

# EXPRESSÕES IDIOMÁTICAS: FRASES FIXAS INTRANSITIVAS DO PORTUGUÊS EUROPEU<sup>1</sup>

GRAÇA FERNANDES  
(Universidade do Algarve)

*ABSTRACT: the main goal of this study is the systematic collection, formal classification and linguistic description of the main syntactic properties of intransitive frozen sentences 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 classified in three main formal classes (CPI, CPN and CPP). This research adopted the theoretical and methodological framework of Lexicon-Grammar, based on the harrissian transformational operator-grammar. The main syntactic 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 experiment was conducted in order to apply the linguistic information encoded in the matrices to a large size corpus of journalistic text, by means of finite-state transducers, and was subsequently evaluated.*

*KEYWORDS: frozen sentences; idioms; European Portuguese; Syntax; Lexicon; Natural Language Processing/Computational Linguistics.*

## 1. Introdução

Este artigo apresenta de forma sucinta um estudo sobre frases fixas intransitivas do Português Europeu (Fernandes, 2007), mais precisamente as construções verbais fixas com um ou mais complementos preposicionais e sem complementos directos, tais como, por exemplo:

- (1) O Pedro deu à sola
- (2) A Maria não chega aos calcanhares da Ana
- (3) O Mário passou de cavalo para burro

O significado global destas frases pouco ou nada tem a ver com o significado individual dos seus elementos constitutivos nem com as propriedades sintácticas que estes verbos apresentam nas suas construções livres. Nelas o

---

<sup>1</sup> Uma primeira versão deste trabalho foi apresentada no II Fórum de Partilha Linguística (CLUNL, Julho de 2007).

verbo e os complementos são distribucionalmente fixos, ou seja, não admitem variação semanticamente previsível.

Assim, por exemplo: a expressão *ir para os anjinhos* ('morrer') apresenta-se obrigatoriamente com o complemento no diminutivo (\**O João foi para os anjos*); na construção *não chegar aos calcanhares de* ('ser incomparavelmente inferior a alguém em dado aspecto') não é possível substituir o nome parte-do-corpo por outro (\**A Maria não chega aos tornozelos / joelhos / canelas /... da Ana*); e, finalmente, na frase *passar de cavalo para burro* ('piorar de condição') não é possível permutar os dois complementos (\**O Pedro passou para burro de cavalo*) nem colocar qualquer desses nomes no plural (\**O Pedro passou de cavalos para burros*), nem sequer empregar o verbo numa construção transitiva (\**A Maria passou o Pedro de cavalo para burro*). Observa-se, por vezes, que o verbo de uma construção fixa pode apresentar diferenças estruturais relativamente ao número e tipo de complementos com que se constrói habitualmente em frases livres. Assim, por exemplo, o verbo *dar*, na sua construção de predicado de transferência (equivalente a 'oferecer') exige um complemento directo; contudo, na construção fixa, *dar à sola*, não é possível inserir nenhum grupo nominal nessa posição sintáctica: \* *O Pedro deu (isso + essa coisa) à sola*. Noutros casos, o verbo apenas apresenta apenas restrições distribucionais, por vezes surpreendentes, mas mantém no geral a estrutura sintáctica da construção livre; assim acontece, por exemplo, no caso de *Neg chegar aos calcanhares de*, em que a combinação com o *nome-parte-corpo* é única, mas o complemento assemelha-se formalmente ao complemento locativo do verbo de movimento *chegar*: *A Maria não chega à prateleira de cima*. Como é evidente, todas estas restrições decorrem da combinatória particular que constitui cada uma destas frases fixas.

A classificação das construções fixas intransitivas foi feita na perspectiva da elaboração de um Léxico-Gramática da língua portuguesa, isto é, com o objectivo de alistar de forma sistemática e, tanto quanto possível, exhaustivamente este tipo de construções, descrevendo formal e explicitamente as suas principais propriedades sintácticas, com referência ao Português Europeu (Baptista, Correia & Fernandes, 2004, 2005, no prelo). Adopta-se o quadro teórico-metodológico do Léxico-Gramática, desenvolvido por M. Gross (1975, 1982, 1989, 1996) e baseado nos princípios da gramática transformacional de operadores de Z. S. Harris (1976, 1988, 1991).

### 1.1. Recenseamento e critérios de selecção

As frases fixas deste estudo foram obtidas a partir de várias fontes e posteriormente organizadas em classes, com base na sua estrutura formal. As frases fixas foram recolhidas essencialmente a partir de dicionários especializados em expressões idiomáticas (por exemplo, Santos, 1990), e dicionários de língua geral. Para complementar as frases já registadas, consultou-se ainda Vale (2001), do qual se seleccionou apenas as frases fixas comuns ao Português Europeu. Recorreu-se ainda a livros, jornais e revistas, a motores de busca, bem como à nossa competência de falante. Foi possível

constituir, com recurso a todas as fontes acima referidas, uma listagem inicial das frases verbais fixas intransitivas.

Devido ao extenso material recolhido, delimitou-se o conjunto das expressões a descrever, retendo apenas as expressões consideradas de uso corrente. Decidiu-se não integrar frases fixas com uma construção transitiva directa, por exemplo:

(4) O Zé bateu as botas

(5) O Pedro comeu gato por lebre

uma vez que as mesmas estão a ser objecto de um estudo autónomo (Correia, em prep.), as construções com sujeito fixo como se observa em:

(6) Ainda a procissão vai no adro

bem como as construções completivas:

(7) O Mário bradava aos quatro ventos que ali não havia golfinhos

Na próxima secção, apresentaremos sucintamente o quadro geral da classificação das frases fixas proposto por M. Gross (1982a, 1989) e os critérios de classificação das construções preposicionais com um ou dois complementos.

