

## **PRONTUÁRIO MULTIMÉDIA: DICIOMÉDIA**

### **1. Introdução**

O proutuário Multimédia da língua portuguesa, no domínio dos Verbos (“DICIOMEDIA”), desenvolve-se no âmbito de um projecto de I&D financiado pela JNICT (PLUS/C/PDP/1175/95). Nesta primeira fase, pretende-se apenas a elaboração de um protótipo, que, depois, será alargado a todos os verbos da língua e, numa fase posterior, a todas as categorias da língua portuguesa.

São várias as equipas que participam neste projecto e pertencem a diversas instituições:

- Faculdade de Letras da Universidade do Porto, responsável pela componente linguística,
- Departamento de Modelação e Simulação de Processos do Instituto de Tecnologias de Informação do INETI e Departamento de Informática da Universidade Nova de Lisboa, responsáveis pela componente informática,
- Associação Portuguesa de Tradutores e União Latina, responsáveis pelo acompanhamento do desenvolvimento do protótipo e pela elaboração de recomendações relativas à funcionalidade da interface à medida que ela vai sendo construída.

O trabalho de desenvolvimento informático tem ainda proporcionado temas de projecto para alunos finalistas dos cursos de Informática das Faculdades de Ciências da Universidade de Lisboa e da Universidade Nova de Lisboa. O trabalho directamente ligado à linguística serve igualmente de tema e campo experimental para Mestrandos e Mestres da Faculdade de Letras da Universidade do Porto.

## 2. Produção do protótipo

O diagrama de produção do protótipo DICIOMEDIA apresenta-se esquematicamente na figura 1.

