilizados nas edições do CLEF. Para mais informações pode-se consultar a página da colecção CHAVE.

Na avaliação do marcador de nomes próprios foram utilizados 7 textos de 100 perguntas cada do CLEF. Com este textos foram feitos 8 testes com variação das pontuações atribuídas aos candidatos a nomes próprios e às interpretações de frases representáveis por estruturas sintácticas.

#### Testes com corpus do CLEF

Os resultados obtidos nos 8 testes com o conjunto de textos do CLEF podem ser vistos na tabela 5.3.

Teste	Prec	Cob	Med-F	Val-H
Teste1	0,9779	0,9796	0,9788	102,96
Teste2	0,9707	0,9739	0,9723	16,87
Teste3	0,9315	0,9335	0,9325	92,59
Teste4	0,8949	0,9029	0,8989	96,77
Teste5	0,9779	0,9796	0,9788	102,98
Teste6	0,9746	0,9768	0,9757	101,90
Teste7	0,9779	0,9796	0,9788	102,95
Teste8	0,9779	0,9796	0,9788	102,46

Tabela 5.3: Testes com corpus do CLEF

Na Figura 5.15 pode ver-se um gráfico com as médias de precisão, cobertura e média-F para o corpus do CLEF nos 8 testes.

Na Figura 5.16 pode ver-se um gráfico com a média dos valores de heurística para o corpus do CLEF nos 8 testes.

Na Tabela 5.3 e no gráfico 5.15 pode ver-se que os testes que apresentaram os melhores resultados, foram os testes 1, 2, 5, 6, 7 e 8 com valores próximo dos 97% de precisão, cobertura e medida-F. Em relação aos testes 3 (não acede à Wikipédia) e 4 (não valoriza o tamanho dos nomes próprios), os resultados baixaram 4% e 8% valores respectivamente, em termos de precisão, cobertura e medida-F.

Na Figura 5.16 pode ver-se que os testes que obtiveram os valores mais elevados de heurística foram os testes 1, 5, 6, 7 e 8, com valores próximos de 102 de heurística.



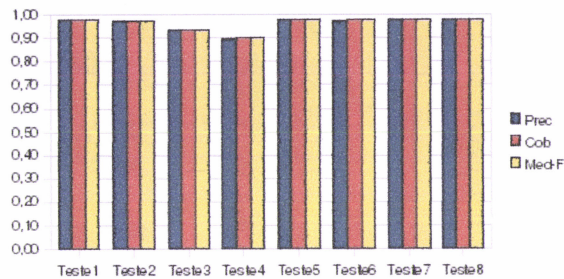


Figura 5.15: Testes com o corpus do CLEF

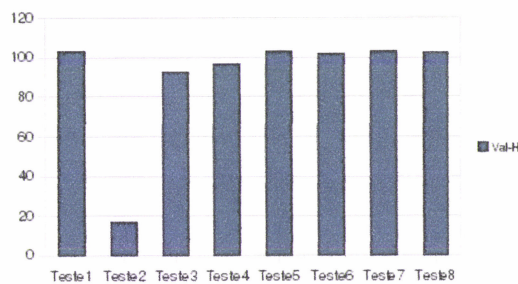


Figura 5.16: Testes com o corpus do CLEF

Em relação aos testes 2 (não pontua a análise sintáctica), 3 e 4, os resultados baixaram 86, 10 e 5 valores de heurística, respectivamente.

#### Análise das falhas no 1º teste com o corpus do CLEF

No 1º teste, o corpus que apresentou os melhores resultados foi o corpus do CLEF. Neste teste, as falhas foram nas seguintes perguntas:

- 65 - Quando é que o príncipe Carlos e Diana se casaram?
- 109 - Que tufão nas Filipinas causou mais de 1000 mortos ou desaparecidos?
- 201 - Qual o título do filme de Stephen Frears com Glenn Close, John Malkovich e Michelle Pfeiffer?
- 295 - Que prémio partilharam Yasser Arafat, Shimon Peres e Yitzhak Rabin?
- 322 - Que instituição constituíram o Brasil, Portugal, Angola, Moçambique, São Tomé e Príncipe, a Guiné-Bissau e Cabo Verde em Brasília?
- 409 - Quem recebeu o prémio Kafka da Literatura em 1995?
- 468 - Que fabricante de móveis gere um museu em Weil am Rhein?
- 515 - Que quantia exige o FC Sevilha de Diego Maradona?
- 553 - Onde aterrou o Airbus da Air France após a tomada de reféns em Argel?

- 555 - Quando se casaram Federico Fellini e Giulietta Masina?
- 593 - Quem é o Presidente da Câmara de Lisboa?
- 594 - Quem é o Presidente da Câmara de Lamego?
- 629 - Em que ano ocorreu o 25 de Abril?
- 631 - Qual o símbolo de liderança da Volta a Itália?
- 665 - Quem encerrou a 6ª Conferência Mediterrânica da Cruz Vermelha e do Crescente Vermelho?

As perguntas 65, 109, 201, 295, 322 e 555 falham devido à gramática não cobrir a coordenação de frases.

A pergunta 109 falha também devido à gramática permitir que números sejam marcados como datas. A resolução deste problema passa por não permitir que numa frase os números seguidos de substantivos sejam marcados como datas.

As perguntas 409, 553, 593 e 594 são marcadas correctamente. No entanto, como na avaliação apenas é considerada uma hipótese de marcação e a hipótese é diferente da marcada pelo sistema, então o sistema falha. A resolução passa por permitir na avaliação do sistema mais que uma solução (suporte à vagueza).

A pergunta 468 falha, porque o sistema foi implementado apenas para a língua portuguesa, não reconhecendo a palavra “am” como ligação de nomes próprios.

A pergunta 515 falha, porque o sistema no teste 1 promove nomes próprios que não se encontram na Wikipédia e que são constituídos por várias palavras.

A pergunta 629 falha, porque o sistema prefere datas a nomes próprios. Nesta frase, as palavras “25 de Abril” são um acontecimento e não uma data.

A pergunta 631 falha, porque o sistema não reconhece a palavra “a” como sendo um elemento de ligação de nomes próprios.

A pergunta 665 falha, porque o sistema não permite num nome próprio, dois elementos de ligação de nomes próprios juntos que no caso são a sequência de palavras “e do”.

### 5.2.3 Público

O Público é um jornal *online*, no qual são colocadas notícias sobre a actualidade.

Na avaliação do marcador de nomes próprios foi utilizado um corpus de 3 textos de 100 notícias cada do jornal, Público. Com este corpus foram realizados 8 testes com variação das pontuações atribuídas aos candidatos a nomes próprios e às interpretações representáveis por estruturas sintácticas.

#### Testes com corpus do Público

Os resultados obtidos nos 8 testes com o corpus do Público podem ser vistos na tabela 5.4.

Na Figura 5.17 pode ver um gráfico com as médias de precisão, cobertura e média-F para o corpus do CLEF nos 8 testes.

Na Figura 5.18 pode ver um gráfico com a média dos valores de heurística para o corpus do CLEF nos 8 testes.

Na Tabela 5.4 e na Figura 5.17 pode ver-se que os testes que obtiveram os melhores resultados, foram os testes 1, 2 e 8 com valores próximos dos 93%, 95% e 94%

Teste	Prec (%)	Cob (%)	Med-F (%)	Val-H
Teste1	0,9304	0,9547	0,9424	60,53
Teste2	0,9292	0,9547	0,9418	48,56
Teste3	0,8931	0,9238	0,9082	35,29
Teste4	0,6945	0,7750	0,7326	42,36
Teste5	0,8743	0,9547	0,9127	60,50
Teste6	0,9003	0,9259	0,9129	56,73
Teste7	0,9003	0,9259	0,9124	56,76
Teste8	0,9304	0,9547	0,9424	60,50

Tabela 5.4: Testes com corpus do Público

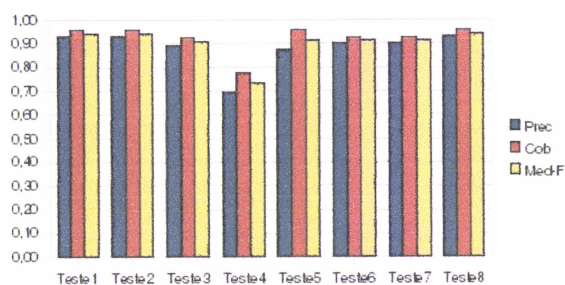


Figura 5.17: Testes com o corpus do Público

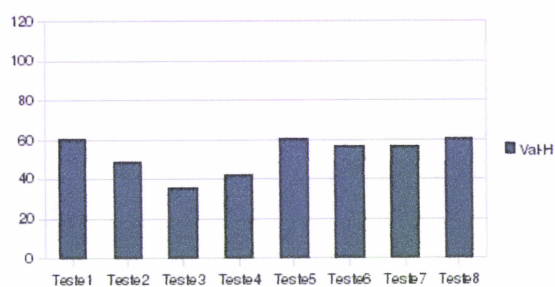


Figura 5.18: Testes com o corpus do Público

em termos de precisão, cobertura e medida-F. Em relação aos testes 3 (não acede à Wikipédia), 6 (não pontua as expressões entre as aspas) e 7 (não pontua nomes próprios marcados como datas), os valores desceram 3% em termos de precisão, cobertura e medida-F.

No teste 5 (não pontua números isolados), os resultados desceram 6% e 3%, em termos de precisão e medida-F.

No teste 4 (não considera o tamanho dos nomes próprios), os resultados caíram 24%, 18% e 21%, em termos de precisão, cobertura e medida-F.

Na Figura 5.18 pode ver-se que os testes que obtiveram os valores mais elevados de heurística foram os testes 1, 5 e 8 com valores próximos dos 61. Em relação aos testes 6 e 7, os valores de heurística baixaram cerca de 4 valores. Nos testes 2 (não pontua a análise sintáctica), 4 e 3, os valores de heurística desceram 12, 18 e 25 respectivamente.

### 5.2.4 HAREM

O HAREM é uma avaliação conjunta de sistemas de REM em português, organizada pela Linguateca. Este evento pretende avaliar o sucesso na identificação e classificação de nomes próprios na língua portuguesa. Mais detalhes sobre o HAREM podem ser encontrados na subsecção 2.4.5.

Na avaliação do marcador de nomes próprios foi utilizado um corpus de 2 textos de 100 frases cada do HAREM. Com este corpus foram feitos 8 testes com variação das pontuações atribuídas aos candidatos a nomes próprios e às interpretações de frases representáveis por estruturas sintácticas.

#### Teste com corpus do HAREM

Os resultados obtidos nos 8 testes com o corpus do HAREM podem ser vistos na tabela 5.5.

Teste	Prec (%)	Cob (%)	Med-F (%)	Val-H
Teste1	0,7659	0,8827	0,8202	33,53
Teste2	0,7559	0,8827	0,8144	21,63
Teste3	0,6684	0,7194	0,6929	20,03
Teste4	0,7126	0,8397	0,7710	25,62
Teste5	0,7630	0,8827	0,8185	34,05
Teste6	0,7582	0,8757	0,8127	32,11
Teste7	0,7573	0,8827	0,8107	33,82
Teste8	0,7659	0,8827	0,8202	32,33

Tabela 5.5: Testes com corpus do HAREM

Na Figura 5.19 pode ver-se um gráfico com as médias de precisão, cobertura e média-F para o corpus do CLEF nos 8 testes.

Na Figura 5.20 pode ver-se um gráfico com a média dos valores de heurística do corpus para o CLEF nos 8 testes.

Na Tabela 5.5 e na Figura 5.19 pode ver-se que os testes que obtiveram os melhores resultados, foram os testes 1, 2, 5, 6, 7 e 8 com valores próximos dos 76%, 88% e 82% em termos de precisão, cobertura e medida-F. No teste 4 (não considera o tamanho dos nomes próprios), os resultados desceram cerca de 5%, 4%, 5% em termos de precisão,

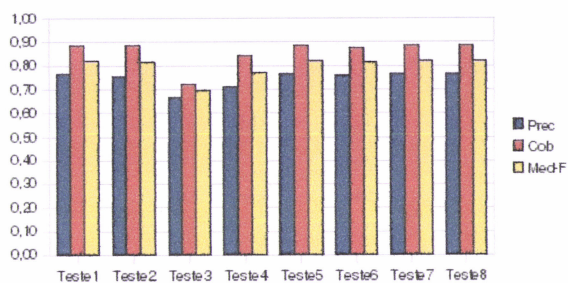


Figura 5.19: Testes com o corpus do HAREM

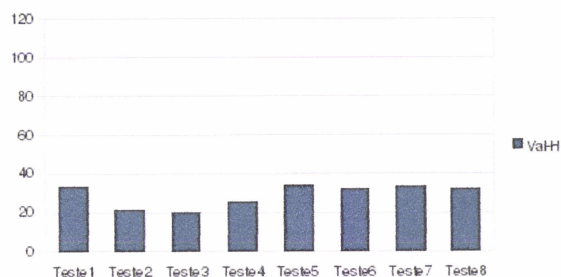


Figura 5.20: Testes com o corpus do HAREM

cobertura e medida-F. No teste 3 (não acede à Wikipédia), os resultados descenderam cerca de 10%, 16%, 13% em termos de precisão, cobertura e medida-F.

Na Figura 5.20 pode ver-se que os testes que obtiveram os valores mais elevados de heurística foram os testes 1, 5 e 7 com valores próximos dos 34. Em relação aos testes 6 (não pontua as expressões entre as aspas) e 8 (não pontua as datas marcadas como nomes próprios), os valores de heurística baixaram cerca de 1 valores. Nos testes 4, 2 e 3, os valores de heurística descenderam 8, 13 e 14 valores respectivamente.

## **Capítulo 6**

# **Conclusões e Trabalho Futuro**

Este capítulo encontra-se dividido em duas secções. Na secção 6.1 são dadas a conhecer algumas das conclusões retidas. Na secção 6.2 são feitas algumas sugestões para o melhoramento do analisador sintáctico/semântico e do marcados automático de nomes próprios.

## 6.1 Conclusões

Nesta dissertação implementou-se um analisador sintáctico/semântico e um marcador automático de nomes próprios para o processamento de frases interrogativas em sistemas de pergunta-resposta.

O analisador sintáctico/semântico pode utilizar um *parser* Top-down e um outro *parser* Bottom-up do tipo Chart Parser. Ainda, o analisador pode recorrer ao uso do marcador de nomes próprios com o intuito de reduzir o número de análises sintácticas e semânticas numa frase.

O analisador sintáctico/semântico sem recorrer ao marcador de nomes próprios e utilizando o Top-down mostrou-se mais eficiente do que com o Chart Parser na análise de 100 perguntas do CLEF (Texto 1 do CLEF). Na subsecção 4.2.1 pode ver que o analisador com o Top-down falhou análise em 3 frases e com o Chart Parser falhou análise em 5 frases. Em relação à percentagem de DRS's incorrectas, a analisador com o Top-down apresentou uma percentagem de 62,35% e com o Chart Parser de 62,42%.

Para avaliar os números de estruturas sintácticas e DRS's geradas com o analisador sintáctico/semântico sem recurso e com recurso ao marcador de nomes próprios foram utilizados 3 corpus (7 textos de 100 frases cada do CLEF, 3 textos de 100 frases cada do Público e 2 textos de 100 frases cada do HAREM) de diferentes fontes.

