

Problèmes formels concernant la traduction des adverbes composés (espagnol/portugais)

Dolors CATALÀ

Université Autonome de Barcelona / LexSem

Jorge BAPTISTA

Université de l'Algarve,

L2F Spoken Language Laboratory, INESC-ID Lisboa (Portugal)

Cristina PALMA

Université de l'Algarve (Portugal)

Résumé

L'objectif de cet article est de présenter les difficultés formelles concernant la traduction automatique des adverbes de l'espagnol (ES) vers le portugais (PT). Nous ne nous intéressons pas directement aux questions d'implémentation, notre but étant de trouver les clés qu'il est nécessaire d'inclure dans un module de transfert ES>PT pour qu'un système de traduction automatique ou assisté puisse choisir, parmi les équivalents donnés, la meilleure traduction. Nous prenons appui sur le dictionnaire des adverbes composés de l'espagnol (Català, 2003) et celui de Palma (2009) et aussi sur une base de données relationnelle contenant les équivalences entre les expressions des deux langues. Notre approche consiste à décrire pour chaque adverbe de l'espagnol ses équivalents portugais. Pour la sélection des expressions équivalentes, nous nous sommes efforcés, autant que possible, de maintenir les correspondances lexicales et structurelles, mais en prenant compte aussi d'autres informations (diasystématiques, diachroniques, diafréquentatives ou diastratiques), quand elles sont disponibles. Pour la validation empirique de ces ressources, nous avons utilisé un corpus de textes extraits des éditions web de journaux espagnols.

Mots clés : adverbes composés, traduction, dictionnaire électronique, corpus.

Abstract

The aim of this article is to present the formal difficulties concerning the automatic translation of Spanish adverbs (ES) to Portuguese (PT). We are not directly interested in implementation issues, our goal is to find the keys that it is necessary to include in an ES>PT transfer module so that an automatic or assisted translation system can choose, among the equivalents given, the best translation. We rely on the dictionary of Spanish adverbs composed (Català, 2003) and that of Palma (2009), and also on a relational database containing the equivalences between the expressions of the two languages. Our approach consists on describing for each adverb in Spanish its Portuguese equivalents. For the selection of equivalent expressions, we have endeavored, as much as possible, to maintain lexical and structural correspondences, but also taking into account other information (diasystematic, diachronic, diafrequentative or diastratic), when available. For the empirical

validation of these resources, we used a corpus of texts extracted from the web editions of Spanish newspapers.

Keywords: compound adverbs, translation, electronic dictionary, corpus.

Resumen

El objetivo de este artículo es presentar las dificultades formales relacionadas con la traducción automática de adverbios del español (ES) al portugués (PT). No estamos directamente interesados en los problemas de implementación, pretendemos encontrar las claves que es necesario incluir en un módulo de transferencia ES > PT para que un sistema de traducción automática o asistida pueda elegir, entre los equivalentes dados, la mejor traducción. Nos basamos en el diccionario de adverbios compuesto del español (Català, 2003) y el de Palma (2009), y también en una base de datos relacional que contiene las equivalencias entre las expresiones de los dos idiomas. Nuestro enfoque consiste en describir para cada adverbio del español sus equivalentes portugueses. Para la selección de expresiones equivalentes, nos hemos esforzado, en la medida de lo posible, por mantener correspondencias léxicas y estructurales, pero también teniendo en cuenta otras informaciones (diasistemática, diacrónica, diafrecuentativa o diastrática), cuando están disponibles. Para la validación empírica de estos recursos, utilizamos un corpus de textos extraídos de las ediciones web de periódicos españoles.

Palabras clave: adverbios compuestos, traducción, diccionario electrónico, corpus.

1. Introduction

Suite aux travaux de Maurice Gross (1986) sur les adverbes figés du français, bon nombre d'études ont été menées en vue de leur description systématique dans plusieurs langues, comme l'italien (De Gioia, 2001), l'espagnol (Blanco & Català, 1998 ; Català, 2003), le catalan (Català, 2002 et 2004), le grec (Voyatsi, 2006) et le portugais (Palma, 2009). Certains travaux ont porté sur des classes particulières d'adverbes figés (Ranchhod, 1991 ; De Gioia, 1994b ; Baptista, 1999), d'autres se sont penchés sur des ensembles sémantiquement définis, par exemple, les adverbes de temps (Maurel, 1990 ; Baptista & Català, 2002), les adverbes focalisateurs (Baptista & Català, 2009 ; 2011). D'autres études se sont intéressées à la comparaison entre des langues (De Gioia, 1994a ; De Gioia, 2008 ; De Gioia & Català, 2004), notamment sur des problèmes formels concernant la traduction automatique (Baptista & Català, 2003).

Dans cet article, nous nous centrons aussi sur les difficultés formelles concernant la comparaison des adverbes figés de l'espagnol et du portugais en vue de leur traduction automatique.