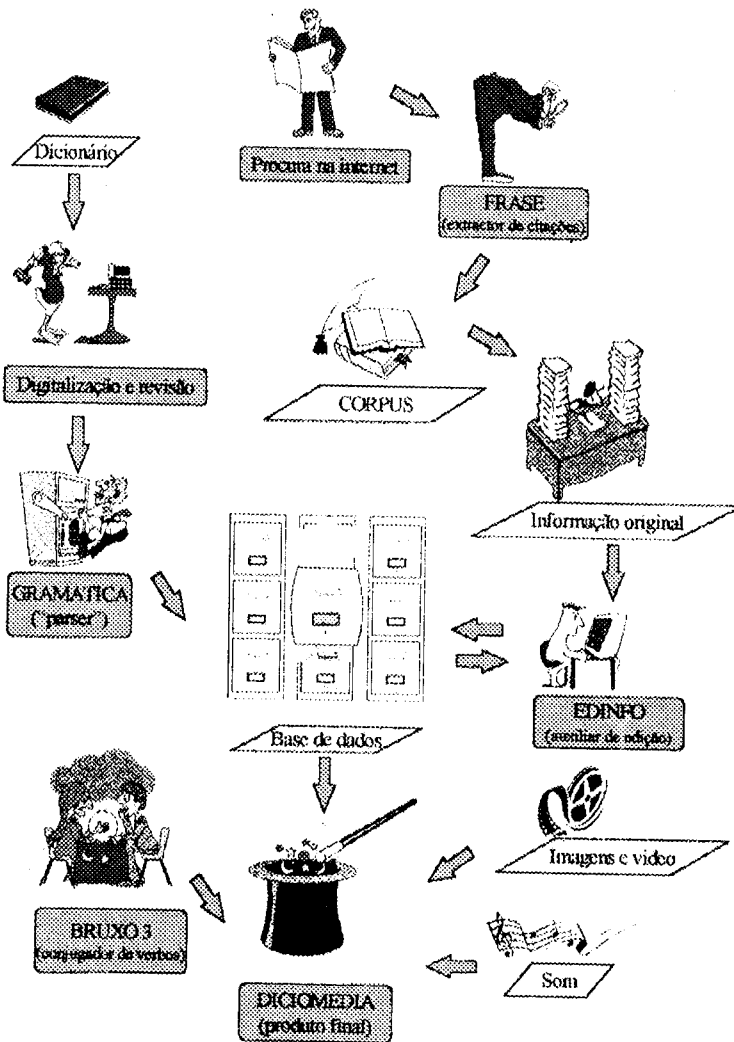


Figura 1 — Diagrama de produção do DICIOMEDIA

O conteúdo é gerado por duas vias: a partir de um Dicionário Básico, publicado por um membro da equipa e sobretudo a partir da informação primária obtida e tratada pelos especialistas da linguística. Os passos percorridos ( e a percorrer) são os seguintes:

- a primeira via inclui a:
  - digitalização
  - revisão por aplicação de um *OCR* (Omnipage)
  - revisão por inspecção visual com vista a eliminar caracteres incorrectamente digitalizados e entradas não verbais
  - compilação por um *parser* (GRAMATICA)
  - exportação para uma base de dados, previamente desenhada em MS Access, que constitui a estrutura de dados da aplicação multimédia;
- a segunda via contempla a/o:
  - desenvolvimento do conteúdo (novas entradas ou completação de entradas do Dicionário Básico) com base em fontes primárias, pelos especialistas da linguística
  - construção de um CORPUS temático (Economia, Ciência, Cultura, Desporto, etc), como complemento das fontes primárias, com base em fontes seleccionadas da imprensa, captadas via Internet
  - extracção automática de citações de verbo(s) e sua ordenação alfabética, por aplicação da ferramenta FRASE
  - carregamento da informação primária e validação dos campos da base de dados através do EDINFO
  - desenvolvimento do protótipo multimédia sobre o ambiente de desenvolvimento ToolBook da Asymetrix
  - geração automática das flexões verbais via BRUXO 3.0, desenvolvido em MS Visual C++
  - selecção e integração de som e imagem.

### 3. Ferramentas Auxiliares

A necessidade de minimizar o esforço de mão-de-obra e de aumentar a fiabilidade da informação inserida na base de dados levou-nos ao desenvolvimento das seguintes ferramentas auxiliares:

FRASE — ferramenta desenvolvida em MS Visual Basic utilizada para a extracção automática de frases que contenham citações de verbos (ou outros *tokens*) pré-definidos e sua ordenação (alfabética por verbo ou outro critério); não apresenta restrições quanto ao número de ficheiros de texto a pesquisar, nem quanto ao número de *tokens*.

GRAMATICA — *parser* (compilador sintáctico) desenvolvido em Lex e Yacc, que reconhece o modelo subjacente à descrição do verbo e que permite a captação das respectivas estruturas e sua exportação e preenchimento automático dos campos da base de dados.

EDINFO — conjunto de *forms* (formulários) em MS Access Basic, para entrada e validação directa de informação na base de dados (ver figura 2).

BRUXO 3 — conjugador universal de verbos da língua portuguesa em *run-time* (em tempo de execução do programa), desenvolvido em MS Visual C++, que integra todos os modelos de flexão conhecidos da língua portuguesa e que pode ser utilizado como corrector ortográfico; na versão 3, é alargada a sua aplicação à conjugação pronomial e reflexa.

### 4. Algumas características da interface de utilização

Na actual fase do trabalho a interface reflecte a estrutura do modelo de verbo tal qual se apresenta no Dicionário Básico.

Os desenvolvimentos que actualmente estamos a fazer, a nível linguístico, apontam para uma significativa complexificação do modelo verbal a adoptar, o que terá reflexos na versão final do protótipo.