O analisador sintáctico/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando o *parser* Top-down obteve a maior quantidade em média de estruturas sintácticas e DRS's para o corpus do Público (28,02 ES's e 11,38 DRS's s/ marcador, e 25,02 ES's e 9,52 DRS's c/ marcador - Tabela 4.12). Enquanto, que com o Chart Parser obteve a maior quantidade em média de estruturas sintácticas para o corpus do CLEF (20,05 ES's s/ marcador e 13,31 c/ marcador) e maior quantidade em média de DRS's para o corpus do Público (11,33 DRS's s/ marcador e 7,66 DRS's c/ marcador - Tabela 4.12).

Visto, que o analisador sintáctico/semântico sem recurso e com recurso ao marcador de nomes próprios apresenta uma quantidade elevada em média de estruturas sintácticas e DRS's para o corpus do Público, é de supor que a maioria das frases deste corpus teve análise sintáctica e semântica. No entanto, pode ver-se pela Tabela 4.13 que a percentagem de frases analisadas pelo analisador sintáctico/semântico sem recurso e com recurso ao marcador de nomes próprios, utilizando os *parser's* Top-down e Chart Parser para o corpus do Público, foi baixa (apenas cerca de 11% das frases obtiveram análises sintácticas e semânticas). Então pode concluir-se, que ter em média um elevado número de estruturas sintáctica e DRS's num corpus não significa que a maioria das frases obtiveram análises sintáctica e semântica.

O uso do marcador de nomes próprios no analisador sintáctico/semântico conseguiu reduzir o número de estruturas sintácticas e DRS's em relação ao mesmo analisador sem utilizar o marcador de nomes próprios, e a percentagem de frases analisadas manteve-se em ambos os casos. No entanto, não se pode concluir se o número de estruturas sintácticas e DRS's correctas se se mantiveram em ambos os casos, porque não foram analisadas as estruturas sintácticas e as DRS's geradas pelo analisador com recurso ao marcador de nomes próprios por falta de tempo.

O analisador apresentou melhores resultados para o corpus do CLEF, relativamente ao corpus do Público e do HAREM. Este resultado deve-se ao analisador ter sido desenvolvido para a análise de frases interrogativas, e visto, o corpus do CLEF ter maior quantidade de frases interrogativas.

O marcador automático de nomes próprios pode recorrer à Wikipédia para pesquisar nomes próprios, além disso, pode utilizar a construção de estruturas sintácticas

em cada interpretação de uma frase para encontrar a interpretação com mais nomes próprios marcados correctamente. Para tal, foi utilizada uma função heurística que atribui pontuações aos nomes próprios marcados e às interpretações de frases que contêm estruturas sintácticas.

Na avaliação do marcador de nomes próprios foram utilizados 3 corpus (7 textos de 100 frases cada do CLEF, 3 textos de 100 frases cada do Público e 2 textos de 100 frases cada do HAREM) de diferentes fontes.

Com o marcador de nomes próprios foram feitos 8 testes com variações das pontuações atribuídas aos candidatos a nomes próprios marcados e às interpretações de frases que contêm estruturas sintácticas.

O Teste 1 foi um dos testes que obteve os melhores resultados. Este teste é visto como um teste de referência para os restantes 7 testes realizados (em cada um destes testes pretendeu-se medir o impacto de um parâmetro de afinação do marcador relativamente ao Teste 1).

Nos restantes testes pretendeu ver-se o impacto:

- Teste 2 - de utilizar a análise sintáctica na marcação de nomes próprios.
- Teste 3 - de utilizar a Wikipédia como fonte para pesquisar nomes próprios.
- Teste 4 - de não valorizar o tamanho dos nomes próprios nomes próprios.
- Teste 5 - de não valorizar negativamente números isolados marcados como nomes próprios.
- Teste 6 - de não valorizar expressões entre aspas marcados como nomes próprios.
- Teste 7 - de não valorizar datas marcadas como nomes próprios.
- Teste 8 - de não valorizar horas marcadas como nomes próprios.

O marcador de nomes próprios com o corpus do CLEF obteve resultados mais baixos nos testes 3 (os resultados baixaram cerca de 4 %, em termos de precisão cobertura e medida-F) e 4 (os resultados baixaram cerca de 8 %, em termos de precisão cobertura e medida-F). Pode-se concluir que não aceder à Wikipédia e não desvalorizar números isolados marcados como nomes próprios no corpus do CLEF baixa o desempenho do marcador.

Em relação aos valores de heurística, os resultados baixaram nos testes 2, 3 e 4. No teste 2, o valor de heurística baixou significativamente (cerca de 86 valores a menos que no teste 1). O teste 2 mostra que a maioria das frases teve análise sintáctica no corpus do CLEF, porque a diferença de heurística entre o teste 2 e 1 é elevada.

O marcador de nomes próprios com o corpus do Público obteve resultados mais baixos nos testes 3, 4, 5, 6, 7. Nos testes 4 (os resultados baixaram cerca de 24%, 18% e 21%, em termos de precisão, cobertura e mediada-F) e 5 (os resultados baixaram cerca de 6% e 3%, em termos de precisão e mediada-F), os resultados baixaram significativamente em relação ao teste 1. Pode concluir-se que não considerar o tamanho dos nomes próprios e não valorizar negativamente números isolados marcados como nomes próprios no corpus do Público baixa o desempenho do marcador.

Em relação aos valores de heurística, os resultados baixaram nos teste 2, 3, 4, 6 e 7. Nos testes 2 e 4, os valores de heurística baixaram cerca 12 e 25 valores. O teste 2 mostra que a maioria das frases não teve análise sintáctica, porque a diferença de heurística entre o teste 2 e 1 é baixa (cerca de 12 valores de heurística). O teste 3 mostra que a maioria dos nomes próprios são compostos de várias palavras, porque



não valorizando o tamanho dos nomes próprios, o valor de heurística caiu cerca de 25 valores.

O marcador de nomes próprios com o corpus do HAREM obteve resultados mais baixos nos testes 3 (os resultados baixaram cerca de 10%, 16% e 13%, em termos de precisão, cobertura e mediada-F) e 4 (os resultados baixaram cerca de 5%, 4% e 5%, em termos de precisão, cobertura e mediada-F). Pode concluir-se que não aceder à Wikipédia e não considerar o tamanho dos nomes próprios no corpus do HAREM baixa o desempenho do marcador.

Em relação aos valores de heurística, os resultados baixaram nos teste 2, 3 e 4. Nos testes 2, 3 e 4, os valores de heurística baixaram cerca de 13, 14 e 8 valores. O teste 2 mostra que a maioria das frases não teve análise sintáctica, porque a diferença de heurística em o teste 2 e 1 é baixa (cerca de 13 valores de heurística).

Relativamente aos 3 corpus, os valores mais elevados nas métricas de precisão, cobertura e medida-F foram encontrados no corpus do CLEF (cerca de 98% de precisão, cobertura e medida-F), em segundo lugar surgiu o corpus do Público (cerca de 93% de precisão, 95% de cobertura e 94% de medida-F) e finalmente o corpus do HAREM (cerca de 77% de precisão, 88% de cobertura e 82% de medida-F).

Quanto aos valores de heurística dos 3 corpus, o corpus do CLEF apresentou os valores mais elevados (cerca de 103 valores de heurística), em segundo lugar surgiu o corpus do Público (cerca de 61 valores de heurística) e finalmente o corpus do HAREM (cerca de 34 de heurística).

Apesar de um corpus obter em média mais de metade dos valores de heurística de um outro corpus não significa que o mesmo tenha tido mais de metade das frases com análise sintáctica. Exemplo disso, é o corpus do Público que obteve em média cerca de 61 valores de heurística (em média mais de metade do valor obtido com o corpus do CLEF) e o corpus do CLEF obteve em média cerca de 103 valores de heurística. No entanto, o corpus do Público obteve análise em cerca de 11% das frases e corpus do CLEF em cerca de 95% (mais 84% em relação ao corpus do Público). Os 61 valores de heurística obtidos com o corpus do Público têm a justificação de este corpus conter mais quantidade de nomes próprios e/ou nomes próprios com maior tamanho.

## 6.2 Trabalho Futuro

A maioria das frases no analisador sintáctico/semântico falhou foi devido à falta de coordenação de frases. Esta é uma regra importante a implementar nos seus *parser's* (Top-down e Chart Parser).

Os números de estruturas sintácticas e DRS's obtidas com o analisador sintáctico/semântico foram elevados. A utilização do marcador de nomes próprios no analisador conseguiu reduzir o número de estruturas sintácticas e DRS's, no entanto não se teve tempo suficiente para verificar se o número de estruturas sintácticas e DRS's correctas se manteve. Este é um estudo importante a fazer futuramente.

Como as DRS's são as estruturas que mais interessam na representação de uma frase, pode desenvolver-se uma função heurística para escolher de entre as DRS's, a que melhor a representa. Escolhendo-se as DRS's que têm menor número de referentes, de entre estas as que têm mais referentes definidos e menor número de argumentos nos predicados que correspondem à reescrita dos verbos.

Como dito anteriormente, o analisador sintáctico/semântico pode ser melhorado com adição de mais regras gramaticais aos seus *parser's*. Ainda em relação ao analisador sintáctico/semântico, será interessante transportar o mesmo para o Inglês.

O marcador de nomes próprios tal como o analisador sintático/semântico, pode ser melhorado com a adição de regras gramaticais ao seu analisador sintático.

No marcador de nomes próprios também podem ser exploradas outras informações presentes nos textos. Exemplo disso, são as palavras próximas do nomes próprios conhecidas por metapalavras. Estas palavras dão alguma informação sobre os nomes próprios. Além das metapalavras, também se podem utilizar as abreviaturas que se encontram próximas dos nomes próprios e dão algumas informação acerca dos nomes próprios.

Quanto ao marcador de nomes próprios será interessante transportar o mesmo para o Inglês e incluir a classificação de entidades mencionadas.



# Bibliografia

- [AA08] Marcelo Adriano Amancio and Sandra Maria Aluísio. Explicitação de entidades mencionadas visando o aumento da inteligibilidade de textos em português. Technical report, Universidade de São Paulo, Agosto de 2008.
- [Bic00] Eckhard Bick. Ensino de sintaxe portuguesa na internet. In *Anais do II Encontro Internacional de Português-Língua Estrangeira*. Universidade de São Paulo, 2000.
- [Car] Nuno Cardoso. Rembrandt - reconhecimento de entidades mencionadas baseado em relações e análise detalhada do texto. In Cristina Mota and Diana Santos, editor, *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*, volume 3. Linguatca.
- [Cho56] Noam Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, pages 113–124, 1956.
- [FG92] C. FUCHS and P. LE GOFFIC. *Les linguistiques contemporaines*. Paris: Hachette: Repères théoriques, 1992.
- [Gaz82] G. Gazdar. *Phrase Structure Grammar: The Nature of Syntactic Representation*. Reidel, Dordrecht, 1982.
- [GGS85] G. PULLUM G. GAZDAR, E. KLEIN and I. SAG. *Generalized Phrase Structure Grammar: The Nature of Syntactic Representation*. Basil Blackwell, 1985.
- [JM92] Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Prentice Hall Series in Artificial Intelligence, 1992.
- [KAY79] Martin KAY. Functional grammar. pages 142–158. Fifth Annual Meeting of the Berkeley Linguistic Society, 1979.
- [KR93] Hans Kamp and Uwe Reyle. *From Discourse to Logic: An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: D. Reidel, 1993.
- [Mar04] Daniel Martineschen. Analisador gramatical por deslocamento e redução para gramáticas categoriais. Master's thesis, Universidade Federal do Paraná, Fevereiro de 2004.
- [Nad05] David Nadeau. Balie - baseline information extraction. Technical report, School of Information Technology and Engineering University of Ottawa, Ottawa, Canada, January 2005.

- [OG08] Hugo Gonçalo Oliveira and Paulo Gomes. Utilização do (analisador sintáctico) pen para extrecção das definições de um dicionário. Technical report, Linguateca - pólo de Coimbra, Nobembro de 2008.
- [PS94] C. Pollard and I. A. Sag. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford: CSLI Publications, 1994.
- [QRPV06] Paulo Quaresma, Irene Rodrigues, C. A. Prolo, and R. Viera. Um sistema de pergunta-resposta para uma base de documentos. volume 41, pages 43–63. *Letra de Hoje - Revista da Pontifícia Universidade Católica do Rio Grande do Sul*, Junho de 2006.
- [SC07] Diana Santos and Nuno Cardoso, editors. *Reconhecimento de entidades mencionadas em português*. Linguateca, 2007.
- [Shi84] Stuart M. Shieber. The design of a computer language for linguistic information. In *ACL-22: Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics*, pages 362–366, Morristown, NJ, USA, 1984. Association for Computational Linguistics.
- [VdL01] Renata Vieira and Vera Lúcia Strube de Lima. Linguística computacional: princípios e aplicações. In Ana Teresa Martins and DÍbio leandro Borges, editors, *SBC - Jornadas de Atualização em Inteligência Artificial (JAIA)*, volume 3, pages 47–86, Fortaleza - CE, 2001.
- [Woo70] W. A. Woods. Transition network grammars for natural language analysis. *Commun. ACM*, 13(10):591–606, 1970.

## **Capítulo 7**

### **Anexos**

## CLEF

### Textos

Em que ano foi atribuído o prémio Nobel a Thomas Mann? Quem é o administrador-delegado da Fiat? Em que cidade se encontra a prisão de San Vittore? Qual a unidade de medida da frequência? Qual o aumento da população mundial por ano? Qual o nome próprio do juiz Borsellino? Quantos guarda-costas foram mortos no atentado ao juiz Falcone? Quem é Shimon Peres? Que cargo ocupa Ariel Sharon? Quando foi atribuído o Prémio Nobel da Paz a Yasser Arafat? Qual é a cidade santa dos judeus? A quanto se eleva o número de refugiados palestinianos no Líbano? Em que ano rebentou a Intifada? Como se chamavam os pilotos suicidas japoneses? Que nação é a principal fornecedora de armas a países do Terceiro Mundo? Em que ano foi lançada a bomba atómica sobre Hiroshima? A que partido pertencia Hitler? Que língua se fala na Alemanha? Quantos milhões de judeus foram mortos durante a Segunda Guerra Mundial? Onde era o campo de concentração de Auschwitz? Quem é o maior operador de telemóveis do mundo? Em que ano terminou a Segunda Guerra Mundial? Qual é o primeiro nome de Hitler? Que moeda se usa na Alemanha? Qual é a capital da Rússia? Onde ocorreu uma catástrofe ecológica? De que nacionalidade eram os dois petroleiros que causaram a catástrofe ecológica junto a Trinidad e Tobago em 1979? Onde reside Arafat? Quando foi introduzido o alfabeto cirílico? Quantos desempregados há na Europa? Como se chama o presidente da Greenpeace Internacional? Qual é a capital do Japão? Que seres vivos são capazes de absorver dióxido de carbono? Por que é coberta a Antártida? Por que é causado o buraco do ozono? Que pode causar o cancro do pulmão? Quantos anos tem o Papa? Quantas religiões monoteístas há no mundo? Quantos são os católicos no mundo? Quantos judeus existem no mundo? Em que cidade se encontra a Basílica de São Pedro? Em quantos estados americanos existe a pena de morte? Quantas execuções houve nos Estados Unidos em 1993? Em que ano foi abolida a apartheid na África do Sul? A quanto se eleva a população dos EUA? Quantos anos de prisão cumpriu Nelson Mandela? Como se chama o partido separatista curdo? Quantos detidos há no Corredor da Morte na Califórnia? O que é a UNICEF? Em que idade termina a escolaridade obrigatória em Itália? Quando se demitiu Nixon? Quem impôs o embargo contra o Iraque? Mencione um cetáceo. De quem tiraram o nome os zapatistas? Quantos genes humanos há? Quem são os amish? Quem é o imperador do Japão? Quantas pessoas falam gaélico na Escócia? Quando obteve a Letónia a independência? Onde é a mesquita de al-Aqsa? Que animal é capaz de emitir luz? Como se forma o ozono? Onde é o Vale dos Reis? Como se chama a companhia aérea oficial alemã? Quando é que o príncipe Carlos e Diana se casaram? Onde se situa o CERN? Quanto custou o Túnel da Mancha? Onde é o Centro Médico Cedars-Sinai? Em que ano ocorreu o furacão Andrew? O que é a UEFA? Cite uma unidade de radioactividade. Em que país europeu fica Galway? Em que ano foram os Jogos Olímpicos de Barcelona? De que são feitos os implantes mamários? Quem é Yves Saint-Laurent? A que galáxia pertence a Terra? Em que ano foi construída a Estátua da Liberdade? De que são feitos os cabos de fibra óptica? Qual a população da Chechénia? Que forma de governo tem a França? Quantos continentes há? Qual o nome da mulher de Kurt Cobain? Indique um jogo de tabuleiro. Em que ano teve lugar a guerra do Yom Kippur? Quantas pessoas morrem devido a minas anti-pessoais anualmente? Qual o vulcão activo mais alto da Europa? Que rio banha Dublin? Quantos bávaros são católicos? O que é a Amnistia Internacional? Quem é Silvio Berlusconi? O que embrulhou o artista Christo? O que faz um luthier? Cite uma razão para suicídios

de adolescentes. Quando foi lançada a sonda espacial Ulisses? Quem são os Simpsons? O que é a maçonaria? Onde fica a Praça Vermelha? Que celebridade foi fotografada por Man Ray? Em que ano se tornou o Papa João Paulo II o Sumo Pontífice? Quem libertou a cidade de Saint-Mère-Église no Dia D?

