

Cr terios e op  es lingu sticas no desenvolvimento do
Palavroso,
um sistema computacional de descri  o morfol gica do
portugu s

Anabela Barreiro Maria de Jesus Pereira Diana Santos

Grupo de Linguagem Natural do INESC

INESC, Dezembro 1993

Relat rio INESC n.  RT/54-93

Introdução

O Palavroso é um sistema computacional desenvolvido pelo Grupo de Linguagem Natural do INESC que caracteriza morfologicamente qualquer palavra do português, entendendo por palavra qualquer conjunto de caracteres gráficos (letras e hífen) separado por espaços.

Os sistemas linguísticos em geral podem ser descritos através de duas entidades que se completam: o léxico e um número finito de regras.

O Palavroso foi desenhado de forma a que o tamanho do léxico fosse variável (com um mínimo determinado pela língua e que corresponde ao conjunto das palavras gramaticais e das formas de verbo francamente irregulares). Obviamente, não tendo qualquer informação sobre uma dada palavra, o programa sobreanalisa. Mas a intenção foi a de que o programa desse uma resposta inteligente quando não tivesse conhecimento sobre o item lexical que lhe é apresentado. A motivação desta arquitectura do Palavroso e as suas características específicas foi descrita em Santos et al. (1992), Medeiros (1992) e Medeiros et al. (1993).

Neste relatório descrevemos a adaptação do Palavroso para utilização num corrector ortográfico. As características de uma tal aplicação obrigam a que

- 1 - assintoticamente todas as palavras do português estejam no dicionário
- 2 - o sistema não pode sobreanalisar, sob pena de aceitar formas incorrectas

Estes requisitos obrigaram a um trabalho considerável de normalização e a decisões sobre assuntos (e itens lexicais) não documentados na literatura (quer científica quer simplesmente dicionarística).

É sobre todos os esses assuntos que este relatório reza. Para os abordar, foi, além disso, necessário explicitar a forma como o conhecimento linguístico é expresso no sistema. Com este texto, pretende-se essencialmente fazer um levantamento dos problemas que foram surgindo durante o preenchimento do dicionário e da construção das regras, descrever os critérios e apresentar algumas das soluções adoptadas para a sua resolução. Aproveitamos o ensejo para apresentar uma análise quantitativa do sistema, bem como os critérios de adaptação ao novo acordo ortográfico.

NOTA:

À primeira vista, poder-se-ia argumentar que o desenho do sistema (concebendo a língua como sistema aberto e dando maior ênfase às regras em detrimento do léxico) é avesso ao tipo de aplicação aqui tratada, e que um desenho baseado em paradigmas (como o dos sistemas DIGRAMA (Ranchod 198.) e Lince (Andrade et al. 1993) seria mais adequado. No entanto, é fácil constatar que a maior parte dos problemas que descrevemos neste relatório se traduz directamente, nesse tipo de sistemas, pela incerteza do paradigma a que um determinado conjunto de palavras pertence, e, portanto, são independentes da arquitectura.

Por outro lado, falhará por defeito um sistema baseado em paradigmas onde o nosso falhará por excesso, e pois tendencialmente tenderão ambos para o mesmo desempenho. O que está portanto em causa é a cobertura que cada sistema de facto consegue, e não o modo como o consegue.

Cr terios Utilizados no Tratamento de Nomes e Adjectivos

Relativamente aos nomes e aos adjectivos, a informa  o do dicion rio diz respeito   categoria gramatical da palavra (nome, adjectivo ou nome/adjectivo), ao g nero (masculino "M", feminino "F" ou invari vel "I") e ao n mero (singular "S", plural "P" ou invari vel "I").

As regras respeitantes a nomes e adjectivos s o:

a) regras de g nero (permitem obter o masculino a partir da forma feminina dos adjectivos);

b) regras de n mero (permitem obter o singular a partir da forma plural dos nomes e dos adjectivos);

c) regras de tratamento de diminutivos e de aumentativos

d) regras de grau dos adjectivos (para obter o grau normal a partir do superlativo absoluto simples).

Tratamento quanto ao N mero

Tanto no caso dos nomes como no dos adjectivos, s  list mos no dicion rio a forma singular,   excep  o de alguns nomes que s  existem no plural e de outros que s o invari veis, ou seja, t m a mesma forma no singular e no plural.

Exemplos:

*arco- ris*¹: M I

cais: nome M I

cal as: nome F P

ourives: nome I I

parab ns: nome M P

O Palavroso inclui,  l m disso, regras que permitem determinar, dada uma forma plural, qual o singular a que correspondem.

Exemplos:

 guas-furtadas: nome F P Comp (* gua-furtada*, 100.00)

casas: nome F P (*casa*, 100.00)

cata-ventos: nome M P Comp (*cata-vento*, 100.00)

homens: nome M P (*homem*, 100.00)

portugueses: nome/adj M P (*portugu s*, 100.00)

postais: nome M P (*postal*, 100.00)

adj I P (*postal*, 100.00)

¹Sobre as palavras compostas iremos falar mais adiante.

Plural das Palavras terminadas em *-ão*

As palavras terminadas em *-ão* têm um comportamento diferente e, por isso, merecem um tratamento especial. De acordo com as gramáticas tradicionais, as palavras terminadas em *-ão* podem ter vários tipos de plural:

1. podem terminar em *-ões*. Exs: *leões, limões, opiniões, portões, sugestões*
+ aumentativos (*carrões, grandalhões*).
-ães. Exs: *cães, capitães, alemães, pães, sacristães*.
-ãos. Exs: *cidadãos, grãos, irmãos, pagãos, corrimãos*
+ paroxítonos (*acórdãos, órgãos, sótãos*).
2. podem ter um plural duplo: *-ães/-ões*. Exs: *guardiães, guardiões*;
bastiães, bastiões;
alazães, alazões.
-ãos/-ões. Exs: *vilãos, vilões; hortelãos, hortelões*;
verãos, verões.
-ães/-ãos. Exs: *sacristães, sacristãos*.
3. podem ter três plurais: *-ões, -ães* e *-ãos*. Exs: *aldeões, aldeães* e *aldeãos*;
piões, piães e *piãos*;
vulcões, vulcães e *vulcãos*.

Perante esta variedade de flexões nominais, e para obter os resultados práticos correctos, decidimos seleccionar para o nosso programa apenas uma regra morfológica, a mais produtiva e que abrange maior número de palavras: *-ão ---> -ões*. Todas as outras palavras foram listadas num ficheiro à parte com a informação acerca do radical correcto. Desta forma evitamos problemas de plurais mal formados, como por exemplo: **cidadões* e **cidadães* para *cidadão* e obtemos os seguintes resultados:

Exemplos:

alemães: nome/adj M P (alemão, 100.00)

cidadãos: nome M P (cidadão, 100.00)

Tratamento quanto ao Género

Palavras com género definido

Adjectivos

De acordo com o funcionamento global do programa, a grande maioria dos adjectivos são apresentados sob a forma do masculino singular. Quase todas as entradas lexicais referentes a palavras que contêm apenas a categoria gramatical de adjectivo estão listadas no masculino.

Exemplos:

gelatinoso: adj M S
mimado: adj M S
maculado: adj M S
tardio: adj M S

Há, no entanto, alguns adjectivos que escapam a esta regra:

adjectivos invariáveis quanto ao género. Ex: *inteligente: adj I S*

adjectivos invariáveis quanto ao género e quanto ao número. Ex: *pires: adj I I*

Todos estes adjectivos constituem entradas lexicais no dicionário, por razões que iremos ver de seguida.

As regras que utilizámos para calcular a forma masculina correspondente ao feminino são aplicadas apenas à categoria dos adjectivos (como podemos ver nos exemplos abaixo), uma vez que todos os nomes vêm atestados no dicionário.

Exemplos:

corajosa: adj F S (corajoso, 100.00)
gil-vicentina: adj F S Comp (vil-vicentino, 100.00)
melancólica: adj F S (melancólico, 100.00)
orgulhosa: adj F S (orgulhoso, 100.00)
sério-cómica: adj F S Comp (sério-cómico, 100.00)
simpática: adj F S (simpático, 100.00)

Um facto relevante diz respeito à formação do feminino de adjectivos terminados em *-ão*, devido ao seu carácter irregular. Alguns destes femininos são formados através do sufixo *-ã*, outros através do sufixo *-ona*, outros ainda através de *-oa*:

1. adjectivos terminados em *-ã*. Exs: *alemã; anã; cristã; pagã; sã*.
2. adjectivos terminados em *-ona*. Exs: *bonacheirona; chorona; comilona; glutona; mandriona*.
+ aumentativos (*grandona; pesadona*)
3. adjectivos terminados em *-oa*. Exs: *ladroa; varoa, capitoa*.

Perante três sufixos diferentes, e à semelhança do que fizemos com os plurais, decidimos optar pela regra *-ão --> -ona*, visto ser aquela que regista um maior número de exemplos, e registámos num ficheiro à parte todas as formas adjectivais femininas terminadas em *-ã* e *-oa*, com a informação de qual o seu radical. Deste modo, conseguimos obter a forma feminina correcta para todos os casos.