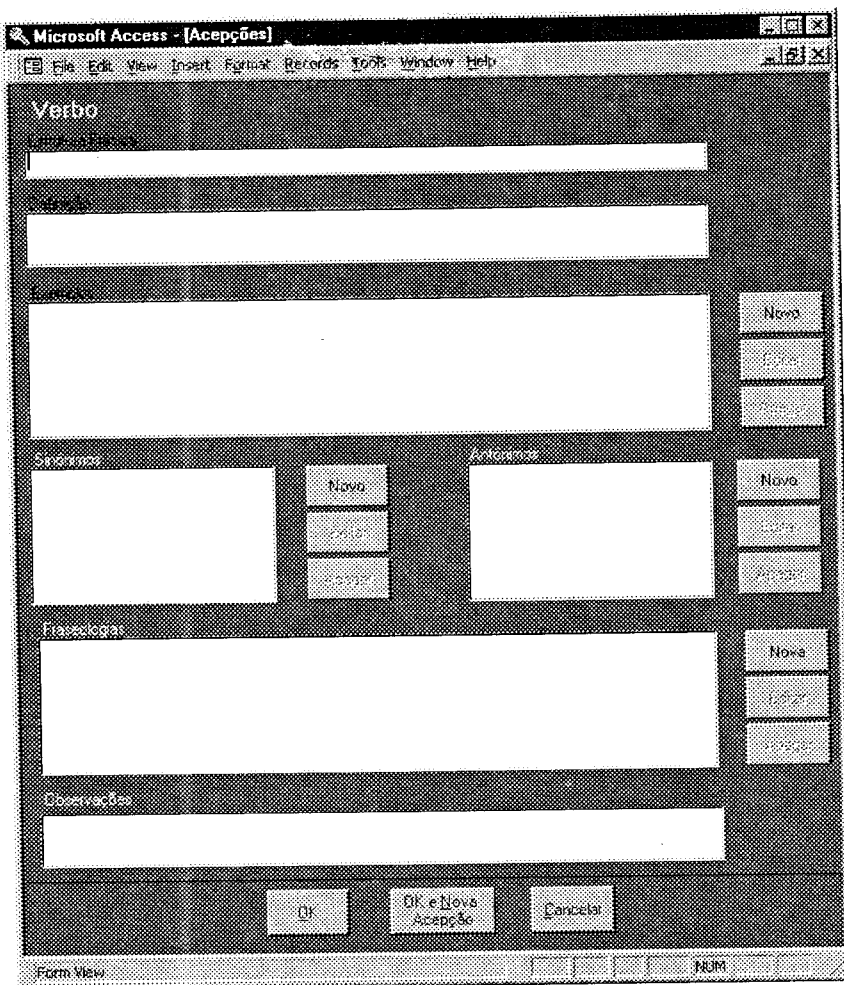


Figura 2 — Ilustração do EDINFO

Na figura 3 apresenta-se a versão actual da interface geral de utilização. Algumas das suas características tornam-se evidentes a uma simples observação. Assim, a cada verbo estão associados sentidos e cada sentido está conectado com exemplos de utilização e uma explicação. É possível obter subconjuntos da lista global de verbos, por aplicação de um “filtro” de categorias.

