

---

This is the **published version** of the bachelor thesis:

Sastre Alaiz, Judith; Blanco Escoda, Xavier. Dictionnaire électronique de catalan coordonné avec le français : le module NooJ. 2007.

---

This version is available at <https://ddd.uab.cat/record/45775>

under the terms of the  license



TREBALL D'INVESTIGACIÓ DE TERCER CICLE  
DOCTORAT EN LLENGÜES I CULTURES ROMÀNIQUES

DICTIONNAIRE ÉLECTRONIQUE DE CATALAN  
COORDONNÉ AVEC LE FRANÇAIS  
LE MODULE NOOJ  
Judith Sastre Alaiz

Dirigit pel Dr. XAVIER BLANCO ESCODA

Departament de Filologia Francesa i Romànica  
Facultat de Filosofia i Lletres  
Universitat Autònoma de Barcelona  
Març 2007



# Table des matières

<b>Table des figures</b>	<b>iv</b>
<b>Liste des tableaux</b>	<b>vii</b>
<b>I Introduction et cadre de la recherche</b>	<b>1</b>
1 Préliminaires	3
2 Lexicologie, Lexicographie et Dictionnairique	5
2.1 Le système DELA (LADL)	6
2.1.1 Les AEF en description linguistique	7
2.1.2 Le lexique-grammaire	8
2.2 Les dictionnaires du LLI	10
2.2.1 Les classes d'objets	10
2.3 Le dictionnaire explicatif et combinatoire	12
2.3.1 Le format du DEC	12
2.3.2 La Théorie Sens $\Leftrightarrow$ Texte	16
<b>II Dictionnaire Électronique Coordonné du Catalan</b>	<b>21</b>
3 Présentation du dictionnaire	23
4 Les catégories grammaticales	25
5 La flexion	29
5.1 Les substantifs	32
5.2 Les adjectifs	41
5.3 Les verbes	42

<b>6</b>	<b>Les traits syntactico-sémantiques</b>	<b>47</b>
6.1	Abstrait vs Concret . . . . .	48
6.2	Abstrait . . . . .	51
6.3	Concret . . . . .	58
6.4	Végétal . . . . .	60
6.5	Humain vs Animal . . . . .	61
6.6	Locatif . . . . .	62
6.7	Les noms collectifs . . . . .	66
6.7.1	Propriétés syntaxiques . . . . .	66
6.7.2	Propriétés morphologiques . . . . .	70
6.7.3	Morphologie et sémantique . . . . .	72
<b>7</b>	<b>Les classes d’objets</b>	<b>75</b>
7.1	Classes d’objets du trait <i>Veg</i> . . . . .	76
7.2	Classes d’objets du trait <i>Anl</i> . . . . .	77
7.3	Classes d’objets du trait <i>Conc</i> . . . . .	80
<b>8</b>	<b>La plausibilité</b>	<b>83</b>
<b>9</b>	<b>La traduction</b>	<b>87</b>
<b>III</b>	<b>Le module NooJ</b>	<b>89</b>
<b>10</b>	<b>Présentation du module</b>	<b>91</b>
<b>11</b>	<b>Les grammaires locales</b>	<b>93</b>
11.1	Les grammaires lexicales . . . . .	93
11.2	Les grammaires syntaxiques . . . . .	98
11.2.1	Un exemple de grammaire locale . . . . .	99
11.2.2	Un exemple de grammaire de désambiguïsation . . . . .	104
<b>12</b>	<b>L’application à des corpora</b>	<b>109</b>
<b>IV</b>	<b>Conclusions et Perspectives</b>	<b>113</b>
<b>13</b>	<b>Conclusions et Perspectives</b>	<b>115</b>

<b>V</b>	<b>Annexes</b>	<b>117</b>
<b>A</b>	<b>Tableaux récapitulatifs</b>	<b>119</b>
<b>B</b>	<b>Graphes de flexion</b>	<b>125</b>
<b>C</b>	<b>Fichier de flexion des verbes</b>	<b>137</b>
	<b>Remerciements</b>	<b>165</b>
	<b>Bibliographie</b>	<b>167</b>
	<b>Index</b>	<b>171</b>



# Table des figures

2.1	Niveaux de représentation linguistique de la TST . . . . .	19
5.1	Graphe avec le logiciel NooJ . . . . .	30
5.2	Graphes avec le logiciel Intex . . . . .	31
5.3	Extrait du dictionnaire fléchi . . . . .	41
5.4	Variables communes de la grammaire flexionnelle . . . . .	42
5.5	Variables spécifiques de la grammaire flexionnelle . . . . .	43
5.6	Extrait de la grammaire flexionnelle . . . . .	43
5.7	Les codes de flexion utilisés pour les verbes . . . . .	44
6.1	Répartition des NColl en traits syntactico-sémantiques . . . . .	68
6.2	NColl dérivés et NColl non-dérivés . . . . .	71
6.3	Distribution des suffixes des NColl . . . . .	71
11.1	Grammaire des pronoms . . . . .	95
11.2	Grammaire des adverbes en <i>-ment</i> . . . . .	95
11.3	Grammaire des substantifs en <i>-ment</i> . . . . .	96
11.4	Sous-graphe : <i>cio</i> . . . . .	97
11.5	Grammaire des substantifs en <i>-ció</i> . . . . .	98
11.6	Grammaire locale des dates : MAIN . . . . .	100
11.7	Grammaire locale des dates : la setmana passada . . . . .	101
11.8	Grammaire locale des dates : el segle passat . . . . .	102
11.9	Grammaire locale des dates : dilluns 1 de desembre . . . . .	103
11.10	Grammaire locale des dates : A quarts de set . . . . .	103
11.11	Grammaire locale des dates : A les dues en punt . . . . .	104
11.12	Grammaire de désambiguïsation : les articles . . . . .	105
11.13	Grammaire de désambiguïsation : Les pronoms vs les articles . . . . .	106
11.14	Analyse de la phrase : Què els veus, els nois? . . . . .	107
11.15	Application du graphe de désambiguïsation . . . . .	108





# Liste des tableaux

4.1	Système de codage des parties du discours . . . . .	25
4.2	Les catégories grammaticales . . . . .	27
5.1	Nombre de graphes de flexion avec Intex et avec NooJ . . . . .	30
6.1	Homonymie et polysémie dans le trait Veg . . . . .	61
6.2	L’endroit et l’évènement (i) : locatif et abstrait . . . . .	64
6.3	L’endroit et l’évènement (ii) : locatif et abstrait . . . . .	65
6.4	Distribution des NColl . . . . .	70
8.1	Plausibilité . . . . .	84
8.2	La plausibilité dans des cas d’ambiguïté . . . . .	85
12.1	Reconnaissance de la grammaire locale des « dates » . . . . .	109
A.1	Codes de flexion des adjectifs . . . . .	119
A.2	Codes de flexion des verbes . . . . .	121
A.3	Codes de flexion des substantifs . . . . .	123



# **Première partie**

## **Introduction et cadre de la recherche**



# 1 Préliminaires

Ce travail s'inscrit dans le cadre des recherches du Laboratoire de Phonétique, Lexicologie et Sémantique (*fLexSem*) de l'Université Autonome de Barcelone. Il s'agit d'un projet de lexicographie informatique et lexicologie comparée catalan-français, dont l'objectif pratique est l'élaboration d'un dictionnaire électronique.

Le nombre croissant de documents informatisés rend indispensable le traitement automatique des langues, en vue de la gestion de l'information qu'ils contiennent, de leur structuration et d'en faciliter l'accès lors de procédures de recherche. Ce qui est toujours vrai pour les langues largement majoritaires l'est d'autant plus pour une langue moins répandue comme le catalan, qui génère de plus en plus de textes informatisés de tout type, y compris sur internet où l'on observe que le nombre de pages web en langue catalane est en croissance continue. Ainsi, si nombreuses entreprises et équipes de recherche continuent de produire des ressources linguistiques pour des langues telles que le français ou l'anglais, il s'avère urgent d'élargir les ressources pour le traitement informatique du catalan, qui sont encore insuffisantes.

Ainsi, comme nous l'avons signalé, notre travail a pour but pratique la réalisation d'un dictionnaire permettant des applications diverses, dans les différents domaines de la linguistique appliquée. Pour réaliser cette étude, nous nous sommes inspirée des travaux réalisés, à l'époque, dans le Laboratoire d'Automatique et de Documentation Linguistique de Paris 7 (LADL), ainsi que des travaux menés à terme au sein du Laboratoire de Linguistique Informatique de Paris 13 (LLI).

Notre dictionnaire a été conçu principalement en suivant les principes d'analyse et de description du lexique-grammaire (M. Gross et son équipe) et de la théorie des classes d'objets (G. Gross et son équipe). Nous avons également considéré pertinent d'introduire certains concepts et formalismes propres à la théorie Sens  $\Leftrightarrow$  Texte (I. Mel'čuk), comme les fonctions lexicales ou les significations grammaticales.

Notre travail est structuré en trois grandes parties. Dans la première, nous décrivons les différents types de dictionnaires issus de différentes approches

théoriques et les théories linguistiques qui les sous-tendent. Dans la deuxième partie de ce travail, nous présentons le dictionnaire électronique du catalan, tel que nous l'avons réalisé, en précisant les critères que nous avons utilisés pour son élaboration. Nous abordons les différentes problématiques qui se posent lors de la description des entrées, ainsi que le type d'information à fournir et la façon dont elle doit être codée pour que le dictionnaire puisse être manipulé par un ordinateur, et, donc, susceptible d'être intégré à de futures applications.

Dans la troisième partie, nous nous intéressons à une application directe de notre étude. En effet, nous avons adapté notre dictionnaire au système de développement linguistique NooJ (M. Silberztein) et nous l'avons complété par une série de grammaires, dont nous présentons également le fonctionnement. Le module NooJ du catalan a été mis à disposition libre des utilisateurs et permet de réaliser, entre d'autres, des recherches sur des corpora ou d'étiqueter linguistiquement un texte catalan. Dans la conclusion, nous récapitulons, comme il est d'usage, les principaux résultats obtenus lors de notre travail et présentons quelques pistes pouvant s'avérer intéressantes pour de futures recherches sur le sujet.

## 2 Lexicologie, Lexicographie et Dictionnairique

**Qu'est-ce qu'un dictionnaire électronique?** Le terme « dictionnaire » est normalement utilisé pour désigner des dictionnaires d'usage ou grand public conçus principalement pour l'utilisation/consultation par un humain. Ces dictionnaires peuvent être présentés sur différents supports ; sur support papier tels que *Le dictionnaire Le Robert*, *Diccionari de la Llengua Catalana* ou *Diccionari de la Real Academia de la Lengua Española*, ou sur support magnétique, c'est-à-dire sur CD-ROM, cartes de mémoire... tels que *la version informatisée de Le Robert* ou *L'hiperdiccionari de la llengua catalana*.

Cependant, en vue du traitement automatique des langues (TAL), les dictionnaires doivent pouvoir être entièrement interprétés par un ordinateur. En effet, pour qu'un dictionnaire puisse être utilisé par une machine il est indispensable tout d'abord qu'il y ait une forte consistance des données. Les informations attribuées doivent être cohérentes avec la structure globale du dictionnaire. Il ne peut pas y avoir, par exemple, de définitions circulaires ou d'informations implicites. Il faut que chaque code utilisé lors de la description soit univoque. Un dictionnaire visant le traitement automatique ne peut pas compter sur l'intuition du lecteur ni sur ses connaissances préalables. Un dictionnaire pour le TAL doit faire en sorte que toutes les informations nécessaires pour discriminer une lexie d'une autre soient présentes dans le dictionnaire. Or, les dictionnaires dont nous parlions précédemment, tout en étant informatisés, présentent ce type d'inconsistances formelles qui les rend inutilisables pour des applications en TAL. Mais il existe des dictionnaires conçus spécialement pour que l'utilisateur principal soit une machine ou plus exactement des logiciels, ce sont les « dictionnaires électroniques ». Les dictionnaires électroniques présentent donc les caractéristiques décrites ci-dessus et d'autres que nous verrons plus en détail dans ce chapitre. Tout au long de ce travail nous réserverons le terme « dictionnaire » pour désigner les dictionnaires électroniques, par opposition aux dictionnaires papier ou informatisés.



Dans ce chapitre, nous verrons différents exemples de dictionnaires électroniques tels que les dictionnaires électroniques du Laboratoire d'Automatique Documentaire et Linguistique (cf. §2.1) et les dictionnaires électroniques élaborés dans le Laboratoire de Linguistique Informatique (cf. §2.2). Finalement nous verrons un exemple de « dictionnaire théorique » qui est un tout autre type d'ouvrage lexicographique servant principalement de modèle expérimental, le Dictionnaire Explicatif et Combinatoire [Mel'čuk, 1984] (cf. §2.3).

Nous ferons référence, également, aux théories linguistiques qui sous-tendent ces dictionnaires et sur lesquelles repose la présente étude.

## 2.1 Le système DELA (LADL)

Vers la fin des années 80 la lexicographie prend une nouvelle orientation avec la croissante popularisation de l'informatique. Ainsi pour une science axée sur la description et le recensement de données, telle que la lexicographie, l'informatique ouvre un nouvel éventail de possibilités<sup>1</sup>. Vers les années 90 l'équipe de Maurice Gross au Laboratoire d'Automatique Documentaire et Linguistique (LADL) (Paris 7) comptait déjà avec tout un ensemble de ressources électroniques de la langue française sous forme de dictionnaires. L'ensemble de ces dictionnaires constitue le système DELA :

- Le dictionnaire des mots simples (DELAS)
- Le dictionnaire des mots composés (DELAC)
- Le dictionnaire phonémique (DELAP)

La répartition des données dans les différents dictionnaires a été effectuée d'après la définition formelle de mot, qui impose une distinction typographique : « *Les mots simples sont des séquences de lettres comprises entre deux séparateurs consécutifs. Les mots composés sont des séquences qui incluent au moins deux mots simples (et donc au moins un séparateur)* » [Silberztein, 1993].

A cet ensemble de dictionnaires, il faut ajouter une série de graphes et d'automates. Pour développer et appliquer ces dictionnaires directement sur des textes, ils ont été intégrés à la plateforme d'ingénierie linguistique Intex (M. Silberztein). Ce logiciel, qui était fondé sur la technologie de machines à états finis<sup>2</sup>, permettait d'une part, la génération automatique de toutes les formes fléchies et d'autre part l'alimentation semi-automatique du dictionnaire. Intex

---

1. Ce changement explique en partie l'adaptation des termes *lexicologie*, *lexicographie* et *dictionnaire* proposée par B. Quemada en 1987.

2. cf. §2.1.1.

a permis, par ailleurs, d'augmenter considérablement la couverture lexicale du système de dictionnaires DELA.

**DELAS** Le DELAS<sup>3</sup> était le dictionnaire de mots simples du français. Dans un premier temps, les mots du DELAS ont été puisés de différents dictionnaires de langue française (*Petit Robert*, *Larousse*...) ainsi que de textes en langue standard, notamment de *Le Monde* [Courtois, 1990]. Chaque mot du DELAS apparaît sous sa forme canonique : masculin singulier pour les noms et les adjectifs, les verbes à l'infinitif, etc. Toutes les entrées sont accompagnées de la catégorie grammaticale et d'un code décrivant leur modèle de flexion. Ces codes font appel à des automates qui permettent de fléchir automatiquement toutes les entrées du dictionnaire et de créer ainsi le dictionnaire de mots simples fléchis (DELAFS). En outre, chaque entrée est dotée d'un numéro qui indique l'appartenance à une couche lexicale donnée. Finalement, les noms sont suivis d'une marque sémantique (Humain, Végétal...) et les verbes sont associés à des tables de constructions syntaxiques (cf. Lexique-Grammaire §2.1.2).

**DELAC** Le DELAC était le dictionnaire électronique des mots composés du français. Dans le DELAC chaque entrée est accompagnée d'un code qui indique la flexion de chacun des composants du mot (en reprenant les mêmes codes utilisés dans le DELAS). Pour chaque entrée est précisé, également, le type morphologique : *NA* - *NDN* - *NAN* - *AN* - *NN* - *PN* - *VN*. Comme dans le DELAS, l'ensemble de formes fléchies des mots composés donne lieu au DELACF.

**DELAP** Le DELAP était le dictionnaire phonémique. Il a les mêmes entrées que le DELAS mais pour chaque entrée il précise la représentation phonémique de sa prononciation.

Le Laboratoire d'Automatique Documentaire et Linguistique (LADL) disposait, au cours des années 90, d'un dictionnaire de mots simples du français (DELAS) avec plus de 80 000 entrées, d'un dictionnaire de mots composés (DELAC) avec plus de 100 000 entrées, d'un dictionnaire phonémique (DELAP) et d'un dictionnaire de noms propres.

### 2.1.1 Les AEF en description linguistique

Parallèlement à l'élaboration de dictionnaires électroniques, le linguiste doit souvent rendre compte de régularités de la langue ou de variations formelles dont le nombre est beaucoup trop important pour en faire une description individuelle.

---

3. Le système DELA reposait en quelque sorte sur le DELAS. En effet, toute entrée du DELAP ou du DELAC devait être répertoriée dans le DELAS.

On peut représenter efficacement ce type de régularités ou de variations par le biais d'automates à états finis. Normalement la représentation d'un automate peut se faire à travers un graphe. Il est constitué d'un nœud initial (le point de départ) et d'un nœud final (point d'arrivée). Entre ces deux nœuds il peut y avoir des nœuds intermédiaires. Le nœud initial et le nœud final doivent être directement ou indirectement reliés entre eux.

D'un point de vue linguistique, un automate sert essentiellement à regrouper des éléments équivalents du point de vue du sens mais différents au niveau de la forme. Les automates rendent donc possible la représentation de phénomènes divers, tels que :

- Les variations orthographiques
- La génération automatique des formes fléchies en les associant à la forme canonique
- La description de phénomènes dérivationnels
- La levée d'ambiguïtés

Finalement un automate à états finis peut servir non seulement à la reconnaissance ou de façon beaucoup moins performante à la génération de textes, mais aussi pour alléger la saisie de données du dictionnaire.

### 2.1.2 Le lexique-grammaire

À la suite des travaux de Z.S. Harris sur les grammaires transformationnelles, Maurice Gross et son équipe du LADL ont commencé à développer, vers les années 70, une nouvelle approche formelle de la linguistique fondée sur l'observation et la classification systématique d'un grand nombre de propriétés syntaxiques du lexique français [Gross, 1975]. Selon M. Gross, la formulation de théories ne peut se faire que sous la vérification d'une analyse exhaustive d'un nombre appréciable de données. Ce souci empirique qui, d'ailleurs, l'a éloigné de ses contemporains, a fait naître la méthodologie du lexique-grammaire.

Les premiers travaux réalisés au LADL se centrent sur la syntaxe du verbe français, notamment, sur les verbes à constructions complétives [Gross, 1975] et les verbes à constructions transitives et intransitives.

Ainsi plus de 10 000 verbes ont été décrits et classés par rapport à ces propriétés syntaxiques en suivant une démarche formelle très précise qui s'applique encore de nos jours. Les données sont recensées dans des tables booléennes où les lignes correspondent aux termes à décrire, et les colonnes aux différentes propriétés transformationnelles et distributionnelles

À partir de ces études, M. Gross a constaté qu'il existe des groupes de verbes avec des propriétés syntaxiques communes qui sont aussi relativement homogènes du point de vue du sens [Gross, 1975, p. 218]. Par exemple, dans la classe des

verbes se construisant avec un sujet ( $N_0$ ) et d'une complétive infinitive ( $V_{inf}$   $\Omega$ ) [Gross, 1975, p. 175] on trouvera essentiellement des verbes de mouvement.

Les propriétés syntaxiques ont pour Gross « *une fonction classificatoire* » [Gross, 1975, p. 225] très pertinente du point de vue sémantique. Gross considère que « *les propriétés syntaxiques sont déterminées par des propriétés sémantiques* » [Gross, 1975, p. 221]. L'analyse des propriétés syntactico-distributionnelles devrait donc permettre d'atteindre le sens. D'ailleurs, Gross met en évidence qu'une analyse beaucoup plus exhaustive des propriétés transformationnelles et distributionnelles montrerait « *qu'il n'existe pas deux verbes qui ont les mêmes propriétés syntaxiques* » [Gross, 1975, p. 214]. Selon Gross, cela prouve qu'une description de la langue par des règles et des exceptions n'est pas adéquate au traitement automatique des langues, car, au moins, tous les verbes constituent des exceptions. Il est donc nécessaire de procéder à une description détaillée des propriétés de chaque élément du lexique en tenant compte des contextes où il peut apparaître et des transformations qu'il peut subir dans ces contextes.

Pour le lexique-grammaire, l'unité de sens élémentaire est la phrase simple. Une phrase simple doit être constituée d'un prédicat et des arguments que celui-ci sélectionne. Ce prédicat, qui est le noyau de la phrase, peut être un verbe, un adjectif, un nom ou une préposition. Les études du lexique-grammaire portant sur des parties du discours autres que le verbe ont forcé l'intégration d'une notion déjà avancée par Z.S. Harris, *les verbes supports*. Ce terme rebaptisé par A. Daladier en 1978, désigne des verbes sémantiquement vides qui accompagnent un prédicat et peuvent exprimer des valeurs aspectuelles. En effet, lorsque le prédicat est un adjectif ou un nom, il doit être associé à des verbes qui actualisent la phrase. Les verbes supports se distinguent des prédicats verbaux par un grand nombre de propriétés syntaxiques. Nous ne donnons ici que quelques exemples :

- Le verbe support n'admet pas l'extraction en *c'est... que* :
  - *Nil fait une promenade*
  - *\*C'est une promenade que Nil fait*
- Le verbe support peut être effacé sans que le sens de la phrase soit modifié. La phrase qui résulte de l'effacement est du type *N de N* :
  - *Nil a fait un voyage*
  - *Le voyage de Nil*

Le LADL compte avec une description de plus de 25 000 verbes. D'autres analyses fondées sur le lexique-grammaire se sont poursuivies sur les différentes parties du discours, e.g. les adverbes, les conjonctions de subordination, les adjectifs ou les substantifs.

## 2.2 Les dictionnaires du LLI

Par ailleurs, au sein du Laboratoire de Linguistique Informatique (LLI) (Paris 13) ont été élaborés des dictionnaires électroniques. Les dictionnaires du LLI incorporent la notion de classes d'objets dans les dictionnaires, en développant ainsi tout un système de description sémantique des unités lexicales (cf. §2.2.1). Ces dictionnaires comportent 4 grands types d'informations différentes : morphologiques, syntactico-sémantiques, diasystématiques et de traduction. Les différentes rubriques sont ainsi organisées<sup>4</sup> :

1. Les champs morphologiques
  - (a) Le champ G qui indique la catégorie grammaticale de l'entrée.
  - (b) Le champ F qui décrit le type de flexion des adjectifs, des noms et des verbes. Chaque entrée est suivie d'un code qui fait appel à un graphe qui génère automatiquement le mot fléchi. Le graphe décrit les transformations formelles que doit subir le mot pour générer les formes fléchies qui lui correspondent.
  - (c) Le champ V qui attribue à chaque entrée les variantes orthographiques si elles existent.
2. Les champs syntactico-sémantiques
  - (a) Le champ T précise les traits syntactico-sémantiques.
  - (b) Le champ C indique la classe d'objets (cf. §2.2.1).
  - (c) Le champ W qui est réservé aux prédicats établit le nombre et le type d'arguments sélectionnés par le prédicat ; il explicite, donc son *schéma d'arguments*.
3. Domaines et registres de langues
  - (a) Le champ S précise le domaine.
  - (b) Le champ R précise le niveau de langue de l'entrée : *familier, populaire, littéraire, vieux, régional*.
4. La traduction

### 2.2.1 Les classes d'objets

Comme nous venons de le voir, dans les travaux du lexique-grammaire la syntaxe prime sur tout autre aspect de la langue. La sémantique est pratiquement

---

4. Pour élaborer notre dictionnaire nous avons adopté un format analogue à celui des dictionnaires du LLI. Puisque nous détaillerons chacun des champs tout au long de la deuxième partie de ce mémoire nous ne donnons ici qu'un bref aperçu extrait directement de l'article [Gross, 1992].

absente si ce n'est que par quelques critères conceptuels utilisés dans la description des prédicats (e.g. Nom Humain (*Hum*). . .)

