
Medievalismo en Extremadura

Estudios sobre Literatura y Cultura Hispánicas
de la Edad Media

Jesús Cañas Murillo
Fco. Javier Grande Quejigo
José Roso Díaz (Eds.)

Medievalismo en Extremadura
Estudios sobre Literatura y Cultura Hispánicas
de la Edad Media



Cáceres
2009

MEDIEVALISMO en Extremadura : Estudios sobre Literatura y Cultura Hispánicas de la Edad Media / Jesús Cañas Murillo, Fco. Javier Grande Quejigo, José Roso Díaz (Eds.). — Cáceres : Universidad de Extremadura, Servicio de Publicaciones, 2009

XXII, 1310 pp. ; 17 × 24 cm.

ISBN 978-84-7723-879-9

1. Literatura medieval-historia y crítica. I. Cañas Murillo, Jesús (Ed.). II. Grande Quejigo, Javier (Ed.). III. Roso Díaz, José (Ed.). IV. Título. V. Universidad de Extremadura, Servicio de Publicaciones, ed.

82.09"04/15"

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra sólo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.



© Jesús Cañas Murillo, Fco. Javier Grande Quejigo y José Roso Díaz, de la edición, 2009

© De los autores, 2009

© Universidad de Extremadura-Grupo "Barrantes Moñino", para esta 1.ª edición, 2009

Ilustraciones de cubierta: miniaturas de cancioneros del siglo XIII (Biblioteca Vaticana y Biblioteca Nacional de Francia) recogidas en el libro de Martín de Riquer, *Vidas y retratos de trovadores. Textos y miniaturas del siglo XIII*. Barcelona, Círculo de Lectores-Galaxia Gutenberg, 1995.

Edita:

Universidad de Extremadura. Servicio de Publicaciones

Plaza de Caldereros, 2. 10071 Cáceres (España)

Tel. (927) 257 041; Fax (927) 257 046

E-mail: publicac@unex.es

<http://www.unex.es/publicaciones>

I.S.B.N.: 978-84-7723-879-9

Depósito Legal: M-52.674-2009

Impreso en España - *Printed in Spain*

Impresión: Dosgraphic, s. l.

ULISSES – UM AMBIENTE INTEGRADO DE DESENVOLVIMENTO (IDE) PARA ANOTAÇÃO E PROCESSAMENTO DE CORPORA

Natália Albino Pires
Universidade de Coimbra

1. INTRODUÇÃO

No âmbito da tese de doutoramento, propusemo-nos estudar um *corpus* romancístico constituído por 1721 textos que correspondem a versões de cerca de cem romances da tradição oral moderna portuguesa recolhidos em diferentes regiões e publicados por vários editores entre 1828 e 1960.

O nosso trabalho de investigação teve como objectivos principais a comprovação da especificidade do léxico e a constituição de um dicionário/vocabulário de formas flexionadas com a indicação do romance/versão em que ocorre cada forma. Consequentemente, necessitámos anotar todo o *corpus* uma vez que a comprovação das particularidades do seu léxico implicaria o estudo das classes de palavras Nome, Adjectivo, Verbo, Advérbio e Quantificador.

Assim, nesta comunicação pretendemos, por um lado, dar conta de como a especificidade dos nossos objectivos e as singularidades do nosso *corpus* nos obrigaram a construir uma nova aplicação informática que respondesse às necessidades do nosso trabalho de investigação e, por outro lado, descrever a aplicação *Ulisses*, o IDE (*Integrated Development Environment* – Ambiente Integrado de Desenvolvimento) desenvolvido especificamente para a anotação e análise do *corpus*.

1.1. Especificidade do *corpus* romancístico

As singularidades do nosso *corpus* advêm, em grande parte, das características específicas do género literário romance: texto poético com tiradas de versos maioritariamente monorrimos com rima assonante e cuja estrutura sintáctica apresenta inversões que resultam em topicalizações dos referentes e dos actantes.

