

# Dicionário aberto de Calão e Idiomáticas

J. João Dias de Almeida      Alberto Simões  
Pedro Rangel Henriques

2000

## Conteúdo

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Motivação e características da área</b>	<b>2</b>
<b>3</b>	<b>Ambiente de desenvolvimento usado – dpl</b>	<b>3</b>
<b>4</b>	<b>Abreviaturas e funções</b>	<b>4</b>
<b>5</b>	<b>Definição orientada ao conceito</b>	<b>6</b>
<b>6</b>	<b>Meta-informação</b>	<b>8</b>
<b>7</b>	<b>Criação duma CGI de consulta</b>	<b>9</b>
<b>8</b>	<b>Conclusões</b>	<b>11</b>

Nesta secção, descreve-se uma experiência que tem vindo a decorrer no âmbito do projecto Natura, envolvendo a criação de um dicionário aberto de calão e de expressões idiomáticas (DAC), consultável e extensível via Internet. Trata-se de procurar um compromisso aceitável entre o gosto de coleccionar e partilhar, com uma vontade de obter qualidade, equilíbrio e coerência da informação.

## 1 Introdução

O desenho de um dicionário é uma tarefa reconhecidamente complexa, quer devido à enorme quantidade de informação que tem que ser processada, quer ao nível da delicada coerência e equilíbrio, tão difícil e trabalhosa de conseguir.

Um projecto de construção de um dicionário pode corresponder a situações e objectivos muito díspares.

No projecto DAC considerou-se importante o permitir que o dicionário estivesse *no ar* logo desde o início, de modo a prestar um serviço e funcionar como ponto de partilha e intercâmbio com contribuições externas. Ou seja, decidiu-se proceder à construção de um dicionário *em directo, à vista*.

Sabe-se que um dicionário não está (nem nunca vai estar!) acabado e defende-se que um dicionário que esteja em construção, com elevadas assimetrias e entradas de completude muito díspares, pode ainda ser um trabalho válido e constituir um recurso útil.

## 2 Motivação e características da área

A área particular das expressões idiomáticas [?] e do calão [?] tem um interesse especial e envolve um rico património linguístico cujo registo é complexo.

Saliente-se no entanto que:

- ameaças infantis tão ricas como “*levas um biqueiro no céu da boca que ficas com a tosse nos calcanhares*” são tão contundentes que, por certo, vencem uma disputa antes de esta se iniciar;
- expressões de calão tão fortes como “*falhar por um pentelho seco de velha*”<sup>1</sup> ou “*confundir o olho do cu com a feira de Montemor*”<sup>2</sup>, têm uma riqueza expressiva notável;
- o registo de expressões como estas, usando uma postura arquivística, é uma tarefa de inegável utilidade patrimonial.

---

<sup>1</sup>Usada no Minho, com o sentido de *falhar por muito pouco*.

<sup>2</sup>Centro-Sul de Portugal.

Esta área constitui um domínio muito rico, um filão inesgotável mas ao mesmo tempo apresenta um conjunto de dificuldades e desafios ligados a:

- envolver termos multipalavra, com variantes flexionadas
- os seus elementos serem pouco normalizados, tanto pela existência dum grande número de variantes, como pelas dificuldades na grafia de termos que raramente aparecem escritos.
- haver um grande número de equivalências semânticas
- haver um grande dependência do contexto:
  - quem os usa
  - situação de uso, época
  - nível de linguagem
- envolver grandes dificuldades de catalogação.

### 3 Ambiente de desenvolvimento usado – dpl

Dadas as condicionantes atrás referidas, houve que otimizar o processo de construção do dicionário que passou a ser definido/mantido na linguagem de programação de dicionários – dpl descrita na Secção ??.

Esta linguagem é orientada ao conceito (definido através de conjuntos de sinónimos) e com possibilidade de usar notação específica para abreviar e evidenciar partes da informação e fenómenos tão variados como:

- origem etimológica
- origem geográfica
- exemplos e paráfrases
- classes a que pertence
- variantes (especialmente importantes em relação a termos multipalavra)
- adivinhas e provérbios
- trocadilhos fonéticos
- etc

Como referido atrás, o `dpl` permite: fazer definições de conceitos independentemente do seu uso; dispor de geração de vários tipos de saídas; funcionar de um modo *declarativo*; manter informação *etiquetada* ou facetada; definir e usar funções.

