

UNIVERSIDADE DO ALGARVE
Faculdade de Ciências Humanas e Sociais
Departamento de Letras Clássicas e Modernas
Mestrado em Linguística
Especialização em Sintaxe

LÉXICO-GRAMÁTICA
das
FRASES FIXAS
do
PORTUGUÊS EUROPEU
Construções Intransitivas

MARIA DA GRAÇA MADEIRA AMARAL FERNANDES

Faro, 2007

UNIVERSIDADE DO ALGARVE
Faculdade de Ciências Humanas e Sociais
Departamento de Letras Clássicas e Modernas
Mestrado em Linguística
Especialização em Sintaxe

LÉXICO-GRAMÁTICA
das
FRASES FIXAS
do
PORTUGUÊS EUROPEU
Construções Intransitivas

Maria da Graça Madeira Amaral Fernandes

Dissertação de Mestrado, apresentada à
Faculdade de Ciências Humanas e Sociais da
Universidade do Algarve para obtenção do grau de
Mestre em Linguística, especialização em Sintaxe

Faro, 2007

*“Se consegui ver mais longe do que outros
é porque estive sobre ombros de gigantes”*

Isaac Newton

Ao Jorge,
minha cara-metade e minha força

À Diana,
minha alegria

Aos meus pais,
meus impulsionadores de sempre

NOME: Maria da Graça Madeira Amaral Fernandes
DEPARTAMENTO: Letras Clássicas e Modernas
Faculdade de Ciências Humanas e Sociais
Universidade do Algarve
ORIENTADOR: Jorge Baptista
JÚRI: Dolors Català-Guitart, Manuel Célio Conceição, Lucília Chacoto
DATA: 18.04.2007
TÍTULO DA DISSERTAÇÃO: *Léxico-Gramática das Frases Fixas do Português Europeu*
– *Construções Intransitivas*

RESUMO:

O objectivo principal deste estudo consistiu no recenseamento sistemático, na classificação formal e na descrição das propriedades sintácticas das construções verbais fixas intransitivas (sem complemento directo) do Português Europeu. Procedeu-se ao recenseamento de cerca de 900 expressões idiomáticas deste tipo, a partir de diversas fontes lexicográficas de natureza geral e especializada, bem como de *corpora* e com recurso a introspecção, para, de seguida, se delimitar o objecto de estudo e organizar os dados em três classes formais: CP1 (*O Pedro jogava pelo seguro*), CPN (*O Pedro não chega aos calcanhares do João*) e CPP (*O Pedro puxou pelos cordões à bolsa*). Adoptou-se o quadro teórico e metodológico do Léxico-Gramática (M. Gross 1982, 1989, 1996), baseado na gramática transformacional de operadores (Z. S. Harris 1991). Descreveram-se as principais propriedades sintácticas (distribucionais, estruturais e transformacionais) destas construções. A informação linguística foi formalizada em matrizes léxico-sintácticas para sua utilização em diversas aplicações. Nesse sentido, procedeu-se ainda a uma experiência de aplicação das matrizes a um corpus de texto jornalístico de grandes dimensões (c. 180 M palavras), recorrendo a técnicas de transdutores de estados finitos, tendo-se procedido a uma avaliação dos resultados.

Palavras-chave:

frases fixas, expressões idiomáticas, Português Europeu, Sintaxe, Léxico, Processamento Computacional de Linguagem Natural.

NOME: Maria da Graça Madeira Amaral Fernandes
DEPARTAMENTO: Letras Clássicas e Modernas
Faculdade de Ciências Humanas e Sociais
Universidade do Algarve
ORIENTADOR: Jorge Baptista
JÚRI: Dolors Català-Guitart, Manuel Célio Conceição, Lucília Chacoto
DATA: 18.04.2007
DISSERTATION TITLE (in English): *Lexicon-Grammar of Frozen Sentences of European Portuguese – Intransitive Constructions*

ABSTRACT

The main goal of this study is the systematic collection, formal classification and linguistic description of the main syntactic properties of intransitive (i.e., without direct object), frozen sentences (or verbal idioms) of European Portuguese. About 900 sentences of this type were collected from several lexicographic sources, both general and specialized dictionaries of idioms, as well as from corpora and from introspection. They were then classified in three main formal classes: CP1 (*O Pedro jogava pelo seguro* 'Peter does not take chances'), CPN (*O Pedro não chega aos calcanhares do João* 'Peter is not a match to John') and CPP (*O Pedro puxou pelos cordões à bolsa* 'Peter was reluctant to make some expenses'). This research adopted the theoretical and methodological framework of Lexicon-Grammar (M. Gross 1982, 1989, 1996), based on the harrissian transformational operator-grammar (Z. S. Harris 1991). The main syntactic (distributional, structural and transformational) properties of these constructions were described and formalized into lexical-syntactic, binary matrices, in view of several applications, namely on natural language processing. For this, an experience was conducted in order to apply the linguistic information encoded in the matrices to a large size corpus of journalistic text (c. 180 M words), by means of finite-state transducers, and was subsequently evaluated.

Key-words:

frozen sentences, idioms, European Portuguese, Syntax, Lexicon, Natural Language Processing/Computational Linguistics.

Agradecimentos

Gostaria de, em breves palavras, agradecer a algumas pessoas que me ajudaram a percorrer este caminho até ao fim.

Em primeiro lugar, quero agradecer ao Professor Doutor Jorge Baptista com profundo reconhecimento pelo crescimento que me proporcionou neste processo. Agradeço ter-me dado a descobrir esta 'nova' teoria da linguística, de que só tomei conhecimento nesta fase; agradeço-lhe ainda todo o encorajamento, confiança e amizade, e todas as oportunidades que criou para eu me desenvolver e poder contactar com outros especialistas na área. Fez-me crescer intelectual e pessoalmente. Obrigada por tudo.