Por seu turno, dado que «as regiões romancísticas não correspondem a entidades administrativas» (Araújo, 1998: 222), parte da especificidade do nosso *corpus* reside no facto de os textos originários de zonas fronteiriças nos surgirem recitados em castelhano, em galego, em mirandês, em português, em castelhano/português, em galego/português, em castelhano/mirandês, em mirandês/português e inclusivamente em castelhano/galego/português ou em castelhano/mirandês/português. Finalmente, outra das características fundamentais do *corpus* estudado diz respeito ao facto de nele se encontrarem versões de romances recolhidas em diferentes épocas e regiões do país, editadas ao longo de 132 anos com distintos critérios.

1.2. Anotação do *corpus*

Os traços que pretendíamos anotar relacionam-se directamente com os objectivos definidos para a investigação (constituição de um dicionário/vocabulário e comprovação da especificidade do léxico do *corpus*) e com a especificidade do género literário objecto de estudo. Assim, por forma a cumprir os objectivos investigativos definidos, tornou-se imprescindível:

- i. manter informações extra-textuais como a proveniência dos textos, o nome do seu primeiro editor, um código de classificação temática, um código que indique se o texto se encontra ou não contaminado, um código que indique o número de versão e um código para o romance;
- ii. atribuir uma anotação a cada *token* em fim de verso;
- iii. anotar a língua a que pertence determinado *token*, dando conta da sua ambiguidade uma vez que há palavras que graficamente são iguais em todas as línguas de contacto presentes nos textos: castelhano, galego, mirandês e português;
- iv. uma ferramenta de *full tagging*.

2. PROCURA DE FERRAMENTAS E APLICABILIDADE DOS ANALISADORES MORFOLÓGICOS AO CORPUS ROMANCÍSTICO

Nos últimos dez anos, aumentou significativamente o número de projectos e de aplicações no âmbito do Processamento de Linguagem Natural (PLN), contudo, de que tenhamos conhecimento, até ao momento, encontram-se desenvolvidos para PE unicamente quatro analisadores morfológicos ou morfossintácticos: o Jspell¹, o Palavroso², o VISL³ e o LX-Suite⁴.

Antes de optarmos pela arquitectura de uma nova aplicação, aceitando todos os riscos que tal decisão encerra, contactámos com os responsáveis por três projectos⁵ e testámos com um texto de amostra cada um dos três analisadores disponíveis para o PE. Por problemas de ordem pessoal, o director do projecto Jspell não pôde continuar a ajudar-nos e, embora os responsáveis pelos outros dois etiquetado-

¹ Nascido na Universidade do Minho e integrado no Projecto Natura, destaca-se por ser um projecto pioneiro e por ser o único *tagger* para o PE totalmente disponível *online* em *open source*. Na nossa óptica, os maiores contras do Jspell residem no facto de, para além de ser um projecto descontinuado, o seu dicionário ter sido construído com base no *corpus* do CETEMPúblico, desenhado a partir de excertos de textos do *Jornal Público*, e no facto de correr unicamente em Linux.

² Desenvolvido pelo Grupo de Linguagem Natural do INESC e, segundo as suas autoras (Barreiro et al. 1993), desenhado para atribuir maior ênfase às regras do que ao léxico, acabou por perder a função para a qual foi aquitectado, a de *tagger*, passando a funcionar actualmente como corrector ortográfico e não se encontrando a versão do *tagger* disponível *online*.

³ Desenvolvido na Noruega (<http://visl.sdu.dk/>).

⁴ Desenvolvido por uma equipa conjunta constituída por membros do Departamento de Informática da Faculdade de Ciências da Universidade de Lisboa e por membros do Centro de Linguística da Faculdade de Letras da Universidade de Lisboa. Segundo os seus autores, é um projecto ainda em desenvolvimento que futuramente possibilitará uma etiquetagem morfossintáctica e estará disponível para *download*.