Notre cadre théorique de référence est celui du lexique-grammaire tel qu'il a été élaboré par Maurice Gross (1975, 1986). Dans ce cadre basé sur la théorie transformationnelle de Z.S. Harris (1976), la phrase élémentaire est considérée comme l'unité minimale de sens. La combinaison de phrases élémentaires permet d'aboutir, après diverses opérations de réduction, de permutation et de modifications morphologiques aux énoncés de la langue. Les adverbes venant s'adjoindre à ces phrases élémentaires entretiennent des relations complexes avec les éléments de ces dernières. Maurice Gross a fourni de nouveaux postulats théoriques et méthodologiques pour leur étude. Il démontre qu'une classification des formes, à partir des propriétés syntaxiques, est essentielle pour donner une définition de la notion de classe grammaticale. Il applique la notion d'adverbe à des structures syntaxiquement différentes de la terminologie traditionnelle comme les adverbes proprement dits, les compléments circonstanciels, les propositions subordonnées circonstancielles, et d'autres formes encore, telle que les exclamations et les conjonctions ayant des propriétés adverbiales. Il présente une classification des formes des adverbes et non de leur fonction.

2. Les dictionnaires d'adverbes figés

2.1 Le dictionnaire d'adverbes de l'espagnol (Català, 2003)

Nous avons rassemblé presque 6.000 adverbes espagnols. L'établissement de ce corpus s'est fait en deux étapes. Dans un premier temps, nous avons dépouillé les trois grands dictionnaires de la langue espagnole : le *Diccionario de la Real Academia Española* (DRAE), le *Diccionario General de la Lengua Española* (DGLE) et le *Diccionario de Uso del Español* (DUE) dans les éditions indiquées en bibliographie. À ces dictionnaires généraux de l'espagnol, il faut ajouter d'autres ouvrages spécialisés comme le *Diccionario Espasa de dichos y frases hechas* de Buitrago (1997). Ce premier dépouillement manuel nous a fourni un nombre important d'adverbes que nous avons confrontés à des corpus journalistiques, ce qui nous a permis d'en élargir la liste. Dans un deuxième temps, nous avons introduit des entrées qui n'étaient pas attestées dans les dictionnaires mais qui appartiennent clairement, d'après l'intuition linguistique de plusieurs locuteurs hispanophones natifs, au système de la langue aussi bien orale qu'écrite. Il s'agit d'expressions

comme par exemple *a bultuntún, a bañalomo, ser más macho que Pancho Villa...* Ces unités adverbiales composées ont été incorporées au DELACs (dictionnaire des mots composés de l'espagnol).

En suivant les bases théoriques et méthodologiques proposées pour le français par Maurice Gross (1986), nous avons construit le dictionnaire d'adverbes de l'espagnol, qui est structuré en 16 tables correspondant à différents types d'adverbes.

Le tableau suivant (Tableau 1) rend compte de l'état actuel du dictionnaire. En regard de chaque nom de classe, figurent un ou deux exemples et le nombre des adverbes de la classe.

Tables	Structures	Exemples	Effectifs
PADV	Adv	<i>actualmente</i>	1563
PC	Prép C	<i>en abundancia, sin ambages</i>	869
PDETC	Prép Dct C	<i>al contado, bajo ningún concepto</i>	585
PAC	Prép Adj C	<i>a primera hora, sin previo aviso</i>	157
PCA	Prép C Adj	<i>a ciencia cierta, a brazo partido</i>	291
PCDC	Prép C de C	<i>por arte de magia, a cuerpo de rey</i>	168
PCPC	Prép C Prép C	<i>de cabo a rabo, con las manos en alto</i>	149
PCONJ	Prép C Conj C	<i>en cuerpo y alma, por far o por nefas</i>	131
PCDN	Prép C de N	<i>en virtud de, a condición de</i>	233
PCPN	Prép C Prép N	<i>de espaldas a, en dirección a</i>	51
PV	Prép V W	<i>sin querer, sin mediar palabra</i>	127
PF	P (phrase figée)	<i>que yo sepa, como diría el otro</i>	169
PECO	(más + como) Adj que C	<i>más claro que el agua, sordo como una tapia</i>	797
PVCO	(V) como C	<i>beber como una esponja, arder como una tea</i>	532
PPCO	(V) como Prép C	<i>cambiar como del día a la noche</i>	46
PJC	Conj C	<i>y no se hable más, y viceversa</i>	91
		TOTAL	5959

Tableau 1. Classes des adverbes en espagnol

2.2 Le dictionnaire d'adverbes figés du Portugais (Palma, 2009)

Nous sommes partis des dictionnaires de langue généraux et des dictionnaires spécialisés d'expressions figées (voir bibliographie), et nous avons recensé, jusqu'à présent, plus de 2.500 adverbes figés. Il s'agit donc d'un travail en cours, ce qui devra permettre d'accroître ces chiffres de façon significative.

