

## **Velhas Palavras, Novos Recursos**

*Maria Francisca Xavier  
Maria de Lourdes Crispim  
Graça Vicente*

Centro de Linguística da Universidade Nova de Lisboa

### **Velhas Palavras ...**

O objectivo de elaboração de um dicionário histórico de uma língua (que dê conta da origem e dos diferentes tipos e momentos da evolução das palavras que a constituem) não pode ser concretizado no âmbito de uma pequena equipa, com recursos humanos e materiais limitados. Para um trabalho dessa dimensão, além de outros meios, é ainda imprescindível uma longa tradição de estudos em múltiplas perspectivas, com a intervenção de várias gerações de investigadores, e com o recurso a um volume significativo de dados linguísticos.

É nesta perspectiva que se situa o projecto de elaboração de um dicionário do português medieval, que se assume como um contributo para um projecto de maior fôlego que será um dicionário histórico da língua portuguesa.

Um primeiro trabalho realizado neste âmbito foi a constituição do CIPM – Corpus Informatizado do Português Medieval –, instrumento indispensável para a elaboração de qualquer dicionário de fases antigas da língua, para cujo tratamento foram utilizadas e em alguns casos desenvolvidas diversas ferramentas informáticas para processamento e extracção de informação.

A partir de dados do CIPM, está em curso a elaboração de um *Dicionário de Verbos do Português Medieval* (DVPM), cujo primeiro produto foi divulgado em suporte de papel, o *Dicionário de Verbos do Século 13*, publicado em 1999. O DVPM contém actualmente entradas dos séculos XII e XIII/XIV, já disponíveis “on-line” em <http://cipm.fcsh.unl.pt>. Trata-se de um projecto em curso, de um trabalho sempre provisório, que irá sendo revisto, actualizado e alargado até abranger todos os verbos de um subcorpus do CIPM com textos dos séculos XII a XV.

### **... Novos Recursos**

A disponibilidade de um corpus informatizado e de ferramentas informáticas auxiliares permite-nos encarar o trabalho de constituição de um dicionário com a certeza de que os dados textuais podem ser amplamente tratados sem a necessidade de um investimento humano que, de outro modo, estaria acima das capacidades existentes.

O CIPM é, neste momento, constituído por vários subcorpora, tendo alguns deles sido objecto de tratamentos informáticos diversificados.

Um subcorpus do CIPM – cerca de 500.000 palavras, o que constitui apenas uma amostra do conjunto – encontra-se etiquetado, permitindo a identificação morfosintáctica das palavras<sup>1</sup>. Apresentam-se no Quadro 1 as etiquetas utilizadas:

Quadro 1 – Etiquetas morfosintácticas

N	substantivo	PCL	pronome clítico
NC	substantivo comum	POS	possessivo
NP	substantivo próprio	PI	indefinido
V	verbo	PD	demonstrativo
VINF	verbo no infinitivo	QU	pronome relativo/palavra-QU
VN	verbo no gerúndio ou no participio presente	AD	artigo definido
VPP	verbo no participio passado	AI	artigo indefinido
A	adjectivo	CARD	numeral cardinal
P	preposição	CARDR	numeral cardinal romano
ADV	advérbio	ORD	numeral ordinal
C	conjunção	!	interjeição
PES	pronome pessoal	[[ ]]	amálgama de palavras
			Ex. [[o]]A=_P=O_AD

Exemplifica-se no Quadro 2 o resultado da etiquetagem de um fragmento de um texto do CIPM, extraído da História do Galego-Português, de Clarinda Azevedo Maia:

Quadro 2 – Texto não etiquetado e etiquetado

<p>Texto não etiquetado</p> <p>((L001)) Inno nome de D(eu)s, am(en). Conoszuda cousa segia a tod(os) aq(ue)lles q(ue) este sc(r)ipto uire~ y oyre~ com(m)o eu ((L002)) Aras Diaz, filo de Diago La'a', por mi~ &amp; por miña uoz &amp; eu Marina Rod(r)iguit, fila de Ma(ri)na Diaz ((L003)) q(ue) foy da Regueyra, por nos &amp; por nossas uozes, a uos, ffrey Pa'a'yo, vestiario de Subrado i~ uoz &amp; i~ nome ((L004)) de Don Joh(an) P(ere)z, abbade de Subrado, &amp; du conue~to desse meesmo lugar, dam(os) &amp; offeremos au d(i)to ((L005)) abbade &amp; conuento &amp; a D(eu)s &amp; ad S(an)ta Maria &amp; ad suas uirtudes &amp; a'a's out(r)as q(ue) sam dus outros s(an)tos &amp; s(an)tas ((L006)) eno subred(i)to moesteyro q(u)anta h(e)rdade &amp; uoz &amp; iur &amp; possissom &amp; sinurio nos auem(os) &amp; a' a'uer d(e)uemos ena ((L007)) igrigia de S(an)ta Maria de Cha~tada &amp; som duas septimas d(e) huna noa.</p>
---