## 1.2. Classificação

A classificação das frases fixas foi feita com base no número e tipo de complementos que as constituem e segue de perto a proposta de M. Gross (1982, 1989) para as construções fixas do Francês. Na sua maioria, as construções aqui estudadas apresentam apenas um complemento preposicional (classe CP1) cuja estrutura interna que pode ser formalizada por:

$N_0$  V Prep  $C_1$

(8) A Ana cantou de galo

e em que o complemento fixo pode ser constituído por um grupo nominal simples, ou, eventualmente, conter um complemento determinativo distribucionalmente livre (classe CPN):

$N_0$  V Prep (C de  $N$ )<sub>1</sub>

(9) O Mário foi na conversa da Ana

Um pequeno número de construções apresenta dois complementos preposicionais (classe CPP) em que,

a) apenas um dos complementos é fixo com o verbo:

$N_0$  V Prep  $C_1$  Prep  $N_2$

(10) O Mário chegou à fala com a Ana

b) ou ambos os complementos são fixos com o verbo:

$N_0$  V Prep  $C_1$  Prep  $C_2$

(11) A Maria deu com a língua nos dentes.

Dado o reduzido número de expressões com dois complementos que foi encontrado, decidiu-se não desdobrar esta classe formal, indicando-se a natureza fixa ou livre dos complementos nas propriedades da respectiva matriz.

Na tabela 1 apresenta-se a classificação adoptada para este estudo e resume-se o estado actual do recenseamento até ao momento.

Classe	Estrutura	Exemplo	Efectivos
CP1	$N_0 V \text{ Prep } C_1$	O Pedro bateu com a porta 'abandonar'	624 (69%)
CPN	$N_0 V \text{ Prep } (C \text{ de } N)_1$	O Pedro foi aos cornos do João 'bater em alguém'	90 (11%)
CPP	$N_0 V \text{ Prep } C_1 \text{ Prep } C_2$	O Pedro foi de cavalo para burro 'ficar pior do que se estava de início'	199 (20%)
Total			913

Tabela 1  
Classificação das frases fixas intransitivas do Português Europeu

Nesta tabela, a primeira coluna apresenta o código convencional que indica a classe formal das frases fixas; na segunda coluna, representa-se a estrutura sintáctica destas frases fixas; na terceira coluna um exemplo ilustrativo; e, finalmente, o número actual de efectivos da classe.

Após a recolha e classificação das frases fixas, observou-se que as construções com um complemento preposicional representam cerca de 80% das frases registadas, 69% em CP1 e 11% em CPN. Os restantes 20% são construções com dois complementos das quais 60% apresentam ambos complementos fixos. Evidentemente, estes números são provisórios, na medida em que estas listagens estão permanentemente a ser actualizadas, mas permitem desde já ter uma ideia aproximada da importância relativa de cada uma destas estruturas.

## 2. Propriedades Sintácticas

De seguida, apresentam-se as principais propriedades sintácticas que foram sistematicamente estudadas para este tipo de construções. Trata-se essencialmente de propriedades distribucionais, estruturais e transformacionais.

### 2.1. Propriedades distribucionais

As propriedades distribucionais dizem respeito às restrições lexicais, sintácticas e semânticas quanto ao preenchimento lexical das posições livres das frases fixas. Estas restrições são definidas da mesma forma como habitualmente caracterizamos as posições argumentais das frases livres (M. Gross 1982a, b; 1989). Para descrever as propriedades distribucionais relativas aos GN livres, seguimos a metodologia proposta por M. Gross (1975).

O conceito de nome humano – *Nhum* – (M. Gross, 1975: 47) é essencialmente de natureza semântica, mas pode ser associado a propriedades formais. Definimos as restrições distribucionais sobre uma posição sintáctica como *Nhum* quando nela podemos inserir nomes próprios como Mário. Pelo contrário, os *N-hum* serão todas as posições sintácticas em que tal não sucede.

O sujeito de todas as frases fixas deste estudo é distribucionalmente livre e pode ser lexicalmente preenchido:

a) por um nome humano (notado *Nhum<sub>0</sub>*):

(12) O Mário faltou à palavra dada

(13) O Mário bateu com o nariz na porta

b) e por um nome não-humano (*N-hum<sub>0</sub>*):

(14) O projecto veio por água abaixo

(15) O barco deu em seco

É interessante verificar que a notação *Nhum* não permite distinguir entre posições sintácticas preenchidas por um nome estritamente humano e aquelas em que nela podem ocorrer extensões de nome humano como, por exemplo, o nome de instituições:

(12a) (O Mário + a empresa) faltou à palavra dada

(13a) (O Mário + ?\*a empresa) bateu com o nariz na porta

(16) (O Mário + ?/?\*a empresa) foi desta para melhor

na medida em que se trata de distinções distribucionalmente muito finas, não as levamos em consideração neste momento.

Portanto, existem frases que admitem tanto nomes humanos como nomes não humanos na posição de sujeito sem que o seu significado se modifique ou sem que a frase se torne agramatical, por exemplo:

(17) (O Mário + o projecto da Câmara + a Economia) vai de mal a pior

Embora admita uma grande variedade de nomes, esta frase não admite completivas (M. Gross, 1975), infinitivas e/ou infinitivas introduzidas por *o facto de*:

(17a) \* (Que isto aconteça + fazer isto + o facto de a Maria ter feito isso) vai de mal a pior

Num reduzido número de casos, a oposição humano/não-humano não parece pertinente para caracterizar as restrições distribucionais sobre uma dada posição argumental.

A natureza distribucional do sujeito pode determinar o significado da construção fixa e levar à necessidade de proceder a desdobramentos lexicais. Tal pode verificar-se no caso de *N<sub>0</sub> falar por si* que pode apresentar duas construções fixas:

(i) *Nhum<sub>0</sub> falar por si* ('falar por iniciativa própria sem ter de prestar contas a mando de ninguém')

(18) O Mário fala por si

(ii) *N-hum<sub>0</sub> falar por si* ('é evidentemente bom')

(19) O currículo da Ana fala por si

Relativamente às posições livres pós-verbais, um grupo nominal livre também pode ser lexicalmente preenchido por um *Nhum* exclusivamente:

(20) A Ana dava à língua com (a Maria + a mãe + o colega)

(20a) \*A Ana dava à língua com (o telefone + a vida)

Com certas frases, os complementos livres que são preenchidos por nomes locativos – notados *Nloc* – (Guillet & Leclère, 1992: 8-11), isto é, trata-se de complementos locativos, na medida em que respondem adequadamente às interrogativas com (Prep) onde?:

(21) O Mário deu com os costados no chão

P: Onde deu o Zé com os costados?