Pour le classement, nous avons aussi suivi les critères généraux proposés par M. Gross (1986). Dans ce travail, nous ne tiendrons pas compte de la classe PADV qui est réservée aux adverbes simples ; nous

n'aborderons pas les adverbes comparatifs (classes PVCO, PECO et PPCO), qui ont fait déjà l'objet d'une étude systématique (Ranchhod, 1991), suivie d'une comparaison avec l'italien (Ranchhod & De Gioia). Nous ne traiterons pas non plus les adverbes des classes PF et PV.

L'état actuel du dictionnaire sans les classes citées est donné par le tableau ci-dessous (tableau 2) :

Tables	Structures	Exemples	Effectifs
PC	Prép C	<i>a olho, de fugida,</i>	430
PDETC	Prép Dét C	<i>de uma sentada, à socapa</i>	430
PAC	Prép Adj C	<i>a duras penas, com segundas intenções</i>	130
PCA	Prép C Adj	<i>por uma unha negra, à vista desarmada</i>	240
PCDC	Prép C de C	<i>a talho de foice, em mangas de camisa</i>	130
PCPC	Prép C Prép C	<i>às duas por três, por amor à camisola</i>	150
PCONJ	Prép C Conj C	<i>à grande e à francesa, em carne e osso</i>	90
PCDN	Prép C de N	<i>a mando de, do ponto de vista de</i>	150
		TOTAL	1750

Tableau 2. Classes des adverbes en portugais

3. Comparaison des adverbes

3.1 Taille relative des classes

Bien que l'état actuel du dictionnaire du portugais soit encore assez incomplet, il est déjà possible de dégager quelques observations intéressantes. D'abord, la taille *relative* de chaque classe est à peu près la même dans les deux langues comme nous pouvons le voir dans le tableau ci-dessous (tableau 3) :

Classe	Es	%	Pt	%	%diff
PC	869	0,32	430	0,25	0,08
PDETC	585	0,22	430	0,25	-0,03
PAC	157	0,06	130	0,07	-0,02
PCA	291	0,11	240	0,14	-0,03
PCDC	168	0,06	130	0,07	0,01
PCPC	149	0,06	150	0,09	-0,03
PCONJ	131	0,05	90	0,05	0,00
PCDN	333	0,12	150	0,09	0,04
total	2683		1750		

Tableau 3. Comparaison des classes des adverbes en espagnol/ portugais

La comparaison avec les dictionnaires du catalan, du français et de l'italien montre aussi que les langues romanes apparentées ont recours à des processus de formation d'adverbes composés d'importance homologue.

Cette remarque devra, néanmoins, être nuancée par le fait que bon nombre d'adverbes de quelques classes ne présentent pas de façon stricte la structure définitoire de la classe. Par exemple, on notera les cas où la préposition qui introduit l'adverbe est absente :

- es : *todo el santo día* [PAC]
- pt : *todo o santo dia* [PAC]
- fr : *toute la journée* [PDETC]
- es : *acto continuo* [PCA]
- pt : *acto contínuo* [PCA]
- ‘immédiatement après’

ainsi que les cas où il n'y a pas de conjonction de coordination dans des suites à parallélisme interne mais qui sont, néanmoins, incluses dans la classe PCONJ :

- es : *día sí, día no* [PCONJ]
- pt : *día sim dia não* [PCONJ]
- fr : *un jour oui, un jour non* [PCONJ]

On remarquera aussi les cas des adverbes qui ne peuvent pas avoir un élément nominal en position C (tête du groupe nominal figé, constitutif de l'adverbe composé) :

- pt : *sem mais aquelas* [PDETC]
- ‘(sans celles-là) sans prendre en compte l'opinion des autres’
- es : *con todo esto* [PDETC]
- ‘avec tout ça’

3.2 Correspondance entre classes des adverbes de l'espagnol et du portugais

En vue de la traduction des adverbes composés des deux langues étudiées (ES[pagnol] > P[or]t[ugais]), nous avons indiqué pour chaque adverbe du dictionnaire de l'espagnol un équivalent portugais. Nous nous sommes efforcés de trouver les équivalents les plus proches, du point de

vue lexical (le même mot-clé, habituellement l'élément C), du point de vue formel (même classe formelle), ou encore, du point de vue sémantique et stylistique (diasystématique). Pourtant, il y a des cas où seule une traduction libre dans la langue cible pourrait permettre d'exprimer approximativement le même sens de l'expression de la langue source, bien qu'il soit difficile de maintenir les mêmes conditions pragmatiques :

- es : *a bote pronto*
- pt : *taco a taco*
- fr : *du tac au tac*

Dans d'autres cas – vraiment «idiomatiques», au sens originel de « propres à la langue » / intraduisibles – nous n'avons pas trouvé d'équivalent de traduction :

- pt : *as bochechos* (dire/payer)
- fr : *par petites quantités*

Au vu des données actuelles, ces cas restent cependant résiduels par rapport à l'ensemble des adverbes de l'espagnol car il est souvent possible de trouver un bon équivalent portugais pour les entrées du dictionnaire source et, en outre, avec une remarquable correspondance lexicale et formelle. Ceci n'est pas surprenant pour des langues très proches, comme c'est le cas du PT et de l'ES.