Às minhas colegas de Mestrado, a Dra. Cristina Palma, a Dra. Anabela Correia e a Dra. Paula Santos, pelos apoios, pelo encorajamento, pela partilha de opiniões.

Quero agradecer ainda a todos os linguistas (sobretudo os Professores das Universidades do Algarve, de Alicante e de Barcelona) com quem contactei, discuti e aprendi diferentes, inovadoras e novas perspectivas, que me fizeram pensar no meu trabalho e melhorá-lo.

Aos meus amigos e familiares agradeço o incentivo e a paciência, principalmente, nesta fase final e mais complicada.

Àqueles que, de alguma forma, contribuiriam para a realização deste projecto, o meu muito obrigada a todos.

Graça Fernandes

Faro, 5 de Fevereiro de 2007

ÍNDICE

Agradecimentos	vi
Índice	vii
Símbolos e convenções	ix
1. Introdução	1
1.1. As principais propriedades das frases fixas	13
1.2. Frases Fixas e Provérbios	20
1.3. Recenseamento	23
1.4. A Classificação das frases fixas	28
1.4.1. Frases fixas com um complemento preposicional	28
1.4.2. Frases fixas com dois complementos preposicionais	30
1.5. Problemas de Classificação e Representação	32
1.5.1. Classe CP1	32
1.5.2. Classe CPN	42
1.5.3. Classe CPP	44
2. Propriedades Sintáticas	48
2.1. Propriedades Distribucionais	48
2.2. Construções Pronominais (Vse)	53
2.3. Construções com negação obrigatória (NegObrig)	56
2.3.1. Diferentes expressões da negação obrigatória	56
2.3.2. Expressões superficialmente afirmativas	58
2.3.3. Aspectos quantitativos da Negação Obrigatória	66
2.4. Propriedades Transformacionais	69
2.4.1. Pronominalização (Pron)	71
2.4.2. Reestruturação Dativa (Rdat)	79
2.4.3. Construções Simétricas (Vsim)	81
2.5. Relações entre construções	84
2.5.1. Relações entre construções verbais fixas	84
Pseudo-conversão	88

2.5.2. Construções verbais fixas e construções verbais livres	90
2.5.3. Complementos Adverbiais	93
2.6. A Representação Matricial	98
3. Reconhecimento de frases fixas em textos: algumas experiências	101
3.1. Métodos	102
3.1.1. Matrizes léxico-sintáticas e dicionários electrónicos	102
3.1.2. Grafos de referência	103
3.1.3. Corpus de trabalho	106
3.1.4. Experiência	108
3.2. Resultados	112
3.3. Discussão	113
3.4. Conclusão	122
4. Conclusão e perspectivas futuras	123
Bibliografia	127

APÊNDICES

Índices

1. Índice Geral: Ordenação alfabética pelo verbo	134
2. Índice Geral: Ordenação alfabética pelos elementos C	142

Matrizes Léxico-Sintáticas

Classe CP1	152
Classe CPN	175
Classe CPP	179

Símbolos e Convenções

<i>Adj</i>	adjectivo
<i>Adv</i>	advérbio
<i>Det</i>	determinante
<i>Prep</i>	preposição
<i>Modif</i>	modificador
<i>C</i>	constante lexical
<i>C₀</i>	sujeito (fixo)
<i>C₁, C₂, C₃</i>	primeiro, segundo e terceiro complemento (distribucionalmente fixo)
<i>E</i>	elemento vazio numa comutação
<i>GN</i>	grupo nominal
<i>Modif-D</i>	modificador à direita da constante
<i>Modif-E</i>	modificador à esquerda da constante
<i>N</i>	nome ou grupo nominal livre
<i>N₀</i>	sujeito (distribucionalmente livre)
<i>N₁, N₂, N₃</i>	primeiro, segundo e terceiro complemento (distribucionalmente livre)
<i>Nop</i>	nome operador
<i>V</i>	verbo
<i>Vop</i>	verbo operador
<i>Vse</i>	construção intrinsecamente pronominal
<i>Vsup</i>	verbo-suporte
<i>w</i>	sequência não especificada de complementos
=	senal que indica equivalência sintáctica e semântica (i.e., transformacional) entre duas frases ou estruturas
=:	senal que especifica a realização lexical de uma categoria ou de uma estrutura
≡	paráfrase; sinónimo de (não implica uma relação transformacional)
≠	senal que indica a não equivalência sintáctico-semântica (i.e., transformacional) entre duas frases ou estruturas
≈	senal que indica que a expressão tem um significado semelhante
?	marca de aceitabilidade duvidosa
?*	marca de aceitabilidade muito duvidosa
*	marca de inaceitabilidade
°	senal que indica que a expressão apresenta apenas um significado literal (ou composicional)
(...)	ligam elementos separados por '+' que podem comutar entre si numa dada posição sintáctica
[...]	especificam a operação sintáctica (transformação) aplicada ou a aplicar
[<i>Rdat</i>]	Reestruturação dativa
[<i>Pron</i>]	Pronominalização
[<i>Pron. Obl.</i>]	Pronominalização pelas formas oblíquas do pronome pessoal
[<i>Pron. Pos.</i>]	Pronominalização pelos pronomes possessivos
[<i>Pass</i>]	Passiva
[<i>X z.</i>]	redução a zero (zeroing) do elemento <i>X</i>

Outras notações serão indicadas *in situ*.