Cependant, il s'avère que certains prédicats sélectionnent non seulement le nombre et la construction de leurs arguments, mais aussi l'ensemble sémantique auxquels ils appartiennent :

- *Nil mange un bonbon* (N<sub>0</sub>:Hum)
- (?)*La cafetière mange un bonbon*

Ainsi, le prédicat *manger* impose certaines contraintes sémantiques sur son premier argument (N<sub>0</sub>) qui ne peut pas être un inanimé concret (Conc). La caractérisation des arguments en traits syntactico-sémantiques<sup>5</sup> permet de désambiguïser un certain nombre d'emplois :

- *punaise* : Animal
- *punaise* : Concret

Mais souvent, cette classification en traits syntactico-sémantiques demeure imprécise, notamment dans les cas de polysémie :

- *Les enfants jouaient à la balle dans le grand jardin de la maison.*
- *Un soldat a alors tiré une balle dans la tête de S. depuis une distance de 15 mètres*<sup>6</sup>.

G. Gross [Gross, 1992] et M. Mathieu-Colas [Mathieu-Colas, 1994] constatent qu'il existe des classes sémantiquement et syntaxiquement homogènes au sein des noms ayant un même trait syntactico-sémantique. Lorsque ces noms sont des arguments, ces classes sont appelées *classes d'objets*.

Une classe d'objets se définit par les prédicats qui lui sont spécifiques, les prédicats appropriés à la classe, c'est à dire un ensemble de prédicats qui peut sélectionner en position argumentale tous les constituants de la classe.

- Exemple de plusieurs prédicats pour caractériser une classe

- Les voies de communication (<voie>) [Mathieu-Colas, 1998] :  
*autoroute, chemin, tunnel, boulevard, passage. . .*
- Nhum *aller* quelque part par <voie>
- Nhum *prendre* <voie> (pour aller quelque part)
- <voie> être *impraticable*

---

5. Abstrait, Animal, Concret, Humain, Locatif, Végétal, Temps.

6. Dans ces deux exemples, le mot *balle* a le trait syntactico-sémantique *Conc* et pourtant il ne s'agit pas de la même unité lexicale dans l'un et l'autre exemple. En effet, dans le premier exemple, on fait référence à un 'jouet' tandis que dans le deuxième exemple on fait référence à un 'projectile'.

- Exemple d'un seul prédicat pour caractériser une classe
  - La classe des imprimés (<s-i>) [Le Pesant, 1994] : *affiches, chèques, livres, passeports...*
  - *imprimer* <s-i>

Ce même principe est utilisé également pour le classement des prédicats. Une classe de prédicats ou classe sémantique peut se définir en fonction des verbes supports ou de la classe d'arguments qu'elle sélectionne.

Le dictionnaire électronique doit mettre en relation chaque classe d'objets ou classe de prédicats avec tous les éléments du lexique qui les définissent. Les classes d'objets, créées au départ comme un système formellement efficace pour la désambiguïsation, ont révélé de multiples applications en vue du traitement automatique des langues et notamment pour la traduction automatique.

## 2.3 Le dictionnaire explicatif et combinatoire

### 2.3.1 Le format du DEC

Dans le cadre des recherches menées par le Groupe de Recherche en Sémantique, Lexicologie et Terminologie (GRESLET) de l'Université de Montréal, Igor Mel'čuk et son équipe ont élaboré un nouveau modèle de dictionnaire issu de la Théorie Sens $\Leftrightarrow$ Texte (TST). Selon cette théorie, un DEC doit fournir toutes les informations nécessaires pour qu'un locuteur (ou un ordinateur) puisse construire tous les énoncés possibles à partir d'un sens donné. Il existe à présent, quatre volumes du Dictionnaire explicatif et combinatoire du français contemporain (DECFC) publiés en format papier. Le premier volume a un total de 50 vocables<sup>7</sup> [Mel'čuk, 1984], le deuxième volume 107 vocables (1988), le troisième volume 171 vocables (1992) et le quatrième volume 180 vocables (1999). L'ensemble des volumes compte donc un total de 510 vocables français avec 1 583 définitions lexicales.

Pour assurer la rigueur et le formalisme nécessaires pour devenir un outil informatique, le DEC a une structure préalablement déterminée au niveau de son organisation : la microstructure, qui correspond à la manière dont est organisée l'information à l'intérieur de chaque entrée, et la macrostructure, qui correspond à l'organisation des entrées entre elles.

---

7. Pour une définition de vocable, cf. §2.3.1.2.

### 2.3.1.1 La Microstructure

Un article du dictionnaire explicatif et combinatoire correspond à une seule lexie munie de différentes informations qui la définissent. Une lexie est « *soit un mot pris dans une acception bien spécifique (=lexème), soit encore une locution, elle aussi, prise dans une acception bien spécifique (=phrasème)* » [Mel'čuk et al., 1995, p. 16]. La microstructure du dictionnaire correspond au type d'information fournie dans chaque article du dictionnaire et à la manière dont cette information est organisée. Nous pouvons distinguer cinq zones d'information :

1. zone phonologique : transcription phonétique de la lexie
2. zone sémantique : définition lexicographique
3. zone de combinatoire
  - (a) morphologique : parties du discours et morphologie flexionnelle
  - (b) stylistique : marques d'usage qui indiquent le contexte dans lequel la lexie peut être insérée
  - (c) syntaxique
  - (d) lexicale restreinte
4. zone d'exemples
5. zone phraséologique

De la microstructure du DEC nous ne décrivons que les zones 2, 3c, 3d et 5, puisque c'est là que repose l'originalité et surtout la difficulté du DEC.

**La zone sémantique** Bien que partant d'une théorie<sup>8</sup> où les représentations sémantiques sont décrites moyennant des réseaux (arcs, flèches et nœuds), dans le DEC, la définition (i.e. la représentation sémantique) des lexies est linéaire à cause de la difficulté qu'un tel réseau entraînerait pour la lecture, ainsi que pour l'introduction des données. Cependant, et malgré cette linéarité, les relations sémantiques que chaque lexie entretient avec les autres sont bien mises en évidence dans la définition. En effet, pour définir une lexie, le sens est paraphrasé au moyen d'unités lexicales considérées sémantiquement plus simples. Le sens d'une lexie *x* est considéré comme une composition de plusieurs lexies qui incluent la lexie *x* mais pas à l'envers, i.e. qui ne sont pas incluses par elle<sup>9</sup>. Dans le cas des lexies prédicatives la définition est une proposition à variables<sup>10</sup> (représentées par des lettres majuscules) qui renvoient aux actants sémantiques du prédicat.

8. cf. La théorie Sens $\Leftrightarrow$ Texte §2.3.2.

9. Ce principe poussé à l'extrême se veut une étude des primitifs sémantiques de la langue cible. Pour plus d'information cf. Wierzbicka, 1994 apud [Mel'čuk et al., 1995, p. 91].

10. Appelée dans [Mel'čuk et al., 1995] « *forme propositionnelle* ».



**La zone de combinatoire syntaxique.** Dans la zone de combinatoire syntaxique, un schéma de régime de la lexie décrit le comportement syntaxique de tous ses arguments (appelés actants sémantiques). Ce schéma est présenté sous forme de tableau où chaque actant sémantique est associé à un actant syntaxique qui précise la façon dont il peut être exprimé, soit par la forme de surface qu'il doit adopter (infinitif, syntagme prépositionnel avec telle préposition. . .) soit par la relation syntaxique qu'il entretient avec l'actant sémantique (ce type de relation est représenté dans le DEC par des chiffres romains et correspond *grosso modo* à ce que la grammaire traditionnelle appelle sujet, complément. . .). Par ailleurs, la zone de combinatoire syntaxique donne une liste de toutes les contraintes de cooccurrence des actants syntaxiques, ainsi qu'un exemple pour chaque cas.

**La zone lexicale restreinte.** De la même façon que sont spécifiées les relations entre la lexie et ses actants, le DEC précise également les relations qu'une lexie  $x$  entretient avec d'autres lexies. Cette information apparaît dans la zone de combinatoire lexicale. Pour exprimer certains sens, comme par exemple l'intensité, la collectivité. . . une lexie se combine de façon restreinte avec d'autres lexies. Pour représenter ce type de combinaisons, Mel'čuk a introduit la notion de fonction lexicale (dorénavant FL). D'un point de vue formel, une FL correspond à une fonction mathématique

$$f(x)=y$$

où  $f$  est la fonction,  $x$  est « le mot clé ou argument » de la fonction et  $y$  sa « valeur ». Cependant, contrairement à la fonction mathématique, une FL est multivoque, i.e. on peut associer plus d'une valeur à un seul mot-clé. Par exemple, dans la fonction « intensificateur » qui exprime le sens de 'très', 'intense' de  $x$ , pour le mot-clé *amour* on obtient deux valeurs :

$$\text{Magn}(\text{amour}) = \text{ardent} ; \text{fou}$$

Il existe deux types majeurs de FL : les fonctions lexicales standard et les fonctions lexicales non standard. Une FL standard diffère d'une FL non standard parce qu'elle s'applique à un grand nombre de lexies et possède des valeurs distinctes en fonction du mot-clé. Les FL standard ont en plus un caractère universel et leur nombre est plus ou moins limité<sup>11</sup>.

Ainsi, pour chaque lexème, le DEC donne toutes les valeurs de toutes les FL dont le lexème peut-être le mot clé, et toujours dans un ordre préétabli pour préserver l'homogénéité du dictionnaire (e.g. les FL standard précèdent les non standard). Voici quelques exemples de FL du lexème *brouillard* (cf. [Mel'čuk, 1984, p. 42]) :

$$\text{Syn}_C(\text{brouillard}) : \text{brume}$$

---

11. Des informations complémentaires sur la notion de fonction lexicale sont données dans la section § 2.3.2.

**Syn**<sub>∩</sub>(brouillard) : *brouillasse, bruine, crachin, embrun*

**Gener**(brouillard) : *phénomène atmosphérique*

**Figur**(brouillard) : *rideau, voile [de ~]*

**Magn**(brouillard) + **Figur**(brouillard) : *écran de [de ~], < mur [de ~]*

**Sing**(brouillard) : *banc, filet, nappe [de ~]*

**Magn**(brouillard) : *épais, dense, opaque* ; **fam** *à couper au couteau* **fam** *purée de pois*

**AntiMagn** : *fin, léger*

**Oper**<sub>0</sub> : *[il] y a [du ~]*

**La zone phraséologique.** Finalement, le DEC inclut une dernière zone où apparaissent toutes les locutions figées ou semi-figées<sup>12</sup> dont la lexie fait partie ; c'est la zone phraséologique. Nous rappelons que ce type de locutions, qui sont des lexies, constituent une entrée de dictionnaire à part entière. Dans cette zone, il n'y a donc pas de définitions. Il s'agit uniquement d'une référence à d'autres entrées du dictionnaire.

### 2.3.1.2 La Macrostructure

Comme nous venons de le voir, la microstructure du DEC comprend quatre grands blocs d'information : les informations sémantiques, les informations syntactico-combinatoires, les informations lexico-combinatoires et les exemples. Toutes ces informations sont soumises à des règles qui déterminent non seulement la présentation formelle mais aussi l'ordre d'apparition. De la même manière, une structure rigide détermine l'ordonnancement de l'ensemble des entrées entre elles, c'est la macrostructure. Un article du DEC correspond à une unité lexicale. Cependant, il s'avère que souvent des lexies ayant un même signifiant présentent des liens sémantiques entre elles. Le DEC représente ces liens sémantiques au niveau de la macrostructure. En effet, dans le DEC, ces lexies sont regroupées sous une même entrée commune appelée *vocab*. Le vocab regroupe donc des lexies en relation de polysémie. Les homonymes, par contre, sont regroupés sous des vocables différents. Ainsi, le vocab « *n'est pas un article de dictionnaire mais un ensemble d'articles apparentés par leur sémantisme et leur forme – un superarticle* » [Mel'čuk et al., 1995, p. 156]. À l'intérieur d'un même vocab les lexies sont également ordonnées par des considérations sémantiques (e.g. par la proximité sémantique entre elles, etc.).

---

12. Phrasèmes complets et quasi-phrasèmes cf. §2.3.2.

### 2.3.1.3 Une version électronique du DECFC : le DiCo

Bien que le DEC ait été conçu pour être un dictionnaire électronique, ce n'est que postérieurement à sa première version papier que l'Observatoire de Linguistique Sens  $\Leftrightarrow$  Texte a commencé l'informatisation du dictionnaire. Ainsi, une version simplifiée et formalisée du DECFC a donné lieu au Dictionnaire de Cooccurrences (DiCo). Par rapport au DECFC et comme l'indique son nom, le DiCo se focalise essentiellement sur zone de combinatoire lexicale [Polguère, 2003b]. De la zone sémantique, le DiCo conserve la forme propositionnelle ; et au lieu de la définition lexicographique, dans le DiCo, le sens des lexies est pris en charge par un système d'étiquettes sémantiques.

Toutes les informations d'une lexie sont stockées dans une base de données sous forme de fiches. Comme dans le DECFC les lexies sont regroupées en vocables et c'est donc dans le vocable que sont réunies l'ensemble de fiches pour chacune des lexies appartenant au même vocable. Une fiche du DiCo contient 17 champs d'information qui peuvent être traités de façon indépendante<sup>13</sup>.

La base de données du DiCo prévoit de possibles implémentations pour le traitement informatique des langues ainsi que pour la recherche lexicologique/lexicographique. Cependant, actuellement, le DiCo comporte la description d'environ 800 lexies, soit près de 400 vocables. Le transfert de données du DECFC ne peut pas être fait de façon automatique et il s'avère qu'une telle description exigerait un important coût de temps.

### 2.3.2 La Théorie Sens $\Leftrightarrow$ Texte

Vers 1960 les recherches menées par Igor Mel'čuk, Alexandre Žolkovskij, Juri Apresjan et ses collègues donnent lieu à une nouvelle théorie, la Théorie Sens  $\Leftrightarrow$  Texte (TST). Cette théorie repose sur la construction d'un système de règles permettant la correspondance doublement multivoque entre un sens  $s$  et la totalité des énoncés de la langue véhiculant le sens  $s$  et inversement.

Selon cette théorie, la description d'une langue doit modéliser le processus qui permet à un locuteur de faire correspondre à un sens  $s$  tous les énoncés possibles d'une langue ayant le sens  $s$  et inversement. La TST a donc pour but l'élaboration de modèles représentant ce type de correspondances pour une langue donnée. Toutefois, bien que ces correspondances soient bidirectionnelles (sens  $\Leftrightarrow$  texte), la TST privilégie la synthèse, c'est-à-dire le passage des sens aux textes (la production). Un modèle sens-texte est donc un moteur de paraphrase qui à partir d'un sens (une représentation sémantique) génère toutes les formes d'expression de ce sens (textes).

---

13. Un exemple de fiche peut être consultée en ligne à travers l'interface graphique « DiCouèbe » à <http://olst.ling.umontreal.ca/dicouebe/>

À cet effet, un Modèle Sens  $\Leftrightarrow$  Texte est organisé en différents niveaux de représentation (cf. fig. 2.1 page 19).

À l'exclusion du niveau sémantique, tous les autres niveaux ( $N$ ) sont divisés en  $N$  profond et  $N$  de surface. Le  $N$  profond est orienté vers le sens ; il met donc en évidence les caractéristiques sémantiques propres à son niveau, tandis que le  $N$  de surface, étant orienté vers le texte, fait ressortir les caractéristiques formelles. Chaque représentation possède une structure interne et une représentation formelle particulière (réseaux sémantique pour la RSém, des arbres de dépendance pour la RSyn, etc.)<sup>14</sup>. Le passage entre deux niveaux de représentations adjacents est assuré par un ensemble de règles regroupées sous les 6 composantes<sup>15</sup>. Une composante possède deux types de règles, celles qui, comme nous venons de le dire, assurent la transition entre deux niveaux<sup>16</sup> différents et adjacents et celles qui agissent à l'intérieur du même niveau : « les règles de transformation structurale » [Polguère, 1998]. L'application de ces règles dépend en grande mesure du dictionnaire explicatif et combinatoire. Par exemple, pour passer de la RSém à la RSynP, il faut prendre en compte quelques règles de transition de la composante sémantique. Dans une représentation sémantique, qui a la forme d'un graphe, les nœuds sont étiquetés par des sémantèmes. Un sémantème « *correspond grosso modo au sens d'une unité lexicale* » [Mel'čuk, 1993, p. 50] Par contre, dans la RSynP les nœuds sont des unités lexicales. Ce passage est possible grâce à l'intervention du DEC. Si le sémantème est élémentaire, le transfert est presque direct à l'aide du DEC, puisqu'il s'agit directement d'une lexie. Par contre, si le sémantème qui apparaît dans la RSém est complexe, c'est-à-dire qu'il est constitué de plusieurs sémantèmes plus simples, il faut chercher dans la zone sémantique du DEC (cf. §2.3.1.1) le sens qui s'ajuste le mieux à cet ensemble de sémantèmes pour réaliser ainsi le passage d'un niveau à l'autre. Nous rappelons brièvement que dans la zone sémantique du DEC les définitions se font par décomposition du sens en unités lexicales plus simples.

Parmi les règles de transformation structurales de la composante sémantique il y a deux types de règles qui servent à effectuer les paraphrases sémantiques<sup>17</sup> : D'une part, les règles sémantiques proprement dites, qui déterminent les différents

14. Étant donné que ce n'est pas là notre objet d'étude, pour plus de détails sur les représentations de la TST cf. [Mel'čuk, 1984], [Mel'čuk, 1997] et [Polguère, 1998].

15. Ces six composantes et le DEC constituent la grammaire TST.

16. Ces règles sont appelées « règles de dérivation », mais nous n'utiliserons pas ce terme pour ne pas porter à confusion.

17. Il existe dans les langues naturelles deux types de paraphrases :

- Les paraphrases syntaxiques qui ont les mêmes unités lexicales mais diffèrent par leur structure syntaxique :
  - *Les étudiants qui ont une autorisation peuvent sortir*
  - *Les étudiants ayant une autorisation peuvent sortir*

choix lexicaux pour réaliser un type de paraphrase sémantique ; d'autre part, les règles sémantiques de paraphrase, qui réalisent un autre type de paraphrase sémantique, en effectuant des substitutions lexicales en fonction des liens existants entre les lexies. Pour procéder à ce dernier type de paraphrase, l'utilisation du DEC s'impose ; les substitutions lexicales, ne peuvent être réalisées sans une analyse exhaustive des lexies et tout particulièrement sans une description en termes de fonctions lexicales. En réalité, c'est ce type de paraphrases qui a motivé la création de cette nouvelle notion linguistique, les fonctions lexicales.

Comme il a été dit dans la section §2.3.2, une fonction lexicale est une représentation formelle qui sert à encoder des informations non prédictibles sur la combinatoire des lexies. Il y a deux types majeurs de FL, les fonctions lexicales standard qui ont un caractère universel et les fonctions lexicales non standard qui sont particulières à la langue. Cependant, les FL peuvent être également classées par rapport à l'information qu'elles véhiculent :

- *Les fonctions lexicales paradigmatiques* qui nous donnent des informations sur les rapports sémantiques entre le mot-clé et d'autres lexies de la langue (les synonymes, les hyponymes, les antonymes. . .)
- *Les fonctions lexicales syntagmatiques* qui nous donnent des informations sur la combinatoire restreinte du mot-clé (collocations).

Quelques exemples de FL sont proposés tout au long de ce mémoire, et une sélection d'exemples de [Mel'čuk et al., 1995] est faite dans les annexes.

Comme nous venons de le montrer le modèle Sens $\Leftrightarrow$ Texte n'est donc pas concevable sans une description particulière de chaque unité lexicale, d'où l'importance d'élaborer des dictionnaires formels et informatissables.

- 
- Les paraphrases sémantiques qui n'ont pas exactement les mêmes unités lexicales :
    - *Elle aime le chocolat*
    - *Le chocolat lui plaît*

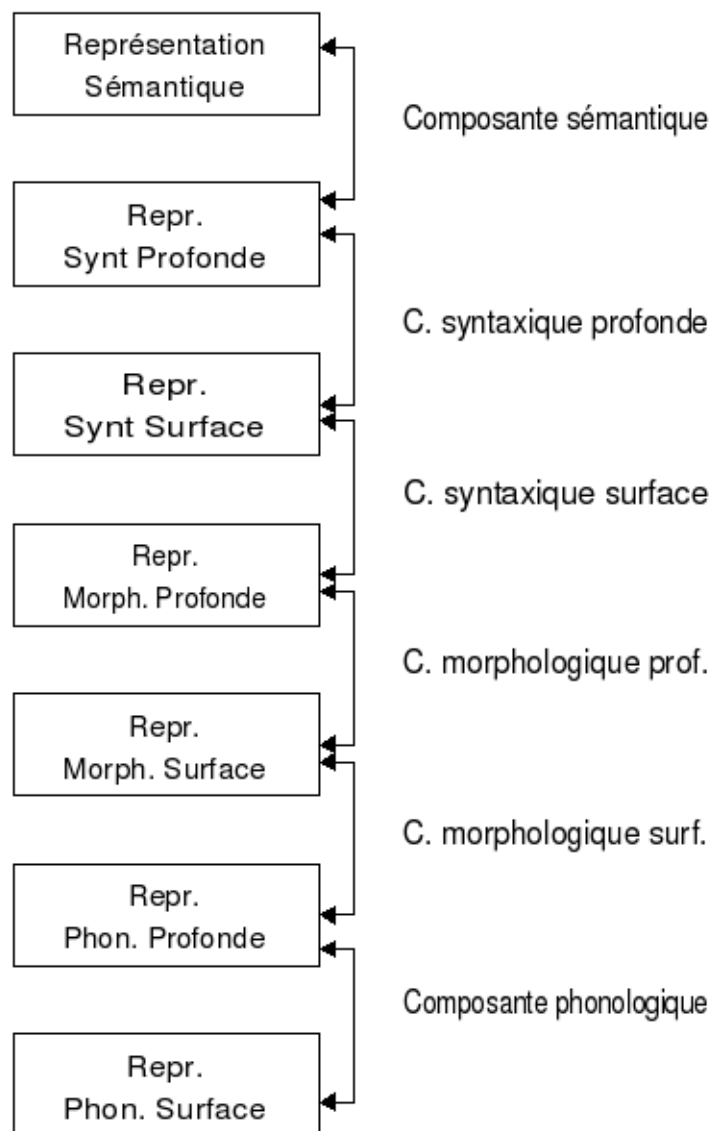


FIG. 2.1: Niveaux de représentation linguistique de la TST



**Deuxième partie**

**Dictionnaire Électronique**  
**Coordonné du Catalan**





### 3 Présentation du dictionnaire

Dans le cadre du système de dictionnaires électroniques du Groupe *fLexSem* (Phonétique, Lexicologie et Sémantique) de l'Université Autonome de Barcelone, nous avons élaboré un prototype de dictionnaire électronique du catalan coordonné avec l'espagnol et le français. Le dictionnaire comporte à l'heure actuelle 39 207 lemmes qui ont été puisés de textes catalans procédant de différentes sources en catalan standard, notamment d'un corpus de 15 Mo de texte provenant de pages webs en catalan<sup>18</sup>. À partir de ce corpus et à l'aide du système de développement Intex/Nooj (cf. §2.1) nous avons dressé une première liste de mots-formes qui ensuite a été complétée à l'aide de dictionnaires en langue catalane [GDL, 1998].