⁵ Jspell (<http://natura.di.uminho.pt/~jj/pln/pln.html>), LX-Suite (<http://lxsuite.di.fc.ul.pt/>) e VISL (<http://visl.sdu.dk/>).

res⁶ tenham mostrado total disponibilidade para nos ajudar na etiquetagem dos textos, ambos *taggers* revelaram problemas técnicos que nos obrigaram a não poder utilizá-los com o nosso *corpus*. Quer o LX-Suite, quer o VISL foram desenhados para serem usados por linguistas e, portanto, não permitem a manutenção de informações extra-textuais que, no nosso caso, são fundamentais, pois sem elas não poderíamos constituir *subcorpora* para comparar os dados sobre o léxico. O LX-Suite apresenta, além disso, um outro problema. Tratando-se de uma ferramenta de *shallow processing of portuguese* não atribui qualquer informação morfossintáctica aos *tokens*, limita-se a atribuir-lhes apenas a categoria gramatical. Assim, excluímos imediatamente a possibilidade de o aplicar uma vez que, dadas as características do *corpus* romancístico e os objectivos do nosso trabalho de investigação, necessitávamos de uma aplicação de *full tagging*.

Do teste de etiquetagem que fizemos aos três etiquetadores⁷ por forma a avaliar a sua proficiência, detectámos um outro problema técnico: os etiquetadores possuem um dicionário preparado para reconhecer exclusivamente PE e, apesar de estudarmos a tradição oral moderna portuguesa, do nosso *corpus* não fazem parte apenas textos em português.

3. DESIGN E OBJECTIVOS DO ULISSES

Concluímos, por isso, que necessitávamos de uma aplicação que, por um lado, nos permitisse manter todas as informações extra-textuais de catalogação dos textos de que viríamos a necessitar após a etiquetagem e que, por outro lado, etiquetasse os textos semi-automaticamente para podermos, de forma manual, corrigir os erros de etiquetagem dos *tokens*. O Ulisses nasce, assim, da necessidade de criar uma aplicação integrada que, além de permitir *tokenizar* e etiquetar morfossintacticamente o *corpus*, nos permitisse:

- i. construir uma base de dados onde se pudessem catalogar e codificar os textos;
- ii. cruzar informação/dados do *corpus* e construir diversos *subcorpora* (organizados, no nosso caso, por romance, por tema, por área geográfica e por editor), a partir dos quais se poderia comprovar a existência ou não de léxicos particulares dentro do léxico geral;
- iii. criar um dicionário/vocabulário de formas flexionadas e de lemas presentes no *corpus* com a respectiva localização; no nosso caso, com a indicação do romance/versão em que surge;
- iv. editar os textos sem perder as anotações já efectuadas;
- v. eliminar ou acrescentar textos, entradas ao *lexicon*, *tags* e quaisquer outros campos considerados pertinentes para a análise do *corpus* em qualquer momento do processo de etiquetagem e sem necessidade de recorrer a um informático;

⁶ António Branco da Faculdade de Ciências da Universidade de Lisboa é o responsável pelo LX-Suite e Eckhard Bick da Universidade de Southern (Dinamarca) o director do projecto VISL.

⁷ Do teste que fizemos a cada um dos três etiquetadores, constatámos que, para o nosso *corpus*, o *tagger* com maior proficiência é exactamente o VISL, paradoxalmente um projecto desenvolvido por investigadores não portugueses.

- vi. escolher autonomamente o conjunto de *tags*;
- vii. corrigir os erros de etiquetagem detectados;
- viii. intergrar novas ferramentas.

Em síntese, na base do projecto, estava o desejo de criar uma aplicação que permitisse utilizar de forma integrada e num mesmo ambiente de trabalho as ferramentas de que necessitávamos: um analisador morfossintáctico, um editor de texto, um módulo de *corpus query*, uma ferramenta para catalogação e atribuição de anotações aos textos e, não menos importante, uma ferramenta que possibilitasse rever os textos para corrigir erros de etiquetagem.

3.1. Arquitectura do Ulisses

O Ulisses caracteriza-se, antes de mais, por ser um IDE (*Integrated Development Environment*) que proporciona um *interface* versátil e sofisticado e que reúne num único ambiente de trabalho todas as ferramentas que permitem ao investigador introduzir, editar, catalogar, anotar, processar e analisar *corpora*. Ancorado numa estrutura modular, admite a integração de novas ferramentas ou funcionalidades que não tenham sido contempladas originalmente, podendo, portanto, adaptar-se facilmente para corresponder às exigências específicas de uma determinada área de investigação.