R: No chão.

Há, pois, que considerar duas construções distintas, provavelmente em relação diacrónica uma com a outra: por um lado, o emprego (fixo) literal de dar com os costados em *Nloc* – ‘cair’, ‘bater com as costas em *Nloc*’; por outro lado, o emprego (igualmente fixo) mas autonomizado do primeiro, em que o segundo complemento apresenta nomes como prisão, hospital, etc. e que significa ‘ir parar a’.

Não obstante, é possível encontrar alguns destes complementos com um preenchimento lexical relativamente restrito (designando ‘instituições’ e/ou ‘locais’), de que resulta uma interpretação global disfórica não composicional:

(22) O Mário deu com os costados (no hospital + na prisão)

P: Onde deu o Mário com os costados?

R: (no hospital + na prisão)

sendo estas frases interpretadas globalmente como ‘O Mário está doente’ ou ‘O Mário está preso’, respectivamente.

Em conclusão, como se poderá facilmente observar por uma leitura cursiva das matrizes, nas classes de construções aqui estudadas predominam, do ponto de vista do preenchimento lexical das posições argumentais livres, os grupos nominais do tipo humano; os grupos nominais não humanos são raros; não se observam casos de nomes não restritos; em alguns casos (raros), observamos complementos locativos.

## 2.2. Propriedades estruturais

### 2.2.1. Construções intrinsecamente pronominais

Tal como acontece nas frases livres, observam-se frases fixas em que o verbo apresenta uma construção intrinsecamente pronominal. Trata-se de construções em que o pronome reflexo não pode ser elidido nem substituído por um grupo nominal da mesma natureza distribucional mas não correferente ao sujeito:

(23) O Mário fez-se ao piso à Clara.

(23a) \*O Mário fez (o Mário + o Pedro) ao piso à Clara

(23b) \*O Mário fez ao piso à Clara

Nos casos como o do exemplo acima apenas sucede que, além da presença do pronome reflexo, há ainda que levar em conta outro complemento fixo com o verbo. Este tipo de construção pode observar-se em todas as classes aqui estudadas:

- (24) O Pedro pôs-se na alheta  
 (25) O Mário riu-se na cara do Pedro  
 (26) O Pedro está-se nas tintas para a Maria

Em alguns casos, a construção intrinsecamente pronominal já se observa nas construções livres do verbo:

- (25) O Pedro riu-se na cara da Rita

cp. (27) O Pedro riu-se (da Rita + E)

na construção fixa (25) o nome parte-do-corpo *cara* é distribucionalmente fixo com o verbo. Este verbo na construção livre (27) não admite um complemento locativo deste tipo, o verbo *rir* é, aliás, um exemplo interessante de conservação da propriedade sintáctica na medida em que constitui um dos raros casos em que, a par da construção pronominal, admitem também a construção sem o pronome reflexo:

- (25)=(28) O Mário riu(-se) na cara do Pedro  
 (27)=(29) O Mário riu(-se) (?do Pedro + E)

O caso de *bater-se em duelo* também constitui uma situação interessante na medida em que, a par da construção pronominal:

- (30) O Pedro bateu-se em duelo com o Mário

encontramos a construção transitiva:

- (31) O Pedro bateu o Mário em duelo

Trata-se de duas construções distintas na medida em que a construção pronominal não admite a substituição do pronome reflexo por um grupo nominal:

- (30a) \*O Pedro bateu (o João + ao João) em duelo com o Mário

mas a construção transitiva não admite a pronominalização reflexa do complemento directo (pois este nunca pode ser correferente do sujeito):

- (31a) \*O Pedro bateu-se (= o Pedro) em duelo

As construções intrinsecamente pronominais, conquanto não constituam uma situação excepcional na sintaxe do verbo, levantam problemas de formalização interessantes na perspectiva do seu processamento automático.

Na medida em que o pronome reflexo não pode derivar da redução de um grupo nominal, ele não deve ser levado em conta na representação dos actantes do predicado. Por outro lado, nas condições sintácticas em que se dá a atracção do clítico para a posição pré-verbal (por exemplo, subordinadas, interrogativas parciais, presença de sujeitos indefinidos, certos advérbios quantificadores, etc.) é necessário levar em conta esta permuta do pronome reflexo aquando da identificação da expressão fixa.

### 2.2.2. Construções com negação obrigatória

Certas frases fixas podem ainda apresentar uma construção com negação obrigatória. Pode-se definir esta propriedade como consistindo no facto de uma frase só poder existir na negativa, geralmente expressa por um advérbio de negação (*não*), sem o qual a frase perde o seu valor idiomático ou, eventualmente, pode até tornar-se inaceitável:

(32) O Mário não passa da cepa torta

(32a) \*O Mário passa da cepa torta

A negação não se expressa apenas com o advérbio *não*, verificando-se que, em muitos casos, podem ocorrer outros advérbios de negação, como por exemplo *jamais*, *nem*, *nunca*, entre outros:

(google, 27.01.06<sup>2</sup>): Um Ferrari jamais chega aos calcanhares de um Porsche em estradas normais. Existe um que o consegue acompanhar e mesmo superá-lo, caso esteja a chover <sic> ou a...

Ext 1438109 (des, 98a): Ficou a festinha dos adeptos benfiquistas, que podem nem perder pela demora.

Ext 467893 (pol, 95a): Talvez lhe valha a tradição parlamentar portuguesa no que toca a comissões de inquérito: nunca dão em nada e absolvem sempre o Governo.

A natureza sintático-semântica complexa da negação faz com que, por vezes, esta se possa exprimir de formas variadas o que tem consequências para a identificação automática deste tipo de expressões em textos.