## Texto2

Qual a área da Selva Lacandona? Quanto tempo demorou Christo a embulhar o Reichstag alemão em Berlim? Quantos astronautas levava a bordo o vaivém Atlantis? Quantos mortos por asfixia houve no metro de Baku? Quantos quilómetros da fronteira mútua foram disputados pela Argentina e pelo Chile? Como se chama o interior de um barco? Que líquido é inodoro e insípido? Onde fica o Reichstag? Que tufão nas Filipinas causou mais de 1000 mortos ou desaparecidos? Que inalou Joseph Bryan Thomson que lhe provocou a morte? O que é que o Barão Marcel Bich inventou? Que complexo desportivo foi inaugurado em Buenos Aires em Dezembro de 1993? Que vaivém americano levou pela primeira vez um astronauta russo a bordo? Em que país fica Essen? Onde se entregam os Óscares? Onde se celebraram os JO de 1992? Sobre que continente está o buraco do ozono? A que altitude está a camada de ozono? Onde fica o arquipélago de Svalbard? Onde fica o Rio de Janeiro? Onde se realizou a Conferência Mundial da Mulher? Onde fica Tarbes? Em que cidade explodiu uma carta-bomba? Que realizador de cinema italiano foi premiado com um Óscar? Quem é o director-geral do Fundo Monetário Internacional? Quem é Arabella Kiesbauer? Quem é Christo? Quem é Bill Clinton? Quem é Leo Reting? O que é o IFOP? Quem é João Paulo II? O que é o Parlamento Europeu? Quem era presidente dos Estados Unidos em 1994? Em que país foi prorrogada a operação das Nações Unidas até 31 de Março de 1995? Quem era o presidente da Coreia do Norte antes de 1994? Quem era o presidente da Coreia do Norte depois de 1994? Que empresas energéticas se fundiram durante 1995? Em que ano se celebrou a Conferência Internacional da Mulher antes de 1995? Quem ganhou o Prémio Nobel da Literatura em 1994? Que equipa venceu a liga de futebol espanhola na época 1994-1995? Que país europeu comprou dois aviões CASA C-212-300 Aviocar em 1994? Que valor se programou no âmbito do IFOP no período 1994-1995 para a renovação da frota espanhola? Qual a melhor maneira de combater as alergias? Como se determina a proporção de deutério característica de cada vinho? Como se garante a cobrança de sanções? Como se pretende levar a cabo o planeamento familiar no Peru? Como começa o cancro, de acordo com o Método Hamer? Como que é que se faz a angiografia com indocianina? Como se eliminam as reacções de rejeição do sistema imunológico? Como se aplica o método Hamer? Em que cidade fica o Museu Van Gogh? Como se chama a filha do líder chinês Deng Xiaoping? Diga o nome de um vulcão japonês. O que é a GM? O que é o FSK? Em que país fica Vukovar? Em que cidade americana se encontra o Museu Warhol? Quem é Andy Warhol? Quem é Keith Richards? Quem descobriu o vírus da sida? O que é a Yves Saint-Laurent? Quem é a ministra do Ambiente alemã? Mencione um bonecreiro. Quem é o presidente da UEFA? Quem é o líder dos republicanos no Senado dos Estados Unidos? Quem foi o fundador da cadeia noticiosa CNN? Em que país fica East London? O que é a MTV? Quando é que houve um golpe de estado em Chipre? Quando nasceu Alberto Giacometti? Como se chama o líder dos cipriotas turcos? Como se chama o filho do escritor e Prémio Nobel Thomas Mann? Quem é o presidente do Iémene do Sul? Quem é o líder do KwaZulu? De que clube de futebol francês foi Bernard Tapie presidente? Qual o nome do presidente da Reserva Federal americana? Quem foi um dos fundadores da Greenpeace? Em que ano foi fundada a



NATO? Onde nasceu Nick Leeson? Quem é João Havelange? Qual era a moeda da Argentina? Qual é a moeda do México? O que significa a abreviatura OUA? Onde fica Hyde Park? Quem assassinou Yitzhak Rabin? Qual o nome próprio do financeiro Soros? De que partido político é dirigente Ian Paisley? Que idade tem o primeiro-ministro russo Viktor Tchernomirdine? Quem é o presidente da Autoridade Nacional Palestiniana? Como se defendeu Vladimir Zhirinovsky contra manifestantes em Estrasburgo? Quem é Ruben Acosta? Que significa a abreviatura AWACS nos aviões AWACS? Por que equipa corre Michael Schumacher? Qual foi o arquitecto do novo aeroporto da cidade japonesa de Osaka? Que organização luta pela libertação do Tamil Eelam? Que gás contribui para o efeito de estufa? Qual a companhia de aviação nacional belga? Qual o maior planeta do sistema solar? Que significa MBA? Que escritor foi o último marido de Marilyn Monroe?

### Texto3

Qual o título do filme de Stephen Frears com Glenn Close, John Malkovich e Michelle Pfeiffer? Que país faz fronteira a norte com o Canadá? Quem foi o primeiro presidente dos Estados Unidos? Qual a profissão de James Bond? Quem interpretava o papel de James Bond nos primeiros filmes da série 007? Quem interpreta o papel de James Bond no último filme da série 007? Onde se realiza o Motorshow? A quem se pede uma hipoteca? Que escola frequentará Guilherme, o filho mais velho do príncipe Carlos? De que grupo musical fazia parte Michael Jackson quando era criança? Como morreu Jimi Hendrix? Quem era o pai de Lisa Marie Presley? Com quem se casou Michael Jackson? Em que género musical se distingue Michael Jackson? Como se chama a companhia discográfica de Michael Jackson? Quem foi o inventor da televisão? Quantos intervalos comerciais num filme são permitidos actualmente pela CEE? O que é a Mossad? O que é o INCB? A que pena foi condenado Pietro Pacciani pelos crimes do monstro de Florença? Quantos crimes são atribuídos ao Monstro de Florença? Qual o diário italiano mais lido? Quantas pessoas sobrem de obesidade nos Estados Unidos? Diga um ingrediente base da cozinha japonesa. Em que ano foi fundada a Greenpeace? Quando foi aprovada a Convenção dos Direitos da Criança? Nomeie uma pessoa acusada de pedofilia. Com que idade são convocados os jovens para a inspecção militar? Quantos analfabetos há no mundo? Quem é o ministro da Saúde francês? Como está o Papa? Que empresa foi acusada de usar trabalho infantil? Que presidente americano não se apresentou ao serviço militar? Quem descobriu a radioactividade? Onde foi fundada a Greenpeace? Há quantos anos funciona o Greenpeace? Qual era o objectivo da primeira acção realizada pela Greenpeace? Quantas guerras estão a decorrer no mundo? Quantas minas anti-pessoais há no mundo? Mencione um factor de risco para as doenças cardiovasculares. Que categoria profissional está mais em risco de cancro do pulmão? Que parte do organismo é atacada pelo vírus Ebola? Quando é que se registou pela primeira vez o vírus Ebola? Mencione um sintoma do vírus Ebola. O que é o Comité Internacional da Cruz Vermelha? Como é transmitido o vírus Ebola? Indique um pesticida. Em que dia calha o Ano Novo chinês? Quem é Jean-Bertrand Aristide? Onde ocorre o El Niño? Em que ano ocorreu a secessão da Bósnia da Jugoslávia? Diga o nome de um jornal francês. Quando ocorreu o ataque bombista ao World Trade Center? Quando é que John Lennon foi assassinado? O que fazem os anti-oxidantes? De que morreu River Phoenix? Em que ano foi visível o cometa Halley? Em que dia calha o solstício de verão? De que é extraído o ácido salicílico? Onde é a Torre Inclinada? Em que unidade é medida a velocidade de processamento de um chip? Quem é Paul Simon? O que são os carabinieri? Onde é o aeroporto de Heathrow? O que

é a Yakuza? Como pode ser tratada uma alergia? De que cor é o monóxido de carbono? Quem é o Presidente da Câmara de Nova Iorque? O que é a FDA? Como é que as plantas convertem a luz em alimento? Onde fica o Museu do Hermitage? Quantos cardeais elegem o Papa? Qual é o primeiro nome de Armani? Como funciona o MRI? Qual é a nacionalidade de Aristóteles Onassis? Quem é Alan Turing? Qual a idade de Beck Hansen? Quem é Arlequim? Em que guerra combateram as Brigadas Internacionais? Quem foi o último czar da Rússia? Mencione o nome de um robot. Como se pode poupar energia? Que animal arrulha? Quantos pandas há em estado selvagem na China? Como funciona a acupunctura? Para que é usado o anidrido acético? De que país é originário o tango? Qual a nacionalidade do Mitubishi Bank? Qual o acrónimo da Amnistia Internacional? Em que país se professa o budismo? Que país exporta arroz? Qual é o país que é o maior produtor de diamantes? Que cargo detinha Silvio Berlusconi? Que país regressou à UNESCO após 38 anos de ausência? Que prémio partilharam Yasser Arafat, Shimon Peres e Yitzhak Rabin? Que percentagem do comércio mundial de diamantes se concentrou em Antuérpia em 1994? Quantas pessoas morreram nas inundações na Holanda e na Alemanha em 1995? Quantos budistas há em Espanha? Qual a distância que é percorrida no rali Granada-Dacar? Quem é o director-geral do Fundo Monetário Internacional?

#### Texto4

Quem recebeu o Prémio Príncipe das Astúrias de Investigação Científica e Tecnológica pelo desenvolvimento da primeira vacina sintética contra a malária? Quem presidiu à reabertura do Museu Sefardita de Toledo? Quando assumiu o cargo o primeiro presidente de Cuba? Em que ano é que o Iraque invadiu o Kuwait? Em que dia foi executado José Rizal? Em que dia se celebra a descoberta da América? Que dia corresponde ao início do século XXI? Quando foi a Guerra do Chaco? Em que dia de Junho de 1995 é que explodiu uma carta armadilhada? Quando é que os manchu invadiram a China? Quando entrou em vigor o tratado da NAFTA? Quando é que se fixou na Colômbia José Celestino Mutis? Refira uma companhia petrolífera. Indique uma companhia de fast-food. De que canal é apresentadora Arabella Kiesbauer? Com que organização palestina anunciou o Vaticano o estabelecimento de relações oficiais? Que exército ocupou o Haiti? Que instituição fundou Andrés Townsend Ezcurra? Que partido venceu as primeiras eleições multirraciais na África do Sul? Que comissão internacional condenou pela primeira vez formalmente o anti-semitismo e a xenofobia? Que igreja aprovou novas regras para a ordenação de mulheres? Que instituição constituíram o Brasil, Portugal, Angola, Moçambique, São Tomé e Príncipe, a Guiné-Bissau e Cabo Verde em Brasília? O que é a NAFTA? O que é o Mercosur? Quem foi Rosa Chacel? Como actua a hormona do crescimento? Quem é Alain Juppé? Quem é Shoko Asahara? Quem é Franz Schönhuber? Onde fica Indiana? Quem é Bruce Morrison? O que são as FARC? Onde é Caltanissetta? Qual a abreviatura do Exército Popular de Libertação do Sudão? Qual é a capital da Venezuela? O que é o ACNUR? Quem é o ministro-presidente bávaro? Esmirna fica em que país? Quem é o ministro da Economia alemão? De que companhia é Christian Blanc presidente? Refira uma cidade finlandesa. Qual a localização de Tipaza? Quantos habitantes tem Bombaim? Quem é o director da AIEA? O que é a AIEA? Onde é a sede da AIEA? Quantas pessoas vivem no Brasil? Em que cidade italiana fica a Fonte de Trevi? Quem é o fundador da Motown? Qual o primeiro nome do líder turco Atatürk? Quem é Andrew Morton? Quantos habitantes conta a África do Sul? O que é a NASA? Quem planeou o Palácio dos Desportos São Jorge em Barcelona? Como se chama a rainha da Dina-

marca? Como se chama a mulher do presidente peruano Alberto Fujimori? Em que país fica Silver Star? Como se chama o primeiro-ministro do Ruanda? Quem é o rei da Bulgária? Que líder do ANC se tornou presidente da África do Sul? Quem é o presidente da Câmara dos Representantes americana? Indique uma província de Espanha. Em que ano foi assassinado o presidente chileno Salvador Allende? Quem é Marvin Minsky? Em que ano nasceu Nelson Mandela? Como se intitula a autobiografia de Nelson Mandela? Como se chama a viúva do falecido presidente de Moçambique, Samora Machel? Quando foi fundada a companhia Netscape Communications? Quem é Juan António Samaranch? Quando é que Pinochet chegou ao poder no Chile? Onde é o CERN? Quem é o líder do partido Likud? O que é a Goldman Sachs? Que país tem uma província chamada Antioquia? Qual a maior companhia de software do mundo? O que é a MSNBC? Quem é Andrew Lack? Quem produz o Invirase? O que é a IUPAC? Onde está preso Hugo Lacour? Qual o valor do Prémio Artístico de Darmstadt? Quando nasceu Marie-Jo Lafontaine? Onde nasceu Marie-Jo Lafontaine? Onde tem lugar a documenta? O que é a ALDI? Como se chama o presidente da associação de comércio do Hesse? Quantas empresas artesanais existem no Hesse? Quantos assinantes tem a MSN? Que produto da Sun vai a Microsoft licenciar? Quando se tornou Butzbach uma cidade de guarnição? Quem é o director dos arquivos municipais de Butzbach? Quando é que o cardeal Raymund Peraudi visitou Butzbach? Onde é a sede do Instituto Monetário Europeu? Quem é Marvin Minsky? O que eram os FTP? Quanto pagou a IBM pelo Lotus? Que empresa tem a sua sede em Armonk? Quando foi preso Hugo Lacour? Que chamou a Saab de volta à fábrica? A quem pertence a Saab Automobiles?

