

A ANOTAÇÃO SINTÁTICA DE TEXTOS MEDIEVAIS PORTUGUESES

SANDRA PEREIRA

Centro de Linguística da Universidade de Lisboa (Portugal)

spereira@clul.ul.pt

RESUMO: Este trabalho tem como principal objetivo descrever o sistema adotado pelo projeto *WOChWEL* para a anotação sintática de textos medievais do português antigo. Paralelamente, pretende-se mostrar as vantagens da utilização deste tipo de anotação para estudar propriedades gramaticais de construções específicas. Como caso ilustrativo, são aqui consideradas as construções com verbos causativos e complemento não finito cuja pesquisa é feita a partir da construção e não do verbo. Para além dos verbos causativos existentes em português europeu contemporâneo (*deixar, fazer e mandar*), os dados revelam a possibilidade de outros verbos (como *achar*, por exemplo) também ocorrerem nestas construções em português antigo.

PALAVRAS-CHAVE: anotação de corpus; português antigo; construções causativas

ABSTRACT: This paper aims to describe the system adopted by the *WOChWEL* project for the syntactic annotation of medieval texts of Old Portuguese. At the same time, we intend to show the advantages of using this type of annotation to study grammatical properties of specific constructions. As an illustrative case, we consider here the constructions with causative verbs and infinitival complements whose search is based on the construction and not on the verb. In addition to the existing causative verbs in contemporary European Portuguese (*deixar, fazer e mandar*), the data shows the possibility of other verbs also occurring in these constructions in Old Portuguese (as *achar*, for example).

KEY-WORDS: Corpus Annotation; Old Portuguese; Causative Constructions

1. INTRODUÇÃO

Este artigo tem como principal objetivo a apresentação do projeto *WOChWEL* (*Word Order and Word Order Change in Western European Languages*) e os sistemas de anotação adotados (POS – *part of speech* – e sintática). Para exemplificar o tipo de pesquisa a efetuar nos textos anotados, são analisadas as construções causativas com complementos infinitivos¹.

São vários os estudos sobre estas construções em português europeu contemporâneo (daqui em diante, PEC), nomeadamente, para o PEC padrão, Gonçalves (1999), Gonçalves e Duarte (2001), Bossaglia (2013), entre outros, e, para o PEC dialetal, Carrilho e Pereira (2010) e Pereira (2012). Há também vários trabalhos sobre estas construções na história do português, designadamente, Silva (1997), Martins (2006), Trannin (2009) e Carvalho (2011).

Referem-se consensualmente os verbos *deixar, fazer e mandar* como os verbos típicos das construções causativas. Em PEC, é possível encontrá-los em construções distintas (*cfr.* Gonçalves e Duarte 2001) : i) a construção de infinitivo flexionado; ii) a

¹ Uma versão preliminar deste trabalho (*cfr.* Pereira 2014) foi apresentada no *CODILIB -Terceiro Colóquio Internacional sobre corpora diacrônicos de línguas ibero-românicas*, Universidade de Zurique, 23-25 de junho.

construção de Marcação Excepcional de Caso (ECM, do inglês, *Exceptional Case Marking*) e iii) a construção de União de Orações (ou *faire-INF*, Kayne 1975). Além destas, Martins (2006) e Bossaglia (2013) identificaram um outro tipo de construção, a construção com infinitivo flexionado e sujeito acusativo, e Pereira (2012), com base nos dados do CORDIAL-SIN, verificou uma outra construção (dialetal), a construção causativa com infinitivo preposicionado. As principais diferenças entre as várias construções encontram-se sobretudo na forma e posição do sujeito da oração infinitiva (o chamado causado) e na existência ou não de flexão no verbo infinitivo.

A **construção de infinitivo flexionado** caracteriza-se por o causado ser um sintagma nominal, que ocorre imediatamente antes do verbo infinitivo, ou um pronome nominativo e o infinitivo exibir flexão, como se mostra em (1):

- (1) a. A mãe deixou **os filhos** brincarem.
b. A mãe deixou **eles** brincarem.

Na **construção de ECM** o causado é igualmente um sintagma nominal, que ocorre entre os dois verbos, ou um pronome acusativo e o infinitivo não flexiona:

- (2) a. A mãe mandou **os filhos** brincar.
b. A mãe mandou-**os** brincar.

Na **construção de União de Orações**, os dois verbos ocorrem em adjacência e o causado pode ser um sintagma nominal ou um pronome acusativo, se o verbo infinitivo for monoargumental (como em (3)), ou um sintagma preposicional ou um pronome dativo, se se tratar de um verbo transitivo (como em (4)):

- (3) a. A mãe mandou brincar **os filhos**.
b. A mãe mandou-**os** brincar.
(4) a. A mãe mandou arrumar os brinquedos **aos filhos**.
b. A mãe mandou-**lhes** arrumar os brinquedos.

A **construção com infinitivo flexionado e sujeito acusativo**, menos frequente, tem como principais características o verbo infinitivo apresentar flexão e o sujeito, quando é pronome, estar em acusativo:

- (5) a. O juiz mandou **os filhos** deixarem a sala.
b. O juiz mandou-**os** deixarem a sala. (Martins 2006: n4)

Na **construção causativa com infinitivo preposicionado**, encontrada em alguns dialetos do PEC, o causado é um sintagma nominal e o infinitivo é introduzido pela preposição *a*:

- (6) Isso mandava sempre a gente **a levantar** ao trabalho. S-EXB18. (Pereira 2012: 344)

A anotação sintática destas construções específicas e a sua pesquisa constitui-se como um dos objetivos deste trabalho.

O artigo está organizado da seguinte forma: na secção 2 apresenta-se o projeto *WOChWEL* mostrando os sistemas de anotação (POS e sintática) que segue; na secção seguinte, destaca-se especificamente a anotação sintática das construções causativas e, na secção 4, a sua pesquisa. Na quinta parte são apresentados os resultados e, finalmente, na secção 6, são enunciadas algumas considerações finais.

2. O PROJETO *WOChWEL*

O *WOChWEL* é um projeto em curso no Centro de Linguística da Universidade de Lisboa (CLUL)². Surge, por um lado, com o intuito de preencher uma lacuna existente no panorama português no que diz respeito à disponibilização de fontes linguísticas para o estudo do português medieval. Nesse sentido, a anotação de textos deste período impõe-se como um dos objetivos fundamentais. Por outro lado, usando as mesmas ferramentas, pretende-se nivelar o português com outras línguas europeias que, para o mesmo estágio, apresentam já *corpora* anotados e pesquisáveis³. Simultaneamente, e no que se refere ao português, o período mais antigo da língua fica representado em textos que são submetidos ao mesmo tipo de anotação de outros textos da história do português, como é o caso dos textos do projeto *Tycho Brahe Corpus*⁴, do *FLY*⁵ e do *CORDIAL-SIN*⁶.