Exemplos:

chã: adj F S (chão, 100.00)
meã: adj F S (meão, 100.00)
ladroa: adj F S (ladrão, 100.00)

Nomes

O género é uma propriedade intrínseca dos nomes. Os nomes podem ter género masculino ou género feminino.

Exemplos:

carro: nome M S
mar: nome M S
mesa: nome F S
porta: nome F S

A entidades animadas, corresponde um género natural (cf. Lyons, 1968), masculino ou feminino, de acordo com o sexo a que a palavra se refere. Nestes casos, muitas vezes não existe qualquer tipo de relação ou identidade gráfica e morfológica entre a "forma" masculina e a feminina.

Exemplos:

cavalo - égua
cão - cadela
genro - nora
homem - mulher
pai - mãe
zangão - abelha

No entanto, há casos em que a forma feminina tem o mesmo radical que a forma masculina. Nestes casos o género é gramatical e pode-se considerar que a palavra deriva do radical masculino, mediante a substituição ou acréscimo de desinências, que são, sem dúvida, marcadores femininos.

Exemplos:

aluno - aluna
orador - oradora
tio - tia

Tal como já referimos anteriormente, todos os nomes femininos (com género natural ou com género gramatical) estão listados no dicionário como entrada lexical. O programa só permite flexionar em género os adjectivos e os nomes, de forma a evitar que *banha* seja identificado como o feminino de *banho*, *bola* de *bolo*, *cigarra* de *cigarro*, *porta* de *porto*, *vinha* de *vinho*, etc.

Em resumo, no que se refere ao grupo de palavras femininas pertencentes a este dicionário há a registar alguns subgrupos distintos:

1. nomes no género feminino independentes de quaisquer outras palavras masculinas listadas no dicionário.

Exemplos:

norma: nome F S

peça: nome F S

flor: nome F S

2. nomes no género feminino que possuem no dicionário um nome masculino correspondente.

Exemplos:

cão: nome M S

cadela: nome F S

matemática: nome F S

matemático: nome/adj M S

pai: nome M S

mãe: nome F S

professora: nome F S

professor: nome M S

3. palavras no género feminino pertencentes às duas categorias gramaticais (nome e adjectivo), enquanto que as palavras listadas no género masculino só são adjectivos.

Exemplos:

vincada: nome F S

vincado: adj M S

culinária: nome F S

culinário: adj M S

cotovelada: nome F S

cotovelado: adj M S

curva: nome F S

curvo: adj M S

Palavras Invariáveis quanto ao Género

Existe um grupo de palavras invariáveis quanto ao género (aproximadamente 7000), que foram listadas no nosso dicionário com a etiqueta I. A distinção entre o masculino e o feminino é exterior à própria palavra e distingue-se através da flexão das palavras do contexto em que ocorre, nomeadamente, no caso dos nomes, pelo género do artigo que a acompanha, e no caso dos adjectivos, pelo nome que estes modificam.

São, geralmente invariáveis os adjectivos terminados em *-a*, *-e*, *-l*, *-ar* e *-or*, *-s*, *-z*, *-m* (cf. Cunha & Cintra, 1987).

Exemplos:

anti-regulamentar: adj I S
celta: nome/adj I S
comum: adj I S
feliz: adj I S
herege: nome/adj I S
ilustre: adj I S
linear: adj I S
simples: nome/adj I S
superior: nome/adj I S
transeunte: nome/adj I S
unilateral: adj I S

Algumas das palavras invariáveis são formadas a partir dos sufixos *-ável*, *-nte*, *-ense*, *-estre* (menos vulgar), *-ista* e *-ita*.

Exemplos:

agradável: adj I S
amante: nome/adj I S
israelita: nome/adj I S
madeirense: nome/adj I S
pedestre: adj I S
pedinte: nome/adj I S
crente: nome/adj I S
pára-quedista: nome I S
optimista: nome/adj I S

Nos exemplos acima, o significado da palavra mantém-se nos dois géneros. No entanto, este grupo engloba também casos em que não existe qualquer relação entre a palavra no masculino e no feminino, como os que podemos ver abaixo:

capital: nome/adj I S
cura: nome I S
final: nome/adj I S

geral: nome/adj I S
guarda: nome I S
lente: nome I S
papa: nome I S

Tratamento de Aumentativos e Diminutivos

Em português os diminutivos são de grande uso e riqueza, e podem ser utilizados com variadíssimas funções (cf. Pedro, 1992). Têm grande força e expressividade e são marcadores de informalidade e afectividade. Em português existe um número ilimitado de palavras que aceitam diminutivos. Os aumentativos, embora usados muito menos frequentemente, têm também um papel de relevo, na medida em que existe um número considerável de palavras que os aceitam.

Tal como nos é referido em Cunha & Cintra (1987), o seu valor é mais afectivo do que lógico. Por conseguinte, não existem critérios rígidos quanto à formação de palavras a partir de sufixos diminutivos e aumentativos. Estas palavras muito raramente fazem parte da lista de entradas lexicais de um dicionário normal.

Começamos por referir os sufixos diminutivos e aumentativos mais produtivos e, por consequência, os mais comuns na língua portuguesa. São eles *-inho(a)/-zinho(a)* e *-ão(ona)/-zão(zona)*. Os primeiros juntam-se não só a nomes (ex: *rapazinho*) e adjectivos (ex: *tristinho*), como também a advérbios, participípios e outras palavras invariáveis (ex: *agorinha*, *estragadinho*, *devagarinho*, *adeusinho*). Os segundos juntam-se a nomes (ex: *garotão*), a adjectivos (ex: *grandão*) e a verbos (ex: *choramingão*). Nenhum destes sufixos altera a categoria gramatical da palavra a que se associam.

No nosso dicionário, tanto nomes como adjectivos aparecem, em grande parte, na sua forma normal, e através de regras, é possível gerar formas diminutivas e aumentativas. Há, porém, algumas excepções.

À semelhança dos critérios utilizados nos dicionários normais, decidimos listar no nosso dicionário aumentativos formados através de consoantes de ligação ou de outros sufixos.

Exemplos:

comilão: S M nome
grandalhão: S M nome
facalhão: S M nome
vozeirão: S M nome

Também listámos todos os casos em que há mudança de género provocada por este processo de sufixação, como acontece com alguns aumentativos masculinos formados a partir de palavras de género feminino.

Exemplos:

perna: S F nome
pernã: S M nome

mulher: S F nome
mulherã: S M nome

sala: S F nome
salã: S M nome

No que diz respeito a palavras terminadas em *-a* que têm género masculino, foram criadas duas regras que permitem reconhecer tanto diminutivos formados a partir de *-inha*, como diminutivos formados a partir de *-zinho*.

Exemplos:

programinha: S M nome Dim (programa, 100.00)
programazinho: S M nome Dim (programa, 100.00)

telegraminha: S M nome Dim (telegrama, 100.00)
telegramazinho: S M nome Dim (telegrama, 100.00)

À excepção dos nomes terminados em *-m*, que exigem as formas *-zinho(a)* e *-zão (-zona)* e dos nomes terminados em *-s* e *-z*, que exigem as formas *-inho* e *-ão* para diminutivo e aumentativo respectivamente, e que o programa trata de forma correcta, também há casos em que nem sempre é fácil indicar as razões que comandam a escolha entre *-inho(a)* e *-zinho(a)* e entre *-ão(-ona)* e *-zão(-zona)*. Por vezes, o uso e formação de palavras com base nestes sufixos é relativamente arbitrário. A selecção está normalmente ligada ao ritmo da frase e à preferência de uma ou de outra forma por parte do falante. No entanto, algumas formas são mais usadas e preferíveis a outras. O programa, neste momento, admite ambas as formas.

Exemplos:

embalagem - embalagenzinha - embalagenzona
homem - homenzinho - homenzão

adeus - adeuzinho - adeuzão
turquês - turquezinha - turquezona
atriz - atrizinha - atrizona
arroz - arrozinho - arrozão

colher - colherinha - colherzinha
colher - colherona - colherzona
cordel - cordelinho - cordelzinho
cordel - cordelão - cordelzão
pasta - pastinha - pastazinha

pasta - pastona - pastazona

Relativamente a adjectivos, os resultados nem sempre são os mais desejados. Se existem adjectivos que aceitam facilmente diminutivos e aumentativos (ex: *bonito - bonitinho/bonitinha - bonito/bonitona*), também existem adjectivos que parecem não aceitar este tipo de sufixos (ex: *fugaz - ?fugazinho/?fugazinha - ?fugazão/?fugazona*).

Neste momento, o programa admite que se crie qualquer diminutivo em *-inho* e qualquer aumentativo em *-ão*, desde que exista no dicionário o adjectivo no grau normal.