### Texto5

O que é a UNICE? Quem é o presidente da UNICE? Quem é o secretário-geral da secção alemã da Amnistia Internacional? Quem foi forçado a demitir-se de governador da Caríntia em 1991? Quando foi fundada a Dreamworks SKG? Quem fundou a Dreamworks SKG? Que companhia investirá na Dreamworks SKG? Quem é Christoph Ransmayr? Quem recebeu o prémio Kafka da Literatura em 1995? Quem é Jorma Ollila? Quem é o presidente da Nokia? Qual é a maior empresa industrial da Finlândia? Qual o lucro do grupo electrónico e de telecomunicações finlandês Nokia em 1994? Quando é que a IBM lançou o primeiro PC? Quantos empregados tinha o grupo HP no final de 1994? Que proporção do seu volume de negócios faz a HP na Europa? Onde é o castelo Esterhazy? O que é o CERN? Onde é a sede do CERN? Quantos estados-membros tem o CERN? Quem é Kevin Mitnick? Onde é a sede da Lindt & Sprüngli? Quando foi criado o CERN? Quem é o director do CERN? Quantas pessoas trabalham no CERN? Qual é a capital estadual do Sarre? Quantos habitantes tem a Eslovénia? Quem é Peter Tulviste? Quando entraram os tanques do exército jugoslavo na Eslovénia? Que firma vendeu a Studer Revox AG? Por quem foi comprada a Studer Revox AG? Quando foi fundada a Studer Revox? Onde é a sede da Harman International Industries, Inc? Como se chama a moeda da Geórgia? Quem é Nodar Dschawachischwili? Em que área aumentou a HP o seu volume de negócios em 1994? Quem é Thomas Klante? Que substâncias serão tratadas na instalação de tratamento definitivo Konrad? O que é Mayr-Melnhof? Quando será Weimar capital cultural da Europa? Quem é Götz Friedrich? O que é o BMFT? Como se chama a norma europeia para comunicações móveis? O que é a Ericsson? O que é o ETSI? Quem produz o carro compacto Smart? Quando aparecerá no mercado o carro compacto Smart? Que empresa será fundada conjuntamente pela Mercedes e pela SMH?

Que produz a empresa Victorinox? Que produz a MCC? Onde é que o Ubilab está representado com o seu próprio endereço? Qual o monte mais alto do mundo? Como se chama a montanha mais alta do Japão? Qual o símbolo mais importante de Paris? Que danificou seriamente o Coliseu em 851? Que foi proibido no Coliseu em 438? Quanto custará o Internet Explorer da Microsoft? De que morreram 232 pessoas no Rajasthan? Que epidemia atingiu a Albânia? Que foi montado em Seibertenrod em Vogelsberg? Quem foi nomeado presidente do grupo das Máquinas Fotográficas Leica? Que produz o grupo tecnológico Leica? Que vende a Oracle? Como ocorreram 22 feridos ligeiros no Empire State Building? Quando será inaugurado o museu Guggenheim em Bilbao? Quem é o arquitecto do museu Guggenheim em Bilbao? O que está Frank Gerry a construir em Bilbao? Que fabricante de móveis gere um museu em Weil am Rhein? Que é exibido no Museu do Design Vitra? Quem é Alberto Fujimori? O que é a OMC? Como se chama o ministro do Interior senegalês? Onde fica Halifax? Quem é o ministro das finanças russo? Quando ocorreu o atentado na estação de metro de Saint-Michel em Paris? Quem é Umberto Bossi? Onde fica o La Scala? Quantas vítimas provocou a vaga de calor na Índia? O que é a Ordem do Templo Solar? Quando morreu Lenine? Quando ocorreu a revolução islâmica no Irão? Como se chamava o ferry que naufragou na Suécia em 1994? Quem cometeu o atentado no metro de Tóquio? Onde teve lugar a Conferência Mundial do Clima de 1995? Onde fica a sede da UNESCO? Quando aderiu a Áustria a Schengen? Como se chama o partido de Silvio Berlusconi? Quantos católicos há em África? Quem é Philippe Biberson? O que é a Eurotunnel? Quando foi o ataque a Pearl Harbor? Como se chama a rainha dos Países-Baixos? Quem é Montxo Armendariz? Quem conquistou a Palma de Ouro em Cannes em 1995? Quem é Paul Feyerabend? Quantos habitantes tem o Iraque? Quantos anos de residência são necessários para obter a nacionalidade suíça? Quem é Butros Butros-Ghali? O que é o IRA? Onde se realiza o Concurso Chopin?

## Texto6

O que é o GIA? Como se chama o presidente do Burundi? Quando teve lugar a primeira volta das eleições legislativas na Hungria em 1994? Qual o cargo de Redha Malek em 1994? Em que dia foi apresentado o plano de reestruturação da Euro-Disney? Onde teve lugar a nona edição da Videoformes? Quem foi eleito presidente do Conselho Geral da Guiana? De que país é Iegor Gaidar primeiro ministro? Qual o nome do primeiro-ministro ucraniano nomeado em Junho de 1994? Quando foram os cannabis despenalizados na Alemanha? De que grupo político é Charles Millon presidente? Quem é o novo presidente do Ruanda? Onde teve lugar a Conferência Internacional sobre a População e Desenvolvimento? Qual a capital do Togo? Que quantia exige o FC Sevilha de Diego Maradona? Como se chama o ministro das Finanças polaco? Onde fica Kishinev? Quantos habitantes tem os Emirados Árabes Unidos? Quem é o presidente da Comissão Europeia em 1995? Como morreu Juvénal Habyarimana? Que significa a sigla EAU? Quem venceu as eleições legislativas húngaras? Que torneio Andrei Medveded venceu? Quem foi derrotado por Andrei Medveded na final do Torneio de Monte-Carlo? Qual a nacionalidade do tenista Sergi Bruguera? Quem venceu o slalom super-gigante em Lillehammer? Qual a data de nascimento de Claude Lévi-Strauss? Onde nasceu Aleksandr Vassilievich Korjakov? Como se chama o único diário independente da Jugoslávia? Qual é a capital da província da Cabília? De que país é Cabul a capital? Qual a esperança de vida em França em 1991? Qual a superfície da República da Chechénia? Quantos habitantes tem a República da Chechénia? Onde decorreu a sétima cimeira da Organização da Conferência Islâmica? De que organização Pierre-

Paul Schweitzer era director-geral? Quem venceu o torneio das Cinco Nações em 1994? Quando foi a inauguração oficial do Túnel da Mancha? Que personalidade inaugurou com François Mitterrand o Túnel sob a Mancha? Onde fica o Lesoto? Onde decorreu o Salão Internacional da Aeronáutica? Quantos habitantes tem o Lesoto? Qual o nome do partido político de Ntsu Mokhele, primeiro-ministro do Lesoto? Quantos filmes foram apresentados no Festival de Cinema Mundial de Montreal? Que filme recebeu o Grande Prémio especial do júri no Festival de Cinema Mundial de Montreal? Que actor recebeu o prémio de interpretação masculina no Festival de Cinema de Montreal? Qual o nome do primeiro-ministro do Ruanda? De que país é a escritora Taslima Nasreen? Quem foi o último rei da Roménia? Quando foi morto Babacar Seye? Qual o cargo de Albert Reynolds na Irlanda? Quantos postos de trabalho prevê a companhia Air France suprimir em 1995? Onde aterrou o Airbus da Air France após a tomada de reféns em Argel? Quantos empregados tem o grupo Warburg? Quando se casaram Federico Fellini e Giulietta Masina? Qual a profissão de Danny Barker? Qual a taxa de desemprego nos Estados Unidos no final de 1994? Qual a companhia cinematográfica mais antiga do mundo? De que escuderia é director Flavio Briatore? Quando tiveram lugar as eleições europeias de 1994? Qual a língua oficial da República do Iémen? Com que idade morreu Fernando Rey? Como morreu Jack Unterweger? Onde fica Luxor? Desde quando está em vigor o embargo ao Iraque? Qual a alcunha de Ho Chi Minh? Onde fica a Esfinge de Gizé? Onde é Izhevsk? Onde fica o Estádio José Alvalade? Onde vive José Saramago? Onde nasceu Nelson Mandela? Qual o maior satélite de Júpiter? Onde fica Turku? Qual a antiga capital da Polónia? Em que distrito fica Paredes de Coura? Onde fica Sosnovy Bor? Em que ilha fica Ponta Delgada? Onde é que nasceu Álvaro Cunhal? Onde é o hospital Júlio de Matos? Qual é o país mais pequeno da União Europeia? Onde fica Gabrovo? Qual o estado mais setentrional dos EUA? Qual é a capital da Bielorrússia? Onde desagua o rio Cubango? Em que cidade o Mosela encontra o Reno? Em que estado do Brasil fica Campo Grande? Onde se situa Tianjin? Onde é a Ilha do Diabo? Quem inventou o saxofone? Quem é o recordista mundial do salto à vara? Quem é o secretário-geral do PCP? De quem é filha Martine Aubry? Quem é o Presidente da Câmara de Lisboa? Quem é o Presidente da Câmara de Lamego? Quem é o embaixador de Portugal em França? Com quem casou Whoppi Goldberg? Quem foi o primeiro presidente dos Estados Unidos? Quem é o ministro-presidente da Renânia-Palatinado? Quem foi o último governador de Timor Leste? Quem era o marido de Vieira da Silva?

### Texto7

Quem é o capitão do FC Porto? Quem é o imã da mesquita de Lisboa? Quem é a ministra sueca do ambiente? Como se chama a rainha da Dinamarca? Quem é o padroeiro de Penafiel? Que grupo matou Aldo Moro? De que grupo é vocalista Teresa Salgueiro? Que equipa venceu a Taça CERS em hóquei em patins? De que clube é treinador Bobby Robson? Que empresa tem uma refinaria em Leça da Palmeira? A que partido pertence Duarte Lima? Quem financia as IPSS? Quantos submarinos tem a marinha portuguesa? Quantos municípios há em Portugal? Qual o comprimento da Ponte do Freixo? Quantos anos tem Inês de Medeiros? Qual a distância de Braga a Guimarães? Qual a altura do K2? Qual o valor da dívida da Eurotunnel? Qual a área da Baixa-Saxónia? Quantos habitantes tem a República Dominicana? Quantos golos marcou Eusébio na sua carreira? A que velocidade viaja a luz? Quando foi a independência de Cabo Verde? Quando foi aprovada a Declaração Universal dos Direitos do Homem? Quando morreu Salvador Allende? Quando morreu Simão Bolívar? Em que dia se

comemora a independência do Brasil? Em que ano ocorreu o 25 de Abril? Em que embateu o Titanic? Qual o símbolo de liderança da Volta a Itália? O que foi erguido em 13 de Agosto de 1961? A que era alérgico Mel Blanc? Que país é campeão do mundo de futebol? Como morreu Pasolini? Como se tornou o Brasil tetracampeão mundial de futebol? Qual foi o primeiro filme sonoro português? Que vende Fausto ao Diabo? Que animal é o símbolo da Namíbia? Qual o pseudónimo de Álvaro Cunhal? Qual é a nacionalidade de Yordan Letchkov? Qual a nacionalidade de Hercule Poirot? O que era Napoleão III a Napoleão Bonaparte? De que material são os frisos do Parténon? Qual a patente de Alfred Dreyfus? De que cor é a neve? Qual é a moeda iraquiana? Qual o endereço da Livraria Barata? Quem é Leonor Beleza? Quem é Arnold Ruutel? Quem é Wim Duisenberg? Quem é Rocha Vieira? Quem é Guilherme da Fonseca? Quem é Fernando Gomes? Quem é Valentina Terechkova? Quem é Jorge Amado? O que é o PC do B? O que é o PSN? O que é o CSKA? O que é a Vigor? Para onde foi Freud viver em 1939? Onde estava a rainha de Inglaterra durante a Segunda Grande Guerra? Em que rio foram encontradas bombas da Segunda Guerra Mundial por explodir? Diga um país atingido por um furacão. Quem encerrou a 6ª Conferência Mediterrânica da Cruz Vermelha e do Crescente Vermelho? Quem foi o chefe do 12º governo de Israel? Quem realizou o filme espanhol "Os piores anos da nossa vida"? Quem escreveu "A bicicleta de Leonardo"? Quem decidiu criar a bolsa "Luis Donaldo Colosio"? Quem venceu o IV Torneio Internacional de Xadrez "Cidade de Pamplona"? Como se fará desaparecer o "Japan Premium"? Em que ano é que o "War Powers Act" foi aprovado pelo Congresso americano? Que banda contribuiu para a banda sonora do filme "O Ponto de Zabriskie"? Quem representa o papel de prostituta em "Taxi Driver"? Quem criou a expressão "encontros imediatos do terceiro grau"? Que afectou o "efeito tequilha"? A que se deveu o "efeito tequilha"? Quantos desaparecidos provocou o furacão "Angela" nas Filipinas? Que petroleiro teve um acidente depois do petroleiro "Mar Egeu"? Que contem o "Chester Beatty"? Em que equipa de basquete joga Shaquille O'Neill? Onde foi fundada a Hewlett-Packard (HP)? Qual a proveniência do Chaos Computer Club (CCC)? O que é Micro Compact Car (MCC)? Onde são produzidas as barras de chocolate "Milka"? Quem é o realizador de "Nikita"? Quando decorreu o 48º festival de Avignon? Quem realizou "Canción de Cuna"? Quem escreveu "O Príncipezinho"? Quem é a "diva dos pés descalços"? Quem realizou o filme "Lisbon Story"? Quando foram criadas as FPLM (Forças Populares de Libertação de Moçambique)? Quando estreia o filme "Lisbon Story"? Quando se tornou "A Portuguesa" hino nacional? A que país pertence o consórcio "Alfa Group"? Onde decorre o festival "Urso de Ouro"? Em que ano decorreu o 48º Festival de Cinema de Cannes? Que acordo foi assinado na 16ª cimeira de chefes de estado da CEI? Em que datas de 1994 se realizou a 51ª Mostra Internacional de Veneza? Quem venceu o prémio de melhor actriz nos 51º Globos de Ouro?