Enquanto IDE e na sua filosofia, pode comparar-se ao projecto GATE (General Architecture for Text Engineering), que se pode consultar em < <http://gate.ac.uk/>>, e ao projecto Ellogon, que parte da proposta do GATE e se pode consultar em < www.ellogon.org/>⁸.

No que se refere a pormenores técnicos, desenvolveu-se em C#, tem como motor de base de dados o SQLite, requiere o .NET Framework 2.0, corre em Windows XP e importa e exporta a informação e a meta-informação do *corpus* em formato XML⁹.

3.2. Módulos já existentes

O Ulisses conta com diferentes módulos: um *tag manager*, um *attributes manager* e um *annotation manager*. Conta com: – um editor de textos; – um *corpus manager*; – com editores para o *lexicon* e para os atributos dos textos. Possui um módulo de *corpus query* e outro de dicionário. Além disso, conta com uma busca simples/avançada, um *tokenizer* e um *tagger* simples, podendo o investigador optar por tokenizar e etiquetar os *tokens* tanto automaticamente como manualmente.

⁸ Não utilizámos nenhum destes projectos porque o GATE não possui um *tagger* que reconheça o português e quando começámos a análise do nosso *corpus* o Ellogon ainda se encontrava em desenvolvimento. Para além disso, as linguagens de programação utilizadas tanto pelo GATE como pelo Ellogon foram um *handicap*, já que o informático que desenvolveu o Ulisses não dominava nem o Java nem o TeL.

⁹ Não dominamos nenhuma das linguagens de programação que nos permitissem a execução do projecto, por isso, contámos com a ajuda de um informático que, gratuitamente, desenvolveu o Ulisses.

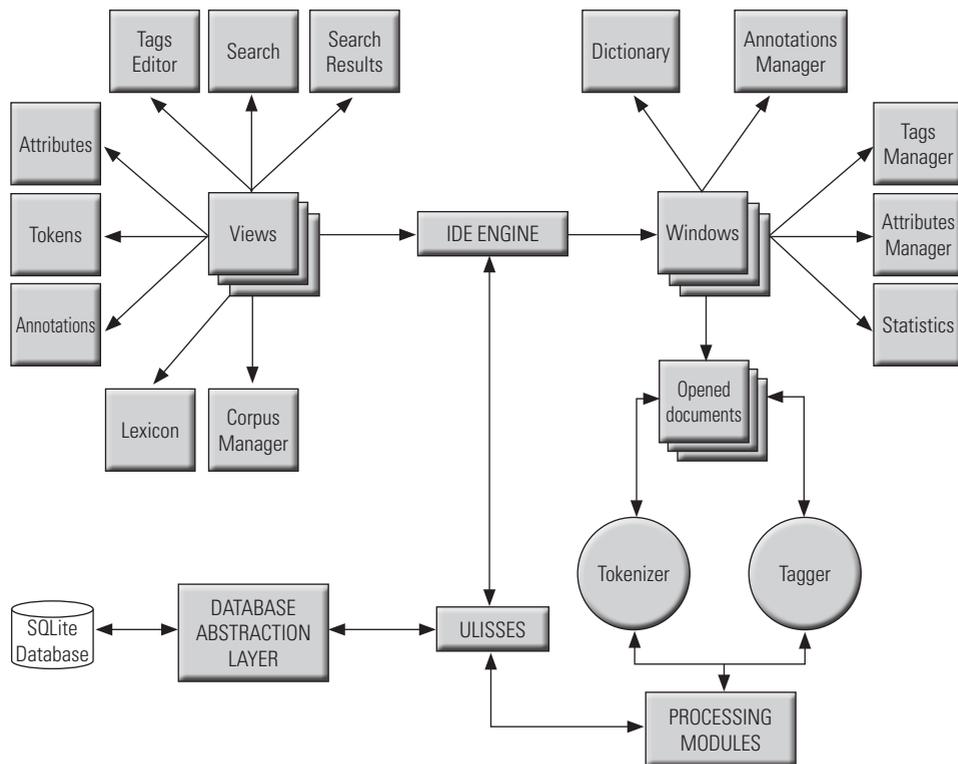


Figura 1. Diagrama funcional do programa.