En nous inspirant du système DELA (cf. §2.1), dans cette première liste de mots-formes nous avons recensé uniquement des mots typographiquement simples, i.e. sans aucun caractère séparateur et sans majuscules.

Le dictionnaire électronique du catalan a été réalisé sous forme de tables d'une base de données. Ce format nous permet à tout moment de corriger, de modifier et d'augmenter le nombre d'entrées du dictionnaire de façon libre et autonome, sans qu'il y ait de répercussions dans les autres parties du dictionnaire. C'est d'ailleurs cette souplesse qui nous permettra à terme d'implémenter les recherches parallèles sur le catalan qui sont menées au sein du groupe *fLexSem*. Actuellement, notre dictionnaire a intégré les équivalents de traduction des noms de professions dressée par S. Fuentes [Fuentes, 2006c] dans le cadre de sa thèse de doctorat en cours ainsi qu'une liste de 1 257 adverbess en catalan réalisée par D. Català [Català, 2004]. Par ailleurs, nous comptons inclure à terme une description des verbes intransitifs catalans réalisée par C. Geiss [Geiss, 2006] à partir d'une première version de notre dictionnaire qui comprenait 6 918 verbes.

La macrostructure et la microstructure du dictionnaire ont été développées d'après le format communément utilisé dans les dictionnaires du groupe

---

18. Ce corpus a été réalisé par Magnus Finkenzeller et Thorsten Vobl dans le cadre du Master PLN de l'Université Autonome de Barcelone (2005).

*fLexSem*<sup>19</sup>. Nous considérons que la macrostructure du dictionnaire doit comprendre uniquement des unités lexicales. C'est à dire que chaque entrée doit correspondre, non seulement à une forme (*a*), mais à un triplet constitué d'une forme (*a*), d'un sens ('*a*') et d'une combinatoire ( $\Sigma a$ ) [Mel'čuk et al., 1995]. Pour ce faire il faut évidemment prendre en compte des aspects propres à la microstructure tels que la combinatoire qui est décrite dans les champs des propriétés syntactico-sémantiques. En effet, macrostructure et microstructure sont fortement imbriquées et c'est en élaborant la microstructure que s'est commencée à ériger la macrostructure du dictionnaire.

Pour atteindre une telle macrostructure et compte tenu de la perspective purement informatique de ce dictionnaire, chaque mot-forme de la liste initiale a été dédoublé dès que l'un des champs de description de la microstructure différait. Tout au long de cette partie nous verrons plus en détail dans quels cas et sous quelles spécificités nous avons dédoublé nos entrées. Le but de ce prototype étant d'obtenir un premier dictionnaire électronique du catalan le plus complet possible du point de vue de la reconnaissance de formes simples, nous nous sommes centrée, dans un premier moment, sur la description de la morphologie flexionnelle.

En deuxième lieu, nous avons introduit des champs d'information syntactico-sémantique et diasystématique et finalement nous avons proposé un équivalent de traduction. Nous développerons par la suite chacune de ces rubriques de façon à ce que les grands champs de la microstructure du dictionnaire correspondent à un chapitre de ce mémoire : les catégories grammaticales (chap. §4), la flexion (chap. §5), les traits syntactico-sémantiques (chap. §6), les classes d'objets (chap. §7), la plausibilité (chap. §8) et le champ des équivalents de traduction (chap. §9).

---

19. Eux-mêmes inspirés des dictionnaires du système DELA et des dictionnaires du LLI.

## 4 Les catégories grammaticales

La première information que nous avons attribuée aux entrées de notre dictionnaire correspond à la catégorie grammaticale. Voici les abréviations que nous avons utilisées pour chaque partie du discours :

<i>A</i>	adjectif
<i>ADV</i>	adverbe
<i>CONJ</i>	conjonction
<i>DET</i>	déterminant
<i>INTJ</i>	interjection
<i>Nm</i>	nom masculin
<i>Nf</i>	nom féminin
<i>PREP</i>	préposition
<i>PRO</i>	pronom
<i>V</i>	verbe
<i>Vp</i>	verbe essentiellement pronominal

TAB. 4.1: *Système de codage des parties du discours*

Les catégories grammaticales regroupent les différentes lexies par rapport à un certain nombre de propriétés grammaticales de nature très hétérogène. Ainsi, l'union de différents critères relevant de la morphologie et de la syntaxe donnent lieu à l'établissement des parties du discours.

Lors de l'étiquetage nous avons essayé de conserver certaines informations pouvant être nécessaires dans des analyses ultérieures. Par exemple, dans le cas des verbes un « p » minuscule indique qu'il s'agit d'un verbe pronominal pur. Ce type d'information a été donné pour 211 verbes. Nous sommes consciente qu'il devrait y avoir un champ spécifique dans la microstructure du dictionnaire pour coder ce type d'information, mais dans cette première version du dictionnaire nous nous sommes centrée principalement sur la description des substantifs. Cette information a été donnée de façon purement supplémentaire et en vue d'une récupération lors d'une analyse plus détaillée de cette catégorie. Puisque, comme

nous venons de le dire, la description des verbes n'a pas encore été réalisée, les mots ayant la catégorie grammaticale V ne peuvent pas être considérés comme des unités lexicales, pour l'instant ce sont uniquement des mots-formes<sup>20</sup>. Nous rappelons brièvement qu'une étude sur une partie des verbes intransitifs de notre dictionnaire a été menée à terme par Chiyo Geiss de l'Université LMU (Munich), dans le cadre du Master PLN de l'Université Autonome de Barcelone. Dans son étude [Geiss, 2006] elle constate que 1 256 mots-formes correspondent environ à 3 660 unités lexicales. Notre dictionnaire, compte à présent un total de 8 324 verbes pouvant appartenir à des ensembles très variés comme :

- des verbes supports: *fer, sentir, practicar...*
- des verbes prédicatifs: *somriure, descansar, menjar...*
- des verbes auxiliaires, modaux ou phasiques: *ser, haver, començar a...*

En ce qui concerne les pronoms et les déterminants, bien qu'ils apparaissent dans la liste d'abréviations (cf. fig. 4.1) ils ne font pas partie de la macrostructure du dictionnaire. Ces deux catégories présentent de nombreuses irrégularités morphologiques, ce qui rend difficile le codage des flexions. En outre, nous croyons que ces éléments relèvent plutôt de la syntaxe, c'est pourquoi ils ont été traités sous forme de grammaires (cf. §11). Ainsi notre dictionnaire est complété par une grammaire des pronoms, des démonstratifs, des articles et des déterminants numéraux.

Un tout autre traitement a été donné aux déterminants nominaux qui pour l'instant ont une place provisoire au sein de la macrostructure. Dans cette première version du dictionnaire les déterminants nominaux ont été classés sous la catégorie grammaticale des noms. Ils portent toutefois une étiquette spécifique: *detnom*. Souvent, ces les déterminants nominaux correspondent selon la terminologie de Mel'čuk à des fonctions lexicales :

EXEMPLE 1 :

1.a : **Sing**(*llana*)=*cabdell, fil*

1.b : **Mult**(*all*)=*rest*

Par ailleurs, nous avons identifié un total de 14 interjections (*bravo, visca, adéu...*), 17 conjonctions (*doncs, ergo, però...*) et 2 158 adverbes. Parmi les adverbes, 1 257 ont été intégrés à notre dictionnaire à partir d'une liste réalisée par D. Català; le reste a été extrait directement du *corpus*. Les adjectifs sont au nombre de 6 310 et parmi les de 22 144 substantifs 11 079 sont des noms masculins et 11 065 des noms féminins, mais la question du genre sera abordée dans le chapitre suivant §5.

---

20. Cela est valable pour toutes les catégories mis à part les substantifs.

Ainsi, à l'heure actuelle, notre dictionnaire compte 39 207 entrées, leur répartition en catégories grammaticales est détaillée dans le tableau 4.2 de la présente page.

ABBRÉV	A	ADV	CONJ	INTJ	N	PREP	V
ENTRÉES	6 310	2 158	17	14	22 144	29	8 535
%	16,09	5,5	0,04	0,04	56,48	0,07	21,77

TAB. 4.2: *Les catégories grammaticales*

Nonobstant, et malgré la tradition qui précède les catégories grammaticales, leur répartition n'est pas exempte de problèmes. Comme dans toute classification, il existe des éléments qui se situent à la frontière de deux catégories. Il est assez fréquent qu'un même mot puisse apparaître dans des contextes syntaxiques différents sans qu'il y ait pour autant un véritable changement de sens, c'est le cas par exemple des dérivations impropres de [2].

EXEMPLE 2 :

2.a : ADJECTIF : *Va arribar un noi nou a l'escola*

2.b : NOM : *Els nens no van deixar d'empipar al nou*

Une autre difficulté repose sur le fréquent chevauchement des catégories. Si nous prenons par exemple, le cas des noms et des adjectifs plus de 1 200 mots de notre dictionnaire admettent la fonction de nom [4] et celle d'adjectif [3] :

EXEMPLE 3 : Adjectifs

3.a : *Un pantaló vermell*

3.b : *Un nou escàndol informàtic*

3.c : *Cal utilitzar un producte detergent i desinfectant, i l'altra que contingui aigua per esbandir*

EXEMPLE 4 : Noms

4.a : *Els daltònics poden confondre el vermell amb el negre*

4.b : *L'informàtic va explicar que aquest tipus de software era gratuït*

4.c : *Una quantitat excessiva de detergent fa que la rentadora treballi més*

Dans cette première version de notre dictionnaire nous n'avons pas dédoublé ce type d'entrées. Pour l'instant elles ont toutes une marque spécifique qui nous permettra dans un deuxième temps de les dédoubler, si nécessaire, ou de décrire leur caractère polycatégoriel, notamment lorsque c'est une propriété commune à toute une classe (c'est le cas par exemple, des 'couleurs', des 'nationalités'). Nous

avons, donc, prévu un champ<sup>21</sup> pour rendre compte de la double catégorie du mot. Ainsi dans les exemples ci-dessus, si l'entrée a été étiquetée comme adjectif, dans le champ *cat* il y aura la marque *n* et si, au contraire, elle a été étiquetée comme nom, il y aura la marque *adj*.

Comme le montre Gaston Gross [Gross, 1994], les critères pour établir s'il s'agit d'une catégorie ou d'une autre sont souvent trop variables. Cependant, d'un point de vue morphologique les catégories grammaticales ont un poids déterminant dans la microstructure du dictionnaire. Comme nous le verrons dans le chapitre §11.1 de nombreux phénomènes dérivationnels peuvent être facilement décrits grâce à un étiquetage préalable des catégories. Les catégories grammaticales rassemblent les lexies non pas par rapport à des fonctions syntaxiques mais par rapport à la possibilité de pouvoir se combiner avec un affixe ou un autre. En effet les transformations morphologiques des mots sont étroitement liées aux catégories grammaticales. Ainsi, la répartition des entrées en catégories grammaticales nous permettra de formaliser et de systématiser de nombreux phénomènes dérivationnels.

---

21. Nous avons appelé ce champ : *cat+FLX*.

## 5 La flexion

Un dictionnaire électronique doit contenir toutes les informations morphologiques liées aux entrées. Nous observons dans les dictionnaires traditionnels<sup>22</sup> que ces informations ne sont données que pour certaines entrées dont le choix est extrêmement arbitraire. Un dictionnaire électronique ne peut se permettre un tel procédé puisque l'ordinateur ne peut déduire ni deviner quelles sont les formes flexionnelles propres à chaque entrée. D'une perspective de génération de texte (synthèse) ou d'une perspective de reconnaissance (analyse), il est indispensable que le dictionnaire électronique dispose des informations nécessaires pour que le système puisse générer et/ou reconnaître toutes les formes d'un mot et l'associer avec l'entrée du dictionnaire qui contient toutes les données.

À notre avis, les automates à états finis constituent la manière la plus adéquate pour traiter les informations flexionnelles dans notre dictionnaire. Les automates nous ont permis de générer automatiquement toutes les formes fléchies correspondant aux entrées de notre dictionnaire sans avoir besoin de recourir à un langage de programmation spécifique. Pour ce faire, nous avons travaillé avec le logiciel Intex ainsi qu'avec sa nouvelle version NooJ. Par le biais de l'éditeur de graphes incorporé à ces logiciels, nous avons réalisé les modèles de flexion qui décrivent les différentes opérations à effectuer pour générer les formes fléchies. Ainsi, chaque entrée du dictionnaire est associée par un code au modèle de flexion qui lui correspond. Nous disposons d'une collection de 193 graphes avec NooJ et de 233 graphes (cf. tableau. 5.1).

La différence du nombre de graphes entre une version et une autre est due aux différents opérateurs intégrés dans l'éditeur de graphes. NooJ, à la différence d'Intex, dispose d'un opérateur qui peut modifier le type d'accent des mots sans qu'il y ait des répercussions sur la voyelle. Pour la description formelle de la flexion d'une langue à accent libre comme le catalan où des modifications diverses

---

22. Désormais nous utiliserons le terme *dictionnaires traditionnels* pour faire référence aux dictionnaires papier et aux dictionnaires sur support électronique directement issus de dictionnaires conçus sur format papier.



<b>INTEX</b>	<b>NOOJ</b>
39 graphes pour les substantifs	29 graphes pour les substantifs
64 graphes pour les adjectifs	50 graphes pour les adjectifs
131 graphes pour les verbes	114 graphes pour les verbes

TAB. 5.1: Nombre de graphes de flexion avec Intex et avec NooJ

peuvent se produire au niveau de l'accent telles que l'apparition, la disparition ou le changement de type d'accent, cet opérateur est d'une grande utilité, car un seul modèle de flexion est valable pour les cinq voyelles. Par exemple :

EXEMPLE 5 :

- 5.a : *cabàs/cabassos*
- 5.b : *accés/accessos*
- 5.c : *anís/anissos*
- 5.d : *arròs/arrossos*
- 5.e : *banús/banussos*

Avec NooJ ces cinq mots correspondent à un seul modèle (cf. Fig. 5.1) tandis qu'avec Intex on a dû créer cinq graphes différents, un pour chaque voyelle (cf. Fig. 5.2).

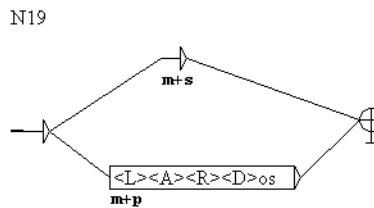
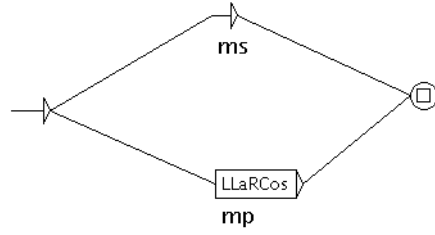
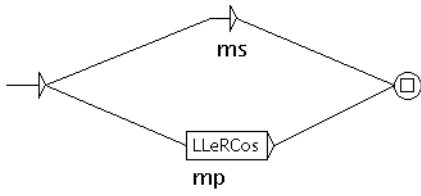


FIG. 5.1: Graphe avec le logiciel NooJ

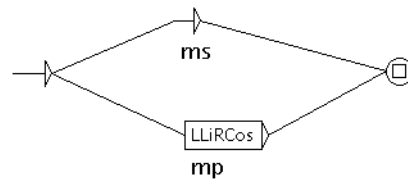
Lorsque nous compilons notre dictionnaire électronique du catalan avec NooJ ou avec Intex, à partir de 39 207 entrées le programme de flexion génère un total de 524 493 entrées. Ces chiffres coïncident en ordre de grandeur aussi bien avec le dictionnaire d'espagnol [Blanco, 2000] qu'avec les études réalisées par IBM sur la langue espagnole, qui montrent qu'un dictionnaire de 35 000 mots simples donnera plus de 400 000 mots fléchis [Bouillon, 1998].



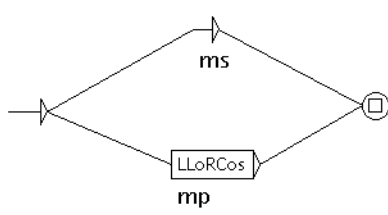
(a) cabàs



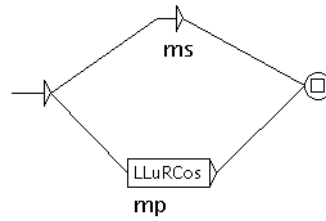
(b) accés



(c) anís



(d) arròs



(e) banús

FIG. 5.2: *Graphes avec le logiciel Intex*

## 5.1 Les substantifs

Le système de codage adopté pour les noms nous donne d'emblée diverses informations : la partie du discours, le genre et bien évidemment le type de flexion. Tous les codes de flexion des substantifs commencent par N (nom), puis un code compris entre 01 et 49 pour les noms masculins et un code compris entre 50 et 99 pour les noms féminins.

**Le Nombre** En ce qui concerne le nombre, il faut dire que le champ morphologique de notre dictionnaire fait référence exclusivement à la formation des formes fléchies. Nous considérons que les valeurs apportées par les marques de flexion relèvent en partie de la sémantique. Nous allons essayer, par la suite, de mettre cela en évidence par des exemples et de montrer par ailleurs les particularités que nous avons rencontrées lors de la description flexionnelle :

EXEMPLE 6 : *rentaplats*

EXEMPLE 7 : *llapis*

EXEMPLE 8 : *tisora/tisores*

EXEMPLE 9 : *pantaló/pantalons*

EXEMPLE 10 : *estovalles*

EXEMPLE 11 : *sal*

EXEMPLE 12 : *fruita*

EXEMPLE 13 : *pare*

EXEMPLE 14 : *amistat*

EXEMPLE 15 : *temps*

EXEMPLE 16 : *dau*

Les deux premiers exemples, [6] et [7], n'ont pas de marque explicite pour indiquer les valeurs de 'pluriel' ou de 'singulier'. Ils peuvent, néanmoins, les exprimer à l'aide des autres éléments de la phrase, en particulier grâce aux déterminants. L'absence de marques formelles, n'implique pas l'absence d'accord en nombre comme on peut le voir dans les exemples ci-dessous.

EXEMPLE 17 :

17.a : *Tinc un rentaplats*

17.b : *Els rentaplats d'aquesta botiga són molt cars*

EXEMPLE 18 :

18.a : *Li vaig deixar el llapis*

18.b : *S'ha comprat uns llapis de colors*

Le graphe de flexion, pour des substantifs comme [6] et [7], aura deux sorties avec la même forme, l'une étiquetée par (genre + singulier), l'autre par (genre + pluriel).

Les exemples [8] et [9] admettent les deux traits, 'singulier' et 'pluriel' et ils ont une marque formelle pour indiquer le 'pluriel'. Par contre dans [8] nous trouvons deux lexèmes différents, la forme qui est au singulier indique 'la lame des ciseaux' et la forme au pluriel indique 'l'instrument de deux lames qui sert à couper...' ; notre dictionnaire compte donc deux entrées différentes. L'exemple [8] ressemble à [6] dans le sens où une seule forme peut faire référence au 'singulier' et au 'pluriel', mais il se distingue de [6] parce que [8] ne peut pas exprimer morphologiquement la notion de 'singulier', son paradigme n'est pas complet :

EXEMPLE 19 :

19.a : *Tens (unes + les) tisores al calaix.*

19.b : *Tens \*(una + la) tisores*

Cette défektivité en [8] est assez systématique pour les noms dénombrables dont le référent est muni d'au moins deux parties ou deux éléments similaires (pinces, ulleres, tenalles...).

Parmi ces noms, l'exemple [9] a deux formes différentes : pantalons qui coïncide avec les propriétés de [8] et pantaló. Puisque ces deux signifiants ont exactement le même signifié, nous avons tendance à croire que la forme défective serait une variante orthographique de la deuxième qui a suivi le même parcours que les noms (pinces, ulleres, calces...) mais sans se distinguer sémantiquement de pantaló.

L'exemple [10] comme dans les exemples précédents, n'accepte qu'une forme, bien que le référent et le signifié soit unique, on doit l'utiliser au pluriel, ce sont des *pluralia tantum*. Nous pouvons retrouver ce même phénomène avec le singulier *singularia tantum* mais ils sont beaucoup plus rares. Pour ces cas là, nous considérons que la forme au singulier ou au pluriel n'existe pas et le graphe qui correspond à ce modèle de flexion a uniquement un nœud ([20]).

EXEMPLE 20 :

20.a : *Les estovalles estan netes*

20.b : *\*L'estovalla / l'estovalles està neta*

Les exemples [11] et [12], lorsqu'on les fléchit en nombre, ont la même marque formelle que les exemples précédents, mais celle-ci attribue non pas une valeur de 'pluriel' mais plutôt de 'variété de' ou 'classes de' :

EXEMPLE 21 :

21.a : *les fruites* = *les différents types de fruits*

21.b : *les sals* = *les différents types de sels*

Cette variante de sens affecte plutôt les substantifs non dénombrables ou continus et, normalement, elle est partagée par tous les constituants de la classe d'objets à laquelle ils appartiennent.

L'exemple [13] présente une autre variété de sens lorsqu'on lui attribue la marque flexionnelle de nombre :

EXEMPLE 22 :

22.a : *El pare del Nil*

22.b : *Els pares del Nil*

Nous constatons que lorsque *pare* est au pluriel, cela ne veut pas dire uniquement 'plus d'un homme du genre masculin'. En réalité la marque de pluriel apporte aussi une valeur de 'parité', c'est à dire 'la/les mère(s) et le/les père(s) de'. Nous retrouvons ce même phénomène avec d'autres lexèmes ayant le trait humain comme par exemple (*els reis*, *els avis*...).

Il existe également des cas où l'adjonction de la marque de flexion nous permet de mieux distinguer l'existence d'unités lexicales qui diffèrent par le trait syntactico-sémantique. Nous avons déjà vu un exemple ([8]), mais il s'agit d'un phénomène régulier pour les noms dont l'entité est composée de deux éléments symétriques. Dans les exemples qui suivent seulement quelques uns répondent à une certaine systématité. La marque de flexion peut indiquer les sens de 'pluriel' comme dans [14] ou pas comme dans [16] :

EXEMPLE 23 :

23.a : *l'amistat / les amistats*

23.b : *el temps / els temps*

23.c : *el dau / els daus*

Dans [14] et [16] la marque de flexion en nombre montre l'existence d'un autre lexème avec le trait syntactico-sémantique *Humain* au lieu d'*Abstrait* :

EXEMPLE 24 :

24.a : *Hi va anar acompanyat de les seves amistats*

24.b : *L'amistat és un dels sentiments més preats*

Dans [15] le singulier peut aussi bien indiquer une notion temporelle qu'une notion climatologique. Par contre lorsqu'il porte la marque de pluriel la notion

climatologique disparaît :

EXEMPLE 25 :

25.a : *Fa un temps plujós*

25.b : ??*Fan uns temps plujosos*

Dans [16] nous avons pu observer une certaine régularité. Les noms de <jeux> ont une tendance à être au pluriel tout en reprenant le même signifiant qui désigne l'objet utilisé dans le jeu, en passant ainsi d'un Concret à un Abstrait :

EXEMPLE 26 :

26.a : *Li agrada jugar a (cartes + daus)*

26.b : *Va perdre una/un (carta + dau)*

Nous avons montré, par ces exemples, que la marque de flexion dépasse souvent le champ purement morphologique et peut avoir des implications sémantiques très marquées. Nous avons vu par ailleurs qu'une même marque de flexion peut apporter différentes valeurs comme celles de 'pluralité', 'classes de', 'parité', etc. I. Mel'čuk [Mel'čuk, 1993] appelle ces marques de flexion des grammèmes, ou, ce qui est la même chose, des significations flexionnelles. Les grammèmes font partie, avec les dérivatèmes, des significations grammaticales d'une langue. Ils s'organisent en catégories de sorte que les grammèmes d'une même catégorie sont en distribution complémentaire (par exemple 'singulier' et 'pluriel' appartiennent à la même catégorie et s'excluent mutuellement). Mais il se peut qu'un même grammème soit polysémique (exemples [11] et [13] entre autres).