As construções verbais intrinsecamente negativas podem, porém, ocorrer (superficialmente) na afirmativa quando acompanhadas de certos advérbios que ‘enfraquecem’ o carácter afirmativo da proposição, como é o caso de certas construções de mal, dificilmente, entre outros:

Ext 410172 (nd, 94b): [...] quando Cerveira Pereira, enfraquecido pelas febres, mal se tendo nas pernas, [...]

Ext1055194 (soc, 91b): [...] os próprios funcionários consideram que dificilmente chegam para as encomendas.

Em rigor, não se trata de construções negativas, na medida em que não envolvem advérbios de negação. Assimilamo-los, contudo, às construções intrinsecamente negativas, na medida em que estas frases se tornam inaceitáveis se não apresentarem qualquer forma de modalização:

Ext 26357 (des, 93b): Estou com gripe, febres altas e quase não me tenho nas pernas.

cp. \*Estou com gripe, febres altas e quase me tenho nas pernas

A expressão da negação pode também fazer-se a um nível mais elevado da estrutura sintáctica da frase, como sucede no exemplo seguinte, em que o valor negativo se encontra expresso por meio da conjunção subordinativa *sem*:

Ext 748104 (soc, 94a): Cinco soldados da Grã-Bretanha e de Hong Kong lá voltaram a ver a cara dos seus, sem terem ganhado para o susto.

---

<sup>2</sup> Todos os exemplos notados *google*, seguidos de data, foram retirados de [www.google.pt](http://www.google.pt) e os exemplos notados Ext foram retirados de [www.linguateca.pt](http://www.linguateca.pt), por esse motivo não foram atribuídos números a estes exemplos.

Outra forma de marcação da negação consiste no emprego de um sujeito ‘negativo’, como sucede no caso dos pronomes indefinidos ninguém ou nenhum:

Ext 28339 (des, 93b): Mas, neste particular, também ninguém chega aos calcanhars do ucraniano Sergey Bubka.

Neste caso, o valor restritivo negativo destes pronomes sobre a determinação do sujeito permite a expressão deste tipo de construções na afirmativa. Situação semelhante ocorre quando o sujeito é preenchido por pronomes com valor restritivo, como único:

Ext 117271 (pol, 93b): O único que fugiu à regra foi Jean-Claude Duvalier.

Além destas expressões, em que as restrições sobre a determinação do sujeito permitem que a construção se apresente na afirmativa, registamos ainda casos em que o sujeito aparece determinado por um advérbio quantificador com valor restritivo semelhante, como sucede em:

Ext 110476 (soc, 91a): Apenas a Sexta-Feira Santa e o Natal fogem à regra.

O advérbio apenas, com um valor determinativo, incide sobre o sujeito da frase, restringindo a sua referência. A frase só é interpretável em contraponto com outros *N* (o conjunto de nomes de <dias feriados>, neste caso) que “não fogem à regra”.

Portanto, o significado destas expressões intrinsecamente negativas mas expressas na afirmativa não é composicional e só se deixa interpretar por referência à construção canónica, com a negação obrigatoriamente expressa.

Em síntese, a expressão da negação pode ou não fazer-se explicitamente. Quando está expressa (o que no corpus consultado constitui o caso mais geral), a negação é introduzida por diversos processos, entre os quais avulta o emprego dos advérbios de negação (sobretudo não), de pronomes indefinidos negativos (ninguém, nenhum) ou de determinantes com valor restritivo na posição de sujeito da construção (apenas, único); tanto o valor negativo, *stricto sensu*, como o valor restritivo sobre a determinação do sujeito podem ainda manifestar-se em operadores de nível superior (por exemplo, conjunções com valor negativo), o que confirma o carácter intrinsecamente negativo da construção idiomática. Um número residual de casos, em que estas expressões aparecem superficialmente na afirmativa, conquanto sejam mais difíceis de analisar, resultam da interacção entre o significado dos operadores de ordem superior e a construção idiomática que sob eles se encontra encaixada, o que tem como resultado, para a interpretação global da expressão, um valor negativo.

Para terminar, referimos ainda um caso especial, o da expressão (*Neg*) *dar em nada*, que tanto ocorre na negativa como na afirmativa sem que o seu significado global se altera:

Ext 35122 (eco, 93b): Tornaremos muito claro aos nossos parceiros que, ou o banco central europeu virá para Frankfurt, ou todo este negócio não dará em nada.

= Todo este negócio dará em nada.

Ext 70741 (soc, 94b): Mas há diferenças: enquanto a CGTP pensa que a concertação dará em nada, a UGT acredita que, se Cavaco Silva quiser, o negócio far-se-á

= A concertação não dará em nada

Nestes casos, a presença ou ausência do advérbio de negação não altera o significado global da expressão. Tal comprova a não composicionalidade do significado da frase fixa.

### 2.3. Propriedades transformacionais

As transformações são relações parafrásticas, de equivalência sintáctica e semântica entre frases (Harris, 1964, 1991), podendo, pois, ser também consideradas como operações formais que ‘deformam’ uma frase produzindo uma outra frase equivalente. De um modo geral, as frases fixas apresentam maiores restrições à aplicação de operações transformacionais do que as frases livres. Contudo, o facto de as transformações poderem ocorrer nas construções fixas demonstra que estas operam sobre estruturas sintácticas independentemente do significado dessas construções (M. Gross, 1989). Nas próximas secções, descrevem-se as principais propriedades transformacionais que se observam nas frases fixas em estudo.

#### 2.3.1. Pronominalização

A pronominalização (ou redução a pronome) é uma propriedade transformacional muito geral das frases livres e que ocorre igualmente em algumas frases fixas. É a operação que substitui um grupo nominal por um pronome adequado, como é o caso da frase:

(33) O Pedro foi à cara da Ana

(33a) O Pedro foi à cara dela

Nesse sentido, apenas os GN livres das frases fixas podem ser objecto de pronominalização, já que a não composicionalidade da combinação verbo-constante lexical implica que esta última não tenha valor referencial.