<sup>1</sup> O etiquetador utilizado é o de Nuno Marques e Gabriel Lopes (1996).

## Texto Etiquetado

[[Inno]]EM=\_P=O\_AD nome\_NC de\_P D(eu)s\_NP, am(en)\_I . Conoszuda\_VPP cousa\_NC segia\_V a\_P tod(os)\_PI aq(ue)lles\_PD q(ue)\_C este\_PD sc(r)ipto\_NC uire~\_V y\_C oyre~\_V com(m)o\_QU eu\_PES Ara\_NP Diaz\_NP, filo\_NC de\_P Diago\_NP La'a'\_NP, por\_P mi~\_PES &\_C por\_P min~a\_POS uoz\_NC &\_C eu\_PES Marina\_NP Rod(r)iguit\_NP, fila\_NC de\_P Ma(ri)na\_NP Diaz\_NP q(ue)\_C foy\_V [[da]]DE=\_P=A\_AD Regueyra\_NP, por\_P nos\_PES &\_C por\_P nossas\_POS uozes\_NC , a\_P uos\_PES, ffrey\_NC Pa'a'yo\_NP, vestiario\_NC de\_P Subrado\_NP i~\_P uoz\_NC &\_C i~\_P nome\_NC de\_P Don\_NC Joh(an)\_NP P(ere)z\_NP, abbade\_NC de\_P Subrado\_NP, &\_C [[du]]DE=\_P=O\_AD conue~to\_NC [[desse]]DE=\_P=ESSE\_PD meesmo\_PD lugar\_NC, dam(os)\_V &\_C offeremos\_V [[au]]A=\_P=O\_AD d(i)to\_VPP abbade\_NC &\_C conuento\_NC &\_C a\_P D(eu)s\_NP &\_C ad\_P S(an)ta\_NC Maria\_NP &\_C ad\_P suas\_POS uirtudes\_NC &\_C [[a'a's]]A=\_P=AS\_AD out(r)as\_PI q(ue)\_C sam\_V [[dus]]DE=\_P=OS\_AD outros\_PI s(an)tos\_NC &\_C s(an)tas\_NC [[eno]]EM=\_P=O\_AD subred(i)to\_VPP moesteyro\_NC q(u)anta\_QU h(e)rdade\_NC &\_C uoz\_NC &\_C iur\_NC &\_C possissom\_NC &\_C sinurio\_NC nos\_PES auem(os)\_V &\_C a'\_P a'uer\_VINF d(e)uemos\_V [[ena]]EM=\_P=A\_AD igrigia\_NC de\_P S(an)ta\_NC Maria\_NP de\_P Cha~tada\_NP &\_C som\_V duas\_CARD septimas\_NC d(e)\_P huna\_CARD noa\_NC .

Neste caso, é possível obter automaticamente listas das formas correspondentes a cada uma das diferentes classes de palavras, e portanto também dos verbos. Ilustram-se no Quadro 3 as listas obtidas a partir do mesmo texto:

Quadro 3: Listagem de palavras obtidas a partir do texto etiquetado

&_C	18	[[dus]]DE=_P=OS_AD	1
Ara_NP	1	[[ena]]EM=_P=A_AD	1
Cha~tada_NP	1	[[eno]]EM=_P=O_AD	1
Conoszuda_VPP	1	a_P	3
D(eu)s_NP	2	abbade_NC	2
Diago_NP	1	ad_P	2
Diaz_NP	2	am(en)_I	1
Don_NC	1	aq(ue)lles_PD	1
Joh(an)_NP	1	auem(os)_V	1
La'a'_NP	1	a'_P	1
Ma(ri)na_NP	1	a'uer_VINF	1
Maria_NP	2	com(m)o_QU	1
Marina_NP	1	conuento_NC	1
P(ere)z_NP	1	conue~to_NC	1
Pa'a'yo_NP	1	cousa_NC	1
Regueyra_NP	1	d(e)_P	1