A partir desta notação, e usando o compilador da linguagem de programação `dpl`, é gerado todo o sistema de consulta por rede que contém a informação necessária à inserção de termos desconhecidos ou à completação da informação associada.

O sistema de pesquisa – baseado no módulo `ddmf` (ver Sec. ??) contempla mecanismos para procurar termos e seus derivados morfológicos.

## 4 Abreviaturas e funções

Um dos aspectos importantes da linguagem `dpl`, é a capacidade de definição de funções, ou seja de poder definir os conceitos metalinguísticos a usar nas definições de termos. Esses conceitos vão desde a simples abreviatura, até construtores que criam mais que um termo.

As funções ajudam à coerência geral e a evidenciar certos fenómenos linguísticos e correspondem de um modo geral a construtores usados na composição de novos termos.

Alguns exemplos de definição de abreviaturas:

```
sub ptn { +{local => 'norte de Portugal'}}
sub n2  { +{nivel => 'coloquial'} }
sub n4  { +{nivel => 'calão carroceiro'} }
sub sm  { +{genero => 'masculino',
           gram    => 'só usado no masculino'} }
sub diabo { +{sem => 'ordem de não aborrecer e de se ir embora',
             cat  => 'interjeição'} }
```

As abreviaturas são, portanto, funções zero-árias que devolvem estrutura de facetas.

Uma função com parâmetros, pode corresponder a uma abreviatura mais complexa, ou pode já descrever conceitos metalinguísticos mais interessantes.

Considere-se o seguinte exemplo de definição de funções com parâmetro:

```
sub seems{ +{ isa    => 'eufemismo por semelhança sonora',
             syn     => shift } }
```

<dict>

```
Júlio++seems(chulo)+"esse? é cá um júlio"++pt;;
</dict>
```

Notas:

**linha 1 2** - **Seems** é uma função que recebe um termo e devolve uma estrutura de facetas em que o parâmetro é posto como sinónimo.

**linha 4** - Na definição do termo *Júlio* é acrescentado a estrutura de facetas referida.

A função **seems** está a constituir uma representação explícita de um construtor linguístico muito usado em calão que corresponde a substituir um termo calão de nível elevado por um outro termo que tem semelhanças a nível fonético e que não é calão. Deste modo, está-se a constituir um novo termo calão de nível menos elevado (e portanto possível de utilizar em situações diferentes).

A escrita de funções como esta, ajuda a clarificar conceitos metalinguísticos e a tornar mais eficaz o processo de descrição dos termos.

Seguidamente, apresenta-se exemplos de entradas usando 3 funções habituais em expressões idiomáticas e calão, mostrando-se em seguida o resultado de consultar o dicionário especificado.

```
+trocfon(gay ; entreguei+{sem = estar cercado de homossexuais});
+adivinha([pi;piolho;olho];
    qual é o animal que tem mais que três olhos e menos que quatro?;
    piolho);
taveirada+"ela só pensa em taveiradas..."++n3++pt
++afterevent(1986?; termo que apareceu após a circulação de um
vídeo clandestino documentando algumas orgias do arquitecto Tomás
Taveira)
: sexo em posições criativas;
```

Notas:

**linha 1** - Uso da função trocadilho fonético.

**linha 2** - Uso da função adivinha.

**linha 6** - Uso da função afterevent.

Exemplo: resultado de procurar *taveirada*

```
* taveirada
  sem -- sexo em posições criativas
  ex -- ela só pensa em taveiradas...
  local -- Portugal
  isa -- Termo nascido de evento
  nivel -- calão
```

```
evento origem --
    termo que apareceu após a circulação de um vídeo clandestino
    documentando algumas orgias do arquitecto Tomás Taveira
data -- aparecimento -- 1986 ?
```

Exemplo: procurar *gay*

```
* gay
. genero -- masculino
  sem -- homossexual masculino
      indivíduo afeminado
  syn --
      maricas; bicha; .....
  en -- homossexual mail
  nivel -- calão carroceiro

. trocadilho --
  name -- entreguei
  isa -- trocadilho fonético
      anedota
  sem -- estar cercado de homossexuais
```

Exemplo: resultado de procurar *entregar*

```
* entreguei
  sem -- estar cercado de homossexuais
  referente -- gay
  isa -- trocadilho fonético

* (encomendar*|entregar) a alma a Deus

* entregar-se de corpo e alma
  sem -- dedicar-se muito a uma tarefa

* (estar*|ficar*) entregue à bicharada

* estar* bem entregue*N
```