Exemplos:

bonito: adj M S (bonito, 100.00)
bonitinho: adj M S Dim (bonito, 100.00)
bonitinha: adj F S Dim (bonito, 100.00)
bonitão: adj M S Aum (bonito, 100.00)
bonitona: adj F S Aum (bonito, 100.00)

fugaz: adj I S (fugaz, 100.00)
fugazinho: adj M S Dim (fugaz, 100.00)
fugazinha: adj F S Dim (fugaz, 100.00)
fugazão: adj M S Aum (fugaz, 100.00)
fugazona: adj F S Aum (fugaz, 100.00)

Para resolver este problema é necessário fazer um estudo mais aprofundado, de forma a definir os critérios que caracterizam uns e outros adjectivos.

No caso de adjectivos invariáveis quanto ao género, geralmente terminados em *-e*, *-l*, *-m*, *-r*, *-s* ou *-z*, o programa reconhece tanto as formas diminutivas e aumentativas masculinas, como as femininas.

Exemplos:

tristinho: S M nome/adj Dim (triste, 100.00)
tristinha: S F nome/adj Dim (triste, 100.00)
tristão: S M nome/adj Aum (triste, 100.00)
tristona: S F nome/adj Aum (triste, 100.00)

vulgarzinho: S M nome/adj Dim (vulgar, 100.00)
vulgarzinha: S F nome/adj Dim (vulgar, 100.00)
vulgarzão: S M nome/adj Aum (vulgar, 100.00)
vulgarzona: S F nome/adj Aum (vulgar, 100.00)

As palavras acentuadas graficamente requerem um processamento mais complicado, visto que os diminutivos e os aumentativos formados a partir delas não têm acento gráfico. No entanto, o programa trata-as de forma correcta, como podemos ver de seguida.

Exemplos:

aguinha: S F nome Dim (água, 100.00)
aguazinha: S F nome Dim (água, 100.00)
buziozinho: S M nome Dim (búzio, 100.00)
cafezinho: S M nome Dim (café, 100.00)
lampadazinha: S F nome Dim (lâmpada, 100.00)
tabuinha: S F nome Dim (tábua, 100.00)
tabuazinha: S F nome Dim (tábua, 100.00)
voozinho: S M nome Dim (vôo, 100.00)

As regras que construímos para a criação produtiva de diminutivos não impedem, evidentemente, a consulta à informação do dicionário. Assim, se pedirmos ao programa a análise das palavras *galinha* e *papelão*, o resultado será:

galinha nome F S (galinha, 100.00) - a partir do registo como entrada lexical

galinha nome F S Dim (gala 100.00) - a partir das regras de formação de diminutivos

papelão nome M S (papelão, 100.00) - a partir do registo como entrada lexical

papelão nome M S Aum (papel, 100.00) - a partir das regras de formação de aumentativos

Diminutivos e aumentativos pouco produtivos como *-aça*, *-aço*, *-acho*, *-alhão*, *-arra*, *-arrão*, *-eco*, *-ejo*, *-elho*, *-etão*, *-eirão*, *-icho*, *-ico*, *-ino*, *-ito*, *-olo*, *-ote*, *-uça*, *-ucho*, etc. (e, em alguns casos, os respectivos femininos) não foram contemplados nas regras. Desta forma, palavras como *soneca*, *burraco*, *pequenino*, *bocarra*, *barcaça*, *dentuça*, etc., foram registadas no nosso dicionário (à semelhança do que acontece nos dicionários em papel).

Finalmente, os diminutivos e os aumentativos de palavras que não são nem nomes, nem adjectivos, foram listados no dicionário, na parte correspondente à sua classificação.

Exemplos:

agora - agorinha
devagar - devagarinho

Tratamento do Grau Superlativo dos Adjectivos

O superlativo traduz, geralmente, um valor elevado da qualidade que o adjectivo exprime (cf. Cunha & Cintra, 1987). Esse valor pode ser representado por meio de um

sufixo, de um prefixo, de um advérbio ou de um artigo antes do comparativo. O superlativo de que vamos tratar é o superlativo absoluto sintético, que é formado por meio de sufixação. Só listámos os adjectivos no grau normal. O superlativo é tratado por meio de regras. Existem três sufixos possíveis: *-íssimo*, *-érrimo* e *-ílmo*, sendo *-íssimo* o mais frequente.

Torna-se, por vezes, difícil determinar quais os adjectivos que podem ou não constituir superlativos, uma vez que a maioria dos superlativos não consta dos dicionários correntes. Porém, é necessário que o sentido do adjectivo admita variação de intensidade ou de grau.

O analisador morfológico apresenta um conjunto de regras que permitem a construção do processo superlativante de uma forma regular para os três tipos. Existe também no nosso dicionário uma lista de superlativos, alguns de étimo latino, outros de formação irregular, que se apresentam como excepções.

Exemplos:

acre - acérrimo
pobre - paupérrimo
doce - dulcíssimo
frio - frigidíssimo

bom - melhor/ótimo
inferior - ínfimo
mau - pior/péssimo
posterior - póstumo

Sempre que analisa um superlativo, o programa recorre primeiro à leitura do ficheiro das excepções e só depois prossegue (se for caso disso) para a leitura das regras. A análise feita pelo Palavroso pode verificar-se nos exemplos seguintes:

amicíssimo: adj M S Sup (amigo, 100.00)
bem-educadíssimo: adj S M Comp Sup (bem-educado, 100.00)
libérrimo: adj M S Sup (livre, 100.00)
humílimo: adj M S Sup (humilde, 100.00)
radicalíssimo: adj M S Sup (radical, 100.00)
riquíssimo: adj M S Sup (rico, 100.00)

Embora haja um conjunto de adjectivos que não sofrem normalmente a superlativização, o programa reconhece e aceita qualquer tipo de superlativo, desde que a regra permita a sua existência.

No Palavroso considerámos que a cada superlativo corresponde um só adjectivo no grau normal. Assim, embora em Cunha & Cintra (1987) *malevolentíssimo* nos surja como o superlativo de *malévolo*, o nosso programa identifica-o como superlativo de *malevolente*, adjectivo sinónimo. Optámos, pois, por um processo regular.

Tratamento das Palavras Compostas

As palavras compostas resultam de um processo no qual se juntam dois ou mais vocábulos, que podem ser de várias categorias, com vista a formarem uma única palavra. Este tipo de palavras pode ser formado por um processo de justaposição ou por um processo de aglutinação. Na justaposição, os elementos ou palavras que entram na formação das palavras compostas (com ou sem hífen, no entanto, mais usualmente escritos com hífen) mantêm a sua ortografia e a sua sílaba tónica. Na aglutinação os elementos ou palavras, que formam a palavra composta, sofrem alteração ortográfica fundindo-se num só vocábulo existindo, assim, apenas uma sílaba tónica (cf. Bergstrom e Reis, 1986).

É importante referir que só as palavras compostas formadas pelo processo de justaposição merecem um tratamento especial no Palavroso. As palavras compostas formadas pelo processo de aglutinação são consideradas palavras de pleno direito tal como todas as outras.

O tratamento das palavras compostas é problemático uma vez que:

Não existe uma uniformização no que se refere à ortografia deste tipo de palavras. Prova disso é o facto de um mesmo dicionário atestar, por vezes, duas formas ortográficas diferentes para uma dada palavra composta, como por exemplo as formas *água-ardente* e *aguardente* que são atestadas simultaneamente (cf. Figueiredo, 1973).

Das palavras consideradas compostas, mas sem qualquer sinal gráfico que o indique (tal como *fim de semana*) não tratamos no nosso dicionário, baseado em palavras gráficas. Relativamente às restantes palavras compostas optámos por aceitar a dupla grafia.

A divergência anteriormente apontada é também encontrada em dicionários ou documentos diferentes. Para uma mesma palavra composta alguns dicionários adoptam a ortografia com hífen outros sem hífen, vejamos por exemplo *boca-aberta* em Moraes (1949/1959) e *boca aberta* em Aurélio (1986).

Este facto tem sido referido por vários autores, veja-se a este respeito Macedo (1992, 276): *Como se pode verificar os critérios seguidos na apresentação e descrição dos nomes compostos não são transparentes e fundamentados.*

No que se refere à flexão em número das palavras compostas, o plural destas é calculado automaticamente através de regras que se encontram no Palavroso, já que as formas atestadas no nosso dicionário, tal como todas as outras, se encontram atestadas no singular.

Os critérios usados na elaboração das regras de plural deste tipo de palavras são os que se seguem (cf. Reis e Bergstrom, 1986 e Luz e Cuesta, 1983), convém referir que estas regras nem sempre são confirmadas pelo dicionário de Aurélio (1986) o qual coloca o plural das palavras compostas:

1. Nas palavras compostas formadas por dois nomes, ambos os nomes se flexionam no plural: *a couve-flor / as couves-flores*.

2. Nos compostos formados por nome e adjectivo, ambos se flexionam no plural: *o capitão-mor / os capitães-mores*.

3. Se a palavra composta é formada por um adjectivo e um nome, ambos são flexionados no plural: *o primeiro-ministro / os primeiros-ministros*.