Les tableaux ci-après permettent de se faire une idée assez précise des correspondances dans les différentes classes. Le tableau 4 présente le nombre d'entrées de chaque classe du dictionnaire source de l'ES qui ont un équivalent en PT et sa distribution parmi les classes de ce dernier dictionnaire. Le tableau 5, quant à lui, fait référence aux pourcentages.

	Equiv	PT											
		V	PAC	PC	PCA	PCDC	N	PCD	NI	PCU	PCPC	PCPN	C
PAC	134	1	84	8	17	12	0	1	1	1	0	7	3
PC	754	24	14	443	23	12	13	11	13	6	176	0	13
PCA	224	0	8	9	179	8	1	2	10	0	13	0	4
PCDC	130	2	8	4	5	90	0	4	4	0	8	0	1
PCDN	167	1	0	0	0	0	163	0	1	1	0	0	0
PCONJ	85	0	1	4	1	1	0	65	5	0	6	0	2
PCPC	92	0	0	2	0	0	0	0	89	0	0	0	0
PCPN	44	0	0	0	0	0	0	0	0	44	0	0	0
PDETC	456	3	11	47	12	13	20	6	13	1	320	0	6
PF	72	0	0	0	0	0	0	0	1	0	0	71	0

Tableau 4. Correspondance entre les classes des adverbes

		Equiv	PT													
			V	PAC	PC	PCA	PCDC	N	PCD	NI	PCU	PCPC	PCPN	C	PDETC	PF
SEI	PAC	134	0,01	0,63	0,06	0,13	0,09	0,00	0,01	0,01	0,01	0,00	0,05	0,02	0,00	0,00
	PC	754	0,03	0,02	0,59	0,03	0,02	0,02	0,01	0,02	0,01	0,23	0,00	0,02	0,00	
	PCA	224	0,00	0,04	0,04	0,80	0,04	0,00	0,01	0,04	0,00	0,06	0,00	0,02	0,00	
	PCDC	130	0,02	0,06	0,03	0,04	0,69	0,00	0,03	0,03	0,00	0,06	0,00	0,01	0,00	
	PCDN	167	0,01	0,00	0,00	0,00	0,00	0,98	0,00	0,01	0,01	0,00	0,00	0,00	0,00	
	PCONJ	85	0,00	0,01	0,05	0,01	0,01	0,00	0,76	0,06	0,00	0,07	0,00	0,02	0,00	
	PCPC	92	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,97	0,00	0,00	0,00	0,00	0,00	
	PCPN	44	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	
	PDETC	456	0,01	0,02	0,10	0,03	0,03	0,04	0,01	0,03	0,00	0,70	0,00	0,01	0,00	
	PF	72	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,99	0,00	0,00	

Tableau 5. Correspondance entre les classes en pourcentage

De ce relevé strictement quantitatif, nous pouvons extraire quelques remarques intéressantes. Bon nombre des adverbes figés ont un correspondant dans la même classe. Par exemple, pour les classes PCPN, PCDC et PCPC, la correspondance concerne la totalité ou presque des adverbes ; pour la classe PF, le taux est de 99 % du total. Par contre, la correspondance pour les classes les plus nombreuses comme la classe PC, PDETC et PAC est moins fréquente quoique élevée.

Certains de ces résultats ne sont pas surprenants. En effet, les adverbes des classes PCPN et PCDC qui correspondent aux locutions prépositionnelles de la grammaire traditionnelle, sont peu figés car souvent ils conservent la syntaxe du nom prédicatif qui en est la tête syntaxique, il est donc normal de trouver un nombre important d'équivalents dans la même classe.

Par contre, les classes PCDC et PCPC, incluses dans bon nombre de « locutions adverbiales » que l'on trouve traditionnellement dans les dictionnaires de langue, présentent un plus haut degré de figement pouvant atteindre le figement total. La mise en évidence du taux élevé de correspondance entre ces classes qui peuvent présenter des phénomènes idiosyncrasiques (historiques, sociologiques) propres à l'évolution de chaque langue ne manque pas d'attirer notre attention.

3.3 Aspects contrastifs

L'équivalence est totale dans bon nombre d'expressions. C'est la situation idéale pour un traducteur, mais ce n'est pas la situation la plus courante. Nous trouvons des exemples dans toutes les classes d'adverbes comme :

- es : *a la letra* / pt : *à letra*
- es : *a duras penas* / pt : *a duras penas*
- es : *sin el menor aviso* / pt : *sem o menor aviso*
- es : *a buen paso* / pt : *a bom passo*

Nous trouvons parfois des variations du déterminant :

- es : *a mano* / pt : *à mão*
- es : *por lo visto* / pt : *pelos vistos* (*pelo visto)
- es : *por las malas* / pt : *por mal* (*pelo mal)
- es : *en un principio* / pt : *em princípio* (*≠ num princípio)

Notamment, nous constatons que les constructions comprenant le neutre *lo*, dans la classe PDETC (47), sont difficilement transposables dans la langue cible, le déterminant est alors un masculin ou encore les équivalents sont des adverbes simples.