Il n'existe pas encore de méthodologie spécifique pour la description des significations grammaticales (bien qu'elles pourraient être décrites par le biais des fonctions lexicales) et, étant donné qu'elles ne sont pas universelles, mais propres à chaque langue, nous ne disposons même pas d'une liste complète des significations grammaticales qui existent en catalan. Nous croyons qu'une telle entreprise mérite une étude spécifique qui échappe au présent travail.

Comme nous l'avons dit précédemment, le champ morphologique de notre dictionnaire ne prétend point décrire les significations grammaticales, mais nous partageons l'idée de I. Mel'čuk [Mel'čuk, 1993] qu'il est toutefois indispensable de donner pour chaque entrée du dictionnaire les formes fléchies qui lui correspondent. L'absence ou présence de la marque du pluriel conditionne, indépendamment du sens véhiculé par cette marque et de façon systémique, tout le schéma de concordances (les accords avec le verbe, les adjectifs, les pronoms...). Par conséquent, toutes les formes de notre dictionnaire ont été étiquetées en tant que *singulier* ou *pluriel*, puisque malgré le caractère déficitaire de ce type d'étiquettes, cet aspect est extrêmement utile du point de vue grammatical.

**Le Genre** En ce qui concerne le genre, la distribution des 31 graphes de NooJ<sup>23</sup> est la suivante<sup>24</sup> :

10 graphes pour le féminin

19 graphes pour le masculin

La flexion du genre ne s'applique qu'à une partie des substantifs, parmi lesquels, nous distinguons deux grandes classes, les noms animés (i.e. noms dont le trait syntactico-sémantique est uniquement *Animal* ou *Humain*) et les noms non-animés. Contrairement aux noms non-animés la différence de genre des noms animés est étroitement liée à la différence de sexe. Il existe une corrélation assez systématique entre l'information sémantique ('de sexe masculin' / 'de sexe féminin') et la flexion du genre des noms animés, absente ou très peu systématique parmi les noms non-animés. C'est pourquoi nous parlerons de « genre naturel » dans le cas des noms animés et de « genre grammatical » dans le cas des noms non-animés.

EXEMPLE 27 : Noms animés : le genre naturel

27.a : *cosí/cosina*

27.b : *porc/truja*

EXEMPLE 28 : Noms non-animés : le genre grammatical

28.a : *el rumor/la rumor*

28.b : *el crin/la crin*

EXEMPLE 29 : Noms non-animés : l'homonymie

29.a : *el clau/la clau*

29.b : *el còlera/la còlera*

29.c : *el llum/la llum*

Dans notre dictionnaire, la formation du genre naturel est considérée comme un procès de dérivation et les mots qui en résultent des unités lexicales à part entière (cf. [Mel'čuk, 2000]). Les affixes servant à exprimer le genre naturel présentent certaines caractéristiques qui les éloignent des significations flexionnelles telles que celle du nombre dans les noms, ou celle du genre dans les adjectifs. L'expression du genre naturel dans les noms, à la différence du genre syntaxique des adjectifs, a une combinatoire très irrégulière. D'un côté, elle n'affecte que les noms qui appartiennent à un groupe sémantique spécifique

23. Dorénavant, pour simplifier la lecture, nous parlerons uniquement des graphes réalisés avec NooJ, au lieu de donner les renseignements à la fois pour NooJ et pour Intex.

24. Tous les modèles de flexion sont donnés dans les annexes A et B.

(*Hum* et *Anl*) et parmi eux nous pouvons même trouver des exceptions comme le montrent les exemples de [30] où la spécification du sexe est absente.

EXEMPLE 30 : Noms animés : Sans spécification du sexe

30.a : *una persona; un bebè; una criatura...*

L'expression formelle du genre naturel des noms est beaucoup plus variable que celle du genre de l'adjectif.

EXEMPLE 31 : Expression formelle régulière

31.a : *avi/àvia*

31.b : *germà/germana*

31.c : *secretari/secretària*

31.d : *llop/lloba*

EXEMPLE 32 : Expression formelle irrégulière

32.a : *gendre/nora*

32.b : *marit/muller*

32.c : *mascle/femella*

32.d : *ase/somera*

Étant donné que les formes dérivées féminines sont des unités lexicales, elles constituent une entrée de notre dictionnaire. Bien que « *les formes féminines conservent toutes les propriétés lexicographiques des formes masculines de départ (le régime, les collocations...)* » [Mel'čuk, 2000] nous croyons qu'il existe certaines particularités qui sont mieux décrites si nous traitons le masculin et le féminin de façon disjointe au lieu de décrire les deux lexèmes par un même article de dictionnaire.

Comme nous avons pu le remarquer dans les exemples [32] certaines formes féminines ne dérivent pas de la forme du masculin et elles ont donc une racine clairement différenciée. Pour traiter ces formes sous une seule entrée nous serions obligée de créer des graphes exclusifs à chaque forme et augmenter considérablement la collection de graphes du dictionnaire. Nous croyons plus convenant de continuer la tradition lexicographique où chacune de ces formes constitue une entrée du dictionnaire.

Il existe, par ailleurs, des formes masculines qui dérivent de la forme féminine :

EXEMPLE 33 : *bruixa/bruixot*

EXEMPLE 34 : *dida/didot*



Mais en ce qui concerne ces formes dérivées, il existe des particularités combinatoires qui diffèrent entre le masculin et le féminin :

EXEMPLE 35 :

35.a : *És una mala bruixa*<sup>25</sup>

35.b : *És un mal bruixot*

EXEMPLE 36 :

36.a : *donar un infant a dida*

36.b : *??donar un infant a didot*

Dans l'exemple [35], la présence de cet adjectif avec la forme féminine rend la phrase ambiguë. Ainsi *mala bruixa* peut vouloir dire 'sorcière méchante' ou bien 'sorcière qui ne maîtrise pas bien les techniques de la sorcellerie'. Par contre, le même adjectif avec la forme au masculin a une seule interprétation 'sorcier qui ne maîtrise pas bien les techniques de la sorcellerie'.

Dans l'exemple [36], nous trouvons également des propriétés de la forme féminine qui ne sont pas partagées par le nom masculin. Il est vrai que nous pouvons toujours considérer *donar a dida* comme un quasi-phrasème mais une grande différence sémantique persiste car *dida* inclut le sens 'qui allaite', absent dans *didot*, comme le prouve l'existence d'une collocation telle que *dida seca*.

Nous avons trouvé d'autres différences dans la combinatoire entre le masculin et le féminin, indépendamment du fait que le féminin dérive du masculin (exemple [37]) ou que les racines des deux genres soit différente (exemple [38]) :

EXEMPLE 37 : *gall/gallina*

EXEMPLE 38 : *bou/vaca*

Lorsque nous décrivons les semi-phrasèmes propres aux exemples [37] et [38] nous nous rendons compte qu'ils ne sont pas identiques. Pour décrire les semi-phrasèmes nous utilisons le formalisme des fonctions lexicales déjà présenté dans §2.3. Dans les exemples ci-dessous nous présentons le verbe ayant le sens 'produire le son typique', c'est la fonction lexicale **Son** :

EXEMPLE 39 :

39.a : **Son**(*gall*) = *cantar*

39.b : **Son**(*gallina*) = *cacarejar*; \**cantar*

39.c : **Son**(*bou*) = *bramar*; *mugir*

39.d : **Son**(*vaca*) = *mugir*; \**bramar*

De ces exemples s'ensuit que la description des paires masculin-féminin peut avoir des particularités qui sont mieux décrites en séparant les entrées

---

25. Nous ne parlons ici que du sens de 'sorcière'.

dans le dictionnaire. Nous avons trouvé d'autres exemples où le féminin des noms animés peut se combiner avec des verbes spécifiques qui ne peuvent apparaître avec le masculin, par exemple les verbes ayant le sens 'accoucher'. Cette fonction lexicale n'est pas précisée dans la liste des FL standards proposée par I. Mel'čuk, nous croyons pourtant nécessaire que ces verbes soient précisés dans la description des lexies, puisqu'il existe différentes valeurs en fonction du mot-clé. Pour décrire ces verbes nous avons postulé la fonction lexicale *Par(x)*<sup>26</sup>. Voici quelques exemples :

EXEMPLE 40 :

- 40.a : **Par**(*gallina*) = *posar ous*
- 40.b : **Par**(*gata*) = *gatar*
- 40.c : **Par**(*dona*) = *donar a llum*
- 40.d : **Par**(*truja*) = *garrinar; godallar*

Mises à part les différences formelles et de combinatoire que nous venons de voir, il existe aussi des variations diasystématiques ou des variations graphiques de la forme féminine et qui, par conséquent, sont mieux décrites si nous disposons d'une entrée distincte pour les noms animés féminins.

EXEMPLE 41 : Variantes diasystématiques

- 41.a : *una eua / una euga*
- 41.b : *una dea / una deessa*
- 41.c : *una poeta / una poetessa / una poetissa*

EXEMPLE 42 : Variantes graphiques

- 42.a : *una estudiant / una estudianta*
- 42.b : *una advocat / una advocada*

Par rapport au nombre dont nous parlions dans le paragraphe précédent, il existe aussi des différences entre le masculin et le féminin. En catalan, le masculin des noms animés peut avoir une valeur 'générique' ou de 'parité' inexistante dans le féminin. Dans le DEC ces différences sont précisées dans la définition lexicographique. Lorsque la mise au pluriel invalide le sens 'personne de sexe masculin' pour devenir 'personnes du sexe masculin et féminin', dans la définition, la valeur de 'sexe masculin' est mise entre parenthèses pour indiquer qu'il s'agit d'une composante faible qui peut se perdre au pluriel :

EXEMPLE 43 : *ROI* : '*Personne (de sexe masculin)<sub>PL</sub> qui...*'

Etant donné que notre dictionnaire n'a pas de définition lexicographique, nous croyons que ce type d'informations doivent être incluses dans un champ

---

26. Tout en respectant l'origine latin des noms des fonctions lexicales proposés par Mel'čuk, nous proposons le nom *Par(x)* qui provient du latin *parere*.

spécifique. Ce champ incluant les informations de la catégorie flexionnelle diffèrent entre la forme du masculin et celle du féminin. Il existe même des cas où c'est le féminin qui s'attribue la valeur 'générique' <sup>27</sup>.

EXEMPLE 44 : Valeur générique dans l'une des paires

44.a : *les gallines: un(s) gall(s) i un(a)(es) gallin(a)(es)*

44.b : *les cabres: un(s) cabró(ns) i un(a)(es) cabr(a)(es)*

44.c : *els professors: un(s) professor(s) i un(a)(es) professor(a)(es)*

Pour résoudre les différences entre la forme masculine et la forme féminine des noms animés, nous avons décidé de dédoubler les entrées en fonction du genre naturel en indiquant dans le champ *genre* le masculin ou le féminin correspondant. Bien que le dédoublement de ces entrées comporte un effort supplémentaire pour le lexicographe, ce procédé nous permet, d'une part, d'assurer l'homogénéité dans la description des unités lexicales ainsi que la cohérence dans l'ensemble du dictionnaire et, d'autre part, de bénéficier de l'autonomie nécessaire pour résoudre les éventuels phénomènes linguistiques qui nous ont échappé dans le présent travail.

Parmi les formes nominales féminines nous pouvons remarquer l'existence d'une polysémie assez régulière, qui concerne surtout les noms 'des humains ayant un titre de noblesse' :

EXEMPLE 45 : La polysémie des formes féminines

45.a : *Duquessa : 'femme ayant le titre de noblesse'*

45.b : *Duquessa : 'épouse d'un duc'*

45.c : *Reina : 'femme ayant le titre de noblesse'*

45.d : *Reina : 'épouse d'un roi'*

Ces formes ont été systématiquement dédoublées dans notre dictionnaire pour rendre compte de l'existence de deux unités lexicales différentes.

En ce qui concerne le genre grammatical (c.f. exemples [28]), l'alternance de genre se produit toujours en préservant la même forme du mot. Dans ces cas, le changement de genre n'implique aucune différence sémantique entre les deux. Cependant, il se peut qu'il existe des différences au niveau de la diasystématique : l'un des deux genres est plus fréquent (diafréquence), ils sont utilisés dans des régions différentes (diatopique), etc.

---

27. Nous n'avons trouvé cette particularité que dans les substantifs ayant le trait *Anl*.

## 5.2 Les adjectifs

Pour fléchir l'ensemble des adjectifs de notre dictionnaire nous avons dû créer 53 graphes au total. À partir de 6 310 lemmes adjectifs, ces transducteurs génèrent 24 897 formes fléchies.

Les nœuds des graphes sont variables. Cinq modèles de flexion possèdent cinq nœuds, puisqu'ils incluent une variante dans l'une des formes. Ces graphes donnent en sortie 5 formes différentes :

EXEMPLE 46 : Modèles de graphes à 5 nœuds

46.a : *basc/basc/bascos/basca/basques*

46.b : *lleig/leigs/lletjos/lletja/lletges*

Bien qu'il existe des formes adjectivales qui, au pluriel ou au féminin, n'ont pas de marque explicite, leur modèle de flexion inclut les quatre nœuds (masculin/singulier ; masculin/pluriel ; féminin/singulier ; féminin/pluriel) de sorte que le transducteur puisse générer les quatre formes avec leurs étiquettes respectives :

EXEMPLE 47 : Modèles de graphes avec absence de marques dans l'une des formes

47.a : *eficaç/eficaços/eficaces*

47.b : *ample/amples/ampla*

47.c : *amable/amables*

Ces modèles de flexion génèrent quatre formes fléchies étiquetées de la façon suivante :

eficaç,eficaç,A+FLX=A39+m+s  
 eficaç,eficaç,A+FLX=A39+f+s  
 eficaces,eficaç,A+FLX=A39+f+p  
 eficaços,eficaç,A+FLX=A39+m+p  
 ample,ample,A+FLX=A42+m+s  
 amplex,ample,A+FLX=A42+f+p  
 ampla,ample,A+FLX=A42+f+s  
 amplex,ample,A+FLX=A42+m+p  
 amable,amable,A+FLX=A45+m+s  
 amable,amable,A+FLX=A45+f+s  
 amables,amable,A+FLX=A45+f+p  
 amables,amable,A+FLX=A45+m+p

FIG. 5.3: Extrait du dictionnaire fléchi

### 5.3 Les verbes

La flexion des verbes, contrairement à celle des noms et des adjectifs, a été réalisée avec un fichier texte<sup>28</sup>. Ce fichier a les mêmes propriétés qu'un graphe, mais tous les états sont décrits linéairement. Lorsque nous devons réaliser un graphe ayant plus de dix nœuds, l'ordre linéaire simplifie la correction et la modification des données. Les opérateurs utilisés dans ce fichier sont les mêmes que pour les graphes. Les nœuds, au lieu d'être représentés par des boîtes sont séparés par 'espace' + 'espace' et les propriétés à l'intérieur d'un même nœud sont séparées par + (sans espaces). Dans ce fichier nous avons utilisé des variables qui renvoient aux terminaisons les plus régulières comme le montrent les figures 5.4 et 5.5.

#### VARIABLES COMMUNES À LA PLUPART DES VERBES

Imperfet = (ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1  
+ íeu/IMP+p+2 + ien/IMP+p+3)  
Futur = (ré/F+s+1 + ràs/F+s+2 + rà/F+s+3 + rem/F+p+1  
+ reu/F+p+2 + ran/F+p+3)  
Condicional = (ria/C+s+1 + ries/C+s+2 + ria/C+s+3 + ríem/C+p+1  
+ ríeu/C+p+2 + rien/C+p+3)

FIG. 5.4: Variables communes de la grammaire flexionnelle

L'utilisation de ces variables nous a permis de minimiser le nombre d'erreurs et de diminuer le volume de la description. Ainsi, dans la figure 5.6 nous trouvons un modèle de flexion qui correspond à la conjugaison de certains verbes réguliers de la première conjugaison comme *cantar*.

Pour fléchir les 8 535 verbes de notre dictionnaire nous avons dû créer 114 graphes au total ayant en moyenne 53 nœuds. Toutes les formes verbales ont été étiquetées par les codes de la figure 5.7. Ce système de codage est uniquement indicatif. Les verbes constituent une mine de significations grammaticales. Une seule terminaison verbale peut regrouper plusieurs sens appartenant à des catégories flexionnelles différentes. Bien que, comme nous l'avons déjà dit, l'étude des significations grammaticales dépasse largement le travail en cours, le système de codage s'avère nécessaire. Ainsi, à partir de 8 535 verbes nous avons généré 452 707 formes conjuguées et toutes ont été respectivement étiquetées.

28. Uniquement avec NooJ; avec Intex, nous avons utilisé l'éditeur de graphes.

VARIABLES SPÉCIFIQUES
ePS = (í/PS+s+1 + eres/PS+s+2 + é/PS+s+3 +...)
eSPR = (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 +...)
eSIMP = (és/SIMP+s+1 + essis/SIMP+s+2 + és/SIMP+s+3 +...)
iPS = (í/PS+s+1 + ires/PS+s+2 + í/PS+s+3 +...)
SPR = (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 +...)
iSIMP = (ís/SIMP+s+1 + issis/SIMP+s+2 + ís/SIMP+s+3 +...)
aIMP = (ava/IMP+s+1 + aves/IMP+s+2 + ava/IMP+s+3 +...)
aPS = (í/PS+s+1 + ares/PS+s+2 + à/PS+s+3 +...)

FIG. 5.5: Variables spécifiques de la grammaire flexionnelle

MODÈLE DE FLEXION POUR LES VERBES COMME <i>cantar</i>
V105=
<E>/INF
+ <B>nt/G
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p)
+ <B2> (o/PR+s+1 + es/PR+s+2 + a/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2> :aIMP
+ <B2> :aPS
+ <B> :Futur
+ <B> :Condicional
+ <B2> :eSPR
+ <B2> :eSIMP
+ <B2>(a/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3)

FIG. 5.6: Extrait de la grammaire flexionnelle

G = gérondif  
 PP = participe  
 PR = présent  
 IMP = imparfait  
 PS = passé simple  
 F = futur  
 C = conditionnel  
 SPR = présent du subjonctif  
 SIMP = imparfait du subjonctif  
 I = Impératif

FIG. 5.7: *Les codes de flexion utilisés pour les verbes*

Dans les modèles de flexion verbale, nous avons décrit uniquement les temps simples, nous considérons que les temps composés doivent être traités par un module syntaxique puisque d'une part il peut y avoir des insertions entre les deux formes qui en empêcheraient la reconnaissance ou la une bonne génération des textes et, d'autre part, parce qu'il faut considérer également la composition avec des verbes modaux :

EXEMPLE 48 : Insertions

48.a : *Segur que ho ets; si no, no hauries **pas** vingut aquí.*

48.b : *Els reis no havien **mai** tingut tant de poder com els reis de Castella.*

48.c : *Aquell dia encara no havia **ni** menjat.*

EXEMPLE 49 : Verbes phasiques et modaux

49.a : *Els cors es **començàven a** accelerar*

49.b : *No volia **anar a** dormir*

De nombreux verbes en catalan ont des variantes graphiques dans leur conjugaison :

EXEMPLE 50 : CRÉIXER

50.a : *creixés / cresqués; creixessis / cresquessis; creixés / cresqués*

EXEMPLE 51 : ESCRIURE

51.a : *escriví / escriguí; escrivires / escrigueres; escriví / escriguí*

Lorsque ces variantes sont présentes uniquement dans un temps de la conjugaison nous les avons inclues dans un même modèle. Cependant, il existe des verbes qui acceptent deux modèles de flexion différents, i.e. des variantes

graphiques dans tous les temps. Pour ces cas, nous avons inclus les deux modèles dans la description du verbe :

EXEMPLE 52 : *júnyer*, V108+V026 ;

Il existe des cas où les variantes graphiques sont reliées à des variantes syntactico-sémantiques des verbes, par conséquent nous avons dédoublé ce type d'entrées :

EXEMPLE 53 : Haver

53.a :	<i>He anat al poble</i>	<i>*Haig anat al poble</i>
53.b :	<i>Haig d'anar a comprar</i>	<i>He d'anar a comprar</i>





## 6 Les traits syntactico-sémantiques

Les deux chapitres suivants (§6 et §7) font référence exclusivement au codage des substantifs. Dans ce chapitre nous expliquerons les critères adoptés lors du classement en traits syntactico-sémantiques.

Selon Z.S. Harris, un prédicat nécessite de ses arguments pour se compléter. Cependant, le choix qu'il opère parmi les arguments n'est pas toujours libre. Le prédicat impose souvent<sup>29</sup> des contraintes sur le caractère sémantique de ses arguments.

EXEMPLE 54 : Restriction sémantique

54.a : *Nil conduit une voiture*

54.b : *conduire(N<sub>0</sub>:Hum; N<sub>1</sub>:Conc<moyen de transport>)*

Pour pouvoir associer à chaque prédicat les arguments dont il a besoin tout en respectant ses exigences sémantiques, nous avons codé tous les substantifs de notre dictionnaire en traits syntactico-sémantiques.

Abst	Abstrait
Abst_Lng	Langue
Anl	Animal
Conc	Concret
Hum	Humain
Loc	Locatif
Veg	Végétal
ConcColl	Concret Collectif
HumColl	Humain Collectif
VegColl	Végétal Collectif

---

29. Il existe des prédicats qui n'imposent aucune restriction sémantique sur leurs arguments, comme par exemple *agradar* qui n'impose aucune restriction sur son deuxième argument : *agradar(N<sub>0</sub>:Hum;N<sub>1</sub> )*.

Ce système de codage sert donc principalement à la description des prédicats. Dans de nombreux cas, seuls les traits syntactico-sémantiques nous permettront de désambiguïser les différents emplois d'un mot :

EXEMPLE 55 :

55.a : *Per tot arreu esclaten les poncelles*

55.b : *Li ha esclatat el globus a les mans*

- esclatar/V/V105/N<sub>0</sub>:Veg/Fr:éclore
- esclatar/V/V105/N<sub>0</sub>:Conc/Fr:éclater

L'application de ces traits se base surtout sur l'intuition du lexicographe, ainsi que sur la définition sémantique des noms. Cependant, il existe parfois des critères formels qui aident à l'attribution des traits sémantiques. Nous verrons par la suite quels sont les critères que nous avons adoptés lors du codage et les propriétés générales de chaque trait.