4. Se a palavra composta é formada por um verbo e um nome ou um adjectivo, só o segundo elemento é flexionado no plural: *o guarda-sol / os guarda-sóis, busca-fundo / busca-fundos, etc.*

5. Se o composto já possui o nome no plural não existe variação entre a forma singular e a forma plural: *o guarda-jóias / os guarda-jóias*.

6. Quando os componentes do composto são ligados por uma preposição, só o primeiro é flexionado no plural: *o fim-de-semana / os fins-de-semana*.

7. Se o primeiro elemento é um prefixo ou um elemento de composição, só o segundo se flexiona no plural: *o ex-marido / os ex-maridos*.

8. Se a palavra é formada por dois verbos, só o segundo verbo se flexiona no plural: *o pisca-pisca / os pisca-piscas*.

Existem outras formas de combinação relativamente à classificação gramatical de cada uma das palavras que entram na formação do composto, para além das apresentadas acima. Estas novas combinações não foram contempladas na bibliografia disponível (artigos e gramáticas) sobre o estudo das palavras compostas, pelo que não encontramos critérios relativos à pluralização dos compostos que deles resultam. Assim, as regras que se seguem são elaboradas com base nos critérios de pluralização das palavras compostas que se encontram atestadas no dicionário de Aurélio (1986).

Os critérios de pluralização criados por nós são os que se seguem:

9. Se o primeiro elemento do composto não tem plural, (advérbio, prefixo, preposições, elementos de composição, etc.) só os elementos seguintes se vão pluralizar: *além-mar / além-mares, abaixo-assinado / abaixo-assinados, mal-estar / mal-estares, ante-véspera / ante-vésperas, vice-primeiro-ministro / vice-primeiros-ministros, vice-secretário-geral / vice-secretários-gerais, etc.* Tornámos, deste modo, mais abrangente a regra 7 proposta por alguns autores.

10. Se ambos os elementos da palavra composta são adjectivos apenas o segundo pluraliza: *físico-químico / físico-químicos, teórico-prático / teórico-práticos, etc.*

Para que todas estas regras pudessem funcionar, isto é, para os plurais em questão poderem ser calculados automaticamente pelo Palavroso, e a fim de se poder

dar conta de eventuais ambiguidades, além da informação categorial da palavra composta no seu todo, foi necessário classificar todas as palavras que entram na formação do composto.

Um dos exemplos de ambiguidade a que nos referimos é o que diz respeito à palavra *guarda*. Esta ambiguidade reside no facto de se considerar esta palavra como um nome ou como um verbo, vejamos o que dizem Luz e Cuesta (1983, 378): *Mas, nos compostos em que uma das partes é constituída pela palavra **guarda**, esta leva ou não a marca do plural consoante seja sentida como substantivo ou como verbo pelos falantes...*

Na tentativa de resolver esta ambiguidade optámos pelos seguintes critérios:

Sempre que a palavra *guarda* designa uma pessoa, e por conseguinte entra numa palavra composta que designa por sua vez uma profissão, *guarda* é classificado como um nome.

guarda-florestal: S I nome nome-adj
guarda-livros: I I nome nome-adj

Se a palavra *guarda* não designa uma pessoa e o composto não designa uma profissão esta palavra é classificada como um verbo.

guarda-chuva: S M nome verbo-nome
guarda-loiça: S M nome verbo-nome

Um outro ponto de interesse é o que diz respeito à flexão do diminutivo, aumentativo e do grau superlativo das palavras compostas. Relativamente a este tipo de flexão existe uma escassa bibliografia é, pois, um estudo pouco trabalhado ainda. Dos poucos trabalhos realizados acerca deste assunto, podemos salientar alguns exemplos de formação desta flexão apresentados por Villalva (1992):

peixe-espada / peixinho-espada
abre-latas / abre-latazinho
dói-dói / dói-dóizinho

Perante estes resultados, concordamos com a autora quando refere que são agramaticais as formas:

**abrezinho-latas*
**dóizinho-dói*

No entanto, não estamos tão seguras de o diminutivo de *peixe-espada* ser *peixinho-espada*, porque não *peixe-espadazinho*?

Como falantes nativos da língua portuguesa sentimos que este é um assunto que merece um estudo mais aprofundado, já que não estamos verdadeiramente certos das

formas preferencias. Assim, e relativamente ao tratamento dos diminutivo, aumentativos e grau superlativo decidimos não os aceitar, além do resultado final bastante estranho, também porque são formas que não são usadas pelos falantes, pelo menos regularmente e em situações normais do discurso. Estamos conscientes de que esta foi a opção mais viável, senão vejamos o resultado aberrante após a aplicação de algumas regras:

?mestre-sábio / mestrezinho-sabiozinho
?couve-flor / couvinha-florinha
?gentil-homem / gentilzinho-homenzinho
?amor-perfeito / amor-perfeitão ou amor-perfeitozão
?abelha-mestra / abelhona-mestrona ou abelhona-mestrazona
?guarda-florestal / guarda-florestalção

Convém também referir, ainda, que existem certos tipos de palavras compostas que não se encontram atestadas no nosso dicionário, uma vez que são formadas com prefixos bastante produtivos, tais como *anti-*, *pré-* e *ultra-*. Como estes prefixos podem ocorrer tanto com nomes como com adjectivos, elaborámos uma regra que permite ao programa gerar e analisar as palavras compostas em questão, palavras essas tais como *anti-sofista*, *anti-hemorrágico*, *pré-universitário*, *pré-amplificador*, *ultra-rápido*, *ultra-realismo*, etc.

Um outro ponto, igualmente importante, é o que se refere ao facto de também não se encontrarem atestadas no dicionário em questão palavras compostas resultantes da junção de dois elementos de composição, tais como *anglo-*, *euro-*, *luso-*, etc., a que Cunha e Cintra (1987) chamaram de **adjectivos pátrios compostos**. Também neste caso recorreremos ao uso de uma regra que permite ao programa gerar e analisar estes mesmos adjectivos através da combinação dos elementos de composição em causa, tais como por exemplo *anglo-americano*, *austro-húngaro*, *teuto-brasileiro*, etc.

Relativamente à classificação gramatical das palavras compostas que se encontram atestadas no nosso dicionário estas dividem-se nas mais diversas categorias, estas poderão ser verbos, advérbios, interjeições, poderão ser unicamente nomes, unicamente adjectivos e poderão ser nomes e adjectivos simultâneamente, à semelhança do que acontece com as restantes formas do dicionário de nomes e adjectivos. Vejamos alguns exemplos:

ab-rogar: verbo *ab-rogar*, *infinit. impes.* Comp (*ab-rogar*, 100.00)

ai-jesus: *ij ai-jesus*, Comp (*ai-jesus*, 100.00)

alto-e-mau: adv *alto-e-mau*, Comp (*alto-e-mau*, 100.00)

água-furtada: nome F S Comp (*água-furtada*, 100.00)

alfaces-do-mar: nome F P Comp (*alface-do-mar*, 100.00)

fónico-rítmico: adj M S Comp (*fónico-rítmico*, 100.00)

extra-escolar: adj I S Comp (extra-escolar, 100.00)

sul-africanos: nome M P Comp (sul-africano, 100.0)

adj M P Comp (sul-africano, 100.00)

vermelho-escuro: nome M S Comp (vermelho-escuro, 100.0)

adj M S Comp (vermelho-escuro, 100.00)

Interacção dos Vários Processos

Finalmente, gostávamos de salientar que todos os processos de análise morfológica interagem correctamente. O programa analisa:

carinhas: nome F P Dim (cara, 100.00)

adj F P Dim (caro, 100.00)

cãezinhos: nome M P Dim (cão, 100.00)

couves-flores: nome F P Comp (couve-flor, 100.00)

guardas-florestais: nome I P Comp (guarda-florestal, 100.00)

importantíssimos: adj M P Sup (importante, 100.00)

mal-asadíssima: adj F S Comp Sup (mal-asado, 100.00)

paupérrimas: adj F P Sup (pobre, 100.00)

postaizinhos: nome M P Dim (postal, 100.00)

adj M S Dim (postal, 100.00)

CrITÉrios quanto à Inclusão no Dicionário de Palavras Problemáticas

Sabe-se que os dicionários de língua corrente contêm omissões e inconsistências que devem ser ultrapassadas na construção de um novo dicionário (cf. Reis, 1993). Um dicionário electrónico necessita, pois, de entradas que não estão atestadas nos outros dicionários. É do conhecimento geral que há palavras que estão de tal forma enraizadas no vocabulário quotidiano, que devem fazer parte de um dicionário deste tipo, ainda que mereçam tratamento especial. Entre estas palavras destacamos as formas femininas das profissões, as palavras terminadas em *-ela*, os estrangeirismos, os neologismos, as

palavras de gíria e calão, as palavras que têm dupla grafia e os participios passados duplos.