Rod(r)iguit_NP	1	d(e)uemos_V	1
S(an)ta_NC	2	d(i)to_VPP	1
Subrado_NP	2	dam(os)_V	1
[[Inno]]EM=_P=O_AD	1	de_P	8
[[au]]A=_P=O_AD	1	duas_CARD	1
[[a'a's]]A=_P=AS_AD	1	este_PD	1
[[da]]DE=_P=A_AD	1	eu_PES	2
[[desse]]DE=_P=ESSE_PD	1	ffrey_NC	1
[[du]]DE=_P=O_AD	1	fila_NC	1

No quadro seguinte podem ver-se os verbos extraídos do mesmo excerto:

Quadro 4: Lista de verbos do excerto etiquetado

a'uer_VINF	1
auem(os)_V	1
Conoszuda_VPP	1
d(e)uemos_V	1
d(i)to_VPP	1
dam(os)_V	1
foy_V	1
offeremos_V	1
oyre~_V	1
sam_V	1
segia_V	1
som_V	1
subred(i)to_VPP	1
uire~_V	1

Em contrapartida, a partir dos textos não etiquetados, a tarefa de compilação das formas de determinada classe de palavras requer uma maior intervenção humana. Para a identificação dos verbos, são extraídas automaticamente listas de todas as palavras. Segue-se uma triagem das formas que são, seguramente, verbais, identificando ainda outras em relação às quais há dúvidas. Tais dúvidas são esclarecidas através da consulta do corpus feita com instrumentos facilmente acessíveis, como processadores de texto disponíveis no mercado<sup>2</sup> e respectivas funções de busca rápida. As formas consideradas verbais são agrupadas sob um lema verbal.

<sup>2</sup> Tem vindo a ser utilizado o Microsoft Word for Windows.

Identificadas as formas, são geradas concordâncias utilizando programas de concordâncias<sup>3</sup>. Procura-se que os contextos à esquerda e à direita sejam suficientemente extensos para, por um lado, decidir se se trata, efectivamente, de formas verbais, e, por outro lado, coligir a informação necessária à construção dos verbetes. As concordâncias de cada verbo são analisadas e limpas de todas as formas erradamente extraídas – por não serem verbais ou por pertencerem a outros verbos.

O resultado é um conjunto de sequências textuais associadas aos verbos a descrever, incluindo todas as formas gráficas atestadas – e deve recordar-se que, neste período, a não havia uma ortografia única –, com registo das respectivas fontes textuais. As referências incluem (i) a data, quando existente, ou o século, na sua ausência, (ii) a sigla adoptada para o texto e (iii) a proveniência geográfica do manuscrito, quando conhecida.

Para a construção do dicionário, foram elaboradas duas fichas. A primeira é constituída por uma grelha destinada ao registo de todas as formas gráficas com a respectiva classificação flexional – modo, tempo e pessoa-número. Quando o verbo surge amalgamado com pronomes clíticos, é registado o conjunto, devidamente assinalado entre chavetas. Regista-se o número de ocorrências, por texto, de cada variante gráfica das formas morfológicas.

A segunda ficha tem carácter lexicográfico: nela se registam (i) a vedeta do verbo, (ii) as diferentes acepções encontradas no corpus, (iii) as correspondentes descrições linguísticas a nível semântico (proposição) e sintáctico (estrutura de subcategorização) e (iv) a atestação mais antiga, acompanhada, eventualmente, de outra(s) mais clara(s) ou mais completa(s).

A informação das fichas é inserida numa base de dados do dicionário<sup>4</sup>, acessível através da Internet aos membros da equipa autorizados a procederem a actualizações da informação.

Um dicionário que visa a descrição diacrónica não pode, obviamente, prescindir de informação sobre a etimologia das palavras que contém. Uma tentativa de incluir informação etimológica levou a um levantamento do que está já consignado em dicionários publicados. Para além de palavras que não se encontram registadas nesses dicionários, e para as quais seria necessário encontrar informação, verificam-se ainda, com frequência, omissões de propostas de etimologia, excesso de indicações “origem obscura” e uma disparidade de étimos e de critérios para a sua selecção que inviabiliza esse “modus operandi”, exigindo uma investigação e uma reflexão mais aprofundadas. Assim, essa informação não é, de momento, introduzida, embora a base de dados esteja já preparada para o seu registo.