Os textos a anotar no âmbito do projeto *WOChWEL* são de dois tipos: i) textos literários pertencentes ao ciclo arturiano (*post-Vulgata*), *O Livro de José de Arimateia* e *A Demanda do Santo Graal* e ii) alguns textos notariais (editados por Martins 1994).

O Livro de José de Arimateia, edição de Castro (1984), é um texto medieval que chegou até aos nossos dias através de uma cópia do século XVI a partir de um manuscrito do século XIII. A anotação deste texto já se encontra concluída⁷ e a pesquisa das construções causativas (V. adiante secção 4) centra-se exclusivamente neste texto medieval.

A anotação da *Demanda do Santo Graal*, texto medieval que é também uma cópia mas do século XV, a partir da transcrição de Toledo Neto (2012-2015), encontra-se, neste momento, em curso⁸.

² O projeto *WOChWEL* (2012-2015) é financiado pela Fundação para a Ciência e a Tecnologia (FCT PTDC/CLE-LIN/121707/2010) e está totalmente disponível em <http://alfclul.clul.ul.pt/wochwel/>.

³ Por exemplo, o inglês (cfr. <https://www.ling.upenn.edu/hist-corpora/>), o islandês (http://linguist.is/icelandic_treebank/Icelandic_Parsed_Historical_Corpus (IcePaHC) e o francês (http://www.voies.uottawa.ca/corpus_pg_en.html) são línguas que apresentam *corpora* anotados que adotam o mesmo sistema de anotação, o dos *Penn Corpora*.

⁴ O *Corpus Histórico do Português Tycho Brahe* (*Tycho Brahe Corpus*), da Universidade de Campinas, contém textos de autores nascidos entre o século XV e o século XIX. Mais informações em: <http://www.tycho.iel.unicamp.br/~tycho/corpus/>.

⁵ O projeto *Forgotten Letters Years 1990-1974* (*FLY - Cartas Esquecidas*), do CLUL, é constituído por cartas privadas do século XX. Mais informações em: <http://fly.clul.ul.pt/>.

⁶ O *Corpus Dialectal para o Estudo da Sintaxe* (*CORDIAL-SIN*), do CLUL, cobre os dialetos do PEC. Mais informações em: <http://www.clul.ul.pt/pt/Investigacao/212-cordial-sin-syntax-oriented-corpus-of-portuguese-dialects>.

⁷ Quer a versão com anotação POS (cfr. Martins e Pereira 2012-2015) quer a versão com anotação sintática (cfr. Martins et al. 2013-2015) estão disponíveis no site do projeto.

⁸ Veja-se, para a versão com anotação POS, Martins e Pereira (2014-2015) e, para a versão com anotação sintática, Martins et al. (2014-2015).

Baseando-se nestes textos, a equipa pretende estudar tópicos relacionados com a ordem de palavras como a ordem Verbo-Sujeito, extraposição, colocação dos pronomes clíticos, negação, entre outros.

2.1. A anotação morfológica

O *WOChWEL* segue os sistemas de anotação POS e sintática implementados pelo *Tycho Brahe Corpus*, fortemente inspirados nos *Penn Corpora*.

A anotação POS ou morfossintática recorre a etiquetas categoriais e flexionais, conforme se explicita no Manual do Sistema de Anotação Morfológica⁹. Assim, são marcadas todas as classes morfológicas: nomes (N), adjetivos (ADJ), advérbios (ADV), verbos (VB), preposições (P), conjunções e complementadores (CONJ, C), etc. Simultaneamente a anotação dá conta da flexão através das subetiquetas flexionais: plural (-P), feminino (-F), passado (-D), etc.

Os textos do *WOChWEL* são submetidos ao *tagger* de Kepler (2005) que está inserido na ferramenta Edictor¹⁰, desenvolvida pela equipa do *Tycho Brahe Corpus*. Esta ferramenta, além de facilitar a edição e modernização dos textos, permite igualmente o processo de etiquetagem morfológica. Uma vez que estes textos medievais específicos não apresentam uma grafia modernizada, o grau de erro do *tagger* é maior. Veja-se em (7) e (8) uma frase antes e depois de ser submetida:

(7) Quando ela isto ouvio, teve dele mayor dor que dantes e disse-lhe:

(8) Quando/C ela/PRO isto/DEM ouvio/VB-D ,/, teve/TR-D dele/P+PRO mayor/ADJ-R-G dor/N que/C dantes/P+ADV e/CONJ disse-lhe/VB-D+CL :/. JAR109,2.62/ID.

A cada palavra e a cada sinal de pontuação é dada uma etiqueta e é atribuído um código de identificação a cada frase. A versão dos textos que contém apenas anotação POS é disponibilizada para buscas mais simples, essencialmente lexicais e morfológicas, centradas na palavra e/ou na etiqueta. Assim, para quem estiver interessado, por exemplo, em participios passados basta pesquisar pela etiqueta POS *VB-PP* e obterá todas as palavras que foram anotadas como tal (*louvado, feita, reteudo*, etc). Do mesmo modo, um estudioso de morfologia pode pesquisar, por exemplo, palavras terminadas em *-ura* combinadas com a etiqueta *N* e no resultado estarão todos os nomes terminados em *-ura* que ocorrem no texto (*vestedura, direitura, ventura*, entre outras).

2.2. A anotação sintática

A anotação sintática complementa a anotação morfossintática, na medida em que torna possível pesquisar informação mais específica: «Syntactically parsed treebanks are even more useful than POS tagged corpora in linguistic research, as they not only provide part-of-speech information for individual words but also indicate constituent types and membership» (McEnery et al 2006).

No sistema de anotação seguido, o do *Tycho Brahe Corpus*, a versão com anotação POS serve também de *input* para a anotação sintática. São apenas necessários alguns

⁹ Disponível em: <http://www.tycho.iel.unicamp.br/~tycho/corpus/manual/tags.html>.

¹⁰ A ferramenta está acessível no endereço: <http://www.tycho.iel.unicamp.br/~tycho/apps/>

procedimentos no sentido de separar contrações existentes. Esta tarefa é feita automaticamente através de macros concebidas para esse efeito. Assim, a frase (8) acima tem de ser transformada em (9) antes de ser submetida ao *parser*:

- (9) Quando/C ela/PRO isto/DEM ouviu/VB-D ./, teve/TR-D **de@/P @ele/PRO** maior/ADJ-R-G dor/N que/C **de@/P @antes/ADV** e/CONJ **disse@/VB-D @lhe/CL** :/.

Como se pode verificar, as contrações são separadas (*dele* = *de@ @ele*) e o sinal «@» marca a existência prévia de uma contração em cada um dos elementos. O código de identificação da frase é igualmente retirado e, posteriormente, um novo código de identificação é atribuído a cada frase com anotação sintática.

Este tipo de anotação pretende ser teoricamente descomprometida, codificando fronteiras de constituinte e dependências estruturais, fornecendo simultaneamente informação oracional, na medida em que identifica o tipo de oração envolvida. Na Tabela 1 apresentam-se os vários tipos de orações adotados no âmbito do sistema de anotação do *Tycho Brahe Corpus*.