A selecção destas palavras foi feita de acordo com os nossos próprios critérios, tendo em conta "corpora" constituídos por excertos de diversos livros, revistas, jornais, gramáticas e textos publicitários.

Feminino de profissões

Em relação a palavras que designam profissões, e que só estão registadas nos dicionários de língua corrente sob a forma masculina, optámos por considerar as formas femininas possíveis na língua e registámo-las no nosso dicionário.

Exemplos:

cantoneira: nome F S

latoeira: nome F S

leiteira: nome F S

tanoeira: nome F S

Relativamente a palavras designativas de profissões e que aparecem geralmente nos dicionários como masculino, mas que também são, do conhecimento geral, palavras do género feminino, optámos por listá-las no dicionário como género invariável.

Exemplos:

dentista: nome I S

gerente: nome I S

guarda-florestal: nome I S

guarda-costas: nome I S

guarda-livros: nome I S

intérprete: nome I S

jornalista: nome I S

pediatra: nome I S

Um outro aspecto relevante é o que diz respeito a algumas profissões que vêm atestadas nos dicionários de língua portuguesa tanto na forma masculina como na forma feminina, mas que são usadas diariamente com um feminino diferente daquele que os dicionários atestam. Exemplos: o feminino de *lavrador*, que os dicionários atestam como *lavradeira* e que usualmente é realizado como *lavradora*; *procuradeira* que os dicionários atestam como feminino de *procurador* mas que usualmente se realiza como *procuradora*. Neste caso optámos por registar ambas as formas do feminino.

Tratamento de palavras terminadas em -ela

O sufixo *-ela* surge-nos frequentemente em palavras tais como *engraxadela*, *mordidela*, *picadela*, *varredela*, que constituem exemplos de morfologia derivacional (formação de nomes a partir de verbos). Apesar de não estarem atestadas em alguns dicionários que confrontámos, estas palavras foram incluídas no nosso dicionário, uma vez que fazem parte do uso quotidiano da língua portuguesa. Primeiro pensámos criar regras para a sua formação, mas acabámos por listá-las, uma vez que o sufixo apresenta pouca vitalidade e existe um grande número de palavras terminadas em *-ela* que não correspondem a sufixos.

Exemplos:

cadela
cancela
fivela
panela
tigela
vela

A lista de palavras terminadas pelo sufixo *-ela* que estão atestadas no nosso dicionário é a seguinte:

amolgadela
apalpadela
assobiadela
bagatela
besuntadela
clientela
cidadela
engraxadela
enrascadela
ensaboadela
enxaguadela
escovadela
esfoladela

lambidela
limpadela
lixadela
magrizela
parentela
penteadela
picadela
pintadela
pisadela
ruela
sacudidela
viela

Estangeirismos

Muitos vocábulos estrangeiros já foram lexicalizados na língua portuguesa, mas existe ainda um grande número que não está atestado nos dicionários. Alguns deles fazem parte do vocabulário de uso corrente, continuando a manter a forma gráfica da língua de que foram importados. Embora ocupem um lugar próprio destacado no nosso dicionário, devido ao seu comportamento diferente, considerámos importante o seu registo, uma vez que constituem vocábulos do dia-a-dia.

Exemplos:

<i>after-shave</i>	<i>leasing</i>
<i>barman</i>	<i>lingerie</i>
<i>bâton</i>	<i>maionaise</i>
<i>biberon</i>	<i>marquise</i>
<i>buffet</i>	<i>menu</i>
<i>cachet</i>	<i>pub</i>
<i>check in</i>	<i>rallye</i>
<i>collants</i>	<i>snack-bar</i>
<i>dossier</i>	<i>soufflet</i>
<i>édredon</i>	<i>soutien</i>
<i>gigolo</i>	<i>stop</i>
<i>hamburguer</i>	<i>topless</i>
<i>jeans</i>	<i>T-shirt</i>
<i>ketchup</i>	

Neologismos

No nosso dicionário foi feita a integração de palavras ainda não atestadas nos dicionários normais, mas que são já de uso corrente. Estas palavras estão geralmente associadas a áreas de trabalho específicas (cf. Reis, 1993).

Exemplos:

desfasamento
implementação
reconhecedor

Gíria e Calão

A gíria e o calão são um conjunto de expressões de tipo popular, mais usuais na linguagem corrente e despretensiosa, em certos meios especiais, sobretudo frequentes na linguagem oral, familiar ou profissional (escolas, prisões, locais de trabalho, etc.).

Embora a gíria e o calão sejam empregados com uma certa parcimónia em textos escritos, não quisemos deixar de registar vários vocábulos susceptíveis de aparecerem nos mesmos.

Exemplos:

carraspana
charro
corrécio
marado
otorrino

piorio
regabofe

Dupla Grafia

Em geral, listámos todas as formas ortográficas possíveis de uma mesma palavra, desde que estejam atestadas quer em dicionários correntes quer em corpora, mesmo que sejam estrangeirismos, regionalismos, etc.

Assim, o contraste dos ditongos alternativos **ou** e **oi** é atestado no nosso dicionário.

Exemplos:

duradouro - duradoiro
louro - loiro
ouro - oiro

Relativamente às palavras terminadas em *-ina* e *-ine* e em *-ão* e *-on*, todas foram listadas, umas como palavras de pleno direito do português, outras como estrangeirismos.

Exemplos:

biberão - biberon
vitrina - vitrine

O mesmo se passa em relação às palavras compostas, como foi atrás referido.

Particípios Passados Duplos

Há verbos que têm dois particípios passados - um regular e outro irregular. Tomemos como exemplo o verbo *salvar*. O seu particípio regular é *salvado* e o seu particípio irregular é *salvo*. O particípio regular usa-se mais frequentemente com o auxiliar *ter*, nos tempos compostos da voz activa.

Exemplos:

tem ganhado,
tem elegido,
tem entregado,
tem juntado,
tem limpado,
tem aceitado,
tem pagado,
tem imprimido.

O particípio irregular usa-se com os auxiliares *ser* e *estar*, quase sempre na voz passiva.

Exemplos:

*está ganho,
foi eleito,
é entregue,
está junto,
foi limpo,
foi aceite,
está pago,
foi impresso.*

Verifica-se, contudo, certa tendência, particularmente com alguns verbos, para a uniformização do particípio, generalizando-se a construção irregular mesmo com o verbo *ter*.

Exemplos:

*tinha ganhado - tinha ganho
tinha elegido - tinha eleito
tinha entregado - tinha entregue
tinha juntado - tinha junto
tinha limpado - tinha limpo*

No Palavroso a forma regular é gerada pelas regras construídas para a sua formação. A forma irregular encontra-se registada num ficheiro de exceções em relação à conjugação dos verbos. De momento, o programa apenas aceita a forma irregular, se esta existir.

Cobertura Quantitativa do Palavroso

Introdução

Neste ponto do relatório apresentaremos os resultados quantitativos no que respeita à "cobertura" do dicionário por nós elaborado, relativamente às várias classes gramaticais da língua portuguesa (convém referir que este dicionário e, por conseguinte, esta contagem, não inclui a lista de estrangeirismos mencionada anteriormente).

Para este fim elaborámos um levantamento do número total de nomes, adjectivos, verbos, palavras compostas, advérbios, "palavras fechadas" (incluem-se neste último grupo artigos, contracções, preposições, interjeições, pronomes, conjunções e os advérbios não terminados em *mente*), que constituem por si só entradas lexicais deste dicionário.

Apresentamos também o resultado quantitativo da aplicação das regras genéricas de análise morfológica pertencentes ao Palavroso, obtendo, deste modo, para cada entrada lexical, a sua forma flexionada. Para o caso dos nomes e adjectivos a formação do plural, diminutivo, aumentativo, do grau superlativo e feminino (estes dois últimos apenas se aplica aos adjectivos, uma vez que são os únicos que podem sofrer estas formações), para o caso dos verbos a flexão em pessoas, tempos e modos.

Contagem do Número Total de Nomes e Adjectivos

O dicionário de nomes e adjectivos contém um total de 34885 entradas lexicais. Destas entradas lexicais 14549 são registadas unicamente como nomes, 15297 são registadas unicamente como adjectivos e 5039 registadas simultaneamente como nomes e como adjectivos.

Somando estas 5039 formas, quer à classe dos nomes, quer à classe dos adjectivos, obtemos um total de 19588 nomes e 20336 adjectivos, num total de 39924 unidades lexicais. Destes nomes 18297 possuem género definido (registados no masculino e no feminino) e 1291 possuem género invariável (uma única forma dá conta do género masculino e feminino). Relativamente à classe dos adjectivos, 14316 são adjectivos com género definido (registados apenas no masculino) e 6020 com género invariável.