## 5 Definição orientada ao conceito

O facto de se fazer definições orientadas ao conceito permite uma grande economia de descrição. O simples facto de se juntar um conjunto de termos que têm

uma acepção comum, é suficiente para definir essa mesma acepção. Considere-se os seguintes exemplos que incluem a palavra *cachimónia*:

```
mona++n2|
  cachimónia++n2|
  tola++fon(tóla)++n2|
  bestunto++n2+"puxar pelo bestunto"=pensar|
  bestunteira++n2|
  cornos++n3+"tens de meter nos cornos essa matéria"|
  cabeça|
  caixa dos pirolitos;;

dar*-lhe na (cachimónia|cabeça|mona|veneta)++n2|
dar*-lhe na (real|) gana++n2
  :lembrar-se de;

faltar*-lhe*PN (um parafuso|parafusos)++n2|
ter* um parafuso a menos++n2|
ter* a rosca moída++n2|
não bater* bem da (bola|mona|cachimónia|tola)++n2|
não ser* bom da cabeça
  :ser ou parecer maluco;
```

A definição estruturada por conjuntos de sinónimos é especialmente útil em situações em que haja grande quantidade de representantes do mesmo conceito, como é o caso do dicionário aberto de calão.

Cada conceito dá origem a várias entradas no dicionário descrito.

Exemplo: resultado de procurar *cachimónia*

```
* cachimónia
  syn --
    mona
    tola
    bestunto
    bestunteira
    cornos
    cabeça
    caixa dos pirolitos
  nivel -- coloquial

* dar*-lhe na (cachimónia|cabeça|mona|veneta)
  syn -- dar*-lhe na (real|) gana
  sem -- lembrar-se de
  nivel -- coloquial
```

```

* não bater* bem da (bola|mona|cachimônia|tola)
  syn --
    faltar*-lhe*PN (um parafuso|parafusos)
    ter* um parafuso a menos
    ter* a rosca moída
    não ser* bom da cabeça
  sem -- ser ou parecer maluco
  nivel -- coloquial

```

## 6 Meta-informação

Como é sabido, um dicionário é algo mais que um conjunto de definições de termos. Para além da definição dos termos, qualquer dicionário contém um conjunto de informação, como seja o título, as notas prévias, os autores, etc.

No dicionário aberto de calão, a metadata associada ao dicionário aparece explícita e consultável (internamente é guardada como se de um termo normal se tratasse).

```

<meta>
title:Dicionário aberto de calão e expressões idiomáticas;
latex:{package="\usepackage{a4wide}"};
author:{name = José João Almeida;
        email = jj@di.uminho.pt;
        project = Natura;
        org = Universidade do Minho, Departamento de Informática};
introduction:
  /{ Acreditamos que as expressões idiomáticas e o calão são uma
  parte nobre e rica da Língua Portuguesa. Ao mesmo tempo que
  .....
  colaboração de vários +daci!informantes!agradecimentos! a quem
  muito agradecemos <p> } ;
sources : [/{http://natura.di.uminho.pt/~jj/pln/calao.dic};
          /{http://natura.di.uminho.pt/~jj/pln/proverbio.dic}];
descs:{ syn = sinónimos;
        isa = é um;
        ex = exemplo } ;
skell:{
  syn = sinónimos (separados por ,);
  isa = [termo; frase pitoresca; interjeição; provérbio;
        insulto; termo calão; idiomática; outra; adivinha];
  semantica = significado;
  level = [normal; calão; erudito; coloquial;
          calão carroceiro ;

```



```

        calão muito carroceiro ;
        calão estupidamente carroceiro];
local = [pt; pt norte ; pt centro ; pt sul ; Brasil ; Angola ;
        Madeira; Açores; particular];
outrolocal = origem geográfica;
from = {nome = nome;
        email = email (não fica público)} ;
ex = { frase = exemplo de uso;
      parafr = explicação}
} ;
copyright: Projecto Natura, J.Joao, Licença GNU;
keywords: slang, calão, expressões idiomática;
</meta>

```

Notas:

**linha 2** - O título é uma informação que provavelmente será usada por todos os processadores do dicionário.

**linha 3** - Esta informação vai ser usada apenas pelo processador L<sup>A</sup>T<sub>E</sub>X.

As ferramentas que processam o dicionário fazem uso parcial dessa meta-informação, ou seja tiram partido desta informação de acordo com as suas necessidades.