<sup>3</sup> Têm sido utilizados o OCP (Oxford Concordance Programme), Oxford University Computing Service, e o Concordance 3.0 (1999, 2000, 2002) de R.J.C. Watt.

<sup>4</sup> Aplicação Web que funciona sobre um servidor Tomcat usando uma base de dados Postgres. Trata-se de um sistema em Java, *open source* e *freeware*.

### Critérios adoptados e problemas

Algumas palavras ocorrem no corpus com uma grande frequência. Trata-se, muitas vezes, de palavras ambíguas: a forma *de*, por exemplo, pode corresponder à preposição ou à terceira pessoa do singular do presente do conjuntivo de DAR. A análise exaustiva das concordâncias destas palavras ambíguas com vista à sua identificação e classificação implicaria um trabalho humano desmesurado e injustificado, pelo que se optou por analisar apenas um subconjunto dessas formas ambíguas, não sendo registado, naturalmente, o número de ocorrências.

Sendo a vedeta entendida como uma representação abstracta do verbo, os critérios de escolha da sua forma gráfica não são triviais. Ao iniciar-se o trabalho conducente à elaboração do *Dicionário de Verbos do Século 13*, considerou-se interessante reflectir na vedeta a grafia da época. Por outro lado, projectava-se já o desenvolvimento do dicionário até ao século XV, o que iria implicar outro tipo de critério. Por outro lado ainda, o dicionário destina-se a um público alargado, incluindo estudantes e investigadores não linguistas, o que determina a opção por formas tão simples e próximas da actualidade quanto possível.

Numa primeira fase, no caso de verbos com formas de infinitivo atestadas, colocaram-se duas hipóteses: escolher a forma mais próxima da moderna ou a mais frequente no corpus. A segunda opção teria conduzido, por exemplo, à escolha de SEER (com 64 atestações no primeiro corpus estudado) em detrimento de SER (com 3 atestações).

A decisão que prevalece neste momento é a de ter como base as atestações mais próximas da forma moderna, independentemente de serem ou não de infinitivo. Um exemplo, desta situação é o verbo *QUEIXAR*. Com três ocorrências no corpus analisado em 1999, *q(ue)ysar* (infinitivo), *q(ue)ysar* (conjuntivo futuro) e *q(ue)yxou* (pretérito perfeito), foi registado como *QUEYXAR* no dicionário então publicado. Presentemente, na versão “on-line”, a forma é a moderna, dado que, entretanto, foi introduzido no corpus um conjunto de textos – *Vidas de Santos de um Manuscrito Alcobacense*, editados por Ivo de Castro *et alii* – com duas outras ocorrências: *q(ue)ixar* e *queixavom*. Manteve-se o critério, mas a vedeta mudou.

Decidiu-se ainda simplificar e modernizar a grafia das vedetas, sem as alterar do ponto de vista da identidade do verbo. Foi, assim, elaborado um conjunto de regras de simplificação ortográfica da vedeta, representadas no Quadro 5:

Quadro 5 – DVPM: Regras de simplificação ortográfica da vedeta

<p>&lt;h&gt; com função de marcação de hiato é eliminado</p> <p>grafemas duplos, iniciais ou internos, com o mesmo valor fonético dos correspondentes simples → grafemas simples</p> <p>&lt;j&gt; ou &lt;y&gt; com valor de vogal → &lt;i&gt;</p> <p>&lt;í&gt;, &lt;gi&gt; ou &lt;y&gt; com valor de sibilante sonora palatal → &lt;j&gt;</p> <p>&lt;u&gt; com valor consonântico → &lt;v&gt;</p>
---

<ll> ou <li> com valor de líquida palatal → <lh>  
 <n> ou <gn> com valor de nasal palatal → <nh>  
 <n> ou <~> representando a nasalação da vogal antes de oclusiva labial → <m>  
 <~> representando a nasalação da vogal antes de consoante não labial → <n>  
 <ç> antes de <e> ou <i> → <c>  
 <s> intervocálico com valor de surda → <ss>  
 <r> intervocálico com valor de múltipla → <rr>

Assim, e para exemplificar, *iuramētar* é representado como *JURAMENTAR*.