Passando a uma análise mais minuciosa podemos referir que cada classe categorial, de nomes e adjectivos, regista nove formas diferentes de ocorrência neste dicionário, contemplando todas as combinações de formação do número e do género possíveis: obtemos, assim, as tabelas apresentadas em baixo. Nestas tabelas, na coluna do parâmetro do número, "I" significa "invariável", "S" significa "singular" e "P" significa "plural" e na coluna do parâmetro do género, "I" significa "invariável", "M" significa "masculino" e "F" significa "feminino". Na coluna do "Número de formas

distintas" encontra-se registado o número de formas diferentes que cada palavra pode ter após a aplicação das diversas regras (de número e género) do Palavroso. Nas colunas das "formação do aum. e dim." encontra-se o número total de formas do aumentativo e diminutivo que cada nome e adjetivo pode ter.

Neste ponto, convém explicitar o motivo pelo qual se registam duas contabilizações diferentes tanto para a classe dos nomes como para a dos adjetivos. De acordo com as regras de formação do aumentativo e diminutivo pertencentes ao programa, estas podem ser aplicadas a todos os nomes e adjetivos. Relativamente aos adjetivos, temos também de considerar para cada um deles a forma no grau superlativo. Para o aumentativo e diminutivo existem duas possibilidades de ocorrência para cada nome ou adjetivo: em *-ão* / *-zão* e *-inho* / *-zinho* respectivamente.

Na verdade, se existem palavras que admitem duas flexões deste tipo, tais como *cadeira* que pode ser tanto *cadeirinha* como *cadeirazinha*, outras existem onde este fenómeno não é possível, como por exemplo *cão* que pode ser apenas *cãozinho* e nunca **cãoinho*. O mesmo acontece em relação à flexão em aumentativo que pode ser em *ão* ou em *-zão*, também aqui existem palavras que podem ser flexionadas tanto numa como na outra forma.

Deste modo, é importante referir que os totais referidos na coluna com 3/4 formas em grau, respectivamente normal, diminutivo e aumentativo são totais por defeito, enquanto que os totais referidos na coluna com 5/6 formas são totais por excesso.

No que respeita à classe dos adjetivos, e mais propriamente aos que são invariáveis quanto ao género, como por exemplo *inteligente*, quando no diminutivo, no aumentativo e no grau superlativo passam a ter tanto a forma feminina como a masculina, assim, *inteligentezinho* / *inteligentezinha*, *inteligentezão* / *inteligentezona* e *inteligentíssimo* / *inteligentíssima* bem como com as formas correspondentes em *-inho*, *-ão*. Os resultados obtidos de todas estas formas encontram-se assinalados na tabela dos adjetivos com um asterisco "*".

Assim, o número real de nomes e adjetivos do português coberto pelo programa encontra-se entre os dois valores apresentados.

Tabela dos **nomes**:

Número	Género	Nomes	Nomes / Adjecti- vos	Total	Número de formas distintas	3 formas: normal 1 dim. 1 aum.	5 formas: normal 2 dim. 2 aum.
I	F	1	0	1	1	3	5
I	I	7	4	11	1	33	55
I	M	9	0	9	1	27	45
P	F	32	0	32	1	96	160
P	I	3	4	7	1	21	35
P	M	39	2	41	1	123	205
S	F	8012	32	8044	2	48264	80440

S	I	321	952	1273	2	7638	12730
S	M	6125	4045	10170	2	61020	101700
Totais:		14549	5039	19588		117225	195375

Tabela dos **adjectivos**:

Número	Género	Adjecti -vos	Nomes / Adjecti- vos	Total	Número de formas distintas	4 formas: (= à tab. anterior) + superl.	6 formas (= à tab. anterior) +superl.
I	F	0	0	0	1	0	0
I	I	29	4	33	1	*231	*363
I	M	1	0	1	2	8	12
P	F	2	0	2	1	8	12
P	I	0	4	4	1	*28	*44
P	M	0	2	2	2	16	24
S	F	23	32	55	2	440	660
S	I	5031	952	5983	2	*83762	*131626
S	M	10211	4045	14256	4	228096	342144
Totais:		15297	5039	20336		312589	474885

Resumindo:

O total por defeito de formas reconhecidas pelo programa é 429814 e o total por excesso é 670260.

No entanto, e para um resultado por excesso mais exacto, como sabemos que todas as palavras terminadas em "s" e "z" não podem ter a forma em *-inho* mas apenas em *-zinho*, subtraímos o total destas palavras, que corresponde a 421 formas, atestadas no nosso dicionário, ao total por excesso apresentado. Assim, o total por excesso será de 669839.

Totais de formas reconhecidas	Mínima	Máxima
Total de nomes e adjectivos reconhecidos	429814	670260
Número de entradas registadas no dicionário	34885	-----

Percentagem de entradas registadas face ao número de formas reconhecidas

$$(34761 * 100) / 39308 = \quad 8.12\% \quad 5.20\%$$

Incremento em relação às entradas registadas (factor de expansão)

$$393081 / 34761 = \quad 12.32 \quad 19.21$$

Como podemos verificar pelos resultados aqui obtidos, no que respeita a nomes e adjectivos, o analisador morfológico reconhece entre 12 a 19 vezes mais palavras do que as existentes no dicionário, ou seja do que as que constituem entradas lexicais. Assim, o total de palavras deste dicionário corresponde apenas a 8.12% ou a 5.20% do total de formas reconhecidas pelo programa.

Contagem do Número Total de Verbos

Relativamente à contagem dos verbos que constituem entradas lexicais no dicionário de verbos, estes totalizam 13040 entradas lexicais registadas no infinitivo, todas as formas em pessoa e número são calculadas automaticamente por regras pertencentes ao analisador morfológico. Estas 13040 entradas são seguidamente multiplicadas por 65, que é o número de formas que cada verbo pode ter, no que se refere à conjugação activa e nas formas simples, ou seja:

Considerando as primeira, segunda e terceira pessoas do singular e as do plural, num total de seis pessoas, flexionadas nos tempos Presente, Pretérito imperfeito, Pretérito perfeito, Pretérito-mais-que-perfeito, Futuro imperfeito, num total de cinco tempos do modo Indicativo, obtemos: 30 formas verbais

Essas mesmas seis pessoas nos tempos Presente, Pretérito imperfeito e Futuro imperfeito, num total de três tempos do modo Conjuntivo, totalizam: 18 formas verbais

Novamente seis pessoas no tempo Presente do modo Condicional: 6 formas verbais

A flexão em apenas duas pessoas (segunda do singular e do plural) o tempo Presente do modo Imperativo: 2 formas verbais

Apenas uma forma flexionada no Gerúndio: 1 forma verbal

A flexão em quatro formas no Particípio passado: 4 formas verbais

E ainda uma única forma no Infinitivo Impessoal: 1 forma verbal

A flexão de seis formas no Infinitivo Pessoal: 6 formas verbais

Resumindo, multiplicando o número total de verbos que constituem entradas lexicais, 13040 como já foi referido, pelo número de flexões que cada verbo (neste contexto específico) pode ter, totalizamos agora 860640 formas verbais cobertas por este dicionário.

Convém, no entanto, referir que este total é um total por excesso, já que não estão aqui contemplados os casos dos verbos que não são flexionados em todos estes tempos ou em todas estas pessoas, chamados verbos **defectivos**.

No entanto, como estes verbos somam um número bastante limitado no total de todos os verbos do nosso dicionário, pensamos não ser relevante o total final que daí advém, pelo que ainda não foi contemplado nesta contagem.

Salientamos, ainda, o facto de o analisador reconhecer, também, quer as formas flexionadas da conjugação reflexa, quer as formas flexionadas resultantes da conjugação pronominal.

Contagem do Número Total de Palavras Compostas

No que respeita ao total das palavras compostas, estas constituem 1245 entradas lexicais no dicionário de palavras compostas. Como para cada uma destas entradas é calculado o número, excepto 169 formas as quais são invariáveis quanto ao número, uma vez que são registadas como "I" ou "P", assim, totalizamos 2152 formas.

No que se refere ao cálculo do género, e relativamente às formas que são atestadas unicamente como nomes, estas estão registadas quer no masculino, quer no feminino (à semelhança do que acontece no dicionário de nomes e adjectivos, e pelas razões anteriormente apresentadas). No entanto, se as formas são atestadas unicamente como adjectivos, o feminino destas palavras é calculado automaticamente pelo programa, tal como já foi referido. Deste modo, o número total de palavras registadas como adjectivos e que são unicamente masculinos (não contamos aqui as formas femininas, nem as invariáveis às quais não se pode aplicar esta regra) é de 152, que com as suas formas femininas vão totalizar 304. As formas que são simultaneamente nome e adjectivo totalizam aqui 53 formas, multiplicando-as por dois obtemos 106 formas, contemplando desta forma o género das que são adjectivos e que sofrem a aplicação da regra em causa.

O número total de formas de palavras compostas cobertas pelo nosso dicionário é de 2562 formas.

Contagem do Número Total de Palavras Gramaticais e Advérbios

O número total de palavras gramaticais é de 288 entradas na sua totalidade. Convém referir que as formas deste tipo de palavras, no que se refere ao número e no que se refere ao género, não são calculadas automaticamente pelo programa.