## **Público**

### **Texto1**

Augusto Duarte é um dos implicados detidos ontem pela PJ para primeiro interrogatório judicial, tendo a juíza de instrução Ana Cláudia Nogueira, do Tribunal de Gondomar, aplicado o termo de identidade e residência como medida de coacção. Augusto Duarte saiu do tribunal durante a madrugada de hoje. Por sua vez, o árbitro Jacinto Paixão e os assistentes José Chilrito e Manuel Quadrado só foram ouvidos hoje, sendo

que passaram a noite sob detenção. Já o empresário António Araújo, envolvido nas transferências de alguns jogadores do FC Porto, está, neste momento, segundo a SIC Notícias, a prestar declarações no Tribunal de Gondomar, após ter sido detido também ontem. Hoje, a PJ deteve o presidente do FC Porto, Pinto da Costa, que se encontra desde as 16h15 no Tribunal de Gondomar, onde aguarda o interrogatório no âmbito deste caso. O cardeal Desmond Connell e mais três arcebispos irlandeses são acusados num relatório de 700 páginas de encobrir durante décadas o abuso sexual de crianças por sacerdotes de Dublin. Tareq e Michael Salahi, dois aspirantes a estrelas nesse cintilante universo americano dos reality shows, infiltraram-se num banquete de Estado, na Casa Branca, na passada terça-feira. A cantora pop Shakira vai estar em Portugal nos próximos dias para se reunir com os chefes de Estado da América Latina e discutir formas de colaboração para ajudar as crianças latino-americanas. Por decisão do mesmo magistrado, foram validadas as escutas telefónicas, e a lei que pune a corrupção no desporto foi dada como constitucional, o que suscitou recursos, sem efeitos suspensivos, dos principais acusados. Estes processos foram entretanto avocados por uma equipa liderada pela procuradora-geral adjunta, Maria José Morgado. No processo Apito Dourado de Gondomar, os 24 arguidos são suspeitos de terem montado um esquema para induzir os árbitros a beneficiar clubes locais. A 8 de Fevereiro, o despacho de acusação do principal processo sobre corrupção no futebol português citava um total de 27 arguidos, mais três do que os levados a julgamento. O secretário de Estado Adjunto e da Saúde, Manuel Pizarro, atribuiu hoje à falta de médicos de família o tempo que alguns doentes têm de esperar por uma consulta nos centros de saúde. O Ministério Público está a braços com a investigação de mais de 500 mil processos-crime, muitos dos quais classificados como graves e violentos. Quatro obras da jovem artista sérvia Katarina Zdjelar, que utilizam muitas vezes a música como ponto de partida. Até 28 de Fevereiro no Centro de Artes Visuais de Coimbra. Robin Soderling continua imbatível no Masters de ténis. Depois de ter derrotado Rafael Nadal, o sueco bateu nesta quarta-feira o sérvio Novak Djokovic. O médio brasileiro Fellype Bastos, que estava emprestado ao Belenenses, vai trabalhar a partir de hoje novamente com a equipa do Benfica. Abusos de sacerdotes sobre crianças têm sido noticiados também noutras países e abalaram a Igreja norte-americana. O Papa Bento XVI defendeu que os abusadores devem responder perante a Justiça e no ano passado, durante a sua visita aos EUA, encontrou-se com vítimas. Um ardiloso manual escrito para a CIA com o propósito de ensinar os agentes secretos a usarem truques de magia chegou às bancas nos Estados Unidos. A activista iraniana dos direitos humanos Shirin Ebadi revelou que as autoridades iranianas lhe confiscaram a medalha e o diploma recebidos quando foi premiada com o Nobel da Paz, em 2003. O conselho de governadores da agência decidiu hoje censurar o Irão por ter construído em segredo uma segunda central de enriquecimento de urânio. Uma sondagem de 2004 mostrava que só 44 dos portugueses manifestaram conhecer o coronel Ernesto Melo Antunes e que destes muitos não o consideravam um figura importante. Há um surto de gripe A no plantel do Estrela da Amadora. Dezasseis dos 22 atletas, os dois treinadores, o fisioterapeuta e o roupeiro foram infectados com o vírus H1N1. Por isso, a equipa está de quarentena. A coordenadora do Observatório Permanente da Justiça considera que a reforma penal efectuada em Portugal não veio resolver os principais problemas estruturais da justiça. O Parlamento Europeu aprovou hoje uma resolução promovida pela eurodeputada Edite Estrela a defender a proibição total de se fumar nos espaços públicos fechados. Portugal vai ter uma delegação do observatório, que terá como missão principal alertar para as várias formas de violência de género. O Governo Regional dos Açores vai lançar, ainda este ano lectivo, um projecto pioneiro de saúde escolar, que incluirá a despistagem de

doenças em áreas como a saúde oral, obesidade, visão e audição. Mais de cem documentos do escritor Luiz Pacheco, incluindo cartas, entrevistas e críticas literárias, foram comprados pela Biblioteca Nacional por 12 mil euros em leilão. Mais de seis mil alunos das escolas do primeiro ciclo do ensino básico nos Açores vão beneficiar, a partir de Janeiro, de um programa de distribuição de fruta e produtos hortícolas. A Câmara de Lisboa recebeu até ao momento 120 propostas de cidadãos para o orçamento municipal de 2010, número que fica muito longe das 580 recebidas na edição de 2009. O plano de retirada das tropas estrangeiras no Afeganistão começará a ser definido numa conferência internacional, a 28 de Janeiro, em Londres, revelou o primeiro-ministro britânico, Gordon Brown. O controlo das armas nucleares no Paquistão foi hoje transferido do Presidente Zardari para o primeiro-ministro, no dia em que expirou uma amnistia para casos de corrupção que protege Zardari. O Governo angolano criou este fim-de-semana uma comissão inter-sectorial para a elaboração de relatórios sobre direitos humanos, que será coordenada pelo Ministério das Relações Exteriores. Uma sondagem de 2004 mostrava que só 44 por cento dos portugueses manifestaram conhecer o coronel Ernesto Melo Antunes e que destes muitos não o consideravam uma figura importante. O Governo afirma que se os diplomatas de natureza económica da oposição forem aprovados, gerarão um desequilíbrio nas contas públicas de 2300 milhões de euros. Pela primeira vez desde 2003, ano em que foi criado o Provedor do Cliente das Agências de Viagens, as reclamações caíram quase dez por cento este ano. As quebras no turismo são apontadas como uma das possíveis razões para a descida, que inverte a tendência dos últimos seis anos. Os empresários portugueses com negócios nos Emirados Árabes Unidos estão convictos de que a crise no Dubai é passageira, mantendo os projectos para aquele mercado e reforçando a aposta no emirato de Abu Dhabi. O banco brasileiro Caixa Económica Federal, detido pelo Estado, estará prestes a anunciar a compra de uma participação em pelo menos mais um banco além do Panamericano até ao final do ano, noticia hoje O Estado de São Paulo. O checo Zdeněk Zeman tem criticado frequentemente o trabalho de José Mourinho. O técnico português respondeu-lhe neste sábado e de forma bem dura. Um chegou há dias a Alvalade com a missão de atenuar a crise do Sporting. O outro promete altos voos aos adeptos do Benfica. Em comum, têm um percurso ascendente na Liga. O presidente do FC Porto criticou os sucessivos adiamentos na regionalização. O interrogatório de cerca de oito horas a Armando Vara, no âmbito do processo Face Oculta, terminou à meia-noite, no DIAP de Aveiro. O juiz deverá divulgar na próxima semana as medidas de coacção. O Banco Alimentar Contra a Fome promove hoje e domingo uma nova campanha de recolha de alimentos em supermercados de 17 regiões do país, numa altura em que aumentou o número de pessoas necessitadas. O conselho de administração da REN-Redes Energéticas Nacionais decidiu hoje, em sessão extraordinária, designar Rui Cartaxo como presidente-executivo da empresa. O Ministério da Educação vai eliminar a divisão da carreira docente, confirmou Dias da Silva, da Federação Nacional da Educação. O regime de Teerão comunicou ao Reino Unido que libertou os cinco velejadores britânicos, detidos há uma semana pelos Guardas da Revolução no Golfo Pérsico. Em entrevista ao PÚBLICO, o presidente da Comissão de Ambiente do PE defende que os EUA devem ir além das metas já anunciadas por Barack Obama para a redução de gases com efeito de estufa. Parte hoje para o sul do Líbano uma nova Unidade de Engenharia do Exército Português com 130 militares, que vão integrar a missão das Nações Unidas, presente naquele país há mais de trinta anos. Carlos Carvalhal, treinador do Sporting, não vai poder contar com Abel, Vukcevic, André Marques e Yannick Djalo para o jogo frente ao Heerenveen, todos devido a lesão, com Izmailov a treinar hoje sem vigilância médica. O argentino Lionel Messi venceu hoje a Bola de



Ouro, prémio atribuído anualmente pela revista France Football ao melhor futebolista mundial. Os 20 clubes da Premier League inglesa gastaram mais de 77 milhões de euros em pagamentos a agentes de jogadores entre Outubro de 2008 e Setembro deste ano. Os números foram revelados pela Liga inglesa. A Federação Internacional de Ténis aplicou uma pesada multa à tenista Serena Williams, que poderá ser afastada do Open dos Estados Unidos se voltar a cometer alguma infracção no torneio. Concorde com Jorge Jesus, quando o treinador do Benfica diz que o Sporting está afastado da luta pelo título? A distância dos leões em relação a Braga, Benfica e FC Porto é irrecuperável? Margarida, Tomás, Joana e Gustavo são quatro dos mais de 800 bebés que foram nascer a Badajoz depois do fecho da maternidade de Elvas. Os pais vêem vantagens. A proposta que a presidência sueca da União Europeia queria ver aprovada pelos ministros da Saúde dos 27 foi rejeitada com os votos de Portugal e Espanha. Os sindicatos de professores querem saber qual será a duração da carreira proposta pelo Ministério da Educação, bem como o número de professores em cada escalão. Os médicos Daniel Serrão e João Lobo Antunes receberam sábado à noite, no Casino da Figueira da Foz, o Prémio Pedro Hispano 2009, atribuído pela Academia com o mesmo nome. O imparável e electrizante maestro venezuelano Gustavo Dudamel está de regresso a Portugal. A Orquestra Juvenil Ibero-americana tem hoje a sua estreia mundial na Fundação Gulbenkian. Duas décadas depois do seu fim, festival Cascais Jazz está de regresso esta semana com cinco concertos que recordam um dos mais importantes eventos de divulgação do jazz em Portugal. Um homem de 41 anos matou hoje a tiro a mulher, que se encontrava dentro de uma ambulância, e um militar da GNR em Montemor-o-Velho, no distrito de Coimbra, disse à agência Lusa fonte da corporação. As taxas de execução do QREN estão abaixo do razoável e as verbas deste programa estão a ser mal geridas, afirma a Associação Nacional de Municípios Portugueses. José Luís Oliveira vice-presidente da Câmara de Gondomar, presidente da Direcção do Gondomar SC e presidente da Assembleia-Geral do Sousense, à data dos factos, é o principal arguido no processo, sendo acusado de 26 crimes dolosos de corrupção activa e 21 crimes dolosos de corrupção desportiva activa. O ex-presidente do Conselho de Arbitragem da Federação Portuguesa de Futebol, José António Pinto de Sousa, responde por 26 crimes dolosos de corrupção passiva para acto ilícito. Na óptica do MP, Pinto de Sousa teria, como contrapartida das nomeações de árbitros a pedido, os votos para ser eleito presidente do Conselho de Arbitragem e as influências de Valentim Loureiro junto do poder político. A acusação sustenta mesmo que uma das contrapartidas dos alegados favores de Pinto de Sousa teria sido a sua integração, em Março de 2004, na comitiva oficial que o então primeiro-ministro, Durão Barroso, levou numa visita a Moçambique. Os árbitros, por sua vez, beneficiavam da oferta de artigos em ouro. Carneiro da Silva está também designado para dirigir, igualmente no Tribunal de Gondomar, o julgamento em que o presidente do FC Porto, Pinto da Costa, pede 50 mil euros de indemnização ao Estado por alegada detenção ilegal. Neste caso, que será julgado a partir de 27 de Novembro, Carneiro da Silva não pediu escusa. O maior impacto vai sentir-se no Sul, incluindo em Portugal, diz um estudo do Centro Comum de Investigação da Comissão Europeia. Real Madrid venceu e subiu ao primeiro lugar do Grupo C, enquanto o Manchester United, derrotado em casa pelo Besiktas, adiou a discussão do primeiro lugar do Grupo B para a última ronda. É um professor de atletismo da Faculdade de Desporto da Universidade do Porto que vai orientar Vanessa Fernandes. Paulo Colaço sucede a Sérgio Santos como treinador da triatleta portuguesa. O Presidente dos Estados Unidos vai estar na conferência climática da ONU em Copenhaga, onde será discutido um substituto ao Protocolo de Quioto. Barack Obama participará na reunião de abertura do encontro, a

9 de Dezembro, anunciou a Administração. A Google vai digitalizar 14 mil imagens de artefactos preciosos guardados no Museu Nacional iraquiano e disponibilizá-las online no início do próximo ano. Para comemorar o bicentenário do nascimento do jornalista, político e orador parlamentar José Estêvão, os Artistas Unidos apresentam esta peça a 25 de Novembro na Biblioteca da Assembleia da República, em Lisboa. José Maria Vieira Mendes escreveu a peça para os Artistas Unidos. Ana é uma mulher que procura justificar a sua existência tendo alguém a seu lado. De 26 de Novembro a 13 de Dezembro no Teatro Municipal de Almada. Para comemorar o bicentenário do nascimento do jornalista, político e orador parlamentar José Estêvão, os Artistas Unidos apresentam esta peça a 27 de Novembro no Teatro Aveirense, em Aveiro. Obras inéditas de João Louro que têm o jogo e a sua iconografia como ponto central. Até 28 de Fevereiro no Centro de Artes Visuais de Coimbra. Meio milhão de mensagens de pagers enviadas entre as 3h00 de 11 de Setembro de 2001 e a mesma hora do dia seguinte nos EUA foram publicadas no site Wikileaks, que divulga documentos confidenciais. A Quinta São José Ribamar, antigo convento do século XV, em Algés, no concelho de Oeiras, vai ser transformada num condomínio de luxo, com um investimento de 20 milhões de euros. O património recuperado poderá vir a ser visitado, ou não, pelo público em geral. A TMN negou hoje qualquer envolvimento na escolha da localização da Red Bull Air Race, na sequência do apelo ao boicote promovido pela Associação dos Comerciantes do Porto contra a eventual deslocalização do evento para Lisboa.