Os complementos determinativos livres de *C* podem, naturalmente, pronominalizar-se. Assim, por exemplo, nas frases com a estrutura:

$N_0$  V Prep  $C_1$  de N

de que é exemplo a construção *puxar pela língua de N*:

(23) A Maria puxou pela língua do Paulo

verifica-se que é possível a redução do GN livre *do Paulo* a pronome oblíquo:

[Pron. Obl.] = A Maria puxou pela língua dele

Porém, com a expressão *ir ao pêlo de N*, que apresenta uma estrutura sintáctica semelhante:

(34) O Mário foi ao pêlo do Pedro

verifica-se que essa pronominalização já não pode ocorrer:

(34a) \*O Mário foi ao pêlo dele

Torna-se, pois, necessário descrever para cada construção o conjunto de pronominalizações que sobre elas podem operar ou, inversamente, que lhes estão interditas. A possibilidade da pronominalização de um GN depende de condições formais mas também de factores lexicais, nomeadamente dos elementos lexicais da construção em que este se encontra.

Neste estudo, consideraram-se as seguintes situações de redução a pronome:

- redução de complemento determinativo a pronome possessivo (*de N = Pos*);
- redução de GN a pronome oblíquo (*Obl*).

Os complementos dativos resultantes da reestruturação do GN são sempre pronominalizáveis (pela forma dativa dos pronomes pessoais), razão por que não trataremos deles aqui.

Na classe CP<sub>1</sub> não há complementos determinativos. A estrutura sintáctica e a fixidez dos seus complementos não permitem, pois, este tipo de transformação.

Trataremos em primeiro lugar da redução do complemento determinativo a pronome (*de N = Pos*). É na classe CPN que encontramos a maior parte das construções com complemento determinativo. A maioria das frases inseridas em CPN aceita a redução a possessivo:

(35) A Paula caiu nos braços do Pedro

o complemento determinativo *do Pedro* poderia ser reduzido a pronome possessivo *os seus*:

[Pron. Pos.] = A Paula caiu nos seus braços

No caso da redução a pronome oblíquo dos complementos determinativos livres (já ilustrado acima), verifica-se que apenas 15% das frases não admitem esta operação:

*Nhum<sub>0</sub>* puxar pela língua de *Nhum<sub>1</sub>*

(23) A Maria puxou pela língua da Carla

Relativamente ao preenchimento lexical dos complementos determinativos, na maior parte destas frases fixas ocorrem os nomes humanos.

Na classe CPN observámos alguns complementos em que tanto podem ocorrer nomes humanos como nomes não-humanos. Tal é o caso da construção:

*Nhum<sub>0</sub>* cair na armadilha de *N+/-hum<sub>1</sub>*

que encontramos atestada no *corpus*:

Ext 566247 (clt, 94a): P. – Para cair na armadilha do Borges?

Ext 305811 (nd, 98b): Na defesa e na política externa está o retrato de um governo que se deixou cair na armadilha da gestão corrente.

Ext 329760 (pol, 94a): Cabe-nos preparar a nossa participação nessa revisão do Tratado sem precipitações, e sem cair na armadilha dos debates estereotipados sobre o modelo institucional.

Naturalmente, a redução a pronome só ocorre quando o complemento se encontra preenchido por *Nhum*:

[Pron. Obl.]= Para cair na armadilha dele?

[Pron. Pos.]= Para cair na sua armadilha?

já que não é adequado estabelecer qualquer relação anafórica entre estes pronomes e *GN=N-hum* como os que observámos nos exemplos acima:

[Pron. Obl.]= ?\*Na defesa e na política externa está o retrato de um governo que se deixou cair na armadilha dela.

[Pron. Pos.]= ?\*Na defesa e na política externa está o retrato de um governo que se deixou cair na sua armadilha.

[Pron. Obl.]= ?\*Cabe-nos preparar a nossa participação nessa revisão do Tratado sem precipitações, e sem cair na armadilha deles.

[Pron. Pos.]= ?\*Cabe-nos preparar a nossa participação nessa revisão do Tratado sem precipitações, e sem cair na sua armadilha.

Para tal, a língua recorre a pronomes determinativos como o emprego dos demonstrativos:

[Pron. Dem.]= [...] cair (nessa + naquela + nesta) armadilha # os debates.

[Pron. Dem.]= [...] cair (nessa + naquela + nesta) armadilha # a gestão corrente

Apesar de as pronominalizações só afectarem cerca de 10% das frases fixas deste estudo – essencialmente na classe CPN – é de notar que a grande maioria destas estruturas sintácticas reúne as condições formais para a aplicação das pronominalizações.

Os diferentes tipos de pronominalização dos complementos livres das frases fixas foram sistematicamente investigados, verificando-se que esta propriedade depende em grande medida dos elementos lexicais que formam a combinatoria fixa destas expressões.

### 2.3.2. Reestruturação Dativa

Outra das transformações que pode operar nas frases fixas deste estudo é a Reestruturação Dativa [Rdat]. Considere-se, por exemplo, as frases:

(36) O João foi às trombas do Pedro

[Rdat]= O João foi às trombas ao Pedro

A Reestruturação Dativa – tal como lhe chama Leclère (1995) – é um tipo específico de reestruturação que divide um grupo nominal (as trombas do Pedro) em dois constituintes as trombas e ao Pedro, passando o complemento determinativo (do Pedro) a complemento dativo (ao Pedro), o qual passa então a estar sintacticamente ligado de forma mais directa ao verbo (ir). As frases com o complemento dativo resultante desta transformação são, em geral, mais naturais se este se encontrar pronominalizado:

(36a) O João foi-lhe às trombas

Ora, tal relação não se observa em:

(37) A Maria fala pela boca da Ana

(37a)\*A Maria fala pela boca à Ana

(37b)\*A Maria fala-lhe pela boca

Tal como noutras situações, podemos confirmar que apenas uma análise caso a caso permite delimitar adequadamente o domínio de aplicação desta operação formal.