3.2.1. Tag manager, attributes manager e annotation manager

No Ulisses, o utilizador tem, sempre que o deseje, total liberdade de de:

- i. no *tag manager*, escolher e definir as *tags* e as suas relações hierárquicas, as cores a atribuir (ou não) a cada *tag* e a terminologia linguística que pretende adoptar, podendo acrescentar, apagar ou reorganizar o etiquetário;
- ii. no *attributes manager*, determinar, acrescentar ou apagar os campos de catalogação que considerar adequados ao estudo do seu *corpus*;
- iii. no *annotation manager*, criar, alterar ou apagar todos os campos de anotação que considerar relevantes para o seu *corpus*¹⁰.

3.2.2. O editor de textos

No editor de textos, podem realizar-se todas as tarefas permitidas e necessárias num editor de textos: eliminar, acrescentar e copiar texto ou partes de texto; fazer

¹⁰ Para o *corpus* romancístico que estudámos, considerámos fundamental criar no *annotation manager* campos que nos permitem indicar os lemas de cada palavra léxica (na acepção de Coseriu, 1987), dar conta das formas em final de verso e também das palavras em castelhano, galego e mirandês.

undo e *redo*; seleccionar todo um texto ou partes de texto; imprimir e pre-visualizar a impressão. Além disso, permite visualizar e seleccionar *tokens* ou seleccionar palavras e convertê-las em *tokens*.

3.2.3. Módulo de pesquisa

Tendo como principal objectivo permitir que o processo de desambiguação e de correção sejam mais rápidos, o módulo de pesquisa permite que o investigador procure no *corpus* palavras, *tokens*, lemas, *tags* e outro tipo de anotações. O resultado da pesquisa apresenta-se em forma de listagem com a possibilidade de, a partir dela, se aceder directamente ao texto onde ocorrem as formas procuradas: basta clicar na respectiva palavra que se pretende ver.

3.2.4. O Tokenizer

Como já afirmámos, o algoritmo do *tokenizer* é bastante simples. O *tokenizer* automático separa os espaços entre palavras, os sinais de pontuação e considera também como um único *token* todas as formas da língua que se separam por hífen, com excepção dos pronomes pessoais adjuntos a verbos. O investigador pode, no entanto, optar por tokenizar manualmente um texto ou uma determinada palavra, corrigindo possíveis erros do processo automático.

3.2.5. O Tagger

Cremos importante deixar claro que o nosso objectivo nunca se centrou no desenvolvimento de um *tagger*, mas sim no estudo do léxico do *corpus* com o auxílio de uma aplicação informática, muito embora para tal tenhamos necessitado de um *tagger*.

O analisador morfossintáctico automático do Ulisses baseia-se num algoritmo probabilístico muito simples: a um *token* que possa receber duas ou mais etiquetas, atribui-se-lhe a etiqueta que tenha sido atribuída mais vezes ao longo do *corpus* etiquetado. É óbvio que este algoritmo apresenta um número significativo de erros, mas após a etiquetagem reve-se o texto para corrigi-los.

Com o propósito de que o processo seja mais rápido e para que a interacção com o programa seja o mais prática possível, o utilizador pode aceder através do teclado às funcionalidades mais frequentes para rever todos os *tokens* e toda a etiquetagem, podendo atribuir manualmente a um determinado *token* uma nova etiqueta.

Em relação ao *tagger* e dado que a estrutura do Ulisses se encontra preparada para receber novos módulos, cremos, e desejamos, que rapidamente se lhe pode integrar um novo analisador que apresente um alto nível de acertos.

3.2.6. Módulo de corpus query

O módulo de *corpus query* permite a extracção de informação quantitativa do *corpus* anotado e a sua apresentação em tabelas, a partir das quais é possível consultar directamente a informação ou exportá-la para outros programas, para posterior proces-

samento ou análise (ex. SPSS ou Excel), ou para a elaboração de documentos (ex. em Word).

Este módulo admite que o investigador defina as *queries* para a extracção da informação que considera relevante para o estudo do seu *corpus* e fornece, ainda, um gestor de *queries* que permite criar e manter uma lista de *queries* organizada por categoria e descrição, para a sua fácil e rápida reutilização.