Exemplo: para geração da página Internet de consulta ao dicionário, o título, a nota introdutória e os ficheiros fonte da página inicial, são retirados dinamicamente da referida meta-informação.

Do mesmo modo, quando um termo é desconhecido, a ferramenta vai consultar o campo `skel` da metadata e com base nessa informação constrói-se uma *form* (ver Figura 1) cujo preenchimento alimenta um ficheiro que será posteriormente revisto manualmente e incorporado no dicionário oficial.

Ver também as Figuras ?? e ?? da Secção ??.

## 7 Criação duma CGI de consulta

Na versão actual, a criação do dicionário consultável, envolve compilar o dicionário, construindo uma `DB_File` perl. Esta base de dados permite rápido acesso a partir do termo origem, ao mesmo tempo que permite aceitar uma enorme quantidade de chaves (termos).

```
lpdc dac.dpl
```

Após esta operação está a ser feita a junção (composição paralela) com um outro dicionário constituído por pesquisa sobre uma lista de provérbios (cuja informação tem uma natural ligação às expressões idiomáticas).

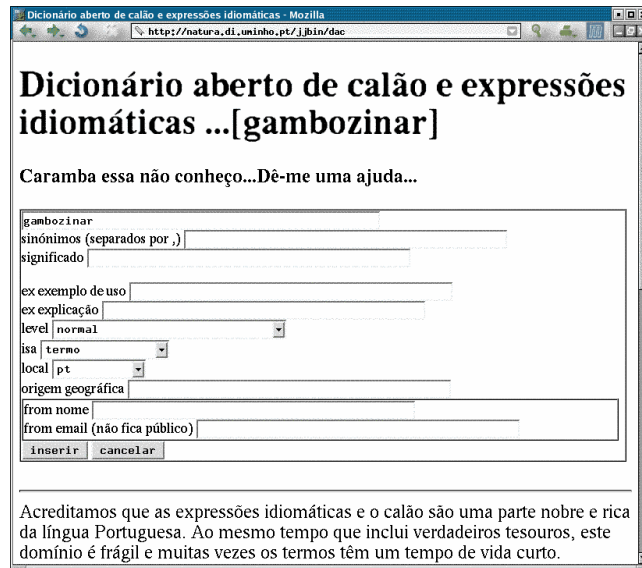


Figura 1: Resultado da pesquisa de *gambozinar*

Para realizar esta tarefa está a ser usado o módulo `ddmf.pm`, que será explicado no próximo capítulo, e que aceita expressões de construção de dicionários como esta:

```
$d1 = mkparcom(
    mkdplddmf("dac.dpl"),
    mkmttab("proverbios",
        { sep => ":", type => "ff", lang => "pt"},
        ["proverbio.txt", "name", "tipo", "result"] ));
```

Notas:

- linha 1** - Fazer a composição paralela dos dois dicionários que se seguem.
- linha 2** - Construir um dicionário a partir dum texto `dpl`. Internamente, é determinado o conjunto dos lemas de modo a permitir pesquisa por lema – ou seja pesquisa a menos de normalização morfológica.
- linha 3** - Construir um dicionário a partir duma tabela textual (contendo os provérbios),
- linha 4** - ... em que o separador de campo é ":", a língua usada é o Português (activa a pesquisa com normalização morfológica),
- linha 5** - ... contidos no ficheiro `proverbios.txt`, sendo cada registo da tabela textual composto por um `nome`, um `tipo` e um `resultado` (utilizado quando o tipo é adivinha).

Alguns detalhes deste módulo serão descritos no Capítulo ??.

## 8 Conclusões

Com a construção / manutenção do dicionário aberto de calão, constatou-se que:

- o campo em causa é muito interessante.
- neste domínio, é indispensável tentar assumir uma postura arquivística, documentalista.
- como exercício de especificação que é, dispor de possibilidade de definir os operadores metalinguísticos a usar nas definições, é algo que ajuda a estudar e entender um domínio.
- o efeito de *disponibilizar tem efeito multiplicativo*, levou a uma grande quantidade de contribuições<sup>3</sup>.

---

<sup>3</sup>Estando a situação presentemente ligeiramente fora de controle!