### 2.3.3. Construções simétricas

Outra propriedade transformacional das frases fixas descritas neste trabalho é a Simetria. Trata-se de construções nas quais se observa que dois constituintes de idêntica natureza distribucional estabelecem com o verbo uma mesma relação semântica. Esta relação implica reciprocidade entre os dois constituintes, que se dizem simétricos, o que faz com que os grupos nominais, que ocupam essas posições estruturais, possam permutar entre si e que possam aparecer coordenados numa mesma posição sintáctica sem que, apesar disso, o significado global da frase se altere. Nas construções com grupos nominais simétricos coordenados, é também possível inserir, facultativamente, uma cópia pronominal (*entre si, um Prep o outro*):

(30a) O João bateu-se em duelo com o Pedro

é equivalente a:

(30a) = (36b) O Pedro bateu-se em duelo com o João

e ambas são equivalentes a:

= (30c) O Pedro e o João bateram-se em duelo

= (30d) O João e o Pedro bateram-se em duelo

Além disso nas construções com *GN* simétricos coordenados, é possível inserir, facultativamente, uma cópia pronominal (*entre si, um Prep o outro*):

(30c')=(30d') (O Pedro e o João + O João e o Pedro) bateram-se em duelo (E + entre si + um com o outro)

A escolha da cópia pronominal (*entre si* ou *um com o outro*) parece depender da construção em causa. Assim, por exemplo, na frase:

(38a) A Maria passou a vias de facto com a Ana

= (38b) A Ana passou a vias de facto com a Maria

= (38c) A Maria e a Ana passaram a vias de facto (\*entre si + uma com a outra)

= (38d) A Ana e a Maria passaram a vias de facto (\*entre si + uma com a outra)

observa-se que a cópia pronominal *entre si* não é aceitável ao contrário do que sucede na construção *bater-se em duelo*.

### 2.3.4. Construções pseudo-conversas

As construções a que chamamos Pseudo-Conversas são estruturas que permitem a permuta de dois constituintes em torno do verbo e do complemento fixo, mantendo a frase o mesmo significado global e sem que se alterem os papéis temáticos desses argumentos, mas invertendo a orientação do processo. Trata-se pois de uma operação semelhante à Conversão, que se observa nas construções predicativas nominais com verbo-suporte, operação descrita por G. Gross (1989, 1996) para o Francês, e, para o Português, entre outros, por Baptista (1997a,b). No caso, porém, das frases fixas aqui estudadas não há nomes. As frases seguintes são ilustrativas desta operação:

- (39) O João deu no focinho do Pedro  
 (39a) = O João deu no focinho ao Pedro [Rdat]  
 (40) = O Pedro apanhou no focinho do João [Pseudo-Conv]

Adaptando a terminologia usada por G. Gross (1989), designámos a construção de tipo ‘activo’, exemplificada em primeiro lugar (a), por construção standard, e à construção de tipo ‘passivo’, exemplificada em (40), chamamos construção pseudo-conversa. Na constante, pode ocorrer um pequeno paradigma de nomes parte-do-corpo (*cabeça, cachola, cara, coirão, focinho, toutiço, trombas, coco*). Os verbos mais frequentes nestas construções são *dar* e *ir*, nas frases standard, de orientação activa; nas frases Pseudo-Conversas, de orientação passiva, observam-se os verbos *comer, apanhar* e *levar*. Note-se que tanto *dar* como os outros verbos *comer, apanhar* e *levar* são, nas construções nominais, verbos-suporte (Baptista, 1997a,b). Porém, nestas expressões não existem nomes predicativos, não sendo, pois, possível atribuir-lhes esse estatuto sintáctico, como se observa:

- (41) O João deu uma bofetada ao Pedro  
 (42) = O Pedro apanhou uma bofetada do João [Pseudo-C]

Contudo, nas frases (39)-(40) não se observa nenhum nome predicativo pelo que não é adequado considerar *dar* e *apanhar* como verbos-suporte. Também não nos parece adequado, porque demasiado *ad hoc*, derivar as construções do exemplo (39)-(40) da redução de um *Npred* na medida em que não é possível prever o conjunto de *Npred* que poderiam ocupar essa posição:

- (39) ?O João deu (uma bofetada + um murro) no focinho do João  
 (40) ?O Pedro apanhou (uma bofetada + um murro) no focinho do João

Ainda assim, o significado global destas frases exprime um acto violento (Baptista, 1997a,b; 2005a,b) e, nesse sentido, o conjunto de nomes que exprimem esse tipo de predicados (sem envolver instrumentos) é relativamente restrito. Por outro lado, note-se que o conjunto de nomes na posição *C* designa, genericamente nomes parte-do-corpo ou são nomes que, nesta construção, adquirem essa interpretação particular (*cornos, focinho, fuças, trombas*, etc.). Essa distribuição é igualmente a que se observa nas constru-

ções com nome predicativo de actos violentos. A análise destas expressões por redução de um nome predicativo poderia, pois, encontrar nestes argumentos alguma força.

Contudo, verifica-se que, para além destas expressões, encontramos frases do tipo standard sinónimas (ou quase) em que o verbo não é *dar* mas sim *ir*:

(43) O Pedro foi a (os cornos + o focinho + as fuças + as trombas) (de + a) o João

(Note-se a melhor aceitabilidade do complemento determinativo nestes casos). Ora, verbo *ir* não é nunca um suporte adequado dos nomes predicativos de actos violentos, o que impede de analisar estas últimas expressões por redução de um nome predicativo.

Dadas as diferentes preposições em jogo, as relações da Pseudo-*Conversão* não foram objecto de um tratamento transformacional, tendo sido classificadas separadamente nas matrizes.

### 3. Experiências sobre um *corpus*

Procedeu-se a algumas experiências de aplicação a um *corpus* das gramáticas formais elaboradas a partir dos dados linguísticos já formalizados. Seleccionaram-se 100 frases fixas, consideradas as mais correntes, da classe CP1 – e aplicaram-se as gramáticas a um (*sub*)*corpus* de texto jornalístico, de grandes dimensões – o CETEMPúblico (aproximadamente 180 milhões de palavras), recorrendo a técnicas de transdutores de estados finitos. Para o efeito, utilizou-se o software UNITEX (Paumier 2004) e os recursos lexicais do português com ele disponibilizados.