As palavras que no nosso dicionário são classificadas gramaticalmente como advérbios totalizam a quantia de 3754 entradas lexicais.

Contagem do Número Total de Formas Cobertas pelo Palavroso

Somando os totais de todas as formas lexicais, que podem ser das mais variadas categorias lexicais, cobertas pelo nosso dicionário obtemos, assim, o total:

	Por defeito	Por excesso
Total de formas verbais reconhecidas:	860640	-----
Total de nomes e adjectivos reconhecidos:	429814	670260
Total de palavras compostas reconhecidas:	2562	-----
Total de palavras gramaticais reconhecidas:	288	-----
Total de advérbios reconhecidos:	3754	-----
Cobertura total de todas as formas:	1260339	1476301

Diferenças Relevantes do Novo Acordo Ortográfico e sua Projectão no Palavroso

Após leitura do *Acordo Ortográfico da Língua Portuguesa*, Decreto do Presidente da República nº 43/91, de 23 de Agosto, Resolução da Assembleia da República nº 26/91, a necessidade de aplicação dessas novas alterações ao Palavroso foi notória. Assim, podemos enumerar e apresentar as diferenças relevantes, relativamente aos critérios de ortografia anteriormente usados, que motivaram as adaptações que se seguem:

O alfabeto português passa a ser formado por 26 letras, uma vez que as letras *k*, *w* e *y* passam a fazer parte do mesmo. Estas letras usam-se em casos especiais, nos casos de palavras que derivam de palavras vindas de outras línguas (tanto topónimos, como antropónimos): *Malawi*, *malawiano*, *Franklin*, *frankliniano*, etc.

a) Assim, passam a fazer parte do nosso dicionário palavras provenientes de outras línguas que contêm as consoantes *k*, *w* e *y*, essas palavras poderão ser tanto nomes próprios como derivações dos mesmos. Estas alterações foram motivadas pela norma ortográfica anterior.

As consoantes *c* e *p* nas sequências *cc*, *ct*, *pc*, *pç* e *pt* ora se conservam ora se eliminam:

No que respeita ao ponto 1, imediatamente a seguir, não houve qualquer alteração relativamente ao que estava feito anteriormente no dicionário, uma vez que se encontra em conformidade com o critério de que:

1. São conservados nos casos em que são realizados foneticamente nas pronúncias cultas da língua, como em *compacto*, *convicção*, *convicto*, *ficção*, *friccionar*, *pacto*, *pictural*, *adepto*, *apto*, *díptico*, *erupção*, *eucalipto*, *inepto*, *núpcias*, *rapto*².

Para se efectuar a conformidade com o ponto abaixo foram retiradas as formas que conservavam a consoante etimológica e substituídas pelas formas correspondentes sem a mesma consoante, obedecendo, deste modo, ao critério seguinte:

2. Eliminam-se nos casos em que nunca têm realização fonética nas pronúncias cultas da língua, como em *ação*, *acionar*, *afetivo*, *aflicção*, *aflito*, *ato*, *colecção*, *coletivo*, *direção*, *direto*, *exato*, *objeção*, *adoção*, *adotar*, *batizar*, *Egito*, *ótimo*, etc.

Nos pontos 3 e 4, e devido à ambiguidade existente em se diferenciar o que na verdade se considera ser e fazer parte de uma **língua culta**, optámos por registar ambas as formas, uma com a consoante em questão outra sem essa mesma consoante, respeitando assim as variedades de pronúncia de cada falante obedecendo, também, aos critérios de que:

3. Conservam-se ou eliminam-se facultativamente quando se produzem nas pronúncias cultas da língua, quer geral quer restritamente, ou quando oscilam entre a produção e o emudecimento, como em *aspecto* e *aspeto*, *cacto* e *cato*, *caracteres* e *carateres*, *dicção* e *dição*, *facto* e *fato*, *sector* e *setor*, *ceptro* e *etro*, *concepção* e *conceção*, *corrupto* e *corruto*, *recepção* e *receção*, etc.

4. Nas sequências consonânticas *mpc*, *mpç* e *mpt* se o *p* se eliminar (de acordo com os parâmetros estipulados), o *m* passa a *n*, ficando respectivamente, *nc*, *nç* e *nt*: *assumpcionista* e *assuncionista*, *assumpção* e *assunção*, *assumptível* e *assuntível*, *peremptório* e *perentório*, *sumptuoso* e *suntuoso*, *sumptuosidade* e *suntuosidade*, etc.

b) Relativamente à conservação ou eliminação no nosso dicionário das consoantes etimológicas, surdas ou não, que antecedem outras consoantes, os critérios agora estipulados são bastante diversificados, tal como pudemos verificar.

As consoantes *b* das sequências *bd* e *bt*, *g* da sequência *gd*, *m* da sequência *mn* e *t* da sequência *tm*, conservam-se ou eliminam-se facultativamente, quando se produzem numa forma culta, geral ou restritamente, ou se há oscilação entre a produção e o emudecimento, assim, *súbdito* e *súdito*, *subtil* e *sutil*, *amígdala* e *amígala*, *amigdalácea* e *amidalácea*, *amigdalar* e *amidalar*, *amigdalato* e *amidalato*, *amigdalite* e *amidalite*, *amigdalóide* e *amidalóide*, *amigdalopatia* e *amidalopatia*, *amigdalotomia* e *amidalotomia*, *amnístia* e *anístia*, *amnístiar* e *anístiar*, *indemne* e *indene*, *indenidade* e *indenidade*, *indenizar* e *indenizar*, *omnímodo* e *onímodo*, *omnipotente* e *onipotente*, *omnisciente* e *onisciente*, *aritmética* e *arimética*, *aritmético* e *arimético*, etc.

c) Os critérios que foram referidos para o ponto 3 e 4, anteriormente, são válidos também para este parâmetro, igualmente aqui foram adicionadas as novas grafias, por forma a estar em conformidade com esta norma.

²Todos os exemplos em itálico são transcritos na íntegra do acordo ortográfico acima referido.

Relativamente à acentuação de algumas palavras oxítonas terminadas em *e* tónico (geralmente provenientes do francês), estas palavras admitem tanto o acento agudo como o acento circunflexo: *bebé* ou *bebê*, *bidé* ou *bidê*, *canapé* ou *canapê*, *caratê* ou *caratê*, *croché* ou *crochê*, *guiché* ou *guichê*, *matiné* ou *matinê*, *nené* ou *nenê*, *ponjê* ou *ponjê*, *purê* ou *purê*, *rapê* ou *rapê*, etc. São também admitidas formas como *judo* e *judô*, *metro* e *metrô*, etc.

No que respeita às palavras paroxítonas que têm na sílaba tónica as vogais *e* ou *o* em final de sílaba, quando seguidas das consoantes nasais *m* ou *n* apresentam algumas oscilações na pronúncia culta da língua, assim, *sémen* e *sêmen*, *xénon* e *xênon*, *fémur* e *fêmur*, *vómer* e *vômer*, *Fénix* e *Fênix*, *ónix* e *ônix*, *ténis* e *tênis*, *pónei* e *pônei*, *gónis* e *gônis*, *bónus* e *bônus*, *ónus* e *ônus*, *tónus* e *tônus*, *Vénus* e *Vênus*, etc.

As palavras proparoxítonas são grafadas com acento agudo ou circunflexo, se apresentam na sílaba tónica uma vogal *e* ou *o* as quais são seguidas das consoantes nasais *m* ou *n*, conforme forem pronunciadas nas formas cultas da língua: *académico* ou *acadêmico*, *anatômico* ou *anatômico*, *cénico* ou *cênico*, *cómodo* ou *cômodo*, *fenómeno* ou *fenômeno*, *género* ou *gênero*, *topónimo* ou *topônimo*, *Amazónia* ou *Amazônia*, *António* ou *Antônio*, *blasfémia* ou *blasfêmia*, *fêmea* ou *fêmea*, *gémeo* ou *gêmeo*, *génio* ou *gênio*, *ténue* ou *tênu*e, etc.

É facultativo acentuar as formas verbais de pretérito perfeito do indicativo, como *amámos* e *louvámos* para as distinguir das formas do presente do indicativo como *amamos* e *louvamos*. No entanto, é obrigatório o acento circunflexo em *pôde* (3ª pes. do sing. do pretérito perfeito do indicativo) para se distinguir de *pode* (presente do indicativo). É facultativo em *dêmos* (1ª pes. do plu. do presente do conjuntivo), para se distinguir de *demos* (pretérito perfeito do indicativo); *fôrma* (nome), que é diferente de *forma* (nome e 3ª pes. do sing. do presente do indicativo ou 2ª pes. do sing. do imperativo do verbo *formar*).

d) Os critérios de acentuação também foram uniformizados neste dicionário, para isso procedemos ao acrescentamento das formas acentuadas que não estavam contempladas nas normas ortográficas anteriores.