- es : *por lo civil* / pt : *por o civil*
- es : *por lo general* / pt : *no + em geral*
- es : *de lo alto* / pt : *de cima*
- es : *en lo antiguo* / pt : *antiguamente*

C'est parfois la préposition qui varie ; cette variation peut aussi exister dans la langue cible, il se peut, donc, qu'une correspondance (partielle ou totale) existe entre des formes équivalentes :

- es : *en oleadas* / pt : *(por + em) vagas*
- es : *de conformidad con* / pt : *(? *de + em) conformidade com*
- es : *(en + dentro) de breves instantes* / pt : *(em + dentro) de breves instantes*
- es : *(en + a) (buena + mala) hora* / pt : *(em + *a) (boa + má) hora*

Si nous examinons dans le détail les équivalents sémantiques, nous observons que certains équivalents n'appartiennent pas à la même classe :

- es : *de buena mañana* / pt : *de manhã (bem) cedo, (bem) cedo de manhã*

Dans ce dernier cas, la valeur intensive de l'adjectif dans la forme ES ne se traduit que par l'apposition de l'adverbe *cedo* à l'adverbe figé du PT *de manhã*. Il n'est pas question ici, pourtant, de ce qu'un adjectif ne puisse pas avoir cette valeur intensive. Nous trouvons les adverbes composés :

- es : *a altas horas de la noche* / pt : *a altas horas da noite*
‘très tard dans la nuit’

où l’adjectif *altas* semble être porteur de cette valeur sémantique.

L’absence de correspondance (ou bien une correspondance imparfaite) est due, parfois à des problèmes de soudure soit dans la langue source soit dans la langue cible. Par exemple, pour les adverbes de la classe PC, nous trouvons les cas de soudure suivants :

- es : *antes de ayer* / pt : *anteontem*
- es : *entre tanto* / pt : *entretanto*
- es : *por ventura* / pt : *porventura*
- es : *de balde* / pt : *debalde*
- es : *a vuelapluma* / pt : *ao correr da pena*

et le cas d’un processus de soudure non encore conclue, c’est-à-dire, où il y a variation entre formes composées et mot agglutiné :

- es : *entre dientes* / pt : *entredentes, entre dentes*
- es : *a manos llenas* / pt : *às mancheias, a mancheias, às mãos(-) cheias*

Ces phénomènes de soudure orthographique (mais pas phonétique, car un accent secondaire persiste dans les formes agglutinées du PT), sont une démonstration pertinente de l’origine générale des adverbes composés dans des groupes nominaux prépositionnels, selon la définition d’adverbe généralisé proposée par M. Gross (1986).

Parfois l’équivalent de la langue cible n’appartient pas à la même classe parce que le modifieur n’est pas situé au même endroit dans la langue cible :

- es : *a la vieja usanza* / pt : *à moda antiga* ‘à l’ancienne mode’
- es : *en semejantes circunstancias* / pt : *em circunstâncias semelhantes*
‘dans des circonstances semblables’

Dans quelques cas, l’équivalent de la langue cible présente un modifieur qui n’est pas présent dans la langue source :

- es : *a sorbos* / pt : *a pequenos tragos* (**a tragos*) ‘à petites gorgées’
- es : *con segundas* / pt : *con segundas intenções* (**com intenções*) ‘avec plein de sous-entendus’

ou le cas inverse, où la forme dans la langue cible ne présente pas de modifieur :

- es : *a espaldas vueltas* / pt : *por as costas* ‘par trahison’
- es : *a teja vana* / pt : *debaixo de telha* ‘sous le toit’

Il existe dans les deux langues un élément équivalent entre les deux formes figées, mais ils n’appartiennent pas à la même partie du discours, quoique morphologiquement (et historiquement) apparentés :

- es : *en resumidas cuentas* / pt : *em resumo* ‘en résumé’
- es : *por vía inhalada* / pt : *por inalação* ‘par inhalation’

Dans ce cas, il ne semble pas possible d’avoir une forme *em resumido* C (avec le même sens) pour établir une équivalence. Nous trouvons aussi des cas où l’élément lexical en position de constante varie sensiblement dans la langue cible :

- es : *por su linda cara* / pt : *pelos seus lindos olhos* ‘pour ses beaux yeux’

L’élément lexical peut exister dans la langue cible mais il n’est pas utilisé dans l’adverbe composé :

- es : *de (buena + mala) gana* / pt : *de (bom + mau) grado* (**de (boa + má) gana*) ‘bon gré mal gré’
- es : *de (buena + mala) (fe + verdad)* / pt : *de (boa + má) (fé + *verdad)* ‘de (bonne+mauvaise) foi’
- es : *hasta (nuevo aviso + nueva orden)* / pt : **até (*novo aviso + nova ordem)* ‘jusqu’à nouvel ordre’