Utilizou-se como *corpus* um fragmento do CETEMPúblico contendo todos os extractos que apresentavam o verbo e o elemento nominal que com ele forma a expressão fixa, permitindo entre ambos uma janela de até 5 palavras. Deste modo, assegura-se que neste *corpus* estão presentes todas as instâncias das expressões idiomáticas seleccionadas, de forma a poder calcular a cobertura / *recall* das gramáticas.

A metodologia aqui seguida segue as propostas inicialmente formuladas por Senellart (1998) e implementadas por Silberztein (2001) no sistema INTEX, e posteriormente implementadas, com modificações, no sistema UNITEX, por Paumier (2004). Com este software, foi construído um grafo de referência (também chamado grafo-padrão, v. fig.1) que faz referência aos elementos lexicais e às propriedades léxico-sintácticas constantes das matrizes por meio de variáveis. Estas variáveis remetem para as colunas respectivas da matriz, onde tais elementos ou propriedades estão formalizados. A intersecção do grafo de referência com a informação presente nas matrizes permite produzir, automaticamente, um transdutor de estados finitos (*fst*). Este *fst* é então aplicado ao *corpus* em modo de fusão (*merge*), servindo, então, para identificar estas expressões em texto, acrescentando a cada instância

de uma expressão a informação linguística que lhe estiver associada nas matrizes.

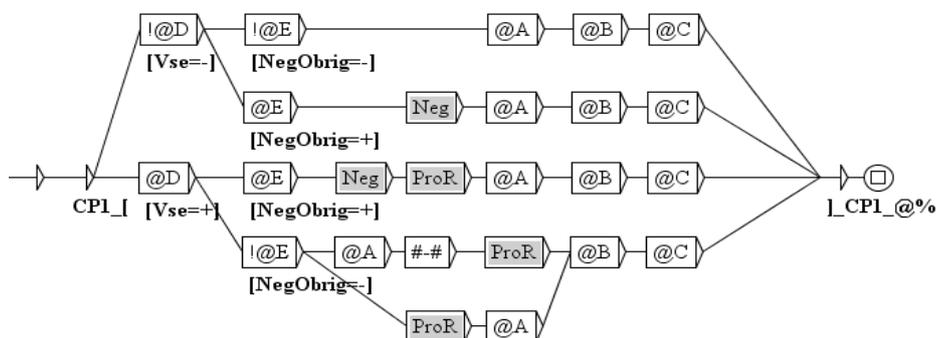


Fig.1. Grafo de referência (simplificado) para a classe CP1

A figura 1 ilustra de forma simplificada a construção do grafo de referência para a classe CP1. O sistema constrói um subgrafo para cada linha da matriz. O conjunto dos subgrafos assim gerados é reunido num grafo geral. Este grafo pode, então, ser utilizado para procurar as expressões em textos. Tratando-se de um transdutor, é possível utilizá-lo para inserir informação linguística nos textos. Por exemplo, a primeira linha do grafo deve interpretar-se da seguinte forma: o grafo reconhece uma sequência constituída pelos elementos lexicais constantes das colunas A, B e C da matriz, representadas pelas variáveis @A, @B e @C, respectivamente, sse as propriedades representadas nas colunas D e E estiverem marcadas negativamente, o que é representado no grafo pela negação das variáveis correspondentes, !@D, !@E. A informação que se encontra fora das caixas a negrito é informação linguística que se pode acrescentar ao texto. Nas caixas cinzentas, ProR apela a um grafo auxiliar com pronomes reflexos, enquanto Neg invoca um subgrafo que contém os advérbios de negação mais frequentes.

Os resultados obtidos nestas experiências apresentam uma média de precisão de 94% e 88% de cobertura, sendo estas semelhantes e elevadas.

Analisando os resultados, por enquanto preliminares, da experiência acima descrita, verificamos que a maioria das expressões idiomáticas não levanta grandes dificuldades ao reconhecimento automático por meio de métodos de estados finitos. O principal factor responsável pelos erros de reconhecimento é, como já se esperava, a inserção de elementos livres, externos à construção fixa (sobretudo os adverbiais). Apesar de se basear num simples (mas mesmo assim já bastante sofisticado) método de reconhecimento de cadeias de palavras, estas gramáticas permitem já uma identificação satisfatória das expressões deste estudo. Provavelmente, concorre para tais resultados, o facto de, na maior parte dos casos, as expressões idiomáti-

cas ocorrerem nos textos sem quaisquer inserções de elementos espúrios. A experiência revelou, porém, algumas limitações próprias deste tipo de representação (inserções, auxiliares, etc.) as quais deverão poder ser ultrapassadas por meio de ferramentas mais complexas, baseadas numa análise sintáctica prévia do texto.

#### 4. Conclusão e perspectivas

Para este estudo, reuniram-se um número considerável de expressões idiomáticas que foram formalmente classificadas e descritas. Julgamos que a sistematicidade e o cuidado posto na selecção deste léxico permite, desde já, considerar que se trata de uma base empírica satisfatória para um estudo mais aprofundado deste tipo de expressões em Português.

A proposta de classificação de M. Gross (1982, 1989) revelou-se adequada à análise das expressões em estudo. Procedeu-se à descrição sistemática das principais propriedades sintácticas (distribucionais, estruturais e transformacionais) encontradas nas frases, o que mostrou a pertinência da metodologia do Léxico-Gramática, na medida em que estas propriedades parecem depender da íntima relação entre o léxico e a sintaxe das expressões linguísticas.

As experiências de aplicação da informação do léxico-gramática codificada nas matrizes a um corpus de textos reais, de grandes dimensões, demonstraram que os métodos de estados finitos, baseados no reconhecimento de cadeias de palavras, permitem uma identificação bastante satisfatória revelando porém algumas limitações (decorrentes sobretudo da inserção de elementos, da presença de verbos auxiliares, etc.), as quais requerem uma análise sintáctica prévia do texto.

Futuramente, pretendemos aplicar a totalidade das gramáticas e os grafos de referência à totalidade do *corpus* do CETEMPúblico, bem como alargar o campo de trabalho ao estudo comparativo de várias línguas. Julgamos ainda ser possível utilizar os recursos já produzidos para verificar a adequação/utilidade de métodos estatísticos na identificação de expressões fixas.