À direita de algumas vedetas encontram-se remissões para outros verbos que apresentam uma clara afinidade morfossintáctica e semântica. É o caso, por exemplo, de *COMENDAR*, *ACOMENDAR* e *ENCOMENDAR*, que remetem uns para os outros através de “links”.

Na entrada lexical de alguns verbos estão registadas expressões multipalavra. Trata-se de um domínio em que é necessária uma investigação aprofundada, mas, neste momento, são registadas várias situações, como: (i) sequências que poderão corresponder a expressões idiomáticas, uma vez que o sentido não parece deduzir-se directamente das palavras que as constituem. Por exemplo, *DAR PAZ* significa ‘saudar, desejando a paz’. (ii) expressões cujo sentido global é diferente do actual, como *FAZER AMOR*, que significa ‘demonstrar amor’. E ainda (iii) outras expressões que parecem lexicalizadas, como *FAZER MESTER*, com o sentido de ‘fazer falta, ser necessário’).

Levantaram-se outros problemas de natureza linguística a que foi preciso dar resposta imediata, ainda que devendo ser sujeitos a reflexão posterior. Foi o caso, por exemplo, de como traçar a fronteira entre as formas participiais de natureza verbal e as de natureza adjectival ou nominal.

Foi decidido incluir nas concordâncias todas as formas participiais constantes das listas das palavras do corpus textual seleccionado. Quanto ao dicionário, no caso dos participios passados, e tendo em conta a dificuldade de distinção entre a função de modificador e a função de predicador, ou entre contextos adjectivais e nominais, foi decidido incluir todas as formas encontradas na ficha relativa ao paradigma flexional, até se proceder a uma análise fina de todos os contextos em que ocorrem os participios. Não havendo outras formas claramente verbais atestadas, a vedeta tem a forma participial, não lhe sendo associada uma descrição sintáctica e semântica.

Como resultado da reflexão sobre a necessidade de rever as entradas correspondentes às formas participiais, decidiu-se registar como verbos – com a vedeta no infinitivo – os participios passados passivos e os participios presentes em contextos claramente verbais – por exemplo, com complemento não preposicionado.

Outro problema é o de poder decidir com segurança o modo/tempo de algumas formas atestadas. Assim, nem sempre é possível determinar se as formas são de infinitivo ou de futuro do conjuntivo. Também não é fácil classificar certas formas

de infinitivo como correspondendo ou não a construções de infinitivo pessoal. Nos contextos que, hipoteticamente, admitem a ocorrência de infinitivo pessoal (INFL) e não pessoal (INF), quando a flexão não é visível – na primeira e na terceira pessoas do singular –, não é possível determinar se se trata de uma ou de outra construção. Nestes casos, e porque a base de dados inicialmente utilizada não nos permitia gerir este problema, optou-se por registar a palavra como não flexionada. Foi decidido generalizar esta solução para todos os casos em que o infinitivo não apresenta flexão visível.

Neste momento, porém, porque a base de dados foi, entretanto, aperfeiçoada, já será possível registar as formas duvidosas com duas etiquetas alternativas. Foi esta a solução já adoptada para um outro caso de dúvida: contextos em que ocorrem verbos de tema em a em que não é possível decidir se se trata de uma forma de primeira pessoa do plural do Presente do Indicativo ou do Pretérito Perfeito. A etiqueta adoptada é, neste caso, IP4/IP4.

Embora conscientes de que a interpretação de alguns passos dos textos antigos é muitas vezes difícil, quando não impossível, apresentamos quase sempre uma proposta de aceção, expressa nos equivalentes modernos dos verbos, quando existentes, ou em breves paráfrases que traduzam uma interpretação dos seus sentidos nos diferentes contextos em que os verbos ocorrem. Contudo, nem sempre é possível essa proposta. Exemplo disso é o verbo *FASTAR*, para o qual não se propõe nenhuma aceção, embora se apresente uma hipótese de descrição semântica e sintáctica, como se ilustra abaixo:

#### Quadro 6 – DVPM: Entrada lexical de *FASTAR*

<b>FASTAR</b>	
ocorrências: 3	
CF3	fastar 2 (CHP) ffastar 1 (CHP)
1 ?	
alguém fasta algo(?)	
[ — SN ]	
1294 CHP048 PEs	(e) q(u)ẽ ssa ffora ffastar q(uø) pøyte. C l(i)br(a)s

Outras vezes, a proposta de aceção é acompanhada de uma interrogação. É o caso, por exemplo, da unidade multipalavra *TER A MÃO*:



## Quadro 7 – DVPM: Entrada lexical de TER A MÃO

TER A MÃO

1 possuir(?)

alguém tem a mão algo

[ – SN ]

1281 HGP055 GOr Do a Santa M(ari)a de Mont(e) de Ramo q(u)anto h(er)dam(en)to ey en Morgade <...> cõ todos llos fforos & con todas las perteenzas & cõ todas las uoentades assi com(m)o eu teño a mão ao dia de meu pasam(en)to de jur & de poder

Na proposição, que representa os argumentos seleccionados semanticamente pelo verbo, optou-se por usar termos genéricos para referir esses argumentos. Assim, por exemplo, “alguém” representa seres animados (pessoas/animais), “algo” inanimados, “que...” complementos proposicionais declarativos temporalizados, e “se...” complementos proposicionais interrogativos temporalizados (interrogativas totais). Os complementos oracionais não frásicos (orações pequenas) são representados pelas expressões “em situação” ou “por função”. Os complementos oracionais infinitivos são representados por “fazer” e “acontecer”, consoante se trate de proposições agentivas ou não.

A estrutura de subcategorização é introduzida no dicionário porque, desse modo, investigadores interessados em estudar a complementação verbal podem, mais facilmente, recolher dados com esse objectivo. Nesse campo do dicionário, refere-se a natureza nominal (SN), preposicional (SP) ou frásica dos complementos, com indicação do modo verbal (Find, Fconj, Q Find, etc.). As orações pequenas são representadas como [ – SN SX ]. Tratando-se de complementos infinitivos ou gerundivos, e tendo em conta a complexidade e a diversidade das construções e das análises possíveis, optou-se por omitir a referência a eventuais fronteiras oracionais entre o verbo em causa e o/s complemento/s subcategorizados: a estrutura proposta é [ – (...) Vinf/ger ], sendo que “(...)” representa aqui constituintes vários, como diversas preposições, SN, SP, etc. A representação é problemática, pelo que deverá ser objecto de revisão.

Nem sempre se encontram representados no corpus os argumentos/complementos que se assume serem seleccionados semântica ou categorialmente pelos verbos descritos. A opção feita neste trabalho foi a de considerar que o objecto da descrição é a estrutura lexical do verbo em causa, e que, por conseguinte, ainda que nem todos os argumentos/complementos se encontrem atestados, devem ser registados no dicionário com base no conhecimento lexical do português. Também esta decisão é problemática sempre que a acepção do verbo é pouco clara e as construções em que ocorre são difíceis de analisar.

Os exemplos que ilustram as acepções de cada verbo do DVPM são excertos seleccionados a partir da análise das atestações presentes nas concordâncias. É sempre seleccionada a atestação considerada mais antiga, mas, se existirem outros

excertos mais completos ou mais claros, ainda que não datados, escolhe-se um segundo exemplo, e, em certos casos, um terceiro.

A questão da abonação mais antiga está sempre pendente, uma vez que muitos dos textos não estão datados e os estudos sobre datações conduzem a frequentes reformulações. É o caso, por exemplo, da *Notícia de Haver*, texto inicialmente identificado como sendo do século XIII e posteriormente localizado no século XII, o que levou a considerar todas as suas formas verbais como potenciais atestações mais antigas do corpus. Idêntica situação se verifica sempre que são editados textos portugueses do século XII, o que tem vindo a verificar-se nos últimos anos. Por outro lado, certos textos incorporam outros mais antigos, ou provenientes de outras regiões, o que obriga a referenciar as abonações dos segundos independentemente dos primeiros.