Il existe des familles d’adverbes figés, fréquemment avec le même mot clé qui sont formées de suites de mots légèrement différentes d’un point de vue syntaxique, mais sémantiquement équivalentes. Ces familles peuvent avoir dans la langue cible soit un seul équivalent, soit aussi une autre famille d’adverbes figés. On se pose, alors, la question du choix de l’équivalent dans l’autre langue, comme dans l’exemple espagnol suivant : *a **grandes** rasgos* de la classe PAC. Et ses possibles équivalents portugais : *em **grandes** traços* PAC ; *a traços largos* PCA ; *em traços largos* PCA ; *em traços gerais* PCA. Nous avons établi comme critère de sélection de garder

l'équivalent le plus proche possible de l'entrée en langue source, donc ici le premier des exemples cités : *em grandes traços*.

4. Perspectives

La construction des dictionnaires coordonnés telle qu'elle a été proposée par Xavier Blanco (1999) présuppose l'établissement en parallèle pour chaque dictionnaire d'une liste d'équivalents pour chaque langue.

Si un mot a plusieurs sens, la solution la plus évidente consiste à dédoubler les entrées dans la langue source dans un dictionnaire coordonné avec les différences lexico-syntaxiques (distribution, transformations) et les différents équivalents.

Nous n'envisageons pas cette situation *a priori*. Il y a peu de cas où la traduction d'adverbes pose des problèmes, la plupart des fois on trouve des équivalents dans la même classe – les adverbes figés ambigus – polysémiques sont très rares. Voici un exemple :

- pt : *ao vivo* <*transmitir*> : transmettre en direct
- pt : *ao vivo* <*descascar laranjas*> = enlever la peau

Ici l'adverbe *ao vivo* s'applique obligatoirement sur le prédicat *descascar* et avec l'argument *laranja* ou toutes ses variétés. Dans d'autres cas, il faut expliciter les propriétés par exemple :

- à base de + N abs =
- à base de + N conc <ingrédient culinaire>

Par contre, dans le cas illustré ci-dessus, quand il y a dans la langue source une forme (ou toute une famille de formes, qui sont des variantes du même adverbe composé) associée dans la langue cible à une forme (ou aussi à une autre famille de formes), l'établissement des équivalents pose des problèmes dans le format des dictionnaires coordonnés. Si on veut garder l'équivalence entre une forme (source) et une famille de formes (cible), il faut dédoubler les entrées en L1 autant de fois qu'il y a des formes différentes dans la famille de l'adverbe équivalent en L2. Or, cela fait exploser de façon artificielle le nombre d'entrées, d'une part, et, d'autre part, pose des problèmes de maniement du dictionnaire.

Une solution alternative (et plus naturelle) serait de représenter les familles d'adverbes, par exemple les noms de temps (figure 1) par des graphes et d'utiliser ces graphes comme des pointeurs pour les formes de la langue cible. Cela présuppose des programmes adéquats, capables d'utiliser l'information des graphes pour décider, parmi des variantes autorisées, la plus adéquate à la traduction comme nous pouvons le voir sur la figure ci-après représentant le graphe auxiliaire nécessaire lorsque nous construisons un transducteur avec les combinaisons *mañana, tarde, noche, manhã, tarde, noite* et les jours de la semaine :