Etiquetas Oracionais	
IP-MAT	oração independente ou coordenada
IP-SUB	oração subordinada
IP-IMP	oração imperativa
IP-IND	oração independente (em CP-QUE e CP-EXL)
IP-INF	oração infinitiva
IP-GER	oração gerundiva
IP-SMC	oração pequena
IP-PPL	oração participial
CP-QUE	oração interrogativa
CP-EXL	oração exclamativa
CP-THT	oração completiva
CP-ADV	oração adverbial
CP-CMP	oração comparativa
CP-REL	oração relativa
CP-FRL	oração relativa livre
CP-CLF	oração clivada
CP-CAR	oração relativa adjunta

Tabela 1: Principais etiquetas oracionais

As etiquetas sintáticas fornecem igualmente informação sobre a natureza categorial (se é um NP, PP, etc.¹¹) e gramatical (SBJ, ACC, DAT, VOC) como se ilustra na Tabela 2.

Etiquetas Sintagmáticas	
ADVP	sintagma adverbial
ADJP	sintagma adjetival
NP	sintagma nominal
NP-SBJ	sintagma nominal (sujeito)
NP-ACC	sintagma nominal (acusativo)
NP-ADV	sintagma nominal (adverbial)

¹¹ Ao contrário das outras categorias, os verbos não projetam, não havendo nenhum nível VP.

NP-VOC	sintagma nominal (vocativo)
NP-DAT	sintagma nominal (dativo)
NP-GEN	sintagma nominal (genitivo)
NP-SE	sintagma nominal (clítico <i>se</i>)
NUMP	sintagma numeral
PP	sintagma preposicional
PP-SBJ	sintagma preposicional (sujeito)
PP-ACC	sintagma preposicional (acusativo)
QP	sintagma quantificador
WXP	sintagma wh (WNP, WPP, WADVP)

Tabela 2: Principais etiquetas sintagmáticas

Alguns constituintes nulos (como os sujeitos) e alguns movimentos (como extraposição, deslocação à direita, subida do clítico, etc.) são também codificados neste tipo de anotação. Vejam-se as principais categorias vazias na Tabela 3.

Categorias vazias	
pro	sujeito nulo referencial
exp	sujeito nulo expletivo
arb	sujeito nulo nas construções causativas
*	sujeito deslocado de uma oração pequena; vestígio deixado pela subida de um clítico
T	vestígio de um operador (por exemplo, nas orações relativas, interrogativas e comparativas)
ICH	vestígio de um constituinte extraposto; vestígio de um constituinte deslocado à esquerda
0	categoria nula; operador nulo

Tabela 3: Categorias vazias

O *parser* usado é o de Bikel (2004). Numa primeira fase, o seu *output* é corrigido semiautomaticamente com o recurso ao *CorpusSearch*¹², um motor de busca concebido para pesquisa e revisão / construção de *corpora* linguísticos que sigam a anotação dos *Penn Corpora*. As funções de revisão automática apresentadas na Tabela 4 são as principais funções usadas pelo *WOChWEL* na correção dos resultados do *parser*.

alterações de etiqueta	
replace_label	substituir etiqueta
append_label	adicionar etiqueta (no final)
prepend_label	adicionar etiqueta (no início)
alterações de estrutura	
add_leaf_before	adicionar ramo (antes)
add_leaf_after	adicionar ramo (depois)
move_up_node	mover nó
move_up_nodes	mover nós
add_internal_node	adicionar nó intermédio
delete_leaf	apagar ramo
delete_node	apagar nó

Tabela 4: Principais funções de revisão automática

¹² <http://corpussearch.sourceforge.net/>

Cada uma destas funções é usada num ficheiro de *query* que corre no resultado do *parser*. Veja-se, por exemplo, o ficheiro apresentado em (10) para alterar etiqueta:

```
(10)
query: ({1}NP* iDoms {2}CL)
      AND ({2}CL iDoms @o|o|@no|no|@lo|lo)
      replace_label{1}: NP-ACC
```

Portanto, neste *query* está especificado o contexto a pesquisar (um sintagma nominal – NP – que domina um clítico *o*) e a ordem a executar (alterar a etiqueta desse NP (**replace_label**{1}: NP-ACC)). Assim, qualquer que seja a etiqueta de um sintagma nominal (NP) que domine um clítico *o* (e suas variantes, *no*, *lo*, em posição proclítica, ou *@o*, *@no* e *@lo*, em ênclise), ela é alterada para NP-ACC.

Veja-se um outro ficheiro para alterar estrutura:

```
(11)
query:(IP-MAT|IP-SUB iDoms !NP-SBJ)
      AND(IP-MAT|IP-SUB
          iDomsFirst
          {1}CONJ|CODE|INTJ|<paren>|*-LFD)
      add_leaf_after{1}: (NP-SBJ *pro*)
```

Neste caso, o contexto de pesquisa é um domínio matriz ou subordinado em que não ocorra um sujeito ((IP-MAT|IP-SUB iDoms !NP-SBJ))¹³ e em que o primeiro constituinte seja CONJ, CODE, INTJ, <paren> ou um constituinte deslocado (-LFD) ((IP-MAT|IP-SUB iDomsFirst {1}CONJ|CODE|INTJ|<paren>|*-LFD)). Encontrado este contexto, é dada a ordem para alterar estrutura e introduzir um sujeito nulo depois do primeiro constituinte (**add_leaf_after**{1}: (NP-SBJ *pro*)).

Os vários ficheiros de *query* são reunidos e ordenados num único arquivo em *perl* que, no momento da sua aplicação, executa todas as alterações descritas nos vários ficheiros.

O resultado do *parser* é um ficheiro de texto (ASCII) que pode ser editado por qualquer editor de texto, apresentando o formato de parentetização indentada, como se mostra seguidamente:

```
( (IP-MAT (IP-MAT (NP-SBJ *pro*)
  (CP-ADV (C Quando)
    (IP-SUB (NP-SBJ (PRO ela))
      (NP-ACC (DEM isto))
      (VB-D ouviu)))
    ( , , )
    (TR-D teve)
    (PP (P de@)
      (NP (PRO @ele)))
    (NP-ACC (ADJ-R-G mayor)
      (N dor)
      (CP-CMP (WADV-1 0)
        (C que)
        (IP-SUB (ADVP-MSR *T*-1)
          (PP (P de@)
```

¹³ O sinal “!” é um operador lógico que exclui o constituinte ao qual se refere.

```

(ADVP (ADV @antes))))))
(CONJP (CONJ e)
  (IP-MAT (NP-SBJ *pro*)
    (VB-D disse@)
    (NP-DAT (CL @lhe))
    (. :))
(...))
(ID JAR109,.63))

```

Além de um editor de texto, o programa *Annotald*¹⁴, concebido no âmbito da anotação do *corpus* do islandês, o *Icelandic Parsed Historical Corpus*, que também adota o formato dos *Penn Corpora*, permite igualmente a edição do *output* do *parser*. Neste formato, a visualização é igualmente sob a forma de indentação, colorida, e a edição é bastante rápida e intuitiva para o utilizador.