Nas palavras compostas por justaposição emprega-se o hífen para a ligação dos vários elementos que formam o composto: *ano-luz*, *arcebispo-bispo*, *arco-íris*, *decreto-lei*, *és-sueste*, *médico-cirurgião*, *rainha-cláudia*, *afro-luso-brasileiro*, etc. Certos compostos, em relação aos quais se perdeu, em certa medida, a noção de composição, grafam-se aglutinadamente: *girassol*, *madressilva*, *mandachuva*, *pontapé*, *paraquedas*, *paraquedita*, etc... Nas locuções de qualquer tipo... não se emprega em geral o hífen, salvo algumas excepções já consagradas pelo uso como: *água-de-colónia*, *arco-da-velha*, *cor-de-rosa*, *mais-que-perfeito*, *pé-de-meia*, *deus-dará*, *queima-roupa* (cf. Acordo Ortográfico da Língua Portuguesa 1991). Emprega-se o hífen para ligar duas ou mais palavras que ocasionalmente se combinam (a divisa *Liberdade-Igualdade-Fraternidade*, percurso *Lisboa-Coimbra-Porto*) e bem assim nas combinações históricas ou ocasionais de topónimos (*áustria-Hungria*, *Tóquio-Rio de Janeiro*, etc.).

O uso do hífen com prefixos ou falsos prefixos é bastante diversificado, só se usa hífen:

1. Nas construções em que o segundo elemento começa por *h*: *anti-higiénico*, *co-herdeiro*, *semi-hospitalar*, etc.

Não se usa, no entanto, com os prefixos *des-* e *in-*: *desumano*, *inábilo*, etc.

2. Em construções em que o prefixo ou pseudoprefixo termina com uma vogal igual à vogal do segundo elemento: *anti-ibérico*, *micro-ondas*, etc.

Com o prefixo *co-* este é aglutinado ao segundo elemento: *coocupante*, etc.

3. Em construções com os prefixos *circum-* e *pan-* e o outro elemento começa por vogal, *m*, ou *n*: *circum-escolar*, *circum-murado*, *circum-navegação*, *pan-africano*, *pan-mágico*, *pan-negritude*, etc.

4. Com os prefixos *hiper-*, *inter-* e *super-*, com palavras iniciadas por *r*: *hiper-requintado*, *inter-resistente*, *super-revista*, etc.

5. Com os prefixos *ex-*, *sota-*, *soto*, *vice-* e *vizo-*: *ex-presidente*, *sota-piloto*, *soto-mestre*, *vice-presidente*, *vizo-rei*, etc.

6. Nas construções com os prefixos acentuados graficamente *pós-*, *pré-* e *pró-*: *pós-graduação*, *pós-tónico*, *pré-natal*, etc.

Não se recorre ao uso do hífen:

1. Nas construções em que os prefixos ou falsos prefixos terminam em vogal e o outro elemento é iniciado por *r* ou *s*, estas consoantes deverão geminar: *antirreligioso*, *minissaia*, etc.

2. Em construções em que o prefixo e o segundo elemento se iniciam por vogais diferentes.

A letra minúscula inicial é usada nos dias, meses, estações do ano.

Nos axiónimos e hagiónimos: *senhor doutor*, *bacharel*, *cardeal*, *santa*, etc.

e) Referentemente às palavras compostas, também nestas se procedeu à uniformização com as normas ortográficas em questão, para isso procedemos à inserção no nosso dicionário de formas como *minissaia*, *contrarrevolução*, etc. e retiramos outras como *mini-saia*, *contra-revolução*, etc. Respeitando deste modo os pontos que acabamos de referir.

Nota: Durante a aplicação das normas ortográficas mencionadas, sempre que nos era difícil decidir se em determinado vocábulo deveria ser retirada ou acrescentada a consoante em causa, optámos por registar ambas as formas.

Contabilização Relativamente ao Novo Acordo Ortográfico

Face ao novo acordo ortográfico da língua portuguesa, o nosso dicionário sofre algumas alterações do ponto de vista quantitativo. Deste modo, existe um número de palavras a ser adicionado e um número de palavras a ser retirado do actual dicionário.

As alterações resultantes da aplicação do novo acordo, mostram os seguintes resultados:

	a adicionar	a retirar
1. nomes e adjectivos	1186	91
2. verbos no infinitivo	127	21
3. advérbios	21	63
4. palavras gramaticais	1	0

Bibliografia

BERGSTRÖM, Magnus e Neves Reis, *Prontuário Ortográfico e Guia da Língua Portuguesa*, Lisboa: Editorial Notícias, 19ª ed., 1988

COSTA, J. Almeida, e A. Sampaio Melo, *Dicionário da Língua Portuguesa*, 6ª edição corrigida e aumentada. Porto: Porto Editora, 1992.

CUNHA, Celso e Lindley Cintra, *Nova Gramática do Português Contemporâneo*, Lisboa: João Sá da Costa, 4ª ed., 1987.

ELISEU, André, Alina Villalva, "Tira-teimas: entre Morfologia e Sintaxe", *Actas do VII Encontro da Associação Portuguesa de Linguística*, Lisboa, Junho, 1991, 116-140.

FERREIRA, Aurélio Buarque de Holanda, *Novo Dicionário da Língua Portuguesa*, Rio de Janeiro: Editora Nova Fronteira, 22ª edição revista e aumentada, 1986.

FIGUEIREDO, Cândido de, *Grande Dicionário da Língua Portuguesa*, Bertrand Editora, 23ª ed., 1987.

KOOGAN LAROUSSE, *Dicionário Enciclopédico*, Lisboa, Selecções do Reader's Digest, 1981.

LYONS, John, *Introduction to Theoretical Linguistics*, London, Cambridge University Press, 1968.

MACEDO, Maria Elisa, "Palavras Compostas: Algumas Observações", *in Actas do VIII Encontro da Associação portuguesa de Linguística*, Lisboa, 1992, 271-277.

MACHADO, José Pedro, *Grande Dicionário da Língua Portuguesa*, Amigos do Livro Editores, 1980.

MALHEIROS-POULET, Eugénia; "A vitalidade dos sufixos comparativos -ão e -inho", *In Palavras*, 9 (1986), 61-67.

MEDEIROS, José Carlos, "Ferramentas de Manipulação de Corpora", *In relatório INESC*, rt/ -91 Dezembro, 1992.

MEDEIROS, José Carlos, Rui Marques e Diana Santos, "Português Quantitativo", *In Actas do 1º Encontro de Processamento da Língua Portuguesa Escrita e Falada, EPLP' 93*, Lisboa, Fevereiro, 1993, 33-38.

REIS, Regina, "Dicionários de Língua Corrente: Algumas Considerações", In *Actas do 1º Encontro de Processamento da Língua Portuguesa Escrita e Falada, EPLP' 93*, Lisboa, Fevereiro, 1993, 141-146.

PEDRO, Emília, "À Volta dos Diminutivos - Uma Análise Contrastiva entre o Português e o Inglês", *Actas do VIII Encontro da Associação Portuguesa de Linguística*, Lisboa, Setembro, 1992, 402-417.

SILVA, Emílio e António Tavares, "Dicionário dos Verbos Portugueses", Porto Editora, 1988.

VILLALVA, Alina, "Compounding in Portuguese", *Rivista di Linguistica* 4, I, pp. 201-219, 1992.

WOLF, E. M., B. P. Narumov, A. S. Vaisbord e M. A. Kosarik, *Dicionário Inverso da Língua Portuguesa*, Moscovo, 1971.

ÍNDICE

Introdução.....	2
Critérios Utilizados no Tratamento de Nomes e Adjectivos	2
Tratamento quanto ao Número.....	3
Plural das Palavras terminadas em <i>-ão</i>	3
Tratamento quanto ao Género.....	4
Palavras com género definido	4
Palavras Invariáveis quanto ao Género	8
Tratamento de Aumentativos e Diminutivos.....	9
Exemplos:.....	13
Tratamento do Grau Superlativo dos Adjectivos	13
Tratamento das Palavras Compostas	14
Interacção dos Vários Processos	18
Critérios quanto à Inclusão no Dicionário de Palavras Problemáticas	19
<input type="checkbox"/> Feminino de profissões	20
<input type="checkbox"/> Tratamento de palavras terminadas em <i>-ela</i>	20
<input type="checkbox"/> Estangeirismos	21
<input type="checkbox"/> Neologismos	22
<input type="checkbox"/> Gíria e Calão	22
<input type="checkbox"/> Dupla Grafia	23
<input type="checkbox"/> Particípios Passados Duplos	23
Cobertura Quantitativa do Palavroso.....	24
Introdução	24
Contagem do Número Total de Nomes e Adjectivos.....	25
Contagem do Número Total de Verbos.....	28
Contagem do Número Total de Palavras Compostas	29
Contagem do Número Total de Palavras Fechadas e Advérbios	30
Contagem do Número Total de Formas Cobertas pelo Palavroso	30
Diferenças Relevantes do Novo Acordo Ortográfico e sua projecção no Palavroso	31
Contabilização Relativamente ao Novo Acordo Ortográfico	35
Bibliografia.....	36