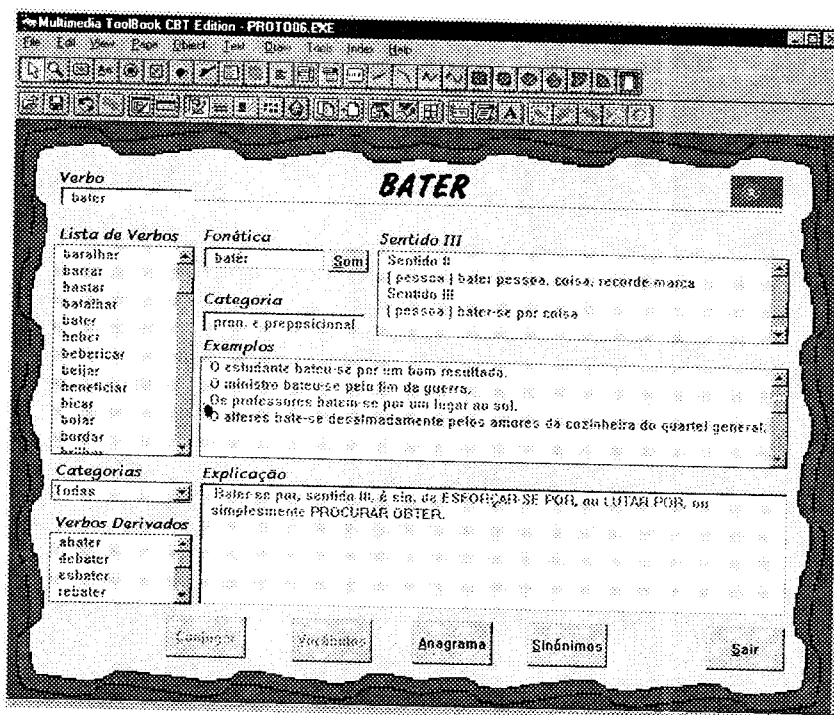


Figura 3 — Interface Geral de Utilização do Protótipo do Prontuário Multimédia de Verbos (DICIONEMEDIA)

Está igualmente prevista a indicação da listagem de verbos derivados por afixação e a complementarização da explicação por meio de sinónimos. Eventualmente, apresentar-se-ão vocábulos associados e anagramas.

A conjugação verbal é visualizada noutra interface, cujo modelo se apresenta na figura 4.

A transcrição fonética é visualizada (na figura 3, os símbolos apresentados não foram corrigidos após digitalização) e associada a um “botão” que possibilita a vocalização da leitura.

A variante do português do elemento em análise, relativamente ao País lusófono, é identificada pela bandeira respectiva, no canto superior direito do écran.

Multimedia ToolBook CBT Edition - PROTOBS.EXE

Arq. Edit. Ferr. Ajuda

## BEIJAR

*Indicativo*

	Presente	Passado imperfeito	Passado perfeito	Passado mais perfeito	Futuro Imperfeito
Eu	beijo	beijava	beije	beijara	beijarei
Tu	beijas	beijavas	beijaste	beijaste	beijarás
Ele / Ela	beija	beijava	beijou	beijara	beijará
Nós	beijamos	beijávamos	beijamos	beijamos	beijaremos
Vós	beijais	beijáveis	beijastes	beijastes	beijareis
Eles / Elas	beijam	beijavam	beijaram	beijaram	beijarão

*Condicional*

Eu	beijaria
Tu	beijarias
Ele / Ela	beijaria
Nós	beijaríamos
Vós	beijaríeis
Eles / Elas	beijariam

*Conjuntivo*

	Presente	Passado imperfeito	Futuro Imperfeito
Eu	beijasse	beijasse	beijaria
Tu	beijasses	beijasses	beijarias
Ele / Ela	beijasse	beijasse	beijaria
Nós	beijássemos	beijássemos	beijaríamos
Vós	beijásseis	beijásseis	beijaríeis
Eles / Elas	beijassem	beijassem	beijariam

*Imperativo*

Eu	
Tu	beija
Ele / Ela	beije
Nós	beijemos
Vós	beijai
Eles / Elas	beijem

*Infinitivo*

	Pessoal	Impessoal
Eu	beijar	beijar
Tu	beijares	
Ele / Ela	beijar	
Nós	beijarmos	
Vós	beijardes	
Eles / Elas	beijarem	

*Participio Passado*

	Pessoal	Impessoal
Eu	beijado	

*Gerúndio*

beijando
----------

Voltar

Figura 4 — Interface de Conjugação de Verbos

De acordo com critérios bem definidos, estarão associadas ilustrações gráficas e sonoras (os respectivos botões de acesso não se encontram ainda incluídos) a sentidos de verbos ou explicações.

INETI / DMS e Faculdade de Letras (U.P.), 8.Julho.1997

Augusto Q. Novais / Mário Vilela