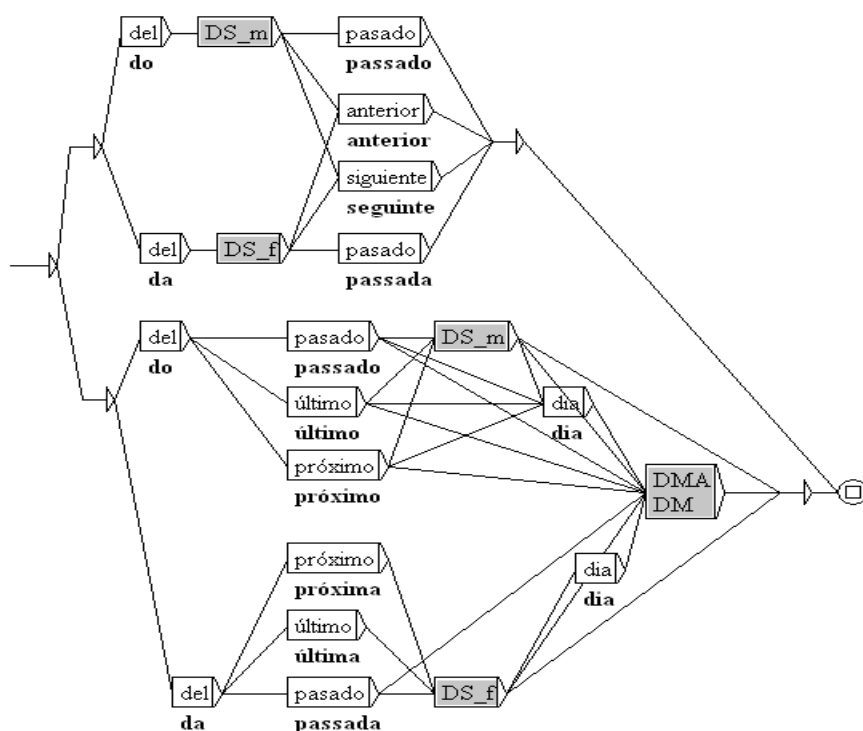


Figure 1. Graphe auxiliaire de MTN - Modifieurs des jours de la semaine

Une autre solution complémentaire aux dictionnaires coordonnés serait de relier les entrées de ce type moyennant une table de transfert qui ne prenne pas les suites comme des pointeurs, mais des codes arbitraires, disons des chiffres qui identifieraient de façon univoque les entrées des dictionnaires.

Nous sommes en train de réfléchir sur la potentialité de cette méthode et d'observer la taille des cas affectés par cette situation. Il faudrait aussi extrapoler ce cas pour le traitement de la synonymie. Cette solution aurait des avantages, surtout si on pense à des environnements multilingues, car elle permettrait de construire des mécanismes de

recherche des équivalents pour lesquels il n'y a pas de forme disponible dans le dictionnaire coordonné, en recherchant dans le voisinage sémantique dans certains mots sources les formes qui pourraient exister dans la langue cible. Bien sûr, il faudrait distinguer les graphes de familles d'équivalents formels d'un même adverbe des graphes d'équivalent de type synonymique.

En ce qui concerne l'implémentation des données, nous travaillons actuellement avec NOOJ (Silberztein, 2015). A titre d'exemple, nous avons automatisé la recherche des adverbes figés de l'espagnol sur un corpus de 8,2 millions de mots correspondant au journal "El Mundo" (1995) et nous avons trouvé 360.000 occurrences d'adverbes composés dont 2.429 formes différentes dans le dlc du texte. Pour le portugais, avec un corpus de 9,6 millions de mots obtenus à partir du journal "Público" (premier fragment du corpus CETEMPUBLICO : www.linguateca.pt/CETEMPUBLICO/), nous avons trouvé 60.080 occurrences d'adverbes composés dont 1.100 entrées du dlc du texte. Ces données, pourtant ne sont pas directement comparables car les tailles des dictionnaires d'adverbes composés sont différentes.

Nous souhaitons, dans l'avenir, regarder de plus près les équivalences déjà établies, en vue d'évaluer la qualité de la traduction ainsi obtenue et découvrir d'autres problèmes n'ayant pas encore été repérés. Si ces traductions sont en général de qualité suffisante, nous envisageons aussi l'utilisation des dictionnaires bilingues d'adverbes composés comme des outils nécessaires à l'alignement de textes pour la construction de corpora parallèles.

Références bibliographiques

- BAPTISTA, J., Manhã, tarde, noite, analysis of temporal adverbs using local grammars, *SEMINARIOS DE LINGÜÍSTICA*, 1999, **3**, 5-31.
- BAPTISTA, J., CATALÀ, D., Compound Temporal Adverbs in Portuguese and in Spanish, *Advances in Natural Language Processing*, Berlin, Springer, 2002, 133-136.
- BAPTISTA, J., CATALÀ, D., Disambiguation of focus adverbs in Portuguese and Spanish, *Bulag n°33, International Symposium in Data and Sense Mining, Machine Translation and Controlled Languages, and their applications to emergencies and safety critical domains*, 2009, 31-37.
- BAPTISTA, J., CATALÀ, D., Adverbes focalisateurs et analyse syntaxique automatique de groupes nominaux, *Passeurs de mots, passeurs d'espoir, Lexicologie, terminologie et traduction face au défi de la diversité, AS, actes des VIII journées scientifiques du Réseau LTT*, Lisbonne, 15-17 octobre 2009, 2011, 97-110.