#### Referências

- Baptista, Jorge 1997a. Sermão, tarefa e facada: Uma classificação das construções conversas dar-levar. In *Seminário de Linguística* 1. Faro: Universidade do Algarve – UCEH, pp. 5-37.
- Baptista, Jorge 1997b. Conversão nomes parte-do-corpo e reestruturação dativa. In Castro, Ivo. (ed.). *Actas XII Encontro da Associação Portuguesa de Linguística* I. Lisboa: APL/Colibri, pp. 51-59.
- Baptista, Jorge 2005a. Instrument nouns and fusion. Predicative nouns designating violent actions. In Leclère, Christina *et al* (eds.). *Lexique, Syntaxe et Lexique-Grammaire. Papers in honour of Maurice Gross. Linguisticae Investigatio-*

- nes Supple-menta* 24. Amsterdam / Philadelphia: John Benjamins Pub-Co, pp. 31-40.
- Baptista, Jorge 2005b. Construções Simétricas: complementos e argumentos. In Rio Torto *et al.* (org.). *Estudos de Homenagem ao Prof. Doutor Mário Vilela* 1. Porto: FLUP, pp. 353-367.
- Baptista, Jorge, Anabela Correia & Graça Fernandes 2004. Frozen Sentences of Portuguese: Formal Descriptions for NLP. *Workshop on Multiword Expressions: Integrating Processing, International Conference of the European Chapter of the Association for Computational Linguistics*. Barcelona (Spain). 26 de Julho de 2004, Barcelona, ACL, pp. 72-79.
- Baptista, Jorge, Anabela Correia & Graça Fernandes 2005. Léxico-gramática das frases fixas do português europeu. Breve presentación. *Cadernos de Fraseoloxía Galega* 7. Santiago de Compostela: Xunta de Galicia, pp. 41-53.
- Baptista, Jorge, Anabela Correia & Graça Fernandes (no prelo). Frozen Sentences of Portuguese: a preliminary survey. *Proceedings of the 1st Iberian Workshop on Contrastive Grammar*. Faro, Portugal, Novembro de 2005.
- Correia, Anabela (in prep). *Léxico-Gramática das Frases Fixas do Português Europeu – Construções Transitivas*. (Tese de Mestrado). Faro: Universidade do Algarve-FCHS.
- Fernandes, Graça 2007. *Léxico-Gramática das Frases Fixas do Português Europeu*. (Tese de Mestrado). Faro: Universidade do Algarve-FCHS.
- Fernandes, Graça & Jorge Baptista, 2007a. Reconhecimento automático de expressões idiomáticas em *corpus* – algumas experiências. *Proceedings of TIL'2007 – V Workshop em Tecnologia da Informação e da Linguagem Humana*. Rio de Janeiro, Brasil. Junho 2007.
- Fernandes, Graça & Jorge Baptista, 2007b. Frozen sentences on large corpus: an experiment. *26th International Colloquium on Compared Lexicon and Grammar*. Bonifacio (Corse du Sud). Outubro 2007, Paris: Univ. Marne-la-Vallée, pp. 49-56.
- Fernandes, Graça & Jorge Baptista (no prelo). Frozen Sentences with Obligatory Negation. in *Linguistic Challenges for Natural Language Processing*. *Cadernos de Fraseoloxía Galega*.
- Gross, Gaston 1989. *Les constructions converses du français*. Droz: Genève.
- Gross, Maurice 1975. *Méthodes en syntaxe*. Paris: Hermann.
- Gross, Maurice 1982. Une classification des phrases “figées” du français. *Revue Québécoise de Linguistique* 11 (2). Montréal: UQAM, pp. 151-185.
- Gross, Maurice 1982. Simple sentences. *Text Processing, Proceedings of Nobel Symposium* 51. Stockholm: Almqvist Wiksell, pp. 297-317.
- Gross, Maurice 1989. Les Expressions Figées, Une description des expressions françaises et ses conséquences théoriques. *RT* 8. PRC-IL. Paris: Université Paris 7 – LADL.
- Gross, Maurice 1996. Lexicon-Grammar. in Brown, K. & J. Miller (eds.). *Concise Encyclopaedia of Syntactic Theory*. Oxford: Pergamon Press, pp. 224-259.
- Guillet, Alain & Christian Leclère, 1992. *La structure des phrases simples en français 2: les constructions transitives locatives*. Genève: Droz.

- Harris, Zellig 1964. *The Elementary Transformations (=Transformations and Discourse Analysis Papers 54.)* Philadelphia: University of Pennsylvania.
- Harris, Zellig 1976. *Notes du cours de syntaxe* (trad. M. Gross). Paris: Le Seuil.
- Harris, Zellig 1981. *Papers on Syntax*. Dordrecht/ Boston/ London: D. Reidel Pub. Co.
- Harris, Zellig 1988. *Language and Information*. New York: Columbia University Press.
- Harris, Zellig 1991. *A Theory of Language and Information. A Mathematical Approach*. Oxford: Clarendon Press.
- Leclère, Christian 1995. Restructuration dative. *Language Research* 31-1. Language Research Institute. Seoul: Seoul National University.
- Paumier, Sebastian 2004. *UNITEX – Manuel d’utilisation*. Paris: Univ. Marne-la-Vallée.
- Paumier, Sebastian 2006. *UNITEX – Manuel d’utilisation*. <http://ladl.univ-mlv.fr/>
- Senellart, Jean 1998. Reconnaissance automatique des entrées du lexique-grammaire des phrases figées. In Lamiroy, Beatrice (ed.). *Le Lexique-Grammaire. Travaux de Linguistique 37*. Bruxelles: Duculot. pp. 109-125.
- Silberztein, Max 2001. *INTEX (Manuel)*. Paris: ASSTRIL/LADL.
- Vale, Oto 2001. *Expressões Cristalizadas do Português do Brasil: uma proposta de tipologia*. Tese de doutoramento. Araraquara: Universidade Estadual Paulista.
- Santos, António 1990. *Novos Dicionários de expressões idiomáticas*. Lisboa: Edições Sá da Costa.