## Texto2

O Ministério da Educação e organizações sindicais realizam hoje a segunda reunião sobre a estrutura da carreira docente, com os representantes dos professores à espera de obter dados mais precisos para a concretização das propostas do Governo. O Tribunal de São João Novo, Porto, começa hoje a julgar um mega-processo de contrabando de tabaco para Inglaterra e Portugal que envolve 80 arguidos acusados de transacionarem ilegalmente mais de 114 milhões cigarros entre 2005 e 2007. Experimente fazer uma busca nas Imagens Google por Michelle Obama. A primeira fotografia que lhe aparece é uma manipulação ofensiva do rosto da primeira-dama americana. Numa rara manobra, a Google já colocou online uma mensagem distanciando-se da imagem e pedindo desculpa pela sua destacada presença online, sublinhando, porém, que não irá retirá-la. Um grupo de suspeitos rebeldes islamistas lançou um violento ataque suicida contra uma mesquita muito frequentada por militares em Rawalpindi, nos subúrbios de Islamabad, causando dezenas de mortos. O chefe da junta militar que governa a Guiné Conacri, Moussa Dadis Camara, encontra-se num avião do Burkina Faso com destino a Marrocos, sendo assistido a bordo por um médico. Vladimir Putin, já com dois mandatos presidenciais cumpridos antes de assumir a liderança do Governo em Maio de 2008, deixou claro que admite voltar a candidatar-se à chefia do Kremlin em 2012. Madrid pediu hoje formalmente a Rabat que dê um passaporte a Aminatu Haidar, expulsa a 14 de Novembro do Sara Ocidental e em greve de fome no aeroporto de Lanzarote desde dia 16. Uma explosão num hotel da capital da Somália causou a morte a três ministros, num ataque que as autoridades crêem ter tido por alvo directo entidades oficiais do Governo de transição no país. José Sócrates anunciou esta manhã no Parlamento que o Governo vai propor o aumento do salário mínimo nacional em 2010 para 475 euros, o que representa mais 25 euros que o valor do corrente ano. A decisão surgiu quando faltava menos de uma hora para terminar o debate potestativo sobre as iniciativas do Bloco de Esquerda e do PCP para combater a corrupção. O primeiro-ministro, José Sócrates, anunciou hoje o alargamento do prazo para as empresas com

dívidas ao fisco. O Governo vai duplicar, também apenas durante o ano de 2010, o prazo para as dívidas já existentes, permitindo que o limite seja o das 120 prestações mensais. O centro de conferências da Cidade do Cabo, onde esta tarde se realiza o sorteio do Mundial 2010, foi evacuado esta manhã durante algum tempo, depois de uma falsa ameaça de bomba. O anúncio foi feito por Jérôme Valcke, secretário-geral da FIFA, após uma reunião do comité executivo do organismo, na Cidade do Cabo. Domingos Soares Oliveira, administrador-executivo do grupo Benfica, assegura que existem 18 milhões de euros para investir no reforço do plantel na reabertura do mercado, em Janeiro. O Nacional da Madeira foi derrotado hoje por 4-1 em Bremen, contra o Werder, e está fora da fase seguinte da Liga Europa. Os madeirenses ainda assustaram os alemães na segunda parte, mas continuam sem vencer. Os mais de mil controlos realizados pela UEFA ao longo da última época desportiva, a maioria com recolha específica para a despistagem da EPO, resultaram em dois casos positivos, embora apenas um tenha motivado castigo disciplinar. A Estradas de Portugal vistoriou este ano 574 pontes e viadutos e detectou 169 casos em que estas estruturas precisam de uma intervenção nos próximos dois anos. Os militares que compunham a coluna que foi atropelada numa estrada nacional em Tancos, Santarém, não usavam coletes reflectores. O Dia Nacional do Doente com Esclerose Múltipla assinala-se hoje e visa alertar para as dificuldades que os doentes enfrentam no dia-a-dia devido às limitações físicas causadas por esta doença que afecta mais de 5000 portugueses. O Conselho de Escolas vai apresentar propostas para a revisão curricular do Ensino Básico, até final do ano lectivo, baseadas em competências-chave definidas pela União Europeia já em experimentação em algumas escolas em Língua Portuguesa e Matemática. Albert Tuijnman, responsável do Banco Europeu do Investimento, declarou, esta manhã, a disponibilidade da instituição em apoiar programas para o desenvolvimento da literacia em Portugal. O Museu de Arte Popular, em Lisboa, vai reabrir em 2010 no edifício onde originalmente funcionou, na zona de Belém, afirmou hoje a ministra da Cultura, Gabriela Canavilhas. Um astronauta americano, Chuck Backer, parte numa missão para descobrir um planeta distante. Será que está preparado para aquilo que o espera? Podem um extra-terrestre e um astronauta ser amigos? A iluminação das principais artérias de Lisboa custou mais de um milhão de euros. O recurso a lâmpadas de baixo consumo encareceu o projecto, mas a quadra ficou mais amiga do ambiente. Um autocarro que transportava 40 alunos da Escola Profissional de Vouzela embateu esta manhã contra um veículo ligeiro e ficou suspenso numa ravina no lugar de Volta Escura, à entrada da vila de Vouzela, Viseu. Apesar do susto, não houve feridos e o autocarro foi retirado mais tarde com uma grua pela empresa de transportes responsável. O ministro dos Assuntos Parlamentares afirmou hoje estar a aguardar a decisão dos tribunais quanto ao processo do quinto canal de televisão generalista e que para já não pretende avançar com a Lei da Concentração dos Media. A organização que atribui os Prémios Pulitzer de jornalismo informou que modificou os seus critérios, passando a dar mais oportunidade de participação a diferentes tipos de artigos publicados exclusivamente na Internet. Espanha a ferro e fogo. De um lado estão os artistas, que exigem medidas de apoio e o fim definitivo da pirataria. Do outro lado da barricada estão os internautas. Ao sexto clique, poderão acabar-se as borlas. Pode assim ser resumido o anúncio sobre a consulta gratuita de notícias de jornais online através do Google News. As empresas de media há muito que travam uma luta com o gigante das pesquisas, que agrega os conteúdos e os disponibiliza gratuitamente. Por isso mesmo, a Google passou a bola para o campo dos media. Quem quiser cobrar as suas notícias indexadas no Google News, pode começar a fazê-lo. As autoridades iraquianas actualizaram para 127 o número de vítimas mortais dos cinco atentados coordenados com carros armadil-

hados desta manhã em Bagdad. Pela primeira vez um coronel indonésio confirmou a morte deliberada de cinco jornalistas às mãos de soldados do seu país. O objectivo era evitar que relatassem o ataque a Timor-Leste, em 1975. As negociações técnicas começaram hoje no segundo dia da conferência climática de Copenhaga, estimuladas pelo anúncio da Agência de Protecção Ambiental dos EUA sobre a possível regulação das emissões. Os quatro deputados do PSD-M com assento parlamentar votarão contra o orçamento rectificativo caso seja recusado um empréstimo 129 milhões de euros à Madeira. O presidente do Vitória de Setúbal, Fernando Oliveira, confirmou que o clube sadino pretende protestar o jogo de ontem com o Sporting por entender que existiram erros técnicos do árbitro Elmano Santos. O presidente do Benfica, Luís Filipe Vieira, disse que o principal objectivo do clube para esta época continua a ser a conquista do título nacional de futebol, em declarações nos Paços do Concelho da Câmara Municipal da Mealhada. Cristiano Ronaldo é um dos cinco jogadores nomeados ao prémio Jogador do Ano da FIFA. O avançado tem como concorrentes ao galardão Messi, Iniesta, Xavi e Kaka. Com o Brasil como grande favorito, a Costa do Marfim como uma das mais fortes selecções africanas e a Coreia do Norte à espreita, quão favoráveis são as perspectivas portuguesas de ultrapassar a fase de grupos do Mundial 2010? O ciclista espanhol Óscar Pereiro, vencedor da Volta a França em 2006, foi contratado por um ano pela Astana, anunciou hoje a equipa cazaque, que tem como chefe de fila o também espanhol Alberto Contador. Patrícia Amorim, ex-nadadora do Flamengo, tornou-se segunda-feira a primeira mulher a chegar à presidência do clube, que já conta com 114 anos, e prepara-se agora para cumprir um mandato até 2012. O Nacional venceu em Leira por 2-1 e manteve-se no quarto lugar na Liga. O golo da vitória dos madeirenses foi apontado por Felipe Lopes a nove minutos do fim. A co-incineração em Souselas pode ser retomada, de acordo com a decisão do Supremo Tribunal Administrativo, que faz prevalecer a sentença do Tribunal Administrativo e Fiscal de Coimbra. A jornalista Maria João Seixas, de 64 anos, é a nova directora da Cinemateca Portuguesa, iniciando funções a partir de Janeiro. O espólio integral de Fernando Pessoa vai estar disponível online no próximo ano através da Biblioteca Nacional Digital. Na Casa Fernando Pessoa, hoje ao final da tarde, em Lisboa, Caetano Veloso disse que Lula lhe surge como uma figura de grandeza épica. Foi a inauguração do ciclo Livres Pensadores. A poucos dias do final do ano, o produtor e encenador Filipe Lá Féria ainda não garantiu a renovação, para 2010, do contrato de acolhimento que lhe permite utilizar o Teatro Rivoli, no Porto. Vinte e anos depois do primeiro pedido, os habitantes da Ilha da Culatra, no concelho de Faro, têm oficialmente, a partir de ontem, acesso a água e saneamento básico. Cinco dos 16 militares feridos num acidente em Tancos na sexta-feira mantêm-se hospitalizados, informou hoje o porta-voz do Exército, tenente-coronel Hélio Perdigão. A Estradas de Portugal está disponível para reavaliar o traçado previsto para a Auto-estrada 32, na freguesia da Branca, em Albergaria-a-Velha. O primeiro-ministro foi ontem a Viseu fazer um discurso completamente ao lado das expectativas dos participantes no XVIII Congresso da Associação Nacional de Municípios Portugueses. O vereador da Câmara de Odivelas Hugo Martins foi ontem detido pela GNR, depois de ter sido apanhado a conduzir alcoolizado e de agredir os guardas que acompanharam a ocorrência. Miguel Sousa Tavares, actual comentador do Jornal Nacional da TVI, vai deixar a estação de Queluz, confirmou ao PÚBLICO o director de informação Júlio Magalhães. O PÚBLICO foi hoje reconhecido como o melhor site na categoria de Breaking News nos prémios atribuídos pelo Observatório de Ciberjornalismo. As violações de direitos humanos no Irão estão ao seu pior nível dos últimos 20 anos, denunciou hoje a Amnistia Internacional, que denuncia casos de tortura, violação e mortes à margem da lei. Barack Obama recebeu hoje o Prémio No-

bel da Paz e não esqueceu a controvérsia que rodeou a entrega do galardão ao líder de um país envolvido em duas guerras. O multimilionário George Soros afirma ter encontrado os cem mil milhões de dólares para os problemas do clima e desta forma garantir o sucesso da conferência de Copenhaga. O diploma que estabelece o casamento entre homossexuais está em fase final de elaboração e deverá ser apresentado no Conselho de Ministros da próxima semana. O ex-deputado do PS Manuel Alegre disse hoje que o Presidente da República, Cavaco Silva, deverá intervir face a uma instabilidade governativa, cabendo-lhe decidir como e quando o fará. O Parlamento aprovou hoje, por unanimidade, um projecto de resolução do PSD para que seja criada uma comissão parlamentar eventual destinada a analisar medidas de combate à corrupção. Apenas o Bloco de Esquerda destoa e anunciou hoje o seu voto contra. Alargar de oito para dez o número de pilotos pontuados num Grande Prémio é uma das novidades propostas pela FIA, que passa a premiar o vencedor de cada prova com 25 pontos. Hulk, do FC Porto, é o segundo jogador com mais remates da fase de grupos da Liga dos Campeões, com 31 tentativas, menos duas que o uruguaio Diego Forlán, do Atlético de Madrid. Ronaldinho, actualmente no AC Milan, foi nomeado futebolista da década pela prestigiada revista britânica World Soccer, à frente de Messi, eleito o melhor de 2009, e Cristiano Ronaldo. A revelação dos valores milionários que o seleccionador cobrou pela qualificação da França para o próximo Campeonato do Mundo deixou o país em estado de choque. Até a ministra da Economia francesa, Christine Lagarde, não calou a sua indignação. A eventual punição com suspensão de funções de José Luís Lopes da Mota pode não implicar o afastamento imediato do procurador-geral adjunto das funções de presidente da Eurojust. A Escola Superior de Educação de Portalegre vai abrir um processo de averiguações para identificar os alegados autores de praxes violentas a um grupo de alunos. A ministra da Cultura, Gabriela Canavilhas, confirmou hoje a abertura do Museu da Língua Portuguesa em 2010 e anunciou que este projecto não irá ocupar o Pavilhão de Portugal. Um homem de 55 anos morreu esta manhã atropelado na freguesia da Lagoa, em Vila Nova de Famalicão. A Protecção Civil ainda não pagou às corporações de bombeiros do distrito de Lisboa as participações pelo trabalho e despesas que tiveram durante as inundações de Fevereiro de 2008. O administrador delegado da empresa Águas do Norte Alentejano garantiu que os quinze concelhos que compõem o distrito de Portalegre vão ficar abrangidos pelo abastecimento até 2011. O ministro das Obras Públicas vai receber em audiência a Plataforma do Entendimento para tentar travar a introdução de portagens na Scut do Norte Litoral. A Câmara de Lisboa ainda diz que não sabe se assumiu algum compromisso para apoiar a construção de uma igreja em forma de caravela no Restelo, nem se o respeitou ou não. O financiamento a curto prazo para ajudar os países em desenvolvimento a enfrentar as alterações climáticas não é a resposta, denuncia a China. Multiplicam-se os esforços diplomáticos para tentar encontrar uma solução para o caso da activista sarauí Aminatu Haidar, há 26 dias em greve da fome na ilha de Lanzarote, nas Canárias. Os restos mortais de Adolf Hitler foram queimados em 1970 pelo KGB e lançados a um rio na Alemanha, por ordem do então chefe dos serviços secretos, Iuri Andropov. Desconhecidos levaram durante a noite o cadáver do antigo Presidente cipriota Tassos Papadopoulos, que em 2004 colocara o país na União Europeia. O Presidente angolano deu a entender que as primeiras eleições presidenciais pós-guerra, que se dizia que se poderiam realizar em 2009, não devem acontecer antes de 2012. Catarina Sarmento e Castro não foi eleita para o Tribunal Constitucional por apenas cinco votos. A jurista precisava de 144 votos, mas apenas 139 deputados votaram no seu nome. A aposta na renovação da economia regional através das indústrias criativas e da regeneração urbana é o principal eixo do desenvolvimento da Capital Europeia da Cultura de 2012.

O Governo prepara-se para a aprovar em Conselho de Ministros, já na próxima quinta-feira, o diploma que permite o casamento entre homossexuais, apurou o PÚBLICO. O Tribunal da Relação do Porto rejeitou os três recursos do Ministério Público e confirmou a absolvição de Pinto da Costa, e restantes arguidos, no caso do envelope, um dos processos ligados ao do Apito Dourado.

### Texto3

O grupo parlamentar do PS vai dar início à discussão interna sobre a concretização de um referendo à regionalização ainda na presente legislatura. Durante várias horas, Jorge Colombo, designer português radicado em Nova Iorque, registou as pessoas e a Times Square em frente ao museu Madam Tussauds. O Governo vai criar um banco de dados online para divulgar o investimento e critérios de distribuição da publicidade do Estado em todos os órgãos de comunicação social. A cimeira leva já cinco dias, mas é a partir de hoje que os milhares de activistas em Copenhaga se farão ouvir. O ex-embaixador moldavo em Portugal, Mihail Carmezan, foi severamente repreendido pelo Ministério dos Negócios da Moldávia e encontra-se em parte incerta. O Ministro da Administração Interna, Rui Pereira, anunciou hoje que a cidade de Gaia irá acolher a nova sede da Unidade Especial de Policia, actualmente instalada no Porto. As tecnologias verdes podem tornar-se no terceiro sector industrial do mundo até 2020, segundo um estudo publicado pelo Fundo Mundial para a Natureza, à margem da Cimeira de Copenhaga. O chefe do Governo italiano, Silvio Berlusconi, foi hoje atingido no rosto com um murro, caiu por terra, a sangrar, e foi conduzido ao hospital de San Raffaele, em Milão. Um jornalista de uma rádio perguntou a Willy Toledo se ele sabia que os membros da Plataforma de Apoio a Aminatu vão ser acusados de falta de prestação de cuidados. Ele respondeu que não faz mal. O central luso-brasileiro Pepe vai ter de ser operado ao ligamento cruzado do joelho direito, devendo ficar afastado dos relvados durante seis meses. Rio Ave surpreende Marítimo na Madeira, por 1-0, enquanto a Académica voltou às vitórias derrotando em casa o Leixões, por 2-0. A selecção feminina portuguesa de corta-mato revalidou hoje, em Dublin, na Irlanda, o título europeu, totalizando 25 pontos, menos 26 do que a segunda classificada, a Grã-Bretanha. Para o resultado de 104-92 foi decisiva a actuação abaixo do normal da estrela da equipa, Kobe Bryant. O Manchester United desperdiçou uma excelente oportunidade de colar-se ao Chelsea no comando do campeonato inglês. A equipa de Alex Ferguson foi derrotada em casa pelo Aston Villa, depois de no início da tarde o Chelsea ter empatado em Stamford Bridge com o Everton. O português Álvaro Parente vai ser o piloto de testes e de reserva da Virgin, apurou o PÚBLICO junto de fonte ligada ao piloto portuense. A Virgin é uma das novas equipas da Fórmula 1 e terá o alemão Timo Glock e o brasileiro Lucas di Grassi como pilotos titulares. O Serviço de Estrangeiros e Fronteiras deteve 14 pessoas, candidatas à aquisição da nacionalidade por naturalização, por fraude na realização do teste de diagnóstico da língua portuguesa. Cerca de 44 por cento dos brasileiros que vivem em Portugal e 29 por cento dos oriundos da África subsariana sentem-se discriminados e dizem já ter sido alvo de atitudes racistas. A Protecção Civil alerta para o rápido abaixamento das temperaturas previsto para os próximos dias, fenómeno que será particularmente acentuado durante a noite. Orlando Monteiro da Silva prometeu negociar com o Governo uma tabela de comparticipações para que os utentes do Serviço Nacional de Saúde possam recorrer a clínicas dentárias privadas. As obras da Escola Secundária Gil Vicente, em Lisboa, no âmbito do programa de modernização do parque escolar, pioraram as condições do estabelecimento de ensino, diz um parecer da comunidade escolar. A Universidade da Beira Interior está a