A título de conclusão, exemplifica-se, no Quadro 8, uma entrada lexical do Dicionário na sua actual formulação<sup>5</sup>:

#### Quadro 8 – DVPM: Entrada lexical de QUERER

##### **QUERER** (Cf. QUERER DIZER)

ocorrências: 581

IP1	quero 2 (HGP1, VS1) q(ue)ro 10 (FG2, FR1, HGP2, VS5)
IP2	queres 2 (VS) q(ue)res 1 (VS) quer(e)s 1 (VS)
IP3	quer 20 (FR13, TP4, VS3) q(ue)r 15 (FG6, FR5, VS4) (que)r 1 (VS)
IP4	queremos 5 (FR) q(ue)remos 7 (FR) q(ue)rem(os) 2 (FR1, TOX1) qu(er)emos 4 (FR) qu(er)em(os) 1 (FR) querem(os) 3 (FR2, VS1) queremos 1 (HGP) q(uer)emos 1 (VS)
IP6	q(ue)rẽ 3 (FR2, VS1) querẽ 2 (FR1, TP1) q(ue)ren 1 (CHP)
IP11	queria 2 (VS) q(ue)ria 3 (VS) queria 1 (VS) q(ue)rya 2 (VS)
IP13	queria 2 (FR) q(ue)ria 5 (HGP1, VS4) q(ue)ria 2 (HGP) querria 1 (FR) q(ue)rria 1 (CA) q(ue)rya 9 (FR2, CHP4, VS3) q(u)ir[i]a 1 (NT)
IP16	queriã 1 (FR) q(ue)riã 3 (TÓX) queriam 1 (VS) q(ue)riam 1 (VS)
IPP1	q(ui)se 2 (VS)
[...]	

##### **1 querer**

alguém quer algo

[ — SN ]

1260 CA03 PEs & q(ue) muyto p(re)zo & de q(ue) muyto confio & p(er)a quem q(ue)rria muyta d(e) bõã uentura

<sup>5</sup> Por razões de economia de espaço, a lista das formas verbais atestadas e respectiva classificação foi truncada.

alguém quer fazer/acontecer

[ — Vinf ]

1214? NT PMI E sup(er) s'aiud[a] mādóc lidar seus om(éé)s c(ū) Martin J(o)h(a)n(e)s que  
q(u)ir[i]ja d(e)sūr ar sa irmana

alguém quer que...

[ — que Fconj ]

1280? FR PBA Ben soffremos e queremos q(ue) todollos omees sabyā outras leys por seerē  
mays sabedores

## Referências

- MARQUES, N.; G. Lopes (1996) "Using Neural Nets for Portuguese Part-of-Speech Tagging" in *Proceedings of the 5th International Conference on the Cognitive Science of Natural Language Processing*, Dublin City University.
- XAVIER, M.F.; G. Vicente (1997) – "A Problemática de um Dicionário de Verbos do Século XIII" in AM. Brito et al., orgs. *Sentido que a Vida Faz. Estudos para Óscar Lopes*, Porto, Campo das Letras, pp. 897-904.
- XAVIER, M.F.; G. Vicente; M.L. Crispim orgs. (1999) *Dicionário de Verbos Portugueses do Século 13*, Lisboa, Linha de Investigação 1 do Centro de Linguística da Universidade Nova de Lisboa.
- XAVIER, M.F.; G. Vicente; M.L. Crispim (1999) "A Língua Portuguesa Medieval e o Diálogo Homem-Máquina" in I.H. Faria org., Lindley Cintra. *Homenagem ao Mestre e ao Cidadão*, Lisboa, Edições Cosmos e Faculdade de Letras da Universidade de Lisboa, pp. 749-760.
- XAVIER, M.F. (2000) – "Informação Sintáctica e Semântica num Dicionário de Verbos do Português Medieval" in Mário Vilela org., Coimbra, Almedina (no prelo). (Tradução da comunicação "Semantic and Syntactic Information in a Dictionary of Medieval Portuguese Verbs" apresentada na Conferência sobre "Lexicon and Grammar", Lugo).
- XAVIER, M.F.; M.L. Crispim (2001) "O Corpus e o Dicionário do Português Medieval" in *Anais do II Encontro de Estudos Diacrônicos do Português*, Araraquara, Faculdade de Ciências e Letras da UEP, 2002, pp. 313-321.
- XAVIER, M.F.; M.L. Crispim; M.G. Vicente; M.A. Fiéis; M.C. Silva; Maria Lobo; A. Castro; M.F. Martins; J. Loureiro; S. Dias, A. Reis (2001) "Léxico e Sintaxe do Verbo no Português Medieval" in *Actas do XVI Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Colibri, pp. 561-568.