O programa *CorpusDraw*, componente do *CorpusSearch*, mostra as frases em formato de árvore e permite abrir e editar os resultados dos ficheiros de *queries* no formato dos *Penn Corpora*.

3. ANOTAÇÃO SINTÁTICA DAS CONSTRUÇÕES CAUSATIVAS

Nesta secção pretende-se demonstrar a utilidade da anotação sintática para o estudo de construções gramaticais específicas, como é o caso das construções causativas.

Em textos não anotados sintaticamente, sem lematização e com grafia instável, a pesquisa das construções causativas impõe-se como uma tarefa morosa e problemática (*cfr.* Davies 2009), na medida em que tem de se olhar para todas as ocorrências dos verbos causativos (*deixar, mandar e fazer*) seguidos de verbos infinitivos. Ora, baseando a pesquisa em três verbos, encerra-se a possibilidade de outros verbos (que possam entrar em construções semelhantes) aparecerem. Adicionalmente, apenas se encontram os casos em que a ordem de palavras corresponde exatamente à ordem de palavras da pesquisa.

Neste sistema de anotação, a particularidade das construções causativas (e percetivas¹⁵) reside na forma como o sujeito da oração infinitiva é codificado, associado ao facto de o verbo infinitivo exibir ou não flexão.

Assim, o sujeito da oração infinitiva pode aparecer sob as seguintes formas:

i) um sintagma nominal pleno, um pronome nominativo ou um sujeito nulo (*pro*) (e o infinitivo apresenta flexão):

- (12) a. A mãe deixou [_{IP-INF} **os filhos**_{NP-SBJ} brincarem].
 b. A mãe deixou [_{IP-INF} **eles**_{NP-SBJ} brincarem].
 c. A mãe deixou [_{IP-INF} ***pro***_{NP-SBJ} brincarem].

ii) uma categoria vazia (*) coindexada com um pronome acusativo que subiu para o domínio finito:

¹⁴ *Cfr.* <http://annotald.github.io/user.html>

¹⁵ Neste trabalho, são apenas consideradas as construções causativas. Ainda que algumas das construções percetivas sejam estruturalmente semelhantes, elas não são aqui abordadas.

(13) A mãe deixou-**os**_{NP-ACC-1} [_{IP-INF} *_{NP-SBJ-1} brincare**m**.

iii) um sujeito nulo (*arb*):

(14) E fez [_{IP-INF} ***arb***_{NP-SBJ} dar pregão por toda sua terra, que qualquer que pudesse dar saude a seu filho, que lhe daria qualquer dom que lhe pedisse. (JAR21,.13)

iv) um sintagma nominal pleno (e o verbo infinitivo não exhibe flexão):

(15) – Certo, disse o cavaleiro, ainda mais fazia, que fazia [_{IP-INF} ressucitar **os mortos**_{NP-SBJ} (JAR21,.28)

(16) – Mais Jesu Cristo, que não leixa [_{IP-INF} **os seus**_{NP-SBJ} espereçer, antes lhes ajuda cada e quando lhe faz mister, me acorreo e trouxe-me ata esta pena. (JAR76,.38)

v) um sintagma preposicional:

(17) Aquele Esprito Santo fazia [_{IP-INF} **aos apóstoles e aos profetas**_{PP-SBJ} falar de Deos, e nom sabião o que i dezião, quanto ãus sandeos. (JAR33,.21)

(18) Mas ele fez [_{IP-INF} jurar **a todos seus criados**_{PP-SBJ} que seu pai o nom soubesse, que nom queria que ele o soubesse ata que ele lhe nom perguntasse se lhe aprazia. (JAR27,.9)

vi) uma categoria vazia (*) coindexada com um pronome acusativo (cfr. (18)) ou dativo (cfr. (19)) que se encontra realizado no domínio superior:

(19) E assi andaram quatro dias, mais agora **os**_{NP-ACC-1} leyxa a estoria [_{IP-INF} *_{NP-SBJ-1} andar na nao por o mar. (JAR79,.100)

(20) A segunda era porque Joseph lhe dissera que **lhe**_{NP-DAT-1} faria [_{IP-INF} *_{NP-SBJ-1} vencer seus imigos, se o quisesse crer, e que lhe faria ganhar alegria sem fim. (JAR37,.7)

vii) uma categoria vazia (*ICH*, do inglês *Interpret Constituent Here*) coindexada com um sintagma nominal deslocado à esquerda:

(21) **Todas estas cousas**_{NP-LFD-1} fazia o Esprito Santo [_{IP-INF} ***ICH***_{NP-SBJ-1} ficar em aqueles que fee tinham em Deos. (JAR33,.22)

Há ainda o caso da construção causativa com infinitivo preposicionado em que no domínio infinitivo não é anotado nenhum tipo de sujeito.

(22) ... entrou em ãa mata e fez i armar suas gentes e mandou a hoste **ũu escudeiro**_{NP-ACC} [_{PP} a [_{IP-INF} ver o que fazião. (JAR47,.14)

Tendo em conta as diferenças de anotação dos vários tipos de sujeitos, a forma como este constituinte é codificado associada ao facto de o infinitivo exhibir ou não flexão (VB-F ou VB) permite pesquisar estas construções de uma forma mais eficaz, como se demonstra na secção seguinte.

4. PESQUISA DAS CONSTRUÇÕES CAUSATIVAS

A pesquisa destas construções é feita recorrendo ao programa *CorpusSearch*. As principais funções de busca disponíveis estão descritas na Tabela 5.

Função	Descrição
CCommands	nem x domina y, nem y domina x e o primeiro nó ramificado que domina x domina y
Dominates	x domina y
iDominates	x domina imediatamente y
iDomsFirst	y é o constituinte imediato de x mais à esquerda
iDomsLast	y é o constituinte imediato de x mais à direita
iDomsMod	x domina y, e o constituinte que intervem entre x e y é z
iDomsOnly	y é o único constituinte imediato de x
iDomsTotal	x tem um número determinado de constituintes imediatos
iDomsViaTrace	x domina imediatamente t e t está coindexado com um constituinte z
Exists	x existe
HasSister	x e y são imediatamente dominados pelo mesmo nó
iPrecedes	x antecede imediatamente e não domina y
Precedes	x antecede e não domina y
IsRoot	x é o nó raiz
SameIndex	x tem o mesmo índice de y

Tabela 5: Principais funções de busca do *CorpusSearch* (cfr. Carrilho e Magro 2010)

Um ficheiro de *query* de busca especifica um contexto de ocorrência de um determinado fenómeno a pesquisar. Veja-se (23):

(23)

```
query: (IP* idoms IP-INF)
      AND (IP-INF idoms {1}NP-SBJ|PP-SBJ)
      AND ({1}NP-SBJ|PP-SBJ idoms !\*pro\*)
      AND (IP-INF idoms VB|ET|HV|SR|TR)
```

O que este ficheiro vai procurar é:

(i) um domínio infinitivo que seja dominado por um qualquer IP (onde se incluem todos os casos da construção de infinitivo flexionado, ECM e União de Orações), condição essa que é expressa na primeira linha ((IP* idoms IP-INF));

(ii) um domínio infinitivo que domine um sujeito que é um sintagma nominal ou preposicional (AND (IP-INF idoms {1}NP-SBJ|PP-SBJ)) e não pode ser nulo (AND ({1}NP-SBJ|PP-SBJ idoms !*pro*));

(iii) um domínio infinitivo que domine um verbo infinitivo sem flexão (AND (IP-INF idoms VB|ET|HV|SR|TR)), que pode ser qualquer verbo para além dos verbos *estar* (ET), *haver* (HV), *ser* (SR) e *ter* (TR)¹⁶.