desenvolver um tecido com sensores electrónicos para prevenir lesões em doentes acamados ou com mobilidade reduzida. A Feira Viva, organizadora da Terra dos Sonhos, onde ontem 15 pessoas ficaram feridas após a cedência de uma bancada, instalou uma unidade móvel no recinto para prestar apoio psicológico às vítimas. As buscas para encontrar o homem que foi ontem arrastado na praia do Tamariz, no Estoril, foram retomadas hoje cerca das 08h30, disse o Centro Distrital de Operações de Socorro de Lisboa. Um casal foi hoje de manhã trucidado por um comboio, na linha de Cascais, em Belém, confirmou à Lusa fonte dos Bombeiros Sapadores de Lisboa. O primeiro-ministro italiano, Sílvio Berlusconi, passou a noite de ontem para hoje no hospital milânês de San Raffaele, a recuperar de uma agressão. O Partido Trabalhista britânico está a preparar-se para a possibilidade de legislativas no fim de Março, depois de uma sondagem ter diminuído a diferença que o separa dos conservadores. O PSD entregou ao presidente da Assembleia da República um requerimento para a análise da execução económico-financeira das entidades que integram o Serviço Nacional de Saúde pela UTAO. Os deputados do PS rumam hoje a Beja com a regionalização na agenda para umas jornadas centradas no desenvolvimento regional e que dificilmente escapará às polémicas do TGV ou do Orçamento. O ministro da Defesa, Augusto Santos Silva, inicia hoje uma visita aos contingentes de militares portugueses no Kosovo e no Líbano. O médico da selecção portuguesa Henrique Jones considerou hoje que o defesa Pepe apenas estará no Mundial 2010 de futebol se recuperar em cinco meses, prazo mínimo para debelar uma lesão no ligamento cruzado anterior do joelho. O FC Porto venceu o Vitória de Setúbal, por 2-0, e vai à Luz defrontar o Benfica a apenas um ponto da liderança. O avançado Alan Kardec, internacional sub-20 pelo Brasil, vai ser jogador do Benfica a partir de Janeiro, afirmou ao site Maisfutebol José Hamilton Mandarico, vice-presidente do Vasco da Gama. O Benfica, no entanto, não confirma o negócio. O Sporting de Braga empatou hoje a zero bolas com a Naval e perdeu a oportunidade de se isolar no primeiro lugar do campeonato. A equipa do Inter aproveitou as derrotas do Milan e da Juventus, cumpridas 16 jornadas. A partir de 1 de Janeiro os portugueses vão deixar de pagar cinco euros por cada dia de internamento num estabelecimento público de saúde e dez euros por cada acto cirúrgico que não exija internamento. Os portugueses vão riscar duas pessoas das listas de Natal. Tornaram-se os mais cautelosos da Europa e vão comprar apenas 13 prendas, segundo um estudo da consultora Deloitte. O Serviço Nacional de Saúde foi lesado em mais de 700 mil euros em consequência de uma fraude ocorrida entre 2003 e 2007, a qual consistiu na falsificação de centenas de receitas médicas. Pode até parecer absurdo, mas o Instituto do Emprego afastou do serviço um dos seus mais antigos advogados, porque este insistia em manter uma postura de isenção e imparcialidade. O julgamento do processo Casa Pia prossegue numa sessão em que o colectivo de juízes poderá vir a pronunciar-se sobre se aceita ou não mais alterações de factos imputados aos arguidos. A advogada da mãe que iniciou em Junho um protesto para reaver o filho disse que Martim deverá ser entregue aos avós paternos, embora a decisão judicial de quarta-feira seja provisória. A posse de uma licenciatura é um dos requisitos essenciais para frequentar o novo curso de Medicina da Universidade de Aveiro, cuja criação será formalizada terça-feira. Os elementos do Conselho Executivo do Agrupamento de Escolas de Coimbra Inês de Castro conheceram a sentença da acção principal proferida pelo Tribunal Administrativo, que lhes foi favorável. A curadora Helena de Freitas, assessora do Centro de Arte Moderna da Fundação Calouste Gulbenkian, está em negociações com a Fundação Paula Rego para a direcção da Casa das Histórias, em Cascais. Foi há 20 anos que Marcelo Rebelo de Sousa, então candidato à presidência da Câmara de Lisboa, se atirou ao Tejo. Já então chamava a atenção para os elevados índices de poluição daquelas águas. A Câmara

do Porto está a negociar a instalação, em Campanhã, de um grande parque temático. A proposta do município é que ocupe uma área que ronda os 20 a 30 hectares, na zona das Areias. Os motoristas dos transportes públicos de Coimbra estão em greve até ao dia 20, exigindo à administração condições mais favoráveis aos trabalhadores na progressão na carreira. Foi um dia para esquecer em Copenhaga. Ao princípio da noite, as negociações estavam praticamente paradas, em consequência de entraves levantados ao longo do dia. O teste feito pelo regime iraniano de uma versão melhorada do míssil de médio alcance Sajil deve ser respondido com um endurecimento de sanções a Teerão, defendeu Gordon Brown. O Supremo Tribunal do Paquistão anulou ontem a amnistia que tem protegido o Presidente Asif Ali Zardari e alguns dos seus próximos de acusações de corrupção, o que traz de volta a possibilidade de caos político neste país, sempre perigosamente à beira da instabilidade. Israel reagiu com firmeza contra o Governo britânico após as notícias de que um juiz tinha emitido um mandado de captura contra a actual líder da oposição, Tzipi Livni. Os médicos anunciaram hoje que só amanhã é que o Presidente do Conselho italiano, Silvio Berlusconi, poderá ter alta do hospital de Milão onde está a ser tratado dos ferimentos sofridos domingo, ao ser agredido no fim de um comício. O ministro Santos Silva sublinhou hoje que existe um contacto regular entre o Parlamento e Comissão de Contrapartidas, pelo que é garantido o escrutínio da actividade governativa no sector da Defesa. O secretário de Estado das Obras Públicas negou hoje que tenha havido um processo de contratação directa dos computadores Magalhães à empresa JP Sá Couto. O Presidente da República dá hoje posse a sete membros do Conselho de Estado, cinco eleitos pelo Parlamento, o juiz conselheiro Alfredo de Sousa e Vítor Bento, que substituiu Dias Loureiro. O argentino Angel Di María foi suspenso por dois jogos pela Comissão Disciplinar da Liga, depois de ter sido expulso no jogo entre o Olhanense e o Benfica. A invasão de campo e a violência dos adeptos após o final do jogo com o Fluminense na última jornada do Campeonato Brasileiro, que ditou a descida à II Divisão do Coritiba, custou ao clube um castigo pesado. A Espanha terminou o ano na liderança do ranking FIFA, de acordo com a lista hoje divulgada pelo organismo que rege o futebol mundial. Portugal manteve o quinto posto. Os credores do Estrela da Amadora reclamam mais de 50 milhões de euros de dívidas contraídas pelo clube, impedido de se inscrever na Liga de futebol no início da temporada. No auge da investigação do processo Face Oculta e com o processo Freeport muito próximo do fim, o Procurador-Geral da República Pinto Monteiro é recebido amanhã em Belém pelo Presidente da República. A audiência está marcada para as 16h00. A Ordem dos Farmacêuticos diz que ainda não conhece a identidade do farmacêutico acusado, mas informa que vai abrir um inquérito para avaliar a sua conduta profissional. A Polícia Judiciária realizou uma operação no âmbito da luta contra a pornografia de menores na Internet, na qual foram executadas 31 buscas em residências e estabelecimentos. A aplicação do acordo ortográfico nas escolas não vai entrar em vigor no próximo ano, anunciou a ministra da Educação Isabel Alçada. O BE vai propor no Parlamento a suspensão do pagamento de propinas no Ensino Superior e um novo regime de atribuição de bolsas de estudo bem como um novo sistema de financiamento das instituições. Cartografar todo o céu que se vê a partir da órbita da Terra e no espectro dos infravermelhos, que permite detectar objectos muito frios e, por isso, distantes, é a missão do WISE. O terreno de Lisboa onde esteve para surgir aquela que seria a única obra do arquitecto brasileiro Óscar Niemeyer em Portugal está ao abandono vai para uma década. O executivo da Câmara do Porto aprovou com os votos contra de toda a oposição, um conjunto de alterações ao Plano Director Municipal. Contudo, a decisão está longe de ser pacífica. A queda de neve é a causa provável do despiste de um autocarro em Vieira do Minho, esta madrugada, que provocou dois mortos e oito



feridos ligeiros. Quatro grandes jornais britânicos e uma agência noticiosa obtiveram hoje, do Tribunal Europeu dos Direitos do Homem, o reconhecimento do direito dos jornalistas a não revelarem as suas fontes. A Comissão Europeia enterrou hoje as suas queixas contra a Microsoft, acusada de abuso de posição dominante nos browsers de Internet. A Comissão Europeia está a planear fixar nos 85 decibéis o volume máximo permitido nos leitores de mp3. Hoje em dia muitos jovens estão a perder audição de forma precoce por causa do uso constante destes aparelhos no volume máximo. Alguns testes levados a cabo no seio da UE identificaram pessoas a ouvir música nos 120 decibéis, o que equivale ao ruído de um avião a descolar. Uma criança de dois anos foi ontem hospitalizada no Brasil com mais de 40 agulhas espetadas no corpo. Berlusconi saiu do hospital de San Raffaele, em Milão, onde se encontrava internado desde que fora ferido no domingo à noite por um agressor de 42 anos, com um historial de perturbações mentais. Ao início do dia 32 a água e açúcar a activista sarauí Aminatu Haidar foi hospitalizada de forma voluntária. Está agora a ser hidratada apenas com água, não com soro, e não será alimentada. O CDS-PP propôs hoje que seja alargado aos autarcas e dirigentes da administração central e local o regime de declarações de rendimentos a que estão sujeitos os titulares de cargos políticos. A operação ao futebolista internacional português Pepe, realizada hoje no Porto, correu bem, de acordo com o cirurgião José Carlos Noronha. Depois dos resultados e exibições das últimas semanas, quem está mais forte para o clássico de domingo? O Benfica, que ainda não perdeu em casa? Ou o FC Porto, que conseguiu encurtar para um ponto a distância em relação aos primeiros? Joachim Loew vai permanecer ao serviço da selecção alemã até 2012. Aconteça o que acontecer na África do Sul, o treinador orientará a equipa durante a campanha de qualificação para o Euro 2012. A temperatura média em Portugal está a aumentar a um ritmo mais rápido do que no resto do mundo, com um crescimento de 0,33 graus por década, segundo a análise climatológica da última década. A associação CAIS desafia hoje os lisboetas a participarem na Marcha contra a Pobreza, para alertar para um problema que afecta cerca de dois milhões de pessoas em Portugal. Desde 1989 que estão na televisão, nunca envelheceram, antes cresceram e adaptaram-se. Mas há quem pense que o mundo amarelo de Springfield já passou do prazo. Um particular adquiriu por 7200 euros, num leilão promovido em Lisboa pela leiloeira Cabral e Moncada, um Obituário do Mosteiro de Santa Cruz de Coimbra. A discussão das propostas do Governo, do BE e do PEV para a consagração do casamento entre pessoas do mesmo sexo foi agendada para 8 de Janeiro, disse à agência Lusa fonte parlamentar.

## HAREM

### Texto1

#### Fatores Demográficos e Económicos Subjacentes

A no século XVI revolta histórica produz normalmente uma nova forma de pensamento quanto à forma de organização da sociedade. Assim foi com a Reforma Protestante. No seguimento do colapso de instituições monásticas e do escolasticismo nos finais da Idade Média na Europa, acentuado pela "Cativerio Babilónica da igreja" no papado de Avignon, o Grande Cisma e o fracasso da conciliação, assistimos no século XVI ao fermentar de um enorme debate sobre a reforma da religião e dos posteriores valores religiosos fundamentais. Este debate passou completamente ao lado de Portugal, demasiado distante do foco onde surgiram estes pensamentos. A

imprensa, inventada na Alemanha por John Gutenberg, foi importante na divulgação destas ideias. As 95 Teses de Martinho Lutero foram imediatamente impressas e divulgadas por todas as regiões de língua alemã, o que contribuiu para a crescente popularidade de Martinho Lutero. Não menos relevante foi a influência da pressão social exercida pela Contra-Reforma, na qual os Jesuítas tiveram um papel de liderança. A Inquisição e a censura exercida pela Igreja Católica foram igualmente determinantes para evitar que as ideias reformadoras encontrassem divulgação em Portugal, Espanha ou Itália, países católicos. Historiadores assumem geralmente que a incapacidade de reformar (grande número de interesses legítimos, falta de coordenação na coligação dos reformadores) poderia levar a uma grande revolta ou revolução, uma vez que o sistema deverá ser gradualmente ajustado ou então desintegrar-se. O fracasso da conciliação levou à Reforma Protestante do ocidente europeu. Estes movimentos reformistas frustrados variam desde o nominalismo, a moderna devoção, ao humanismo, e ocorrem em conjunção com o crescente desagrado perante a riqueza e o poder da elite clerical, sensibilizando a população para a corrupção moral e financeira da Igreja.

#### A Reforma Religiosa e Política na Inglaterra

Henrique VIII O curso da Reforma foi diferente na Inglaterra. Tinha havido desde há muito uma forte corrente anti-clerical, tendo a Inglaterra já tido o movimento Lollard, que inspirou os Hussitas na Boémia. Mas em cerca de 1520, no entanto, os Lollards não eram já uma força activa, ou pelo menos um movimento de massas.

O carácter diferente da Reforma Inglesa deve-se ao facto de ter sido promovida inicialmente pelas necessidades políticas de Henrique VIII. Sendo este casado com Catarina de Aragão e estando apaixonado por Ana Bolena, Henrique solicita ao Papa Clemente VII a anulação do casamento. Perante a recusa do Papado, Henrique faz-se proclamar, em 1531, protector da Igreja inglesa. O "Ato de Supremacia", votado no Parlamento em Novembro de 1534, colocou Henrique e os seus sucessores na liderança da Igreja: os súbditos deveriam submeter-se ou então seriam excomungados e perseguidos. Apesar de uma certa deriva em direcção ao luteranismo, Henrique reafirma a ortodoxia católica através da "Confissão dos Seis Artigos"(1539).