## 6.1 Abstrait vs Concret

Traditionnellement les noms abstraits ont été opposés aux concrets dans la mesure où les concrets sont les noms qui désignent des 'objets ou des substances' et les noms abstraits tous les autres, i.e. des 'objets non matériels'. Selon cette définition l'opposition entre abstrait et concret repose essentiellement sur la réalité désignée par la langue. Cependant pour les chercheurs du LLI la distinction entre un nom abstrait et un nom concret ne repose pas uniquement sur l'objet désigné par le mot. L'opposition entre un concret et un abstrait est associée aux notions de prédicat et d'argument élémentaire ; ainsi, les noms concrets « *ne peuvent fonctionner dans la phrase que comme des arguments* » [Mathieu-Colas, 1994] alors que les abstraits ont tous une valeur prédicative. Ils considèrent donc que les traits syntactico-sémantiques servent, d'une part, à la description/désambiguïisation des prédicats et, d'autre part, ils contribuent à la distinction des arguments élémentaires ou entités des prédicats nominaux. Dans le cadre des recherches du LLI la notion de prédicat coïncide avec la notion harrissienne d'opérateur mais avec quelques précisions. Selon Z.S. Harris « *les mots qui n'ont aucune contrainte sur leurs arguments sont des arguments élémentaires. Les autres sont des opérateurs, ils appartiennent à divers ensembles* » [Harris, 1976, p. 14]. G. Gross rajoute à cette définition que seuls les prédicats peuvent apparaître dans des constructions à verbe support et que « *un prédicat est caractérisé par son inscription dans le temps (. . .). Les prédicats ont donc une double définition : ils sont définis par leur domaine d'arguments et par leur actualisation* » [Gross, 1996]. C'est, à partir de leur « inscription dans

le temps » qu'est réalisé un deuxième classement des prédicats en trois grandes classes : états, actions et événements.

Nous constatons cependant qu'il existe des divergences entre les chercheurs à ce niveau-là. Si pour G. Gross et son équipe, la notion d'abstrait et de prédicat sont indissociables, pour I. Mel'čuk de ces deux notions découlent deux analyses dissemblables. Est prédicat, selon Mel'čuk, tout mot ayant des actants sémantiques et « *les seules lexies d'une langue quelconque dénuées d'ASém sont les noms propres (personnes, lieux, ethnies...), les noms de certaines substances (liquides, solides, poudres...), de certains phénomènes naturels (éléments géographiques...), et de certaines espèces naturelles (animaux, oiseaux, fleurs...), etc. Il est souvent impossible d'assigner aux sens de tels types de lexies une situation qui présuppose des participants clairement identifiés, de telles situations étant indispensables pour qu'il y ait actants sémantiques* » [Mel'čuk et al., 1995, p. 77]

Selon I. Mel'čuk les noms peuvent donc avoir des arguments (appelés selon sa terminologie des actants sémantiques) indépendamment de leur désignation référentielle.

EXEMPLE 56 : Noms prédictifs concrets<sup>30</sup>

56.a : TÊTE : tête de X= X a une tête

56.b : NEZ : nez de X

56.c : CUILLÈRE : cuillère avec laquelle X mange Y

Même si nous admettions que c'est la notion de prédicat ce qui diffère entre ces deux chercheurs, nous croyons que si l'on dissocie la notion d'abstrait de celle de prédictivité, certaines lexies peuvent être décrites dans un dictionnaire électronique de façon plus opérante. Considérons les lexies suivantes :

EXEMPLE 57 :

57.a : una fotocòpia

57.b : una fotografia

57.c : un biberó

57.d : una almoïna

57.e : els diners

57.f : la sang

57.g : un rebut

57.h : una invitació

57.i : una construcció

57.j : un premi

57.k : una radiografia

57.l : el pipi

---

30. Tous ces exemples ont été tirés de [Mel'čuk et al., 1995, p. 77].

Il est évident qu'elles peuvent toutes être considérées en tant que *concret* dans le sens où elles désignent un objet matériel et tangible. Si nous reprenons la définition de prédicat de G. Gross, dès qu'un nom est susceptible d'apparaître dans une construction à verbe support, celui-ci est considéré comme prédicatif et donc comme abstrait (puisque les concrets ne peuvent être que des arguments élémentaires). Pourtant, ces noms concrets peuvent bien apparaître dans une construction à verbe support, ce qui voudrait donc dire que ce sont des prédicats :

EXEMPLE 58 :

- 58.a : *Fer una fotografia*
- 58.b : *Fer una fotocòpia*
- 58.c : *Fer una radiografia*
- 58.d : *Fer un rebut*
- 58.e : *Fer una invitació*
- 58.f : *Fer una construcció*
- 58.g : *Fer una bombolla*
- 58.h : *Posar una bomba*
- 58.i : *Posar una trampa*
- 58.j : *Posar (una) anestèsia*
- 58.k : *Donar almoïna*
- 58.l : *Donar el biberó*
- 58.m : *Donar diners*
- 58.n : *Donar sang*
- 58.o : *Donar un premi*
- 58.p : *Donar (un) verí*

Il est vrai que parmi les exemples de [57] il y en a qui sont limitrophes entre le trait abstrait et concret car lorsqu'ils se combinent avec un verbe support nous pouvons apprécier une certaine abstraction. Il y en a d'autres, par contre, que nous percevons comme concrets même en présence du verbe support. Pourtant, en associant la notion d'abstrait à la prédicativité, nous serions obligés de dédoubler systématiquement toutes ces lexies, l'une en tant que *Abst* et l'autre en tant que *Conc*. Or, dans l'état actuel du dictionnaire (les classes d'objets n'ayant pas été complètement implémentées), nous croyons que le dédoublement nous ferait perdre des informations spécifiques à ce type de noms.

Si nous dédoublons les noms comme ceux de [57] nous nous engageons à les considérer comme deux unités lexicales différentes, l'une *abstrait* et l'autre *concret*. Si nous comparons leur comportement avec des unités lexicales vraiment différentes (homographes) (cf. [60]), dont l'une est *abstrait* et l'autre *concret*, nous nous rendons compte que leur combinatoire n'est pas la même ; nous ne pouvons pas combiner leurs traits dans une même phrase [60] ; soit nous nous référons à celle qui a le trait *abstrait*, soit nous nous référons à celle qui a le trait

*concret*, mais pas aux deux à la fois. Or, dans les exemples de [59] nous pouvons combiner le trait concret avec le prétendu trait abstrait de façon naturelle car en réalité ce que nous sommes en train de combiner c'est une même lexie en tant que prédicat ou en tant que prédicat de deuxième ordre et non pas deux unités lexicales différentes ayant différents traits sémantiques.

EXEMPLE 59 :

- 59.a : *He trencat la **fotografia** que (li) vaig fer*
- 59.b : *Se m'han embrutat les **fotocòpies** que vaig fer l'altre dia*
- 59.c : *Va ficar en un calaix els **diners** que li havien donat*
- 59.d : *Després de donar-li el **biberó** el va ficar al rentaplats*
- 59.e : *Va posar el **premi** que li havien donat sobre la taula*
- 59.f : *Va netejar el **pipi** que havia fet el gat al mig de la sala*

EXEMPLE 60 :

- 60.a : *\*He perdut la **cadena** que vam formar*
- 60.b : *\*Vaig netejar el **billar** que havia jugat*

Enfin, nous considérons que prédicat n'implique pas abstrait car il s'agit de deux propriétés différentes. Il existe certainement un rapport entre les traits et la prédicativité puisque les noms étiquetés comme *Abst* ou *Hum* sont tous des prédicats, mais nous ne pouvons pas baser l'étiquetage des traits sémantiques sur la prédicativité puisqu'il existe aussi des prédicats parmi les concrets. Nous verrons par la suite comment nous avons classé tous ces exemples.

## 6.2 Abstrait

Dans notre dictionnaire nous avons 8 620 noms *abstrait*s dont 5 134 sont féminins et 3 486 sont masculins. Comme nous venons de dire, les noms *abstrait*s se caractérisent parce qu'ils sont tous des prédicats. Sous ce trait nous retrouvons des noms ayant différentes propriétés non exclusives tels que : des noms déadjectivaux ou déverbaux, des noms dénombrables ou indénombrables, des noms ayant des valeurs aspectuelles ou pas, etc.

EXEMPLE 61 : Noms abstraits déverbaux

- 61.a : *el rescat*
- 61.b : *el/la temor*
- 61.c : *l'admiració*

EXEMPLE 62 : Noms abstraits déadjectivaux

- 62.a : *la tristesa*
- 62.b : *la fluidesa*

62.c : *la modèstia*

EXEMPLE 63 : Noms abstraits dénombrables

63.a : *un viatge*

63.b : *un concert*

63.c : *un bombardeig*

EXEMPLE 64 : Noms abstraits indénombrables

64.a : *la bondat*

64.b : *el fútbol*

64.c : *l'admiració*

EXEMPLE 65 : Noms abstraits ayant des valeurs aspectuelles

65.a : *el concert*

65.b : *el partit*

65.c : *la festa*

EXEMPLE 66 : Noms abstraits sans valeurs aspectuelles

66.a : *el cristianisme*

66.b : *el wolof*

66.c : *l'art*

En ce qui concerne notre codage, même s'il repose sur la base de la valeur sémantique du nom, tout au long de l'élaboration du dictionnaire nous avons essayé d'établir quelques critères génériques (et en tant que tels de prendre en compte leurs respectives exceptions) et formels, pour assurer ainsi la cohérence dans la description des données. Ainsi, en reprenant 3 des 6 cas de figure que nous venons d'exposer, nous avons codé en tant qu'abstrait tous les noms déverbaux ou déadjectivaux.

L'une des caractéristiques<sup>31</sup> de ces noms est que la plupart d'entre eux conservent la valeur prédicative du prédicat verbal ou adjectival dont ils sont issus. En effet, ces substantifs héritent la totalité ou une partie des arguments (exemple [67] : *pes* et *cost*) sélectionnés par les prédicats dont ils dérivent et ils maintiennent les mêmes restrictions sur le sémantisme de leurs arguments. Néanmoins, il faut souligner que les valeurs aspectuelles des verbes ou des adjectifs dont ils sont issus ne sont pas systématiquement préservées.

EXEMPLE 67 :

67.a : *El passeig de X per L / X passeja per L*

67.b : *El rescat de A per X / A rescata a X*

---

31. Nous décrivons ici une caractéristique commune à ce type de noms et non pas le critère par lequel nous les avons codés en tant qu'*abstrait*. Pour la discussion sur abstrait vs. prédicat, voir la section §6.1.

- 67.c : *El descobriment de N per X / X va descobrir N*  
 67.d : *L'admiració de X per A / X admira A*  
 67.e : *La temor de X per N / X tem N*  
 67.f : *El pes \*(de 2 kilos) de la caixa \*(de 2 kilos) / La caixa pesa 2 kilos*  
  
 67.g : *El cost \*(de 7 euros) del llibre \*(de 7 euros) / El llibre costa 7 euros*  
 67.h : *La tristesa de X / X està trist*  
 67.i : *La modèstia de X / X és modest*  
 67.j : *La simpatia de X / X és simpàtic*

Cependant, parmi les noms déverbaux nous retrouvons des noms qui se situent à la frontière du concret et de l'abstrait (cf. section §6.1) comme *construcció*, *invitació*, *traducció*... Dans les dictionnaires traditionnels, ces noms sont souvent définis comme *l'action et l'objet qui en résulte*. Voici un extrait du *Diccionari de la Llengua Catalana de l'Institut d'estudis catalans* [Hip, 1993] :

1. **Invitació**: acció d'invitar; l'efecte.| Imprès escrit amb què s'invita.
2. **Construcció**: acció de construir.| Cosa construïda.
3. **Compra**: acció de comprar; allò que hom ha comprat

Cette ambiguïté entre abstrait et concret n'affecte pas uniquement des mots isolés mais tout un groupe de noms qui 'désignent le résultat d'un procès'. D'autre part, comme nous l'avons vu dans les exemples [59], il est possible de combiner les deux emplois dans la même phrase sans que cela produise un zeugma. Nous considérons donc qu'il s'agit d'une polysémie régulière. Puisque ce phénomène est systématique pour tout un ensemble de mots ayant des caractéristiques sémantiques communes, au lieu de dédoubler toute la classe et de perdre ainsi des informations sur leur nature nous avons décidé de les coder tous sous un même trait, en attendant une étude plus approfondie. Dans le chapitre suivant §6.3 nous exposons les propriétés syntactico-distributionnelles de ce type de noms qui les différencient des noms concrets.

Il y a d'autres noms non dérivés qui posent aussi des problèmes lors du codage à cause de leur caractère *concret* ou *abstrait* ambigu. Les noms tels que *sonata*, *sinfonia*, *cant*, *música*, *poesia*... peuvent être perçus comme des noms concrets, comme le montrent N. Flaux et D. Van de Velde [Flaux and Van de Velde, 2000] :

EXEMPLE 68 :

- 68.a : *Llegeixo poesia*  
 68.b : *Escolto música*



Elles appellent ce type de noms des « *idéautés homogènes* » et elles les classent parmi les *concrets*, malgré le caractère *abstrait* qui s'en dégage dans certains contextes :

EXEMPLE 69 :

69.a : *El Nil fa música*

69.b : *El Nil fa poesia*

Selon N. Flaux et D. Van de Velde, l'une des propriétés de ce type de noms est que « le N du genre<sup>32</sup> est un indéénombrable tandis que le N de l'espèce est déénombrable » [Flaux and Van de Velde, 2000, p. 72]. Nonobstant, en ce qui concerne l'indénombrabilité du genre, nous pouvons constater avec les exemples ci-dessous que ce n'est pas une caractéristique propre à toutes les espèces d'une classe (genre) :

EXEMPLE 70 : Genre : música

70.a : *Escolto música*

70.b : *Escolto jazz*

70.c : *Escolto una simfonia*

70.d : *Escolto una sonata*

Une autre propriété qui fait, d'après ces auteurs, que ce type de noms soit plus proche des *concrets* est la possibilité de pouvoir utiliser des noms de parties tels que *morceau*, *bout* *fragment*.

EXEMPLE 71 :

71.a : *Un tros d'òpera / Un tros d'una òpera*

71.b : *Un fragment de música*

71.c : *Un tros d'una simfonia*

71.d : *Dues hores de música*

Cependant nous ne croyons pas que ces propriétés dévoilent l'appartenance à un trait ou à un autre (abstrait ou concret). Par rapport à la partition en fragments ou en morceaux, nous croyons que c'est une propriété des noms faisant référence à un évènement de perception auditive où interviennent deux actants, 'celui qui explique' et 'celui qui écoute' même si ces deux actants ne sont pas syntaxiquement représentés [72] .

EXEMPLE 72 :

72.a : *Un tros de conversa*

72.b : *Un fragment de diàleg*

72.c : *Només he sentit un tros de l'acudit*

---

32. Par *genre* nous entendons ici l'hyperonyme.

À notre avis, est concret tout ce qui a une extension spatiale ; la ‘musique’ par contre, a une extension temporelle, c’est pourquoi ce type de noms acceptent naturellement la quantification de mesure de temps<sup>33</sup> :

EXEMPLE 73 : *Tres hores de música*

Nous sommes donc d’accord que ces noms sont limitrophes entre l’abstrait et le concret mais selon notre définition de concret, ils relèvent plutôt du trait abstrait.

Mis à part les noms vus jusqu’à présent, nous avons codé en tant qu’abstrait les noms ayant des informations aspectuelles inhérentes. Bien que la charge aspectuelle soit typique des noms dérivés de prédicats comme ceux de [61] et [62], il existe des substantifs non dérivés qui portent en eux des valeurs aspectuelles [65] et [74].

EXEMPLE 74 :

74.a : *El concert d’ahir al vespre*

74.b : *El partit de diumenge*

74.c : *Durant la festa...*

74.d : *Després de 2 hores de gimnàstica...*

74.e : *Durant la varicel·la és recomanable no sortir al carrer*

EXEMPLE 75 :

75.a : *\*(després de + durant) l’arbre*

75.b : *La carretera \*(d’ahir la nit + de la setmana passada)*

75.c : *\*(durant + després de) la taula*

Comme le montrent ces exemples, il existe des différences entre les substantifs de [74] et les substantifs de [75]. Ces derniers ont été étiquetés comme *Végetal*, *Locatif* et *Concret* respectivement. Les substantifs de [75] ne désignent pas le début, la fin ou l’évolution temporelle de l’objet. Par contre, les exemples de [74] font référence à des événements délimités dans le temps. Il faut souligner que l’aspect n’est pas une caractéristique exclusive du trait abstrait, c’est plutôt une caractéristique des prédicats. En réalité l’aspect est fortement lié aux verbes supports puisque ce sont eux qui actualisent les prédicats nominaux.

Finalement nous constatons que la dérivation ne fait référence qu’à une partie des noms abstraits et l’aspect ne peut pas rendre compte uniquement des noms ayant le trait abstrait. Il existe pourtant d’autres ensembles de noms qui ont été codés en tant qu’*abstrait*, mais cette fois-ci en nous basant exclusivement sur leur

---

33. La quantification est à gauche comme dans le cas des noms d’activités : *Dues hores de natació, dues hores de fútbol...*

statut référentiel: ce sont des noms qui n'ont aucune dimension spatiale. Voici une esquisse de ces ensembles de noms.

1. Les noms de religions

el paganisme, el politeisme, l'ortodòxia...

2. Les noms de jeux

el go, el bingo, el burro, la botifarra, el billar...

3. Les noms de sports

l'halterofília, el ciclisme...

4. Les noms de disciplines ou de sciences

la història, la biblioteconomia, la demografia, la deontologia, la ètica...

5. Des noms divers

el destí, el zodíac...

Il existe également d'autres ensembles de noms que nous considérons abstraits mais auxquels nous avons donné un étiquetage plus spécifique, soit à cause du caractère fermé de la classe, soit à cause de leur spécificité. Le codage que nous présentons ci-dessous sert uniquement de repère pour une analyse ultérieure :

1. Abst-Lng: Langues

danès, eslau, ucraïnès...

2. Abst-form: Formes

triangle, cub, corba...

3. Abst-mat: Termes mathématiques

arrel, polinomi, cosinus, integral...

4. Abst-clim: Termes climatologiques

boira, anticicló, vent...

## 5. Abst-Tps : Termes temporels

dilluns, època, instant, vespre...

Etant donnée la grande diversité de noms susceptibles de porter le trait sémantique *abstrait*, il est nécessaire de réaliser un deuxième classement à l'intérieur de ce groupe pour que le codage puisse être opérationnel. Il existe différentes propositions de sous-classement des noms abstraits. Nous ne citerons que l'étude réalisée par N. Flaux et D. Van de Velde et celle de M. Emorine.

N. Flaux et D. Van de Velde [Flaux and Van de Velde, 2000] proposent une classification axée sur 4 aspects : le caractère continu ou discontinu<sup>34</sup> et le caractère extensif ou intensif<sup>35</sup>. Ainsi à partir de la combinaison de ces 4 aspects on obtient le classement suivant :

- Les noms abstraits intensifs :
  - qualités / sentiments / états/
- Les noms abstraits extensifs :
  - indénombrables : activités
  - mixte : manières de faire
  - dénombrables : actions / procès / événements
- Noms abstraits extensifs et intensifs :
  - les facultés

Leur classification se base, donc, fondamentalement sur la détermination mais aussi sur d'autres critères formels tels que le type de verbe ou d'adjectif dont est issu le nom... Ce classement, exige parfois une analyse préalable des prédicats verbaux et adjectivaux.

Par ailleurs, G. Gross et son équipe reprennent un autre classement des prédicats sémantiques (noms abstraits pour G. Gross) réalisé par M. Emorine. En suivant les mêmes principes de classification des prédicats verbaux de l'Aktionsart, M. Emorine présente une taxinomie des prédicats nominaux qui se structure en fonction de leur caractère statique vs. dynamique. Le caractère dynamique d'un prédicat implique une situation présentant des changements, elle en distingue 4 sous-classes : substantifs d'activité, résultatifs, performatifs et une classe mixte. Parmi les statiques (i.e. une situation où il n'y a pas de changements) elle distingue 2 sous-classes : les substantifs de qualité et les états.

---

34. Termes équivalents à *indénombrable* et *dénombrable* respectivement.

35. Avec ces termes kantiens, les auteurs opposent les noms ayant une extension spatiale ou temporelle, les extensifs, de ceux qui à défaut de posséder ces propriétés ne peuvent qu'accentuer la force de leur effet, les intensifs. Exemples de noms extensifs : « *arbre, jardin, lac, parc, tristesse, construction...* » Exemples de noms intensifs « *admiration, intérêt, bonté, sagesse, courage...* ».

Ce classement est donc fondé sur le calcul aspectuel des lexies, mais dans le cas des prédicats nominaux, c'est le verbe support qui est chargé de donner les informations de temps (et les marques de personne et de nombre), i.e. c'est le verbe support qui actualise le prédicat nominal. Selon G. Gross les noms prédicatifs ne sélectionnent pas librement leurs verbes supports : il existe dans le sémantisme du prédicat nominal des restrictions aspectuelles qui empêcheront une libre combinatoire avec les différents verbes supports. C'est pourquoi le classement des prédicats nominaux se fait, essentiellement, en fonction des verbes supports qu'ils sélectionnent.

Les différents classements que nous venons de voir montrent que sous ce trait on regroupe une grande variété de noms. Ce trait est donc tellement générique qu'il ne peut être opératif du point de vue du traitement automatique, il est nécessaire de réaliser au moins une deuxième analyse pour que l'étiquetage soit efficace. Pourtant, dans l'élaboration de notre dictionnaire, des étiquettes aussi génériques sont nécessaires dans la mesure où elles aident à gérer la base de données. Avec une base de données de plus de 30 000 lemmes nous ne pouvons pas commencer notre analyse par des descriptions détaillées, il nous faut au moins un classement préalable, aussi générique soit-il, qui nous permette dans un deuxième temps d'aborder les différentes caractéristiques et problématiques d'un ensemble plus circonscrit d'entrées.

### 6.3 Concret

Notre dictionnaire compte 5 353 noms concrets au total dont 2 222 sont féminins et 3 131 sont masculins. Comme nous l'avons vu dans la section précédente, nous considérons qu'un nom est *concret* à partir du moment où il a une dimension spatiale<sup>36</sup>. Cependant une telle catégorisation ne peut recouvrir l'immense quantité de lexies d'une langue, puisqu'il y a de nombreuses lexies qui se situent à la frontière de différents traits. Nous avons constaté, par ailleurs, que plus un ensemble est élevé dans une taxinomie, moins il a de propriétés communes qui permettent de le définir. Les noms *concrets* tout comme les *abstrait*s, regroupent des lexies très variées entre elles. Cependant, contrairement au trait *abstrait*, le trait *concret* peut servir à la désambiguïsation et à la description de certains prédicats.

EXEMPLE 76 :

76.a : *El Nil pren una aspirina*

76.b : *El Nil pren una dutxa*

---

36. Et peut donc apparaître avec des prédicats de <dimension>, de <couleur>, etc.

Nous présenterons par la suite un aperçu des lexies que nous avons classées sous ce trait, puis les différentes problématiques que nous avons rencontrées.

Il existe un grand nombre de noms pour lesquels le trait *concret* fait consensus :

EXEMPLE 77 :

- 77.a : *armari, taula, cadira...*
- 77.b : *mocador, pantaló, vestit...*
- 77.c : *ampolla, galleda, gibrella...*
- 77.d : *sac, cartera, alforja...*
- 77.e : *trompeta, arpa, piano...*
- 77.f : *pinta, embut, cullera...*
- 77.g : *falç, rastell, pala...*
- 77.h : *aspirador, astrolabi, àbac...*
- 77.i : *ferrocarril, vaixell, cotxe...*
- 77.j : *ploma, pinzell, llapis...*

Par contre, il existe d'autres cas pour lesquels la classification n'est pas claire. Dans notre dictionnaire, le trait *concret* se distingue de *l'abstrait* par le fait d'avoir une extension dans l'espace. Ainsi les noms comme :

EXEMPLE 78 :

- 78.a : *fotocòpia*
- 78.b : *fotografia*
- 78.c : *rebut*
- 78.d : *comprovant*
- 78.e : *factura*

ont été étiquetés comme concrets. En effet, ces noms peuvent parfaitement apparaître avec des prédicats qui sélectionnent uniquement des concrets (*esbarriar*, *embrutar...*), et avec lesquels aucun abstrait ne pourrait apparaître<sup>37</sup>.