- BAPTISTA, J., CATALÀ, D., What glues idioms together may not be just statistics after all ? The case for compound adverbs in Portuguese and Spanish, in PAMIES ET AL. (eds.), *Phraseology and discourse: cross Linguistic and Corpus based Approaches*, Schneider Verlag Hohengehren GmbH, 2012, 19-29.
- BLANCO, X., *Lexicographie bilingue français-espagnol et classes d'objets*, Bellaterra, Universitat Autònoma de Barcelona, 1999.
- BLANCO, X., CATALÀ, D., Quelques remarques sur un dictionnaire électronique d'adverbes composés en espagnol, *LINGVISTICÆ INVESTIGATIONES*, 1998, **XI**, 2, 213-232.
- CATALÀ, D., BAPTISTA, J., Spanish Adverbial Frozen Expressions, *Proceedings of the Workshop on a Broader perspective on Multiword Expressions*, ACL, 2007a, 33-40.
- CATALÀ, D., BAPTISTA, J., Limits to pattern-matching, finite-state techniques on lexical analysis for multiword expressions: the case for compound adverbs in Spanish, *Actes du 26^e Colloque International Lexique-grammaire*, 2007b, 17-24.
- CATALÀ, D., BAPTISTA, J., Aspects contrastifs espagnol-portugais, MEJRI, S. (ed). *A la croisée des mots. Hommages à Taïeb Baccouche*, Université de Sousse, Université Paris 13, Sousse, Villetaneuse, 2007c, 101-118.
- CATALÀ, D., Una classe d'adverbs compostos del català, *CUADERNOS DE FILOLOGÍA*, 2002, **Anejo XLIV**, 47-58.
- CATALÀ, D., *Les adverbes composés : approches contrastives en linguistique appliquée*, Thèse de doctorat, Universitat Autònoma de Barcelona, 2003.
- CATALÀ, D., Formalización lingüística dos adverbios compostos do catalán, *CADERNOS DE FRASEOLOGIA GALEGA*, 2004, **6**, 11-26.
- DE GIOIA, M., Sur quelques comparaisons d'adverbes figés de l'italien et du français, *LINGVISTICÆ INVESTIGATIONES*, 1994a, **XVIII** : 1, 89-120.
- DE GIOIA, M., Chiaro come il sole a mezzogiorno. Una classe di avverbi idiomatici dell'italiano, *THE LINGUIST*, 1994b, **33** : 6, 220-225.
- DE GIOIA, M., *Avverbi idiomatici dell'italiano. Analisi lessico-grammaticale*, Torino, l'Harmattan Italia, 2001.
- DE GIOIA, M., *Carrément. Lexique bilingue français-italien des adverbes figés*, Roma, Aracn, 2008.
- DE GIOIA, M., CATALÀ, D., Étude comparée d'adverbes figés du français, de l'italien et de l'espagnol, *SEMINARIOS DE LINGÜÍSTICA*, 2004, **5**, 17-44.
- GROSS, M., Une famille d'adverbes figés : les constructions comparatives en *comme*, *REVUE QUÉBÉQUOISE DE LINGUISTIQUE*, 1984, **13** : 2, 237-269.
- GROSS, M., *Grammaire transformationnelle du français*, vol 3, *Syntaxe de l'adverbe*, Paris, Asstril, 1986.
- HARRIS, Z., *Notes du cours de syntaxe*, Paris, Le Seuil, 1976.
- MAUREL, D., Adverbes de date : étude préliminaire à leur traitement automatique, *LINGVISTICÆ INVESTIGATIONES*, 1990, **XIV** : 1, 31-63.
- PALMA, C., *Expressões Fixas Adverbiais Descrição léxico-sintáctica e subsídios para um estudo contrastivo Português-Espanhol*, Dissertação de Mestrado, apresentada à Faculdade de Ciências Humanas e Sociais da Universidade do Algarve, 2009.
- RANCHHOD, E., Frozen adverbs-Comparative forms Como C in Portuguese, *LINGVISTICÆ INVESTIGATIONES*, 1991, **XV** : 1, 141-170.

RANCHHOD, E. & DE GIOIA, M., Comparative romance syntax. Frozen adverbs in Italian and in Portuguese, *LINGVISTICÆ INVESTIGATIONES*, 1996, **XX** : 1, 33-85.

SILBERZTEIN, M., *La formalisation des langues. L'approche de Nooj*, London, ISTE Éditions, 2015.

VOYATZI, S., *Description morphosyntaxique et sémantique des adverbes figés en vue d'un système d'analyse automatique des textes grecs*, Thèse de doctorat, Université Paris-Est, 2006.

Dictionnaires

ALMEIDA COSTA, J., SAMPAIO E MELO, A., *Dicionário da Língua Portuguesa 2009 – Acordo Ortográfico*, Porto, Porto Editora, 2008.

ALMEIDA, J. J. *Dicionário Aberto de Calão e Expressões Idiomáticas*, 2020.
<https://natura.di.uminho.pt/~jj/pln/calao/dicionario.pdf>

ALVAR EZQUERRA, M. (dir), *Diccionario General de la Lengua Española*, Barcelona, Vox, 1997.

BOSQUE, I. (dir.), *Diccionario Combinatorio del Español Contemporáneo*, Madrid, S.M. Ediciones, 2004.

BUITRAGO JIMÉNEZ, A., *Diccionario Espasa de dichos y frases hechas*, Madrid, Espasa Calpe, 1997.

COROMINAS, J., *Diccionario crítico etimológico de la lengua castellana*, Madrid, Gredos, 1974.

REAL ACADEMIA ESPAÑOLA, *Diccionario de la lengua española*, 21.^a ed., 1992.

MOLINER, M., *Diccionario del uso del español (DUE)*, Madrid, Gredos, 1989.

PÉREZ RIOJA, J.A., *Modismos del español*, Salamanca, Lib. Cervantes, 1997.

SECO, M., ANDRÉS, O., RAMOS, G., *Diccionario Fraseológico Documentado del Español Actual. Locuciones y Modismos Españoles*, Madrid, Aguilar Lexicografía / Santillana Ediciones Generales, 2004.