A maioria das construções causativas é captada nesta pesquisa, nomeadamente as construções envolvidas nos exemplos (14)-(21) acima. Dadas as especificidades da

¹⁶ No sistema de anotação POS do *Tycho Brahe Corpus*, os verbos *estar*, *haver*, *ser* e *ter* têm uma etiqueta própria.

anotação destas construções, o resultado contém apenas exemplos com verbos causativos (e perceptivos, *cfr.* Nota 15).

As construções com infinitivo flexionado, como as dos exemplos (12)-(13) acima, uma vez que não se encontram com o *query* em (23), têm de ser pesquisadas através de um outro ficheiro, apresentado em (24):

(24)
 query: ({1}IP-INF idoms NP-SBJ)
 AND ({1}IP-INF idoms VB-F|ET-F|HV-F|SR-F|TR-F)
 AND ({1}IP-INF HasSister VB-*)

Neste caso, a pesquisa é feita num domínio infinitivo que contenha um sujeito (({1}IP-INF idoms NP-SBJ)) e um verbo infinitivo com flexão visível (VB-F|ET-F|HV-F|SR-F|TR-F). O IP-INF tem como nó irmão um verbo finito (({1}IP-INF HasSister VB-*)) e, portanto, não é dominado por um PP.

Para além das construções em causa, os resultados incluem ainda outras construções com infinitivo flexionado, obrigando a uma seleção dos exemplos (ainda que limitados e em número mais reduzido do que seriam se a pesquisa fosse feita usando a versão com anotação POS¹⁷).

Apenas os casos de infinitivo preposicionado selecionados por verbos causativos não são captados pelas pesquisas anteriores. É necessário um novo ficheiro de *query* que procure os infinitivos preposicionados:

(25)
 query: ({1}PP idoms IP-INF)
 AND (IP-INF idoms !NP-SBJ)
 AND (IP-INF idoms VB|ET|HV|SR|TR)
 AND ({1}PP HasSister NP-ACC)
 AND ({1}PP HasSister VB-*)

Esta pesquisa tem como alvo sintagmas preposicionais que dominem uma oração infinitiva (({1}PP idoms IP-INF)). Essa oração infinitiva não possui nenhum sujeito (AND (IP-INF idoms !NP-SBJ)) e contém um verbo infinitivo sem flexão (AND (IP-INF idoms VB|ET|HV|SR|TR)). Para os resultados serem mais restritos, foram adicionadas mais duas condições que dizem respeito aos nós irmãos do sintagma preposicional: por um lado, é necessária a ocorrência de um complemento direto (AND ({1}PP HasSister NP-ACC)) e de um verbo finito (AND ({1}PP HasSister VB-*)).

Na secção seguinte são apresentados os resultados das várias pesquisas, realçando-se o tipo de causado e a sua posição.

¹⁷ Em alternativa, é possível elencar no ficheiro *query* ou num ficheiro de definições quais são os verbos causativos e, assim, o resultado conteria apenas os verbos causativos a selecionar domínios com infinitivos flexionados. Optou-se por não se definir, à partida, os verbos causativos e extrair todos os verbos que podem ocorrer neste contexto.

5. RESULTADOS

De acordo com a versão anotada sintaticamente de *O Livro de José de Arimateia* (cfr. Martins *et al.* 2012-2015), ocorrem 420 construções com verbos causativos e complementação não finita: 46 com *deixar* (ou *leixar*, cfr. (26)), 298 com *fazer* (cfr. (27)) e 69 com *mandar* (cfr. (28)); há ainda uma ocorrência com *enviar*¹⁸ (cfr. (29)) e seis com o verbo *achar* (cfr. 30)). O texto mostra que, neste período, mais verbos (como *enviar* e *achar*) podiam selecionar estas construções. Indubitavelmente, o verbo mais frequente é o verbo *fazer* com mais de dois terços das ocorrências.

- (26) E assi andaram quatro dias, mais agora os **leyxa** a estoria andar em@ @a nao por o mar. (JAR79,.100)
 (27) E a todos os outros **fez** queimar. (JAR23,.24)
 (28) Como me enganou aquele de que vós me **mandastes** guardar!. (JAR66,.55)
 (29) E **envia**@ @te dizer por mim o anho, aquele que em tua visão te dava os bons manjares que o lobo te tolhia, que tu venceste o lobo. (JAR66,.79)
 (30) E foi mais adiante e **achou** outrossi os cavaleiros jazer. (JAR59,.21)

Entre as diferentes formas que o causado pode assumir neste texto, uma das mais frequentes é ser sujeito nulo, como em (31) e (32):

- (31) E ele mandou logo [_{arb} armar tendas e tendilhóis, que disse que queria ali folgar ata que visse algũa aventura sobre a demanda de@ @a nao. (JAR74,.139)
 (32) E entam fez Celidonis [_{arb} despir todos seus guarnimentos e fez@ @lhe vestir pobres panos, (JAR78,.12)

Note-se que no primeiro exemplo o sujeito nulo é um sujeito com referência arbitrária. Contrariamente, no segundo exemplo o sujeito nulo da oração infinitiva é um sujeito recuperável em contexto e tem como referência *el-rei*, parafraseável por *E entam Celidonis fez el-rei despir todos os seus guarnimentos*¹⁹. Na oração seguinte, o causado aparece sob a forma de um clítico dativo cujo referente é o mesmo da oração anterior (*el-rei*).

O causado pode estar expresso sob a forma de pronomes, nomeadamente um pronome acusativo (como em (33)), o pronome *se* (como em (34)) ou pronomes de primeira e segunda pessoa que, em português, podem ser acusativos ou dativos (como em (35) e (36) respetivamente)²⁰.

¹⁸ Sobre o comportamento deste verbo na construção causativa em galego medieval, cfr. Sousa Fernández (1999).