Nous pourrions alors penser qu'ils ressemblent plutôt aux noms « résultatifs »<sup>38</sup> dont nous avons parlé précédemment (cf. page 53 : *traducció*,

37. Il est important de souligner qu'il existe des concrets qui sont acceptés uniquement par quelques uns de ces prédicats ; par exemple les noms de parties du corps peuvent apparaître avec le verbe *embrutar* mais pas avec *esbarriar* : \*He esbarriat el ( melic + braç + cama... ). Ce qui marque vraiment la différence, c'est qu'un abstrait ne peut pas se combiner avec ces verbes sans avoir recours à un sens figuré : *embrutar la reputació d'algú*.

38. Bien que *comprovant* puisse dériver de *comprovar*, *rebut* de *rebre* ou *factura* de *facturar*, nous ne les considérons pas des noms résultatifs comme *traducció* ou *compra*. Le résultat de *comprovar* est *comprovació*, de *facturar* est *facturació*, et de la même façon, nous considérons que de *rebre*, l'objet ou ce qui en résulte, ce n'est pas un *rebut*. Nous considérons donc ces trois lexies indépendantes, à ce niveau-là, du verbe duquel elles dérivent.

compra, construcció...). Pourtant, contrairement aux « résultatifs », les noms de [78] n'ont pas de structure événementielle. Par conséquent, et comme le montrent les exemples [79] ces noms ne peuvent pas être inscrits dans le temps sans la présence d'un verbe support.

EXEMPLE 79 :

79.a : *Durant la traducció em vaig trobar molts dubtes*

79.b : *Durant l'edició...*

79.c : *\*Durant la fotocòpia...*

79.d : *\*Durant la fotografia...*

79.e : *\*Durant el rebut...*

Il existe un autre ensemble de noms qui présentent des points en commun avec l'abstrait et le concret mais que nous avons codé comme concrets.

EXEMPLE 80 :

80.a : *llibre*

80.b : *carta*

80.c : *diari*

80.d : *revista*

Contrairement aux « résultatifs », les noms de [80] ne dérivent pas d'un verbe. D'ailleurs, ce type de noms peut apparaître avec des prédicats propres aux NConc comme :

EXEMPLE 81 :

81.a : *Un llibre (vermell + gran + rectangular + de plàstic)*

81.b : *Una traducció \*(vermella + gran + rectangular + de plàstic)*

En effet, l'une des caractéristiques principales des NConc est le fait de pouvoir être sélectionnés par des prédicats de <dimension> (gran, llarg, gruixut, etc.), de <forme> quadrat, rectangular, etc. ou de <couleur> (blau, gris, etc.) ainsi que par des prédicats comme *esbarriar*, *plastificar*, etc. Nous avons essayé de montrer à travers ces exemples que souvent l'identification des traits n'est pas toujours facile à déterminer, mais elle nous permet, par ailleurs, d'attribuer les premiers indices sémantiques à nos lexies.

## 6.4 Végétal

Parmi les substantifs de notre dictionnaire, 664 entrées ont été classées sous le trait syntactico-sémantique *Végétal* (Veg). Parmi eux 345 sont des noms masculins et 319 des noms féminins. Le trait *Veg* constitue une catégorie assez bien définie grâce aux prédicats *plantar* ou *regar*.

<i>Veg et Conc</i>	<i>Veg et Hum</i>	<i>Veg et Anl</i>
escombra	fadrí	gall
escopeta	monja	mula
trompeta		
rellotge		
raspall		

TAB. 6.1: *Homonymie et polysémie dans le trait Veg*

EXEMPLE 82 :

82.a : *NHum regar NVeg*

82.b : *NHum plantar NVeg*

Cependant, cette définition des *Veg* demeure insuffisante car elle ne recouvre pas l'ensemble de noms que nous avons associé à ce trait. En effet, il existe, à l'intérieur des *Veg*, des noms qui ne peuvent pas se combiner avec ces prédicats, ce sont des noms comme *alga* ou *molssa*. Les noms *Veg* sont à la frontière des inanimés (i.e. des *Conc*) et des animés (i.e. des *Anl* et *Hum*). En effet les *Veg* sont acceptés par des prédicats propres aux noms animés comme *créixer* ou *morir* qui ne peuvent pas être accompagnés d'un inanimé. Mais, comme les inanimés, ils ne peuvent pas coapparaître avec de nombreux prédicats, notamment certains prédicats de 'mouvement'.

EXEMPLE 83 :

83.a : *L'atzavara mor després de la floració*

83.b : *\*L'atazavara anava ràpid*

83.c : *La molssa creix al cantó nord dels arbres*

Ainsi, pour rendre compte des éléments de l'ensemble des *Veg* nous ferons appel aussi bien à des prédicats spécifiques comme *plantar*, *regar*, *cultivar*, *fotosíntesi*, *floració* qu'aux prédicats, moins spécifiques, comme *créixer*, *respirar* ou *viure*. La définition des *Veg* deviendra de moins en moins souple lors de l'établissement de sous-ensembles, i.e. de leur répartition en classes d'objets (cf. §7). Le trait *Veg* nous a permis de désambigüiser de nombreuses entrées de notre dictionnaire dont la forme correspond à deux unités lexicales différentes correspondant à deux traits syntactico-sémantiques différents (cf. tableau 6.1).

## 6.5 Humain vs Animal

Notre dictionnaire comprend un total de 4 385 nom étiquetés comme *Humain* (Hum) et 724 noms étiquetés comme *Animal* (Anl). Parmi les noms humains,



2 472 sont des noms masculins et 1 912 sont des noms féminins. Par rapport aux noms ayant le trait *Animal*, 454 sont des noms masculins et 270 des noms féminins. Le trait *Anl* et le trait *Hum* ont en commun le fait de pouvoir apparaître avec de nombreux prédicats qui les différencient, d'autre part, des autres traits :

EXEMPLE 84 :

84.a : *El/L' (Nil + ocell) va veure un escarbat*

84.b : *El (Nil + lleó) beu més d'un litre d'aigua al dia*

84.c : *El (Nil + guepard) és esvelt i prim*

Or, d'un point de vue linguistique chacun d'entre eux est atteint de ces propres spécificités. Nous allons présenter de manière sommaire quelques unes de ces différences [Fuentes, 2006a].

En premier lieu, un *Hum* a la possibilité d'apparaître avec des prédicats qui lui sont spécifiques :

- comunista, solter, honest, etc.
- trucar, opinar, jubilar-se, etc.

Par ailleurs, les *hum* sont les seuls à disposer d'un réseau de termes relationnels aussi complexe que *tia*, *sogre*, *nòvio*, *àvia*...

Du point de vue de la syntaxe, le recours aux pronoms dévoile à nouveau les différences linguistiques entre ces deux traits. En effet, seuls les noms *humains* peuvent apparaître comme référents des pronoms *jo*, *tu*, *nosaltres*, *vosaltres*. D'ailleurs, d'autres pronoms sont aussi réservés aux *humains* :

EXEMPLE 85 :

85.a : *Un escriptor és **algú** que necessita escriure per viure.*

85.b : *\*El gat de la veina és **algú** que no pot. . .*

85.c : ***Qui** és el president de Catalunya?*

85.d : *\***Qui** és el gos del Nil?*

Nous avons montré par ces exemples les principales dissemblances des deux traits qui constituent la catégorie des « noms animés ». En ce qui concerne le trait *Anl* nous donnerons des spécifications sémantiques plus précises dans la section §7.

## 6.6 Locatif

Comme nous l'avons vu jusqu'à présent, la répartition en traits syntactico-sémantiques n'est pas exempte de problèmes. En effet, les traits sémantiques englobent des éléments de nature très diverse ce qui rend difficile l'établissement de caractéristiques communes. Cependant, comme pour les autres traits, dans le

cas des locatifs nous avons essayé de procéder à l'étiquetage en fonction des prédicats spécifiques.

Selon D. Le Pesant [Le Pesant, 2000], l'une des propriétés communes à tous les locatifs est le fait de pouvoir apparaître dans le contexte suivant :

$N_0$ :Hum être à  $N_1$ :Loc

Si nous l'adaptions au catalan, nous obtenons les configurations distributionnelles suivantes :

1.  $N_0$ :Hum ser a  $N_1$ :Loc
2.  $N_0$ :Hum estar a  $N_1$ :Loc
3.  $N_0$ :Hum estar-se a  $N_1$ :Loc  $N_2$ :Tps

EXEMPLE 86 : Ser

86.a : *El director és al seu despatx*

86.b : *El Nil és a la presó*

EXEMPLE 87 : Estar

87.a : *El Nil va estar tota la tarda a casa*

87.b : *El Nil està a l'estació*

EXEMPLE 88 : Estar-se

88.a : *El Nil s'estarà tot el dia a l'hospital*

88.b : *El Nil es va estar quinze dies a la platja*

Ainsi, dans les exemples [86] et [87], nous aurions comme synonyme trobar-se et dans l'exemple 88 nous aurions le synonyme passar-se.

Or, il existe des noms non-locatifs qui pourraient éventuellement apparaître dans les contextes ci-dessus [89] mais, contrairement aux NLoc, ils ne peuvent pas être donnés comme réponses à des questions comme celle de [90] :

EXEMPLE 89 :

89.a : *El Nil està al sofà*

89.b : *El Nil és al cotxe*

89.c : *El Nil està a una reunió*

EXEMPLE 90 : Question: On vas anar el cap de setmana?

90.a : *\*Vaig anar al sofà*

90.b : *\*Vaig anar al cotxe*

90.c : *Vaig anar a una reunió*

Il existe, en effet une différence par rapport au terme reunió puisque non seulement il peut rentrer dans les 3 configurations distributionnelles vues précédemment mais aussi et comme nous venons de le voir comme

réponse à [90]. Cette caractéristique est extensible à de nombreux noms liés à un ‘événement’ comme *reunió*, *circ*, *teatre*, *concert*, *òpera*, *cinema*, *missa*...

En effet, ce type de noms présentent des propriétés du trait *abstrait* ainsi que des propriétés du trait *locatif*. Pourtant, nous avons remarqué des différences parmi ces entrées ce qui nous a conduit à les étiqueter différemment : soit avec le trait *Abst*, soit avec le trait *Loc*.

Si nous observons les phrases suivantes :

EXEMPLE 91 :

91.a : *El Nil va visitar el/la (teatre + circ + cinema + classe + òpera)*

91.b : *El Nil va visitar el/la \*(missa + concert + reunió)*

91.c : *La reconstrucció del/de la (teatre + circ + òpera + cinema)*

91.d : *La reconstrucció del/de la \*(concert + missa + reunió)*

Nous pouvons remarquer que ces noms n’ont pas les mêmes propriétés distributionnelles. En effet, à différence des autres noms, *concert* et *missa* refusent des contextes comme ceux de [91] c’est pourquoi nous ne les avons pas étiquetés comme *locatifs* mais comme *abstrait*s. Nous croyons qu’il s’agit de noms abstraits, des noms événementiels qui sont étroitement liés à un locatif ce qui explique leur adaptabilité à certains contextes propres à des locatifs (cf. tableau 6.2).

Loc	Abst
església	missa
auditori	concert
sala de reunions	reunió

TAB. 6.2: L’endroit et l’évènement (i) : locatif et abstrait

En ce qui concerne les noms *teatre + circ + cinema + classe + òpera* nous considérons qu’il s’agit de noms polysémiques dont l’un est locatif et l’autre abstrait. En effet, ces noms ont une structure événementielle inexistante chez les locatifs. Contrairement aux noms comme *missa*, *concert*, *reunió*, pour *teatre + circ + cinema + classe + òpera* le même signifiant désigne l’évènement et l’endroit où se déroule l’évènement. Ce qui explique l’acceptabilité des constructions de [92]. Ainsi ces entrées ont été systématiquement dédoublées, l’une comme *abstrait* et l’autre comme *locatif*.

EXEMPLE 92 : *La (òpera<sup>39</sup> + classe + circ) ja havia començat i no em van deixar entrar*

39. *òpera* aurait en plus un autre sens correspondant à ‘style de musique’ comme peut l’être *le jazz* ou *le blues*

Il faut remarquer que *teatre* ou *cinema* refuseraient les contextes de [92]. En effet, nous ne considérons pas ces deux noms polysémiques de la même façon que ceux de [92] car pour ces exemples l’endroit et l’évènement correspondent à des mot-formes différents comme le montre le tableau 6.3.

Loc	Abst
cinema	película
teatre	funció
òpera	òpera
circ	circ

TAB. 6.3: *L’endroit et l’évènement (ii) : locatif et abstrait*

Ainsi, parmi ces entrées, nous avons observé trois phénomènes différents :

- Des noms dont la forme de l’évènement (*Abst*) et de l’endroit (*Loc*) sont identiques : *circ*, *òpera*, *classe*.
- Des noms abstraits étroitement liés à un endroit précis : *reunió*, *concert*, *missa* et qui admettent donc certaines constructions propres à des locatifs.
- Des noms locatifs<sup>40</sup> : *cinema*, *teatre*.

A l’intérieur des NLoc, Le Pesant [Le Pesant, 2000] distingue deux grandes catégories :

1. « *Les noms d’étendues de la surface de la terre* »
2. « *Les noms de bâtiments et de pièces* »

Jusqu’à présent nous n’avons parlé que de noms qui correspondent à la deuxième catégorie, i.e. aux noms de bâtiments. En ce qui concerne les noms d’étendues de surface nous avons également rencontré quelques difficultés lors de la classification car ils présentent une combinatoire très proche de celle des noms collectifs. Ce type de noms sera pourtant traité en détail dans la section §6.7<sup>41</sup>.

Finalement, cette grande répartition des NLoc nous permet de donner les principaux prédicats qui vont nous aider à les délimiter dans notre taxinomie.

- Prédicats appropriés aux noms d’étendues et de surface de la terre :  
– *sobrevolar*, *cultivar*, *caminar per*, etc. (N<sub>0</sub>Hum; N<sub>1</sub>Loc)

40. Nous voudrions faire remarquer que ces noms ont également un homographe abstrait qui correspondrait à ‘l’art de’ ou ‘discipline de’ : *A Grècia el teatre es va començar a desenvolupar...*

41. Il existe également des NLoc de la première catégorie avec des caractéristiques communes aux NColl comme *universitat*, *ajuntament*, *ciutat...* que nous traiterons, également, dans la section §6.7

- Prédicats appropriés aux noms de bâtiments et de pièces: *entrar a*, *sortir de*, *anar a*, *viure a/en*, etc. (N<sub>0</sub>Hum; N<sub>1</sub>Loc)

Notons que mis à part cette double répartition, les contraintes de sélection de ces prédicats nous donnent déjà des indices sur l'existence de différentes sous-classes à l'intérieur des locatifs. En effet, dans un deuxième temps il sera nécessaire de réaliser une répartition en classes d'objets de tous les NLoc.

## 6.7 Les noms collectifs

Nous venons de voir les traits sémantiques qui font référence à un individu ou une seule entité, or il existe dans la langue des noms qui dénotent une pluralité interne, des noms qui au singulier font référence à plus d'une entité, ce sont les noms collectifs (NColl). Nous considérons qu'un nom collectif est celui qui désigne un 'ensemble de X' où X est spécifié (*moscam*, *gossada*, *prunerar família...*). Cette définition exclut des noms tels que *sèrie*, *multitud*, *col·lecció*, *munió...* où l'entité désignée n'est pas indiquée<sup>42</sup>.

En vue de l'élaboration d'un dictionnaire électronique, nous croyons qu'il est nécessaire de distinguer les noms collectifs des noms individuels<sup>43</sup> par la spécificité de leur propriétés distributionnelles. Nous exposerons par la suite ces propriétés spécifiques des NColl [Bosque, 1999], [Le Pesant, 2000], dans le cas du catalan [GDL, 1998] et [Solé, 2002].

### 6.7.1 Propriétés syntaxiques

L'une des caractéristiques essentielles des NColl est la possibilité d'apparaître au singulier dans des contextes où, normalement, ne sont acceptés que des noms coordonnés ou des noms au pluriel.

- Un NColl au singulier peut, par exemple, être précédé de la préposition *entre* ou coapparaître avec l'adjectif *nombrós*.

EXEMPLE 93 : *Entre*

93.a : *AnlColl: El carboncle és una malaltia comuna entre el bestiar causada pel bacteri Bacillus anthracis.*

93.b : *ConcColl: Molts dels llibres foren recuperats entre l'equipatge del rei Josep I Bonaparte després de la batalla de Vitòria.*

93.c : *HumColl: En aquest moment hi ha un important malestar entre el professorat per la falta de reconeixement la seva tasca.*

42. Nous traiterons ce type de noms dans la section suivante.

43. Désormais, nous utiliserons ce terme pour faire référence aux noms comme *gos*, *estudiant*, *cadira* que nous avons décrit tout au long des sections précédentes.

93.d : *S'havia amagat entre \*(el gat + el professor + l'arbre + la maleta + ...).*

EXEMPLE 94 : Nombrós

94.a : *Després d'una intensa i acurada preparació durant anys, el 1714 un nombrós exèrcit liderat per Rafael de Casanova es va llençar a la conquesta de Castella.*

94.b : *Si s'urbanitza la Vall causarà un efecte molt pernicios per la nombrosa fauna de l'entorn.*

- Leur pluralité interne, leur permet également d'apparaître au singulier avec des prédicats tels que reunir, agrupar, ajuntar, aplegar, agregar, disgregar, congregar, amuntegar, etc.:

EXEMPLE 95 :

95.a : *La fira de Teatre congrega públic de totes les edats*

95.b : *La Coral reuneix sopranos, contralts, tenors i baixos*

95.c : *(...) és una antiga activitat ramadera que consisteix en ajuntar el bestiar que viu de forma salvatge a la muntanya.*

- Le NColl, qui normalement fait l'accord en genre et en nombre avec le verbe, peut, dans des contextes particuliers<sup>44</sup> faire l'accord avec le verbe, non pas à partir de la forme fléchie du NColl mais à partir du sens pluriel qu'il véhicule ; faire, donc un accord *ad sensum* :

EXEMPLE 96 :

96.a : *La policia va fer vaga*

96.b : *\*La policia van fer vaga*

96.c : *Va avisar la policia, però no va arribar fins al cap de mitja hora.*

96.d : *Va avisar la policia, però no van arribar fins al cap de mitja hora*

96.e : *\*Va avisar el seu germà, però no van arribar fins al cap de mitja hora.*

- Mis à part les prédicats signalés ci-dessus [95], les NColl ont d'autres prédicats appropriés<sup>45</sup> qui rendent compte de la relation méronyme/holonyme présente dans la plupart des NColl :

EXEMPLE 97 :

97.a : *A la solana, la vegetació està formada per matolls, on l'espècie dominant és el garric.*

44. Pour plus d'information sur ce type de contextes cf. 2.4.4.1 [Solà and Rigau, 2002].

45. Des prédicats comme comptar amb ou (format, integrat, constituït, compostat) per ainsi que les formes féminines correspondantes.

97.b : *El vestuari compta amb dues peces negres; una armilla i els pantalons on hi ha estampades flames blanques i blaves.*

Toutes ces propriétés nous ont donc amené à discriminer les NColl des noms individuels dans la microstructure de notre dictionnaire. Les noms collectifs constituent 3,5% des substantifs (soit 779 entrées). Pour leur codage, nous avons utilisé le même principe de classification que celui adopté pour les *noms individuels* mais en ajoutant l'extension Coll. Nous avons, donc, regroupé les NColl en traits syntactico-sémantiques. La figure 6.1 montre la distribution des NColl par rapport à chaque trait ainsi que le système de codage adopté.

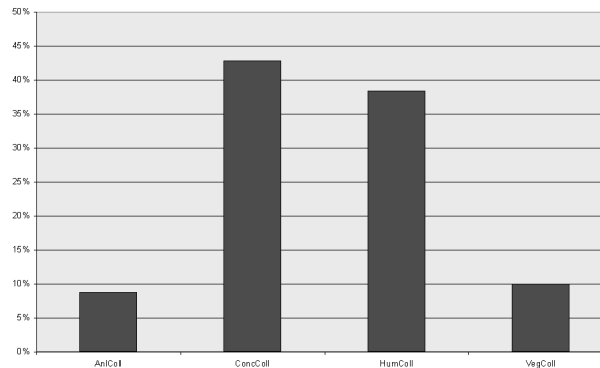


FIG. 6.1: Répartition des NColl en traits syntactico-sémantiques

#### 6.7.1.1 Plus que pluralité

En observant les propriétés syntaxiques à l'intérieur même des NColl nous pouvons remarquer qu'il existe parmi eux des différences remarquables. Comme nous venons de le voir l'une des caractéristiques des NColl est le fait de pouvoir être précédés de la préposition *entre* (cf. exemples [93]). Or, il existe de nombreux NColl qui n'acceptent pas cette préposition :

EXEMPLE 98 :

98.a : *AniColl: \*entre (el formiguer + el vesper + ...)*

98.b : *HumColl: \*entre (la universitat + el parlament + ...)*

98.c : *VegColl: \*entre (el bosc + el prunerar + ...)*

D'autre part, ces mêmes NColl peuvent apparaître dans des contextes où la majorité de noms collectifs ne peut pas :

EXEMPLE 99 :

99.a : *El Nil passeja (pel bosc + per la universitat + ...).*

99.b : *El menjar que les formigues transporten pel formiguer.*

99.c : *El Nil visita la universitat.*

Pourtant, comme tout autre NColl et contrairement aux noms individuels, ils peuvent apparaître dans les contextes suivants [100] :

EXEMPLE 100 :

100.a : *La universitat va aprovar la llei **per majoria***<sup>46</sup>

100.b : *El prunerar estava **format per** pruners del mediterrani.*

En réalité, ces dissemblances sont dues au fait que ces noms ont un sens supplémentaire inexistant dans les autres NColl. En effet, ces noms ne correspondent pas uniquement au sens un ‘ensemble de X’ mais plutôt à des situations du type :

- ‘X est un endroit où il y a beaucoup de Y’
- ‘X est un endroit où l’on cultive Y’
- ‘X est l’endroit où habite un ensemble de Y’

Ainsi, ces noms ont un comportement différent aux autres NColl à cause de cette double dimension dans laquelle ‘la collectivité’ est associée à un endroit précis. Cela explique le fait qu’ils puissent être sélectionnés à la fois et par les prédicats généraux du trait auquel ils appartiennent (*Anl*, *Conc*, *Hum* ou *Veg*) et par les prédicats propres aux locatifs<sup>47</sup> (cf. [99]), comme par exemple *entrar*, *sortir*, *tornar de/a* ou *anar a*. Ainsi, par exemple, les VegColl sont acceptés par des prédicats tels que *créixer*, *marcir* ou *assecar* ; les AnlColl par *viure* ou *alimentar-se* et finalement les HumColl par des prédicats comme *decidir* ou *votar*. Et ils ont tous en commun le fait de pouvoir apparaître dans des configurations propres aux collectifs comme celles de [100] .

Nous considérons que tous ces noms sont donc polysémiques, mais puisque les deux emplois sont partiellement compatibles [101] nous ne croyons pas qu’il s’agisse de deux unités lexicales, mais plutôt d’une propriété inhérente de ce type de lexies ; c’est pourquoi ces mots n’ont pas été dédoublés dans notre dictionnaire, au contraire, ils ont été affecté d’une étiquette particulière pour rendre compte de leur double dimension.