Apesar de as *queries* serem escritas como expressões em linguagem SQL e, portanto, serem difíceis de criar para o investigador que não tenha conhecimentos de informática, o módulo proporciona um conjunto de *queries* pré-definidas que podem servir de inspiração e como uma base para a criação de novas *queries*, alterando simplesmente alguns dos parâmetros.

3.2.7. Módulo de dicionário

A partir dos critérios definidos para a anotação dos *tokens* do *corpus* ou dos atributos de catalogação dos textos, este módulo permite gerar um dicionário/vocabulário de lexemas ou de lemas (tanto por ordem alfabética como por ordem decrescente de frequência) com a sua classificação gramatical e respectivos contextos de ocorrência no *corpus*.

Um interface simples e intuitivo permite ao investigador restringir o conteúdo do dicionário às partes do *corpus* que considerar relevantes, através da especificação de variados critérios de filtragem baseados nos atributos de catalogação dos textos ou nas anotações aplicadas aos *tokens* do *corpus*.

A apresentação do dicionário/vocabulário é feita num formato de fácil leitura, pensado especialmente para a impressão em papel. Através de um processador de texto (como o Word) será ainda possível formatar o dicionário em múltiplas colunas para uma apresentação mais condensada e semelhante a um dicionário convencional. A utilização do dicionário para posterior processamento por parte de outros programas foi também contemplada, podendo o mesmo ser exportado em formato universal XML.

3.3. Ambiente de Trabalho

Uma das particularidades do ambiente de trabalho do Ulisses é a sua interactividade, que resulta do facto de o utilizador poder ver os textos do *corpus* organizados por atributos que define numa estrutura hierárquica de até três níveis, os quais pode voltar a organizar.

Com o objectivo de que o processo de correcção seja mais rápido, o programa permite a visualização das etiquetas atribuídas:

- i. com a coloração do texto e através de outros guias visuais (como o sublinhado ou o negrito dos *tokens* etiquetados);
- ii. através de uma lista (paralela ao *lexicon*) dos *tokens* do texto com o lema e respectivas etiquetas;
- iii. através de uma janela interactiva na qual se podem ver e alterar as etiquetas de um *token* previamente seleccionado no texto.

Da sua interactividade destaca-se, por último, o facto de o utilizador poder organizar as janelas do *interface* com total liberdade, podendo colocá-las na disposição que considerar mais conveniente: minimizando-as, redimensionando-as ou ocultando-as.

4. CONSIDERAÇÕES FINAIS

Do exposto, parece-nos que o Ulisses, comparativamente com outras aplicações disponíveis, se apresenta como uma ferramenta bastante poderosa em virtude de poder utilizar-se tanto no âmbito dos estudos linguísticos como no âmbito dos estudos filológicos, podendo ser aplicado a quaisquer *corpora*.

No entanto, no âmbito dos estudos linguísticos reconhecemos que, actualmente, o maior problema do Ulisses se encontra no tempo empregue na etiquetagem do *corpus* e no tempo gasto na construção do *lexicon* a partir do qual se etiquetará automaticamente o *corpus*, sempre e quando o investigador não opte por importar um *lexicon* já construído por outros investigadores. Por isso, o nosso desejo é o de que algum dia o Ulisses venha a integrar um *tagger* com melhor performance.

Não obstante, cremos que as vantagens do programa Ulisses as encontramos:

- i. na sua aplicabilidade a todo o tipo de *corpus*, permitindo manter todas e quantas informações extra-textuais o investigador considere necessárias para o estudo do seu *corpus*. Ou seja, através do Ulisses pode estudar-se o léxico de um autor, o léxico de vários autores, o léxico de diferentes géneros literários, o léxico de uma ou mais publicações periódicas contemporâneas ou de diferentes épocas, etc.;
- ii. no facto de se tratar de uma estrutura modular à qual, a qualquer momento, se podem acrescentar novos módulos, como um gestor de concordâncias, outro módulo de *POS-Tagging* automático, um *Lemmatizer* automático, um módulo de *Data Mining*, um módulo de análises estatísticas, etc.;
- iii. no facto de ser uma aplicação interactiva que permite que cada investigador defina, modifique, acrescente ou elimine os campos de catalogação, o etiquetário e as anotações, ajustando-as à especificidade do seu *corpus* e aos objectivos do seu trabalho de investigação;
- iv. no facto de que se pode reutilizar um *lexicon* já constituído por outro investigador como construir um novo *lexicon* a partir de um novo *corpus*;
- v. no facto de que permite construir um dicionário/vocabulário (por ordem de ocorrências ou por ordem alfabética) de palavras léxicas e/ou palavras gramaticais com a indicação das respectivas ocorrências em cada texto do *corpus*;
- vi. no facto de o programa possuir um interface muito intuitivo;
- vii. por arquivar toda a informação e meta-informação em base de dados;
- viii. e, por fim, pelo facto de reduzir substancialmente a dependência do investigador em relação a um informático. Ou seja, com esta aplicação o investigador planeia e gere no *interface* todo o seu trabalho, necessitando do informático unicamente para o desenvolvimento de novos módulos.

BIBLIOGRAFIA

- Afonso, Susana; Bick, Eckhard; Haber, Renato e Santos, Diana: «Floresta Sintá(c)tica: um treebank para português», in Anabela Gonçalves e Clara Nunes Correia (org.), *Actas do XVII Encontro da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística, 2002, pp. 533-545.
- Araújo, Teresa: «Casada em Terras Longínquas no Baixo Alentejo em confronto com outras tradições atlânticas e mediterrânicas», *Arquivo de Beja*, Série III, VII/VIII, 1998, pp. 221-227.
- Barreiro, Anabela; Pereira, M^a de Jesus y Santos, Diana: *Critérios e Opções Linguísticas no Desenvolvimento do Palavroso, um Sistema Computacional de Descrição Morfológica do Português*, Grupo de Linguagem Natural do INESC, Relatório INESC n^o RT/54-93, <www.linguateca.pt/diana/download/criterios.ps>, 1993, pp. [1-39].
- Bick, Eckhard: «Automatic parsing of portuguese», in Laura Sánchez García (ed.), *Anais do II Encontro para o Processamento Computacional do Português Escrito e Falado*, Curitiba, CEFET-PR, 1996, pp. 91-100.
- Correia, Margarita: «Terminologia e Lexicografia Computacional», in *Jornada Panllatina de Terminologia*, Barcelona, IULA/Universidade Pompeu Fabra, 1996, pp. 83-91.
- Coseriu, Eugenio: *Gramática, Semántica, Universales*, Madrid, Gredos, 1987.
- Oksefjell, Signe y Santos, Diana: «Breve panorâmica dos recursos de português mencionados na Web», in Vera Lúcia Strube de Lima (ed.), *Anais do 3^o Encontro de Processamento da Língua Portuguesa Escrita e Falada, PROPOR'98*, Porto Alegre, 1998, pp. 38-47.
- Pires, Natália Albino: «O léxico dos romances carolíngios da Tradição Oral Moderna portuguesa editados entre 1828 e 1960: uma amostra», in Ana Sofia Laranjinha e José Carlos Miranda (eds.), *Modelo – Actas do V Colóquio da Secção Portuguesa da Associação Hispânica de Literatura Medieval*, Porto, Faculdade de Letras da Universidade do Porto, 2005, pp. 231-242.
- : «Verbos e tempos verbais nos romances carolíngios da tradição oral moderna portuguesa, editados entre 1828 e 1960», in *Actas do XI Congreso de la Asociación Hispánica de Literatura Medieval*, 2005, no prelo.
- Ranchhod, Elisabete Marques: «Dicionários Electrónicos e Análise Lexical Automática», in Palmira Marrafa e M^a Antónia Mota (org.), *Linguística Computacional – Investigação Fundamental e Aplicações*, Lisboa, APL/Edições Colibri, 1999, pp. 207-233.
- Rocha, Paulo; Simões, Alberto Manuel y Almeida, José João: «Cálculo de frequências para entradas de dicionários através do uso conjunto de analisadores morfológicos, taggers e corpora», in Anabela Gonçalves e Clara Nunes Correia (org.), *Actas do XVII Encontro da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística, 2002, pp. 407-418.