¹⁹ São frequentes os exemplos em que o sujeito da oração matriz ocorre entre os dois verbos (cfr. i) e ii) e (32) acima) ou a seguir ao complexo verbal (cfr. iii):

i) Então fez **el-rey** ribar os templos onde os deoses eram honrados e britoar os idolos e queymar aquelas cousas que eram sinays de@ @a ley pagam (JAR115,. 35).

ii) O nome de Deos seja louvado pera sempre jamais e deixe **Deos** bem viver e bem obrar aquele que o mandou fazer (JAR119,. 62).

iii) Assi lhes mandou fazer **aquele desleal pagão**, que bem cuidou por este restringimento que renegariam a ley de@ @os cristãos e tornariam a sua (JAR99,. 14).

²⁰ No momento de anotação, por defeito são estes clíticos são NP-ACC.

- (33) Aquele que sempre fogia correrá com aquele que **o** fazia fogir e metê@á -lo- em pavor de morte, (JAR44,.29)
 (34) E então **se** fez bautizar e Joseph foi seu padrinho. (JAR27,.8)
 (35) – E por isto que **me**, Senhor, fizeste entender, eu nom comerei cousa que em nenhum lugar ache. (JAR65,.72)
 (36) E eu **te** farei entender como isto é e como aconteceu. (JAR31,.12)

Repare-se que, quando o causado é um clítico, ele não ocorre no domínio infinitivo mas no domínio superior, em adjacência ao verbo finito (exceto nos casos de interpolação, como em (35)).

O sujeito da oração infinitiva pode aparecer ainda sob a forma de um sintagma nominal (antes ou a seguir ao verbo infinitivo, como em (37) e (38)) ou de um sintagma preposicional. O sintagma preposicional é mais frequentemente introduzido pela preposição *a* (como em (39)-(40), na posição pré e pós-verbo infinitivo, respetivamente) mas também foram encontrados exemplos com a preposição *por*²¹ (como em (41)-(42), na posição pré e pós-verbo infinitivo, respetivamente).

- (37) E el-rei Tolomer, quando assi vio os seus fogir, fez **as outras duas batalhas** ir contra os de Sarafes, (JAR49,.49)
 (38) E deixou ir **o diabo** e mandou@ @lhe que nunca enganasse nhum homem que o sinal de@ @a cruz trouxesse. (JAR54,.21)
 (39) Devemos haver algũa piadade de@ @eles e não os leyxemos comer **a@ @as bestas feras**, (JAR80,.66)
 (40) E depois veio a@ @o bordo de@ @a nao e fez **a um de@ @a sua companhia** escrever i letras. (JAR74,.145)
 (41) E, se tu nom meteres em obras as palavras que ele te manda dizer **por mim**, pequena pessoa, tanto mais seras escarnido e mais confundido. (JAR77,.53)
 (42) Pois esto vos manda Deos **por mim** dizer que vades tirar a desleal imagem que todo o dia oulhades. (JAR57,.36)

Em muitos exemplos, não foi possível identificar a construção envolvida devido quer à omissão do causado quer à ambiguidade estrutural do exemplo. Das várias construções descritas anteriormente (*cf.* Introdução) não foram encontrados exemplos de nenhuma das construções com infinitivo flexionado. Ou seja, neste texto não há atestações da construção de infinitivo flexionado nem da construção com infinitivo flexionado e sujeito acusativo. A construção não ambígua mais frequente é a construção de União de Orações, ilustrada nos exemplos seguintes:

- (43) – Certo, disse o cavaleiro, ainda mais fazia, que fazia ressucitar **os mortos**. (JAR21,.28)
 (44) Mas preguntou@ @lhe se achara ir **um cavaleiro soo**. (JAR88,.31)
 (45) E, de@ @este feito que mandava fazer **a Celidones**, haviam gram pesar os que o faziam e haviam gram doo de@ @a morte de@ @o menino, mas nom podiam al fazer. (JAR67,.75)
 (46) E envia@ @te dizer **por mim** o anho, aquele que em tua visão te dava os bons manjares que o lobo te tolhia, que tu venceste o lobo. (JAR66,.79)

²¹ Sobre a construção *Faire-par*, veja-se, por exemplo, Guasti (2006).

- (47) antes foy sempre falso cristão, como aquele que dentro tinha o coração de@ @o diabo que **lhe** nom leixava fazer boas obras, nem o povo nom cuidava que ele dessimulava enganoso. (JAR103,.13)

Nos exemplos (43)-(44) e (45-46) o sujeito da oração infinitiva é um sintagma nominal e um sintagma preposicional, respetivamente, que ocorre depois do complexo verbal. Em (47) o causado é um pronome dativo. Esta construção ocorre com todos os verbos (o único exemplo com o verbo *enviar*, em (46), é um caso de União de Orações e também há um exemplo com o verbo *achar*, que corresponde a (44) acima).

A construção de ECM, apesar de não ser tão frequente, também se encontra neste texto medieval. Vejam-se alguns exemplos:

- (48) – Ainda, disse o cavaleyro, ele faz mais, que faz **os mudos** falar e da intendimento a@ @os que nom entendem. (JAR83,.97)
- (49) E, depois que os de Sarrat forão cristãos, Josefes se foi por todo o reino e mandou **todos os seus cristãos** pregar, afora tres que guardavam a arca onde a santa escudela era. (JAR53,.64)
- (50) – Mais Jesu Cristo, que não leixa **os seus** espereçer, antes lhes ajuda cada e quando lhe faz mister, me acorreo e trouxe@ @me ata esta pena. (JAR76,.38)
- (51) – Senhor, disse um cavaleiro que ante ele estava, tal é o costume de@ @os cristãos, que jamais não achareis **homem** mentir como eles. (JAR76,.41)

Em (48)-(50) o causado é um sintagma nominal no plural que ocorre entre os dois verbos, sendo estes casos considerados, por isso, instâncias da construção de ECM. O exemplo em (51) que, em PEC, seria um caso ambíguo entre a construção de infinitivo flexionado e a construção de ECM por o infinitivo não apresentar flexão na terceira pessoa do singular, foi aqui analisado como um caso de ECM, uma vez que não há, neste texto, evidência para a existência da outra construção. Todos os verbos, exceto *enviar*, aparecem nesta construção, sendo mais frequente com o verbo *fazer* e *achar*²².