EXEMPLE 101 :

101.a : *Aquest pomerar d’unes 40 hectàrees està completament marcit*

101.b : *La universitat va romandre tancada des de que va decidir fer vaga*

46. *per majoria*, *per unanimitat*.

47. D’autres exemples de prédicats appropriés aux locatifs sont disponibles dans la section §6.6.



### 6.7.2 Propriétés morphologiques

D'un point de vue morphologique, nous pouvons distinguer deux grandes sous-classes à l'intérieur des NColl. D'une part, les noms où la notion de 'collectivité' est exprimée par un suffixe et dont la racine indique l'entité qui va être multipliée (*arbreda*, *alumnat*...). D'autre part, les noms, morphologiquement indivisibles<sup>48</sup>, où tout est exprimé par la racine (*vaixella*, *públic*, *família*...). Les premiers constituent 69,1% des NColl, tandis que les seconds 38,1%.

Selon le tableau 6.4 ( de la présente page) nous pouvons constater que les VegColl et les AnlColl se construisent majoritairement par dérivation, tandis que pour les autres traits la différence n'est presque pas remarquable. En effet, la file %*trait* du tableau 6.4 nous indique justement la proportion de noms dérivés et de noms non-dérivés (lexicaux) par rapport à chaque trait.

Par ailleurs, la file %NColl du tableau 6.4 nous indique le poids relatif de chaque NColl en fonction du trait et du type de composition par rapport à l'ensemble de NColl. Ces mêmes données sont plus clairement exposées dans la figure 6.2 page suivante.

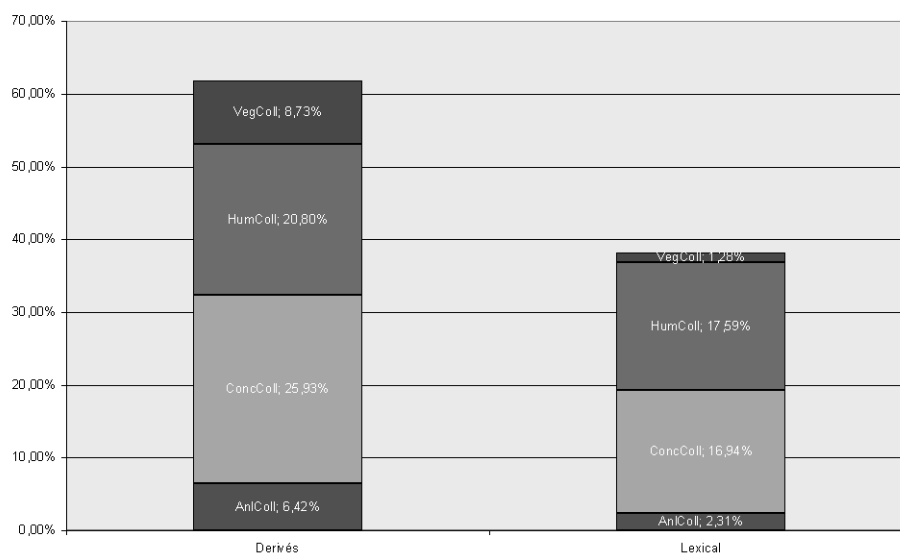
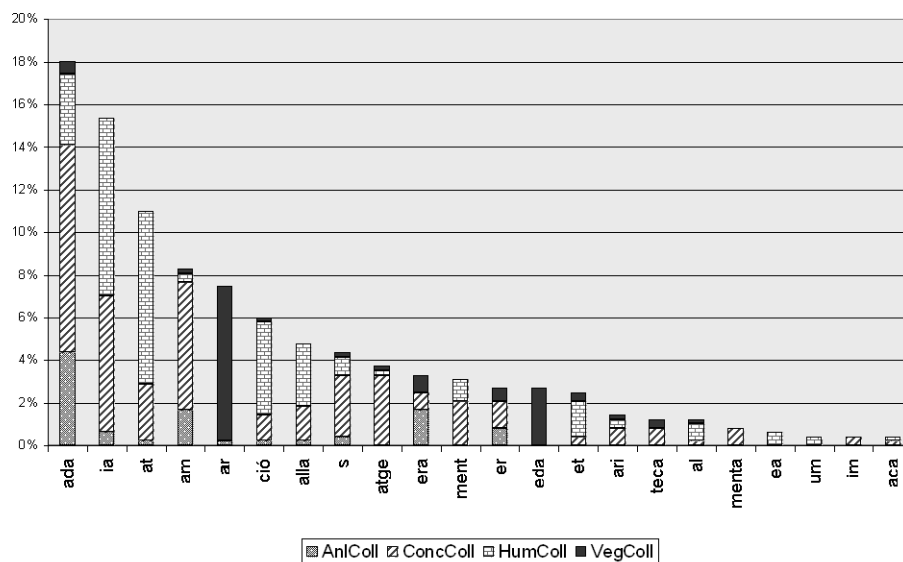
		AnlColl	ConcColl	HumColl	VegColl	Total general
Dérivés	Num	50	202	162	68	482
	% Ncoll	6,4%	25,9%	20,8%	8,7%	61,9%
	% trait	73,5%	60,5%	54,2%	87,2%	61,9%
Lexical	Num	18	132	137	10	297
	% Ncoll	2,3%	16,9%	17,6%	1,3%	38,1%
	% trait	26,5%	39,5%	45,8%	12,8%	38,1%
Total Num		68	334	299	78	779
Total % Ncoll		8,7%	42,9%	38,4%	10,0%	100,0%
Total % trait		100,0%	100,0%	100,0%	100,0%	100,0%

TAB. 6.4: *Distribution des NColl*

Par ailleurs, nous avons réalisé une étude détaillée sur la distribution de toutes ces formes suffixales attachées aux NColl. Pour ce faire, nous avons repris la liste de suffixes catalans proposée par E. Solé [Solé, 2002]. Ensuite, nous avons regroupé les NColl en fonction de leur marques formelles ou, le cas échéant, par leur absence. Ainsi, nous avons obtenu le pourcentage d'apparition de chaque suffixe par rapport aux traits syntactico-sémantiques, comme le montre la figure 6.3.

Comme nous avons vu précédemment, les VegColl avaient un comportement quelque peu différent des autres traits: 87,2% des VegColl étaient formés par dérivation. La figure 6.3 nous montre aussi l'existence d'un suffixe presque

48. Dans le sens où il ne sont composés d'aucun suffixe qui véhicule la notion de 'collectivité'.

FIG. 6.2: *NColl dérivés et NColl non-dérivés*FIG. 6.3: *Distribution des suffixes des NColl*

exclusif aux VegColl et qui compose la majorité des VegColl de notre dictionnaire le suffixe *-ar*<sup>49</sup>. Toujours selon la même figure, nous pouvons voir qu'il y a également d'autres suffixes exclusifs ou presque exclusifs à un trait comme par exemple, *-atge* pour les ConcColl ou *-eda* pour les VegColl, mais leur taux d'apparition est remarquablement inférieur à celui du suffixe *-ar*.

A l'exception de ce suffixe, la figure 6.3 nous aide à voir, d'un côté, la grande variété de suffixes corrélés à la notion de la collectivité et d'un autre côté, l'absence de régularité dans la distribution des suffixes.

Voici les principales caractéristiques de ces suffixes :

- Ils véhiculent et constituent un sens ce qui les rend facultatifs puisqu'ils peuvent être paraphrasés lexicalement :

EXEMPLE 102 :

102.a : *moscam* = *conjunt de mosques*

102.b : *veïnat* = *conjunt de veïns*

- Ils ont une combinatoire assez restreinte puisqu'ils n'atteignent qu'un groupe très réduit de substantifs (3,5% du total de substantifs de notre dictionnaire)
- Il existe plus d'une dizaine de suffixes pour exprimer la collectivité (cf. figure 6.3) dont la distribution se prête peu ou mal à une systématisation.

De par leur nature, nous considérons que ces suffixes nominaux dénominatifs<sup>50</sup> sont des dérivatèmes. Par rapport aux typologies des significations dérivationnelles décrites par I. Mel'čuk, nous présumons que ces dérivatèmes appartiennent aux significations dérivationnelles sémantiques du type P(P) (« *des dérivatèmes qui forment des désignations de participants à partir de désignations de participants*<sup>51</sup> »). Ces dérivatèmes « *servent à désigner une entité en termes d'une situation quelconque impliquant, de façon essentielle, une autre entité (= celle de départ)* » [Mel'čuk, 1994].

### 6.7.3 Morphologie et sémantique

Si nous analysons en détail les entrées de notre dictionnaire étiquetées VegColl, nous nous rendons compte qu'à différence des autres traits, presque la totalité des VegColl dérivés présentent la double dimension avec le trait locatif

49. Nous n'avons trouvé qu'un seul exemple où ce suffixe n'appartient pas aux VegColl mais à un AnlColl : *bestiar*.

50. Le lexème de départ est un nom (= dénominatif); le lexème dérivé est également un nom (= nominal).

51. Le terme *participant* désigne les entités qui ne sont pas des faits [Mel'čuk, 1994, p. 313].

dont nous avons parlé dans la section §6.7.1.1. En effet, parmi les VegColl dérivés il y a seulement 2 cas sur 68 qui ne partagent pas le trait 'locatif' : arbratge et vegetació; et parmi les VegColl non-dérivés des 10 cas aucun ne véhicule le sens de 'lieu' : flora, landa, bardissa, artell, etc. Pour la majorité des VegColl, le sens véhiculé correspondrait aux situations :

- 'X est un endroit où il y a beaucoup de Y' : alzinar, albera, pinar...
- 'X est un endroit où l'on cultive Y' : arrossar, bladar, patatar...

Pour les cas isolés, la paraphrase lexicale correspondrait tout simplement à la situation 'X est un ensemble de Y'.

En ce qui concerne les HumColl, les dérivatèmes *-al*, *-ari* ont sans exception la double dimension de 'locatif' et 'collectif' contrairement à ce qui se passe avec les dérivatèmes *-ada*, *-alla*, *-ció*, *-am* *-ia* et *-at* qui n'impliquent pas ce premier sens. Ainsi, le sens véhiculé par *-al*, *-ari* correspondrait plutôt à 'endroit où il y a beaucoup de' ou bien à un 'endroit où travaillent des Hum'.

Dans le cas des AnlColl le sens 'locatif' est véhiculé exclusivement par les dérivatèmes *-er/-era*. Cette fois-ci le sens véhiculé correspondrait à 'tanière' ou 'nid de l'animal X' :

EXEMPLE 103 :

103.a : *avisper*

103.b : *termiter*

103.c : *formiguer*

Parmi les AnlColl nous trouvons un autre ensemble de NColl dont le sens véhiculé par le dérivatème *-ada* exprime la 'quantité de X nés d'un accouchement' :

EXEMPLE 104 :

104.a : *niuada*

104.b : *llopada*

104.c : *llocada*

104.d : *gossada*

104.e : *gatinada*

104.f : *llodrigada*

104.g : *garrinada*

Comme nous venons de le voir, l'étiquette *Coll* rassemble une grande diversité de noms dont le sens n'est pas toujours très homogène. En effet, les NColl ne sont pas uniquement des noms désignant 'un ensemble de X', il existe à l'intérieur des NColl de nombreuses typologies sémantiques dont les différences sont notoires du point de vue de leurs propriétés distributionnelles. En ce qui concerne la morphologie des NColl nous avons pu constater qu'il n'y a pas une tendance claire vers la composition suffixale des NColl, sauf dans le cas des végétaux (VegColl)

et des animaux (AnlColl). Ainsi la régularité morphologique à l'intérieur d'un des traits collectifs s'explique par l'absence de diversité sémantique ; en effet, la majorité des VegColl fait référence à une même situation sémantique ('endroit où il y a de nombreux X').

## 7 Les classes d'objets

Nous venons de voir les traits syntactico-sémantiques que nous avons utilisés pour coder tous les substantifs de notre dictionnaire. Le classement en traits syntactico-sémantiques nous permet déjà de faire face à l'un des problèmes de base qui se pose lors de la description formelle des langues naturelles, la polysémie. Certains traits syntactico-sémantiques s'avèrent nécessaires pour la description des prédicats puisqu'ils délimitent la nature sémantique des arguments qu'ils sélectionnent et rendent possible une première désambiguïsation :

- El Nil agafa l'autopista
  - agafar ( $N_0$ :Hum; $N_1$ :Loc)
- El Nil agafa un boli
  - agafar ( $N_0$ :Hum; $N_1$ :Conc)

Or, ce classement en traits syntactico-sémantiques demeure insuffisant pour rendre compte de la totalité des emplois d'une lexie. En vue du traitement automatique de la langue, un seul classement en traits syntactico-sémantiques pourrait donner des résultats erronés, puisque les unités lexicales ne sont pas suffisamment désambiguïsées :

- Du point de vue de la génération de textes :

EXEMPLE 105 :

105.a : *El Nil tasta el vi*

105.b : *\*El Nil tasta l'ordinador*

- Du point de vue de la traduction automatique <sup>52</sup>:

EXEMPLE 106 :

106.a : *tallar( $N_0$ Hum;  $N_1$ Conc): El Nil talla un diamant*

106.b : *découper ( $N_0$ Hum;  $N_1$ Conc) / tailler ( $N_0$ Hum;  $N_1$ Conc): Nil (découpe? / taille?) un diamant*

---

52. D'après cet exemple le traducteur ne peut discerner l'unité qu'il devrait choisir

Un sous-classement à l'intérieur des traits, délimitant encore plus la nature syntactico-sémantique des arguments, devrait permettre de mieux rendre compte de l'existence d'unités lexicales différentes. Ce sous-classement correspond aux classes d'objets postulées par Gaston Gross. Une classe d'objets se veut un regroupement de lexies ayant des propriétés sémantiques et syntaxiques analogues. Elle se définit à partir du prédicat ou de l'ensemble de prédicats qui sélectionnent dans une position syntaxique déterminée, un ensemble précis de lexies (i.e. une classe d'objets), ces prédicats sont appelés les prédicats appropriés à la classe. Chaque classe d'objets sera donc décrite par ses opérateurs appropriés ainsi que par ses propriétés syntaxiques. Dans ce chapitre nous indiquerons les différentes classes d'objets que nous avons attribuées à chaque trait syntactico-sémantique en donnant quelques exemples des lexies de notre dictionnaire faisant partie de la classe. Pour des raisons d'espace, nous décrirons uniquement les classes d'objets qui à notre avis sont les plus représentatives. Nous voulons souligner qu'il s'agit ici d'une première version de notre dictionnaire, les classes d'objets que nous proposons ne sont pas définitives, il s'agit uniquement d'un premier classement en vue d'ultérieures modifications <sup>53</sup>.

## 7.1 Classes d'objets du trait *Veg*

Parmi les substantifs ayant le trait syntactico-sémantique *Veg* nous avons distingué 11 classes d'objets :

```
<alga>: coral·lina, actínia, verdet...
<arbre>: alzina, palmera, avet...
<cultiu>: ceba, patata, bròquil...
<flor>: dàlia, badoc, malva...
<fong>: bolet, candela, trompeta...
```

---

53. Dorénavant, les entrées du dictionnaire seront présentées linéairement et chaque champ sera séparé par une barre oblique. Voici l'information correspondante à chaque champ :

1. lexie
2. catégorie grammaticale
3. code de flexion
4. trait syntactico-sémantique
5. classe d'objets
6. fonctions lexicales (s'il y a en a plus d'une elles sont séparées par le signe : +)
7. plausibilité
8. traduction vers l'espagnol
9. traduction vers le français

<fruit>: carxofa, mandarina, avellana...  
 <gramin>: avena, bambú, mill...  
 <herba>: lavanda, userda, poniol...  
 <infusio>: te, valeriana, anís...  
 <planta>: ficus, fonoll, ginesta...  
 <raim>: cabernet, picapoll, xarel·lo...

La classe <arbre> se distingue des autres par le fait de pouvoir apparaître dans les trois contextes suivants :

talar(N<sub>0</sub>:Hum; N<sub>1</sub>:Veg<arbre>)  
 abatre(N<sub>0</sub>:Hum; N<sub>1</sub>:Veg<arbre>)  
 podar(N<sub>0</sub>:Hum; N<sub>1</sub>:Veg<arbre>)

La classe des <gramin> est délimitée par des prédicats divers. Par des prédicats verbaux comme granar, mallar, trillar, batre ou par des prédicats adjectivaux comme orb.

Ainsi chaque classe dispose de son propre « faisceau définitionnel » [Mathieu-Colas, 1998] : des prédicats qui s'appliquent simultanément sur tous les éléments de la classe et seulement sur ceux-ci.

## 7.2 Classes d'objets du trait *Anl*

Parmi les substantifs ayant le trait syntactico-sémantique *Anl*, nous avons distingué 18 classes d'objets :

<amfibi>: gripau, granota...  
 <animal generic>: mascota, animaló, bèstia...  
 <aranya>: taràntula, aranya...  
 <cavall>: euga, cavall, poni...  
 <crustaci>: llagosta, cranc...  
 <fabulos>: ogre, fènix, unicorn...  
 <gat>: mixa, siamès...  
 <gos>: caniche, dàlmata...  
 <inmov>: coral, pòlip...  
 <insecte>: lluernà, paparra, eruga...  
 <larva>: cullerot, larva...  
 <lupus>: llop, guineu, guilla...  
 <microorg>: virus, àcar, ameba...



```

<molusc>:escopinya, petxina, musclo...
<ocell>:òliba, cigonya, perdiu...
<peix>:anxova, piranya, salmó...
<reptil>:colobra, caiman, escurçó...
<mammif>

```

À l'intérieur de la classe d'objets <mammif> nous avons entamé le suivant sous-classement :

```

<mammif>aigua:morsa, balena, dofí...
<mammif>herbivor:girafa, vedella, pècora...
<mammif>camelid:camell, dromedari, cavall...
<mammif>felid:iaguar, pantera, guepard...
<mammif>mustelid:marta, ermini, toixó...
<mammif>porc:garrina, verra, godall...
<mammif>primat:mico, goril·la, orangutan...
<mammif>rossegador:rata, castor, esquirol...
<mammif>terra:elefant, mofeta, gasela...

```

Par ailleurs, nous avons constaté qu'il existe toute une série de prédicats dont l'existence est commune à la plupart des *Anl*, mais dont la forme varie d'une entrée à l'autre même si celles-ci appartiennent à une même classe d'objets. Nous avons décrit certains de ces prédicats en termes de fonctions lexicales. Dans le cas des *Anl* nous avons pris en compte, en premier lieu, la fonction lexicale *Son(x)* pour représenter le cri spécifique des animaux. Par ailleurs, et comme nous l'avons déjà vu dans le chapitre §5, nous avons décrit la fonction lexicale *Par(x)* qui fait référence aux prédicats ayant le sens 'X accouche'.

**<OCELL>** A l'intérieur du trait *Anl* nous retrouvons, par exemple, la classe d'objets <ocell>. Comme nous avons déjà dit, un seul prédicat peut suffir pour distinguer une classe d'objets de tout autre ensemble. Ainsi, par exemple le prédicat *bequetejar* sert à lui seul à établir la classe d'objets <ocell>. Cependant, lors de l'établissement des classes d'objets, il est nécessaire de décrire le plus largement possible tout ce qui rend spécifique la classe. Nous donnerons, par la suite, une série de prédicats appropriés à la classe <ocell>. Bien entendu, la position argumentale de la classe par rapport au prédicat fait également partie de la description. Voici une liste de prédicats verbaux appropriés à la classe <ocell>, lorsqu'elle est en position  $N_0$  :

- Prédicats appropriés :  $N_0$  : *Anl*<ocell>
  - *aletejar/V120/N0*:*Anl*<ocell>/*Es*:*aletear*/*Fr*:*battre des ailes*

- alabatre/V016/N<sub>0</sub>:Anl<ocell>/Es :aletear/Fr:battre des ailes
- bequetejar/V105/N<sub>0</sub>:Anl<ocell>/Es :picotear/Fr:becqueter
- espicassar/V105/N<sub>0</sub>:Anl<ocell>/Es :picotear/Fr:picorer

Il existe également des prédicats appropriés qui sélectionnent la classe <ocell> en position N<sub>1</sub> :

- Prédicats appropriés : N<sub>1</sub>: Anl<ocell>
  - plomar/V105/N<sub>0</sub>:Hum/N<sub>1</sub>:Anl<ocell>/Es:desplumar/ Fr:plumer

La description d'une classe peut comprendre non seulement des éléments prédictifs mais aussi d'autres arguments sémantiquement dépendants de la classe. C'est le cas, par exemple, des parties du corps appropriées à la classe <ocell> :

- Parties du corps appropriées : N de Anl<ocell>
  - ploma/N52/Conc/pc<ocell>/Es:pluma/Fr:plume
  - bec/N01/Conc/pc<ocell>/Es:pico/Fr:bec

**<PEIX>** Une autre classe d'objets à l'intérieur du trait *Anl* est la classe <peix>. Cette classe se caractérise parce qu'elle est sélectionnée en position N<sub>0</sub> par les prédicats suivants<sup>54</sup> :

- Prédicats appropriés : N<sub>0</sub>: Anl<peix>
  - aletejar/N<sub>0</sub>:Anl<peix>/Es:aletear/Fr:battre les nageoires
  - nedar/N<sub>0</sub>:Anl<peix>/Es:nadar/Fr:nager

Il s'avère que la plupart des prédicats appropriés à la classe <peix> appartiennent au domaine de la *pêche*. Ces prédicats ont en commun le suivant schéma d'arguments :

prédicat du domaine de la pêche /N<sub>0</sub>:Hum/N<sub>1</sub>:Anl<peix>

En effet, ces prédicats sélectionnent la classe <peix> en position N<sub>1</sub>, voici quelques exemples :

- Prédicats appropriés : N<sub>1</sub>: Anl<peix>
  - pescar/V119/N<sub>0</sub>:Hum/N<sub>1</sub>:Anl<peix>/Es:pescar/Fr:pêcher

---

54. Notons que la forme *aletejar* est présente aussi bien dans la classe <ocell> que dans la classe <peix>. Il s'agit en réalité de deux unités lexicales différentes, comme le montrent d'ailleurs les équivalents de traduction :

- aletejar/V120/N<sub>0</sub>:Anl<ocell>/Es:aletear/Fr:battre des ailes
- aletejar/V120/N<sub>0</sub>:Anl<peix>/Es:aletear/Fr:battre les nageoires

Ceci constitue un exemple où la description des prédicats par classes d'objets nous permet de lever l'ambiguïté des formes polysémiques.

Finalement, comme dans la classe précédente, la classe <peix> a aussi des parties du corps appropriées à la classe (N de Anl<peix>) :

- Parties du corps appropriées : N de Anl<peix>
  - aleta/N52/Conc/pc<peix>/Es:pluma/Fr:plume
  - escama/N52/Conc/pc<peix>/Es:escama/Fr:écaille
  - espina/N52/Conc/pc<peix>/Es:espina/Fr:épine

### 7.3 Classes d'objets du trait *Conc*

Parmi les substantifs ayant le trait syntactico-sémantique Conc nous avons distingué pour le moment 59 classes d'objets :

```

<aliment>: espagueti, truita, pizza...
<Anl_pc>: cua, trompa, antena...
<aparell>: ordinador, aspersor, estufa...
<arma>: pistola, fusell, ballesta...
<beguda>: vi, licor, vermut...
<calçat>: sabata, espardenya, avarca...
<cinta>: passamà, veta, banda...
<ciri>: ciri, espelma, candela...
<conducte>: xemeneia, canonada, alcaduf...
<contenedor>: caixa, envàs, joier...
<corda>: sogà, corda, bagueta...
<dispositiu>: alarma, fre, motor...
<document>: carta, examen, article...
<eina>: tirabuixó, martell, falç...
<element arqui>: pilar, dentell, arquivolta...
<element quimic>: oxigen, età, fluor...
<tumor>: gra, butllofa, herpes...
<estri>: alfil, arquet, esquí...
<ferida>: gandalla, ferida, xiribec...
<fil>: brodat, ris, xap...
<foguera>: alimara, foguera...
<forat>: esvoranc, bretxa, esquerdà...
<hum_pc>: nas, orella, entrecella...
<ingredient>: allioli, bacó, arròs...