Entre 1540 e 1553, sob Thomas Cromwell, a política conhecida como a dissolução dos mosteiros foi posta em prática. A veneração de santos, locais de peregrinação foram atacados. Enormes extensões de terras e propriedades da Igreja passaram para as mãos da coroa e posteriormente da nobreza e das classes altas. Os direitos adquiridos foram uma força poderosa de apoio às dissoluções.

Houve muitos opositores da Reforma de Henrique, tais como Thomas More e o Bispo John Fischer, que foram executados pela sua oposição. Mas também existiu um partido crescente de Protestantes genuínos que estavam inspirados pelas doutrinas então correntes no continente. Quando Henrique foi sucedido pelo seu filho Eduardo VI em 1547, os protestantes viram-se em ascendente no governo. Uma reforma mais radical foi imposta, com a destruição de imagens e o fecho de capelas, para além de ter sido revogada a "Confissão dos Seis Artigos". Em 1552, é redigido o novo "Livro de Orações" e promulgada a "Confissão de Fé em Quarenta e Dois Artigos", que aproximava doutrinariamente a Igreja de Inglaterra do calvinismo.

Seguiu-se uma breve reacção católica durante o reinado de Maria I (1553-1558). De início moderada na sua política religiosa, Maria procura a reconciliação com Roma, consagrada em 1554, quando o Parlamento vota o regresso à obediência papal. Porém, as perseguições violentas que move aos não católicos e o seu casamento com Filipe II de Espanha sempre se viu como a "eclesia anglicanae", ou seja, "A Igreja cristã na Inglaterra" e não derivação da Igreja de Roma ou do movimento reformista do século XVI.



Como consequência disso sempre existiu dois grandes "partidos" ou "facções": a Igreja Alta (High Church) e Igreja Baixa (Low Church), que refletem a controvérsia histórica sobre as formas de culto e de expressão.

A Reforma Anglicana buscou ser a "via média" entre os extremos romanos e puritanos. Assim aceitam os dois sacramentos do Evangelho: o Santo Batismo, através do qual a pessoa é feita membro da Igreja de Cristo, sendo que tal graça é complementada na Confirmação, e na Santa Comunhão, que une o cristão ao sacrifício de Cristo Jesus que os alimenta com seu corpo e sangue.

Para os anglicanos estes dois sacramentos foram instituídos pelo próprio Senhor Jesus Cristo. Os demais ritos sacramentais da Igreja também são aceitos, apesar de não terem sido instituídos por Cristo, mas são reconhecidos por serem, em parte, estados de vida aprovados nas Escrituras: a Confirmação, Penitência, Ordens, Matrimônio e a Unção dos enfermos.

Embora tenham se afastado de muitas das práticas devocionais medievais com relação aos santos e a Virgem Maria mantiveram um calendário específico para sua comemoração na Igreja, em especial as antigas festas marianas diretamente associadas aos méritos de seu Filho Jesus Cristo (Anunciação, Natividade etc).

Lápis-lazúli, conhecido também como lápis, é uma rocha metamórfica de cor azul utilizada como gema ou como rocha ornamental utilizada desde antes de 7000 aC em Mehrgarh na Índia, situado nos dias de hoje no Paquistão. A sua azul-escuro e opaca, fez com que esta gema fosse altamente apreciada pelos faraós egípcios, como pode ser visto por seu uso proeminente em muitos dos tesouros recuperados dos túmulos faraônicos. É ainda extremamente popular hoje. Trata-se de uma rocha e não de um mineral porque é composto de vários minerais. A primeira parte do nome, lápis, em latim significa pedra. A segunda parte, lazúli, é a forma genitiva no latim, lazulum, que veio do árabe (al)-lazward, que veio do persa lazward, que veio do sânscrito Raja Warta significando anel, vida do rei. Lazúli era originalmente um nome, mas logo veio a significar azul por causa de sua associação com a pedra. A palavra em inglês azure, do azul espanhol e português, e o azzurro italiano são cognatos.

#### História

No antigo Egito o lápis-lazúli era a pedra favorita para amuletos e ornamentos; foi usado também pelos assírios e pelos babilônicos nos selos cilíndricos (locais onde se gravavam pinturas contando a história do povo). As escavações egípcias que datam de 3000 aC continham milhares de artigos como jóia, muitos feitos de lápis. Os lápis pulverizados foram usados por senhoras egípcias como uma sombra cosmética para o olho.

Como inscrito no capítulo 140 do Livro dos Mortos egípcio, o lápis lazúli, na forma de um olho ajustado no ouro, foi considerado um amuleto de grande poder. No último dia do mês, oferecia-se este olho simbólico, porque se acreditava que, nesse dia, um ser supremo colocou tal imagem em sua cabeça. Os antigos túmulos reais sumérios de Ur, situados perto do rio Eufrates no baixo Iraque, continham mais de 6000 estatuetas belamente executadas, de lápis-lazúli, de pássaros, cervos, e roedores bem como pratos, grânulos, e selos de cilindro. Estes artefatos vieram indubitavelmente do material minado em Badakhshan no norte do Afeganistão.

#### Curiosidades

É a pedra oficial do anel de formatura dos psicólogos, assim considerada a partir de 31 de março de 2006, pela resolução Nº 002/2006, do Conselho Federal de Psicologia brasileiro.

Uma alteração num único gene é capaz de transformar um vírus inofensivo no agente causador de uma gripe mortal. A descoberta, segundo pesquisadores da Uni-

versidade de Wisconsin-Madison (EUA), poderá levar à compreensão de como as epidemias da doença surgem e qual a melhor forma de combatê-las.

A equipe liderada por Yoshihiro Kawaoka analisou amostras do H5N1, o vírus influenza que, em 1997, matou 6 das 18 pessoas infectadas, em Hong Kong.

Essas amostras foram inoculadas em camundongos e, de acordo com os sintomas apresentados pelos animais, foram divididas em dois grupos.

O primeiro grupo causava uma forma de infecção letal, matando os camundongos, e o segundo provocava apenas problemas respiratórios como sintomas.

Para descobrir a razão dessa diferença, Kawaoka usou uma estratégia de "engenharia" dos vírus, trocando genes entre as linhagens virais. Desse modo, conseguiu identificar qual deles era o responsável pela maior capacidade de infecção do influenza.

"Uma mudança de apenas uma base [letra] no gene PB2 [que resultou na modificação de um aminoácido na proteína por ele codificada] parece ser a causa da virulência do influenza", explica.

Os cientistas ainda não sabem exatamente qual o papel do PB2, mas ele parece codificar uma enzima responsável pela indução de um número maior de partículas virais nas células infectadas.

O estudo do grupo de Kawaoka está na edição de hoje da revista científica "Science".

P: Então tu passaste uma infância muito sozinha?

Bem, na província, às vezes brincava com outros meninos filhos do . ou doutras pessoas do tribunal, ou da rua, até. Brincava. Ali, em Alenquer. Porque depois viemos, voltámos para Lisboa, o meu pai esteve então, se não me engano, como. chefe de gabinete do Ministro da Justiça, por isso, voltámos, que era o dr Catanho de Meneses. Voltámos para Lisboa e foi quando saímos de Santa Quitéria para a Rua Sampaio Pina. E na Rua Sampaio Pina, que era ao lado do Rádio Clube Português, a casa que ainda existe hoje e faz esquina para a rua Castilho, no primeiro andar, então alugámos essa casa, e aí então é que eu comecei foi a aprender o alemão com uma preceptora alemã, que era. que vinha três vezes por semana, dar-me lição de alemão. Não estava lá em casa, havia preceptoras internas, que estavam dentro, que eram internas, estavam lá, nas casas dos respectivos alunos, mas aquela senhora, a Kaethe Goethe, não. Vinha três vezes por semana dar-me a lição de alemão.

D: E tu gostavas?

Gostava! Tinha uns livros assim muito engraçados, que ainda hoje conheces, que conservo.

D: E ela era professora? Ela era professora mesmo, ou?

Era professora, quero dizer ela veio para Portugal dar aulas. Não sei se veio interna para uma casa e depois deixou, estava.

D: Mas enfim, não é uma senhora alemã que dava umas aulinhas para... para

Não, ela era uma senhora que tinha vindo da Alemanha só para ser professora, e tinha. tinha-se instalado numa casa que era o lar das. das professoras de alemão, que havia nessa altura, na Avenida, perto da Praça da Alegria. Lembro-me de ela lá me ter levado e tenho aí até fotografias de um cão, desse dito lar, a querer abraçar-me, coisa que eu não gostei nada, como continuo a não gostar de abraços de cão.

D: Mas já tinhas o Sirius...

Não, o Sirius veio muitíssimo mais tarde. Não, na Sampaio Pina não tinha o Sirius, não tinha cão nenhum.

D: Então tinhas a tua mãe, como professora de tudo o resto e a senhora

**Texto2**

Pois, o resto, tinha grandes amigas sempre. Constante, até era, mais até que as minhas primas. tinha as minhas primas, e tinha a Joaquina Sampaio e Melo, que era filha da maior amiga da minha mãe, da Lourinhã.

P: Então tu passaste uma infância muito sozinha?

Bem, na província, às vezes brincava com outros meninos filhos do... ou doutras pessoas do tribunal, ou da rua, até. Brincava. Ali, em Alenquer. Porque depois viemos, voltámos para Lisboa, o meu pai esteve então, se não me engano, como... chefe de gabinete do Ministro da Justiça, por isso, voltámos, que era o dr Catanho de Meneses. Voltámos para Lisboa e foi quando saímos de Santa Quitéria para a Rua Sampaio Pina. E na Rua Sampaio Pina, que era ao lado do Rádio Clube Português, a casa que ainda existe hoje e faz esquina para a rua Castilho, no primeiro andar, então alugámos essa casa, e aí então é que eu comecei foi a aprender o alemão com uma preceptora alemã, que era. que vinha três vezes por semana, dar-me lição de alemão. Não estava lá em casa, havia preceptoras internas, que estavam dentro, que eram internas, estavam lá, nas casas dos respectivos alunos, mas aquela senhora, a Kaethe Goethe, não. Vinha três vezes por semana dar-me a lição de alemão.

D: E tu gostavas?

Gostava! Tinha uns livros assim muito engraçados, que ainda hoje conheces, que conservo.

D: E ela era professora? Ela era professora mesmo, ou?

Era professora, quero dizer ela veio para Portugal dar aulas. Não sei se veio interna para uma casa e depois deixou, estava.

D: Mas enfim, não é uma senhora alemã que dava umas aulinhas para... para

Não, ela era uma senhora que tinha vindo da Alemanha só para ser professora, e tinha... tinha-se instalado numa casa que era o lar das... das professoras de alemão, que havia nessa altura, na Avenida, perto da Praça da Alegria. Lembro-me de ela lá me ter levado e tenho aí até fotografias de um cão, desse dito lar, a querer abraçar-me, coisa que eu não gostei nada, como continuo a não gostar de abraços de cão.

D: Mas já tinhas o Sirius.

Não, o Sirius veio muitíssimo mais tarde. Não, na Sampaio Pina não tinha o Sirius, não tinha cão nenhum.

D: Então tinhas a tua mãe, como professora de tudo o resto e a senhora

Pois, o resto, tinha grandes amigas sempre. Constante, até era, mais até que as minhas primas. tinha as minhas primas, e tinha a Joaquina Sampaio e Melo, que era filha da maior amiga da minha mãe, da Lourinhã.

Ali, só me lembro de ser eu. Mas talvez houvesse outros, da família Xara Brasil, que eram as pessoas ali principais, e eram realmente... não me lembro. Agora outros meninos em Lisboa, ou iam para as piscinas, de Algés e Dafundo, e outras coisas assim.

D: Mas conta lá... como é que aprendeste a nadar, que é engraçado

Aprendi a nadar, como digo, com esse velho lobo do mar que se tinha tornado banheiro da praia de S Bernardino; tinha um chapéu mole assim amachucado, um fato de macaco, e uma cabaça à cintura e ele puxav... agarrava-me assim no queixo, com a mão, e depois dizia-me os movimentos que eu devia fazer, nadar é claro, os bálticos, não é? E foi assim que eu aprendi.

D: ele atirava-te para a água com a cabaça?

Como? Não, a cabaça ele andava ali sempre com ela, como banheiro. era bóia, a cabaça servia de bóia

D: então ele atirava-te à água.

Não, ali à borda de água! à borda, primeiro começava-se à borda de água... mas com altura suficiente para eu já poder nadar. E ele pegava-me assim pelo queixo para me ajudar, e nadava de bruços, claro! nadava de bruços

D: Mas depois... depois... tornaste-te uma grande nadadora

Depois... na dita praia de S Bernardino, e depois na Praia das Maças, tornei-me uma grande nadadora é uma maneira um bocadinho... como hei-de dizer? pouco rigorosa de dizer. Eu era uma nadadora de muita resistência. De fundo, digamos. Mas sempre com aquele estilo, que passou de bruços para agulha, e que realmente me fazia avançar muito, mas eu não sabia outro estilo, nunca nadei de braçada, crawl.

D: Até longe!

S Bernardino não ia longe, mas... na Praia das Maças ia. Até tão longe que a minha mãe da praia não me via! Só que havia um barco, uma chata... um bocado... entre o sítio onde eu estava e a praia que era para nos mostrar, a mim e ao Caldeira Pires, que era... que éramos os únicos que tínhamos essa autorização para ir assim longe numa determinada direcção... em S Bernardino, a única avaria foi na Berlenga, também. Parece que estava fadada para as tais aventuras da Berlenga! Porque aí, na Berlenga, fomos numa excursão organizada pelo. Xara Brasil e por um senhor que era governador de Peniche ou qualquer coisa assim... ou presidente da Câmara, para homenagear uma figura nobre que era um arquiduque austríaco... e fizeram

D: Estava lá?

Estava, estava em Portugal, e foi lá parece, casar com uma senhora que era a rainha da Boémia e que... tinha um filho. Mas isto foi muito mais tarde, foi aos quinze anos... ou dezasseis, porqu nós fomos. Não íamos nunca seguidamente para um sítio, não é? íamos um aano para uma praia, outro ano para outra.

D: Ah, nunca iam para o mesmo sítio.

Não. Foi por isso que o meu pai nunca aceitou fazer uma praia na Areia Branca! Porque ele tinha tido a possibilidade de comprar terreno baratíssimo e fazer a segunda casa da Areia Branca! Ele não queria um sítio fixo que o obrigasse a.

HEIDEGGER e o 4-4-2: Foi, ontem, apresentado o livro "Liderança - As Lições de José Mourinho" de Luís Lourenço, baseado na sua tese de mestrado. Cada vez que leio isto (e faço-o várias vezes ao dia, na esperança de o vir a compreender), gosto de imaginar a seguinte conversa entre dois adeptos, à saída de Stanford Bridge:

Eh pá, o Chelsea não merecia ter empatado este jogo.

Pois não. Mas a primeira parte foi tão má que parecia que o Mourinho não tinha lido "O Ser e o Tempo".

Quantas vezes é que o Heidegger avisou que o 4-4-2 só resulta se o trinco compensar a subida dos laterais?

Exacto. Meu amigo, se é para ver a nossa equipa voltar ao pensamento cartesiano, não contes mais comigo para vir ao estádio.

Calma. Também tens que perceber que estava nevoeiro. Se calhar, por isso é que, durante a primeira parte, o Mourinho não conseguiu ver o Todo.

Segundo o autor, José Mourinho "é diferente por partir de um novo paradigma de pensamento, preconizado por Heidegger ou Morin, por exemplo, e o operacionalizar. Trata-se de um novo olhar para o Todo, que põe de parte a divisão e análise das partes e o pensamento cartesiano de há quatro séculos".