A classificação dos seguintes exemplos é problemática na medida em que i) a ordem dos constituintes corresponde à ordem na construção de ECM e ii) o causado é um sintagma preposicional como na construção de União de Orações:

- (52) E depois veio a@ @o bordo de@ @a nao e fez **a um de@ @a sua companhia** escrever i letras. (JAR74,.145)
- (53) Mas depois fez **a um homem de mui santa vida** escrever esta estoria mui alta de@ @o Santo Greal. (JAR70,.57)

Na linha de Martins (2006), estes exemplos são aqui considerados casos de ECM uma vez que, neste período, há evidência para se afirmar que o sujeito da oração infinitiva podia ser introduzido pela preposição *a*. O texto apresenta ainda exemplos

²² Silva (1997: 290) refere que, em textos notariais dos séculos XIII e XIV, com o verbo *deixar* é mais frequente a construção de ECM (ou impermeável, na designação da autora) do que a construção de União de Orações (construção permeável). No texto literário que aqui se analisa, essa tendência não se confirma: há unicamente um exemplo do verbo *deixar* e do verbo *mandar* na construção de ECM. Apenas com o verbo *achar* se encontram mais exemplos inequívocos da construção de ECM do que da construção de União de Orações.

desta construção em que os dois tipos de sujeito (sintagma nominal e sintagma preposicional) co-ocorrem:

(54) ca meu esprito é de tal força que faz os que bem falão mudecer e os que vem cegar, e faz **os mudos** bem falar e **a@ @os cegos** bem ver. (JAR44,.35)

(55) Mas ele tem gram poder que a duro o podia crer, se o nom vira, que ele faz **a@ @os segos** ver e **a@ @os surdos** ouvir e os mancos são e **os coxos** andar. (JAR83,.95)

Em (54) há duas orações não finitas dependentes do verbo *fazer*: na primeira o sujeito da oração infinitiva é um sintagma nominal (*os mudos*) e na segunda é um sintagma preposicional (*aos cegos*). O mesmo acontece no exemplo (55) com as orações infinitivas dependentes de *fazer*: as duas primeiras têm como sujeito um sintagma preposicional e a última, um sintagma nominal.

Veja-se ainda o seguinte exemplo em que há três complementos não finitos a depender do verbo *fazer*:

(56) E fe@ @la encher d' agua e **a el-rei desnudar@ @se** e meter dentro. (JAR78,.53)

Enquanto na primeira oração o causado é um sujeito nulo (arbitrário), nas outras orações ele é partilhado e assume a forma de um sintagma preposicional (*a el-rei*). A segunda oração é parafraseável por *E fez a el-rei desnudar-se*, havendo um sintagma preposicional entre os dois verbos, como nos exemplos (52) e (53) acima, sendo considerada um caso da construção de ECM. Um outro argumento adicional a esta classificação reside no facto de o clítico *se* ocorrer no domínio infinitivo. De acordo com Gonçalves (1999: 321), na construção de ECM os clíticos anticausativos e reflexos são legitimados no domínio encaixado, podendo ocorrer, contrariamente ao que se passa na construção de União de Orações, onde estes clíticos não podem aparecer²³. Nesse sentido, a terceira oração, ainda que nela o causado não ocorra e se trate de um caso de coordenação, foi classificada como construção de União de Orações.

Um outro exemplo cuja identificação é igualmente problemática é o exemplo (42) acima e aqui repetido como (57):

(57) Pois esto vos manda Deos **por mim** dizer que vades tirar a desleal imagem que todo o dia oulhades. (JAR57,.36)

Por um lado, o facto de o causado ser um sintagma introduzido pela preposição *por* indicia tratar-se de um caso de União de Orações; por outro lado, a ordem de palavras (o causado ocorre entre os dois verbos) é semelhante à da construção ECM. Dado que a ordem dos constituintes na frase era mais livre neste período e que aquele sintagma nunca pode receber caso acusativo, este exemplo foi classificado como instância de União de Orações.

²³ É com base nesta diferença entre as duas construções que é possível identificar os exemplos seguintes como casos da construção de União de Orações já que o clítico *se* não ocorre:

- i) Agora faze tomar aquele que o lião matou e fa@ @lo **levantar** diante os teus deoses. (JAR105,.90).
- ii) E foy a eles e fe@ @los **assentar** e pôr manteis. (JAR94,.13).

No texto encontram-se muitos exemplos ambíguos entre a construção de ECM e a construção de União de Orações, nomeadamente os casos em que o causado está sob a forma de um pronome acusativo e o verbo infinitivo é monoargumental (*cfr.* (58)) e os casos em que o pronome é o mesmo para as formas acusativas e para as formas dativas (como é o caso dos pronomes de primeira e segunda pessoa, como em (59) e (60) respetivamente). Os exemplos em que o sintagma nominal está deslocado (*cfr.* (61)-(62)) constituem igualmente casos de ambiguidade.

- (58) Depois chamou Josefes todos os outros çento e çincoenta, ùus depós os outros, e fe@ @los entrar um e um n' aba de@ @a camiza. (JAR91,.69)
- (59) – Senhor, disse ùu de@ @eles, a pena de acharmos o que demandamos **nos** fez ca entrar, que não sabemos se é antre cristãos, se antre paganos. (JAR80,.14)
- (60) – Rei Ebalac, se me tu quiseres crer, eu **te** farei cobrar toda tua terra. (JAR63,.28)
- (61) **Este** faremos nós tornar de@ @o mao siso em que entrou. (JAR76,.44)
- (62) **Todas estas cousas** fazia o Espirito Santo ficar em aqueles que fee tinham em Deos. (JAR33,.22)

Para além da construção de ECM e da construção de União de Orações foram também observadas no texto ocorrências da construção causativa com infinitivo preposicionado, como se ilustra em (63)-(64):

- (63) E, quando el-rei Ebalac chegou a@ @aquela agoa perto de@ @o castelo, entrou em ùa mata e fez i armar suas gentes e mandou a hoste ùu escudeiro **a ver** o que fazião. (JAR47,.14)
- (64) E entam mandaram um mensageiro **a chamar** Josefes e mandaram@ @lhe dizer todo como aviera, que se tornasse o mais presto que podesse, que muito lhes era mister sua vinda. (JAR103,.40)

Apenas o verbo *mandar* aparece nesta construção. Note-se que nos dados dialetais analisados em Pereira (2012) este verbo era igualmente o que ocorria mais frequentemente nesta construção.

6. CONSIDERAÇÕES FINAIS

Com este trabalho demonstrou-se a utilidade da anotação de textos no estudo de propriedades gramaticais específicas. Os sistemas de anotação adotados permitem que os textos do *WOChWEL* estejam em harmonia com outros textos de outros períodos do português (como os do *Tycho Brahe Corpus*) e, simultaneamente, com outros textos de outras línguas (como os do *Penn Corpora*). Se o estabelecimento da anotação POS facilita, indubitavelmente, as pesquisas lexicais e morfológicas, a implementação da anotação sintática representa um avanço na exploração dos textos, abrindo-se potencialidades de pesquisa mais complexas, mais eficazes e menos morosas. Nesse sentido, o *WOChWEL* constitui um recurso especializado para o estudo da sintaxe do português antigo.

O caso de estudo aqui apresentado são as construções causativas, a sua anotação e pesquisa, no texto medieval *O Livro de José de Arimateia*. Para além dos verbos

consensualmente classificados como causativos em PEC (*deixar, fazer e mandar*), estas construções são igualmente possíveis com os verbos *achar* e *enviar* em português antigo.

Quanto ao tipo de sujeito da oração infinitiva, tal como em PEC, ele pode assumir diversas formas, nomeadamente, um sintagma nominal, um sintagma preposicional, um pronome clítico acusativo ou dativo, um sujeito nulo (arbitrário ou referencial). Quanto à posição do causado na frase, ele pode ocorrer entre os dois verbos ou a seguir ao complexo verbal.

Relativamente às várias construções encontradas, os dados mostram que o infinitivo flexionado ainda não aparece nestas construções. A construção mais frequente é a construção de União de Orações, possível com todos os verbos. Nesta construção, para além de outras formas, o causado pode ser um sintagma introduzido pela preposição *a* ou *por*. A construção de ECM também ocorre no texto, com todos os verbos, exceto *enviar*, sendo a construção mais comum com o verbo *achar*. A construção causativa com infinitivo preposicionado foi igualmente encontrada com o verbo *mandar*.

Como trabalho futuro, pretende-se explorar estas construções em outros *corpora* e em outros períodos da língua.

REFERÊNCIAS BIBLIOGRÁFICAS

- BIKEL, Dan (2004): *On the Parameter Space of Generative Lexicalized Statistical Parsing Models*. Diss. PhD: University of Pennsylvania.
- BOSSAGLIA, Giulia (2013): «Inflected/Non-inflected Infinitive Alternation in Causative and Perception Constructions of Contemporary European Portuguese: A Corpus-based Study», *Procedia – Social and Behavioral Sciences* 95 (2013), Elsevier Ltd, pp. 220-230.
- CARRILHO, Ernestina e Sandra PEREIRA (2010): «Causees in European Portuguese dialects: some observations on the properties and the position of the causee in causative constructions in CORDIAL-SIN», comunicação apresentada no *Wedisyn's First Workshop on Syntactic Variation*, IKER. Bayonne, Março 2010.
- CARVALHO, Lorena e Even NAZARETH BRANDIZZI (2011): *Aspectos relativos à sintaxe de construções causativas no português arcaico e no português atual*. Trabalho de Conclusão de Curso (Bacharelado em Letras Português e Respektivas Literaturas)—Universidade de Brasília, Brasília. Ms.
- CASTRO, Ivo (1984): *Livro de José de Arimateia* (Estudo e Edição do COD. ANTT 643). Dissertação de Doutoramento. Universidade Lisboa.
- DAVIES, Mark (2010): «Creating Useful Historical Corpora: A Comparison of CORDE, the Corpus del Español, and the Corpus do Português», in Andrés Enrique-Arias (ed.), *Diacronía de las lenguas iberorromances: nuevas perspectivas desde la lingüística de corpus*. Frankfurt/Madrid: Vervuert/Iberoamericana, pp. 137-66.
- GONÇALVES, Anabela (1999): *Predicados Complexos Verbais em contextos de infinitivo preposicionado do português europeu*. Lisboa: Universidade de Lisboa. Tese de doutoramento orientada por Inês Duarte.
- GONÇALVES, Anabela e Inês DUARTE (2001): «Predicados complexos com verbos causativos e perceptivos do Português europeu», in Clara Nunes Correia e Anabela Gonçalves (eds), *Actas do XVI Encontro da Associação Portuguesa de Linguística*. Lisboa: APL, pp. 227-239.
- GUASTI, Maria Teresa (2006): «Analytic Causatives» in M. Everaert, H. van Riemsdijk, *The Blackwell Companion to Syntax*. Oxford: Blackwell, pp. 142–172.
- KAYNE, Richard (1975): *French Syntax: The Transformational Cycle*. Cambridge, Mass.: MIT Press.

- KEPLER, Fábio (2005): *Um etiquetador morfo-sintático baseado em cadeias de Markov de tamanho variável*. Dissertação de Mestrado. São Paulo: Instituto de Matemática e Estatística da Universidade de São Paulo.
- MARTINS, Ana Maria (1994): *Clíticos na História do Português*. Lisboa: Universidade de Lisboa. Tese de doutoramento.
- MARTINS, Ana Maria (2006): «Aspects of infinitival constructions in the history of Portuguese», in S. Gess Randall / Deborah Arteaga (eds.), *Historical Romance Linguistics: Retrospective and Perspectives*. Amsterdam & Philadelphia: John Benjamins, pp. 327-355.
- MARTINS, Ana Maria e Sandra PEREIRA (2012-2015): *POS-tagged José de Arimateia*. CC licensed: WOChWEL by Centro de Linguística da Universidade de Lisboa. [Last update 15-01-2015] (Disponível em: http://alfclul.clul.ul.pt/wochwel/documents/Jose_Arimateia.txt).
- MARTINS, Ana Maria e Sandra PEREIRA (2014-2015): *POS-tagged Demanda do Santo Graal* (chapters 01-410). CC licensed: WOChWEL by Centro de Linguística da Universidade de Lisboa. http://alfclul.clul.ul.pt/wochwel/documents/DSG01_410.txt [Last update 29-01-2015].
- MARTINS, Ana Maria, PEREIRA, Sandra e Adriana CARDOSO (2013-2015): *Parsed José de Arimateia*. CC licensed: WOChWEL by Centro de Linguística da Universidade de Lisboa. <http://alfclul.clul.ul.pt/wochwel/documents/José%20Arimateia.psd> [Last update 15-01-2015].
- MARTINS, Ana Maria, Sandra PEREIRA e Adriana CARDOSO (2014-2015): *Parsed Demanda do Santo Graal* (chapters 01-225). CC licensed: WOChWEL by Centro de Linguística da Universidade de Lisboa. http://alfclul.clul.ul.pt/wochwel/documents/DSG01_225.psd [Last update 29-01-2015].
- MCENERY, Tony, Richard XIAO e Yukio TONO (2006): *Corpus-Based Language Studies. An Advanced Resource Book*. London & New York: Routledge.
- PEREIRA, Sandra (2012): *Protótipo de um glossário dos dialetos portugueses com informação sintática*. Lisboa: Universidade de Lisboa. Tese de doutoramento orientada por Ana Maria Martins.
- PEREIRA, Sandra (2014): «A anotação das construções causativas em textos medievais», comunicação apresentada no CODILIB -Terceiro Colóquio Internacional sobre corpora diacrônicos de línguas ibero-românicas, Universidade de Zurique, 23-25 de junho.
- SILVA, Maria Cristina Vieira da (1997): «As construções causativas em textos notariais dos séculos XIII e XIV», *Actas do XXII Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: Associação Portuguesa de Linguística, pp. 289-295.
- SOUSA FERNÁNDEZ, Xulio César (1998): *Estudo diacrónico das construccions con mandar como verbo de orde em galego*. Santiago de Compostela: Universidade de Santiago de Compostela. Tese de doutoramento
- TOLEDO NETO, Sílvio (2012-2015): *Transcrição / Edição da Demanda do Santo Graal*. Manuscrito não publicado.
- TRANNIN, Juliana Batista (2009): «A Sintaxe do infinitivo com verbos causativos na história do português europeu», In *Anais do Seta*, n.º 3, pp. 522-531.