```

<instrument music>: flauta, violí, guitarra...  
 <instrument>: xiulet, termòmetre, cullera...  
 <joc>: billar, puzzle, escac...  
 <joguina>: bola, baldufa, diàbolo...  
 <joia>: solitari, perla, diamant...  
 <llibre>: diccionari, calendari, agenda...  
 <maquina>: càmera, afaitadora, assecador...  
 <material>: làtex, pintura, moqueta...  
 <mecanisme>: embragatge, alçavidres, antirobatori...  
 <medicament>: pastilla, aspirina, antibiòtic...  
 <mineral>: ametista, guix, àgata...  
 <monument>: talaiot, obelisc, monòlit...  
 <mt>: helicòpter, vaixell, cotxe...  
 <nus>: barrilet, бага, llaç...  
 <objecte>: present, amulet, regal...  
 <paret>: muralla, ampit, cloenda...  
 <peça>: argolla, clau, nansa...  
 <plec>: arruga, plec, sacsó...  
 <projectil>: bala, míssil, torpede...  
 <sac>: alforja, barriola, banasta...  
 <seient>: cadira, banc, banqueta...  
 <substancia>: benzina, pus, verí...  
 <taca>: pinta, gallada, lentigen...  
 <teixit>: cotó, llana, seda...  
 <tela senyal>: bandera, vela, bandinella...  
 <trampa>: llosa, envescada, parany...  
 <tumor>: abscess, quist, condroma...  
 <vas>: ampolla, olla, gerra...  
 <veg\_pc>: tronc, branca, tija...  
 <vestit capa>: capilot, toga, morrió...  
 <vestit part>: màniga, solapa, cinyell...  
 <vestit gorro>: txapela, còfia, birret...  
 <vestit\_supl>: guant, faixa, calça...  
 <vestit>: xandall, pijama, faldilla...

**<VESTIT>** Tout comme les classes vues jusqu'à présent, parmi les concrets, la classe d'objets <vestit> présente des prédicats qui lui sont appropriés et qui nous servent en même temps à définir la classe :

- Prédicats appropriés : N<sub>0</sub>: Conc<vestit>
  - esquinar/ Vp/ N<sub>0</sub>:Conc<vestit>/Es:desgarrar/Fr:déchirer
- Prédicats appropriés : N<sub>1</sub>: Conc<vestit>
  - afiganyar/ N<sub>0</sub>:Hum/N<sub>1</sub>:Conc<vestit>/Es:arrugar/Fr:froisser
  - atavellar/ N<sub>0</sub>:Hum/N<sub>1</sub>:Conc<vestit>/Es:doblar/Fr:plier
  - planxar/N<sub>0</sub>:Hum/N<sub>1</sub>:Conc<vestit>/Es:planchar/Fr:repasser
  - posar/Vp/N<sub>0</sub>:Hum/N<sub>1</sub>:Conc<vestit>/Es:poner/Fr:mettre
  - portar/N<sub>0</sub>:Hum/N<sub>1</sub>:Conc<vestit>/Es:llevar/Fr:porter
  - plegar/N<sub>0</sub>:Hum/N<sub>1</sub>:Conc<vestit>/Es:doblar/Fr:plier

**<APARELL>** De la même façon, la classe d'objets <aparell> se caractérise parce qu'elle est sélectionnée par les prédicats ci-dessous. Voici, un exemple de prédicats qui sélectionnent cette classe en position N<sub>0</sub> ou en position N<sub>1</sub> :

- Prédicats appropriés : N<sub>0</sub>: Conc<aparell>
  - espatllar/Vp/N<sub>0</sub>:Conc<aparell>/Es:estropeado/Fr:abîmer
  - funcionar/N<sub>0</sub>:Conc<aparell>/Es:funcionar/Fr:fonctionner
  - rutllar/N<sub>0</sub>:Conc<aparell>/Es:funcionar/Fr:marcher
- Prédicats appropriés : N<sub>1</sub>: Conc<aparell>
  - engegar/N<sub>0</sub>:Hum/N<sub>1</sub>:Conc<aparell>/Es:encender/Fr:allumer
  - tancar/N<sub>0</sub>:Hum/N<sub>1</sub>:Conc<aparell>/Es:apagar/Fr:éteindre

## 8 La plausibilité

Il existe dans la langue des mots plus « essentiels » que d'autres. Des mots que nous pouvons utiliser tous les jours, comme par exemple *cuina, pa, ordinador...* ou bien des mots que nous n'utilisons jamais ou presque jamais comme par exemple *maltol, forcípula, gerga*.

Un dictionnaire doit rendre compte de l'existence de ces couches lexicales qui vont des lexies les plus courantes aux lexies les moins courantes.

Il existe de nombreuses méthodes statistiques capables de calculer l'importance relative de chaque lexie par rapport à un ensemble fini de lexies. Ce calcul se base le plus souvent sur la fréquence d'apparition de chaque élément du lexique dans un ensemble de textes. Pourtant, il se peut que les résultats obtenus en appliquant ces méthodes ne coïncident pas avec l'intuition linguistique des locuteurs.

Pour élaborer notre dictionnaire, nous avons retenu l'approche proposée par M. Garrigues [Garrigues, 1992] qui est totalement opposée aux méthodes statistiques. Selon cette approche, les différentes couches du lexique sont extraites à partir des compétences des locuteurs. Elle est fondée sur deux critères, la *représentation* et la *plausibilité*. La *représentation* rend compte du degré de reconnaissance d'un mot-forme par le locuteur. Elle est décrite formellement par un codage binaire :

- R0 : pour coder les mots non reconnus
- R1 : pour coder les mots connus ou reconnus

Le deuxième critère, la *plausibilité*, s'applique uniquement sur l'ensemble R1. La plausibilité sert à « *évaluer intuitivement les possibilités d'apparition du mot dans l'ensemble des énoncés écrits ou oraux auxquels nous sommes habituellement exposée* ». Le tri se fait à partir de 3 codes qui indiquent le degré de plausibilité :

- P1 : plausibilité haute
- P2 : plausibilité basse
- P3 : plausibilité faible

Le croisement de ces deux critères permet de réaliser une partition du lexique en trois couches :

- couche 1 : R1(P1)
- couche 2 : R1(P2)
- couche 3 : R0 et R1(P3)

Ainsi, dans notre dictionnaire, chaque entrée est accompagnée d'un code 1, 2 ou 3 correspondant à chacune de ces couches. Nous avons codé 14 362 entrées en tout (37% du dictionnaire). Ce codage a été réalisé en 4 passages espacés d'environ 3 mois entre l'un et l'autre. Sur cette sélection, la couche 1 représente environ 53% soit 7 606 mots, la couche 2 représente 26%, soit 3 751 mots et finalement, les 21% restants, correspondent à la couche 3 avec 3 003 mots. Cette différence entre la première couche et les deux dernières est sûrement due à l'origine des entrées de notre dictionnaire ; nous rappelons brièvement que la première liste de mots a été puisée d'un fichier texte de pages web en catalan, notamment celles des journaux catalans.

Voici quelques exemples :

ID	LEMME	CAT	CODE	TRAIT	ES	FR	P
3415	astor	Nm	N01	Anl	azor	autour	3
6728	carilló	Nm	N09	Conc	carillón	carillon	3
4520	basilisc	Nm	N04	Anl	basilisco	basilic	3
7212	celleta	Nf	N52	Conc	cejilla	sillet	2
35719	tarongerar	Nm	N01	Loc/VegColl	naranjal	orangerie	2
177	abscés	Nm	N19	Conc	absceso	abcès	2
4155	baldufa	Nf	N52	Conc	peonza	toupie	1
23617	llet	Nf	N51	Conc	leche	lait	1
36356	tija	Nf	N55	Conc	tallo	tige	1

TAB. 8.1: *Plausibilité*

Cette distribution en couches devrait déjà nous permettre de faire face, bien que très sommairement, aux problèmes dus à l'homonymie et à la polysémie .

Lors de l'application de notre dictionnaire sur un corpus, le logiciel NooJ nous permet d'établir différentes priorités aux dictionnaires. Cela veut dire que, dès qu'un mot est trouvé dans le dictionnaire avec la plus haute priorité, les autres dictionnaires ayant une priorité inférieure ne sont plus pris en compte. Si nous divisons notre dictionnaire en trois parties correspondant à chacune des couches nous pouvons ainsi attribuer la priorité la plus haute au dictionnaire ayant les termes de plausibilité 1, une priorité inférieure pour le dictionnaire ayant les termes de plausibilité 2 et finalement la priorité la plus basse pour le dictionnaire des termes de plausibilité 3.

ID	LEMME	CAT	CODE	TRAIT	CO	ES	FR	P
3549	atles	Nm	N24	Conc	<llibre>	atlas	atlas	1
3548	atles	Nm	N24	Conc	<hum_pc>	atlas	atlas	3
4155	baldufa	Nf	N52	Conc	<joguina>	peonza	toupie	1
4156	baldufa	Nf	N52	Conc	<instrument>	taladro	perceuse	3
5686	buc	Nm	N01	Conc	<mt>	buque	navire	1
5687	buc	Nm	N01	Conc	part vaixell	casco	coque	3

TAB. 8.2: *La plausibilité dans des cas d'ambiguïté*

Comme nous le montre le tableau 8.2, lors de la reconnaissance, il ne pourrait envisager que les entrées ayant la plausibilité 1. Ce système est loin d'être infallible et même s'il permet de résoudre certaines ambiguïtés, il s'avère trop contraignant. En effet, il est préférable d'utiliser la plausibilité uniquement lorsque les grammaires locales ne sont pas capables de désambiguïser. Ainsi, lors de l'analyse syntaxique ou « parsing » le champ de plausibilité peut déterminer de quelle lemme il s'agit.





## 9 La traduction

Le dernier champ de notre dictionnaire correspond à la traduction. Les champs vus jusqu'à présent (catégories grammaticales, codes de flexion, traits syntactico-sémantiques, classes d'objets et plausibilité) font référence à la description des propriétés du lemme. Par contre, ce dernier champ a une toute autre fonction ; il sert à relier des dictionnaires en différentes langues, en vue de la traduction automatique. L'équivalent de traduction est une sorte de pointeur vers un autre dictionnaire où est décrite l'autre langue (L2). Évidemment, pour que la coordination soit possible, les deux dictionnaires doivent être conçus sous le même modèle et, par conséquent, il devront avoir non seulement les mêmes caractéristiques formelles, mais aussi les mêmes principes de description.

Cela étant, la correspondance formelle des dictionnaires n'implique pas une correspondance absolue entre les champs des dictionnaires. Comme le montre X. Blanco [Blanco, 1999] il peut y avoir des divergences dans la plupart des champs descriptifs :

- Il est possible qu'il y ait des différences par rapport à la forme et qu'un mot typographiquement simple en catalan corresponde à un mot composé en français.

motxilla/Nf/N52/Conc/<sac>/P1/Es:mochila/Fr:sac à dos

mànega/Nf/N56/Conc/<instrument>/P1/Es:manguera/Fr:tuyau  
d'arrosage

- Il peut également y avoir des différences de genre et du type de flexion en nombre entre l'entrée et son équivalent de traduction.

dubte/Nm/N01/Abst/P1/Es:duda/Fr:doute

llebra/Nf/N52/Anl/P1/Es:liebre/Fr:lièvre

cap/Nm/N01/Conc/<pc>/P1/Es:cabeza/Fr:tête

- Il est possible, également, que l'équivalent de traduction proposé présente des variantes orthographiques. Pour éviter de surcharger le champ des équivalents de traduction, il est préférable que ce type d'information apparaisse dans la microstructure des dictionnaires. Dans notre dictionnaire

nous avons donc inclus un champ *Var* où sont décrites les variantes orthographiques de l'entrée. Ainsi, chaque variante correspond à une entrée de la macrostructure mais elles sont mises en relation par le champ *Var*. De la même façon, lorsque l'équivalent de traduction proposé a d'autres représentations formelles, celles-ci sont rendues évidentes lors de la recherche de l'entrée.

advocat/Nm/N01/Hum/P1/Var:avocat/Es:abogado/Fr:avocat  
apendre/V/V127/P1/Var:apprendre/Es:aprender/Fr:apprendre

Il faut pourtant souligner que, pour qu'il y ait une véritable coordination entre les dictionnaires, il est nécessaire qu'au moins la description des « *champs syntactico-sémantiques (traits, classes d'objets et domaines)* » [Blanco, 1999, p. 16] soient identiques dans l'un et l'autre dictionnaire.

Dans notre dictionnaire, nous avons proposé un équivalent de traduction vers l'espagnol et vers le français. 2 789 entrées sont donc accompagnées d'un équivalent de traduction espagnol et français. Notons que l'équivalent de traduction est également un moyen pour déceler l'existence d'unités lexicales différentes :

agulla/Nf/N52/Conc/1/Es:aguja/Fr:épingle  
agulla/Nf/N52/Conc/1/Es:aguja/Fr:aiguille

Parfois, il n'existe pas d'équivalent de traduction vraiment satisfaisant pour le terme d'origine ; c'est le cas des 'culture-bound terms' [Blanco, 1999] :

rosquilla/Nf/N52/Conc/1/Es:rosquilla/Fr: -  
fuet/Nm/N01/Conc/1/Es: -/ Fr: -

Dans d'autres cas, il se peut que le lemme catalan ne corresponde pas exactement à une unité lexicale de L2 mais à un équivalent ayant une valeur collocationnelle. C'est le cas, par exemple, du verbe *quedar* dont l'équivalent de traduction correspond à la fonction lexicale *Oper(rendez-vous) = prendre*; Ou du verbe *estiuejar* qui, en principe, devrait correspondre au verbe français *estiver* mais dont l'emploi est tellement rare qu'il est préférable de donner à sa place le semi-phrasème *passer l'été*. Finalement, nous avons aussi l'entrée *cadell* qui correspond elle aussi à une collocation en français (*petit de (chien + loup + ...)*).

Il faut souligner que, dans le système de dictionnaires monolingues coordonnés, l'équivalent de traduction sert uniquement de renvoi vers un autre dictionnaire ; tous les champs du dictionnaire de L2 devront être pris en compte avant de traduire, soit par le terme proposé comme équivalent de traduction, soit par un autre terme récupéré dans le dictionnaire de L2.

## **Troisième partie**

### **Le module NooJ**



## 10 Présentation du module

NooJ est un environnement de développement linguistique qui permet de construire et de gérer des dictionnaires et grammaires électroniques à large couverture, afin de formaliser différents niveaux et composantes des langues naturelles : orthographe, morphologie flexionnelle et dérivationnelle, lexique de mots simples, mots composés et expressions figées, syntaxe locale et désambiguïsation, syntaxe, sémantique et ontologies. Les descriptions formalisées peuvent ensuite être appliquées pour traiter des textes et corpus de taille importante. Une application à court terme de notre dictionnaire a donc été possible grâce à ce système de développement. En vue d'une telle application nous avons créé le module NooJ de catalan. Ce module est constitué du dictionnaire électronique du catalan ainsi que de toute une série de ressources linguistiques<sup>55</sup>.

- Le dictionnaire de catalan.
- Un conte de Mercè Rodoreda.
- Le texte de la déclaration universelle des droits de l'homme en catalan.
- Un petit corpus de phrases avec des déterminants et des pronoms pour tester la grammaire de désambiguïsation
- 12 grammaires morphologiques :
  - 2 exemples de grammaires flexionnelles.
  - 4 grammaires lexicales pour les cardinaux, les déterminants et les possessifs et deux pour les pronoms.
  - 6 grammaires pour décrire des phénomènes dérivationnels divers en catalan.
- 2 grammaires syntaxiques :
  - Un exemple très sommaire de grammaire de désambiguïsation des pronoms et des déterminants.
  - Un exemple de grammaire locales des dates en catalan.

Nous décrirons par la suite chacune de ces ressources puis, finalement, nous donnerons quelques exemples d'application du module sur des textes.

---

55. Ce module est disponible dans [www.nooj4nlp.net](http://www.nooj4nlp.net)



# 11 Les grammaires locales

Dans NooJ, le dictionnaire est complété par toute une série de grammaires électroniques servant à décrire différents phénomènes linguistiques. Ces grammaires électroniques se distribuent en deux grands blocs, comme le montre d'ailleurs le propre système de répertoires du logiciel : les grammaires lexicales et les grammaires syntaxiques.

## 11.1 Les grammaires lexicales

Parmi les grammaires lexicales nous pouvons distinguer 3 types :

- Les grammaires flexionnelles
- Les grammaires morphologiques simples
- Les grammaires morphologiques complexes<sup>56</sup>

**Les grammaires flexionnelles.** Nous avons déjà vu dans le chapitre §5 que, pour pouvoir générer automatiquement toutes les formes fléchies de notre dictionnaire, nous avons créé une collection de graphes décrivant les modèles de flexion en catalan. Ces graphes, ou le fichier équivalent que nous avons utilisé pour décrire la flexion verbale (cf. §5.3), font partie des grammaires flexionnelles<sup>57</sup>. L'une des principales caractéristiques de ces grammaires réside dans le fait qu'elles opèrent au niveau du dictionnaire. En appliquant les grammaires flexionnelles au dictionnaire nous obtenons un dictionnaire parallèle ayant toutes les formes fléchies ou dérivées<sup>58</sup> regroupées par rapport à l'entrée de base. Ce regroupement est extrêmement utile, car au moment d'analyser un texte, nous pouvons effectuer des recherches soit par mot-forme, soit par ensemble de formes fléchies correspondant à un lemme.

---

56. Grammaires morphologiques avec des contraintes.

57. Ce type de grammaires sont en réalité des transducteurs à états finis (FST).

58. Ce fichier correspondrait au DELAF avec l'ancienne version Intex (cf. §2.1).



**Les grammaires morphologiques simples.** Le module NooJ du catalan est également constitué de grammaires morphologiques. Ces grammaires peuvent reconnaître des ensembles précis de mots-formes et leur associer des informations diverses<sup>59</sup>. Un exemple de grammaire morphologique simple serait le graphe des chiffres romains ou des chiffres en toutes lettres. Il va de soi qu'un dictionnaire ne peut contenir tous les chiffres de la langue écrits en toutes lettres et, pourtant, il faut bien que ces formes soient reconnues dans un texte. C'est précisément à l'aide des grammaires morphologiques que tous ces ensembles de mots vont être décrits. Une caractéristique importante de ces grammaires est qu'elles sont appliquées directement sur les textes, elles sont donc complémentaires au dictionnaire.

Dans notre module du catalan, nous avons eu recours à ce type de grammaires pour rendre compte de certains éléments propres à la linéarisation comme, par exemple, les articles, les démonstratifs, etc. Nous disposons à présent d'une grammaire pour les démonstratifs, les possessifs, les déterminants, les pronoms (personnels, faibles, etc.) et nous avons adapté également le graphe des chiffres romains élaboré par le Dr. M. Silberztein pour le module NooJ du français.

Dans la figure 11.1, nous pouvons voir l'un des cinq graphes qui constituent la grammaire morphologique des pronoms. Ce graphe reconnaît tous les pronoms à la troisième personne postposés au verbe. Chaque nœud correspond à un pronom qui est, à son tour, associé aux informations morphologiques qui lui correspondent.

Une autre caractéristique des grammaires morphologiques simples est qu'elles sont capables de reconnaître différentes parties à l'intérieur d'un mot, ce qui permet de représenter certains phénomènes dérivationnels. Comme, par exemple, des mots ayant les préfixes *ex*, *anti*...

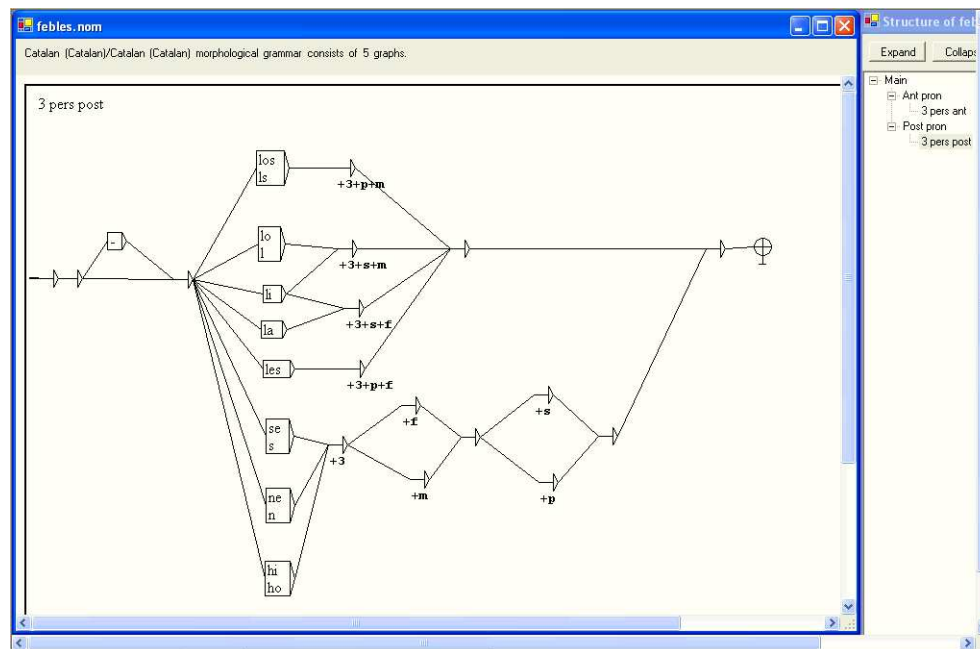
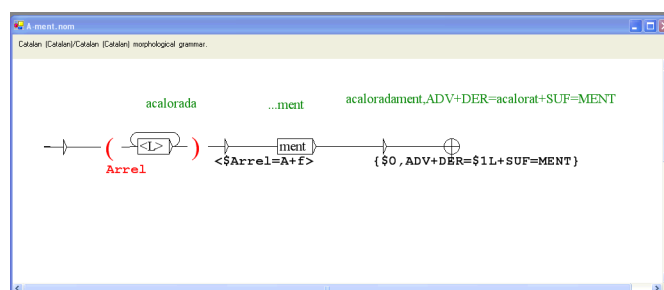
Cependant les processus dérivationnels sont souvent plus complexes et l'utilisation de variables et de contraintes lexicales s'impose.

**Les grammaires morphologiques complexes.** L'inclusion de variables et de contraintes rendent ce type de grammaires extrêmement puissantes pour la représentation de phénomènes dérivationnels. Avec les contraintes nous pouvons imposer non seulement qu'une partie du mot analysé apparaisse dans la macrostructure du dictionnaire, mais imposer aussi des conditions particulières pouvant faire référence à toute information contenue dans la microstructure. Ainsi, avec ces outils nous pouvons décrire avec précision la différence entre des mots déverbaux nominaux tels que *enrichissement*, *ensorrement*... et des mots déadjectivaux adverbiaux tels que *acaloradament*, *sorollosament*...

Dans le premier nœud de la figure 11.2, NooJ cherche lettre après lettre jusqu'à ce qu'il retrouve exactement la forme suffixale *-ment*. La séquence de

---

59. Ce sont des transducteurs ou RTNs [Silberztein, 2006].

FIG. 11.1: *Grammaire des pronoms*FIG. 11.2: *Grammaire des adverbes en -ment*

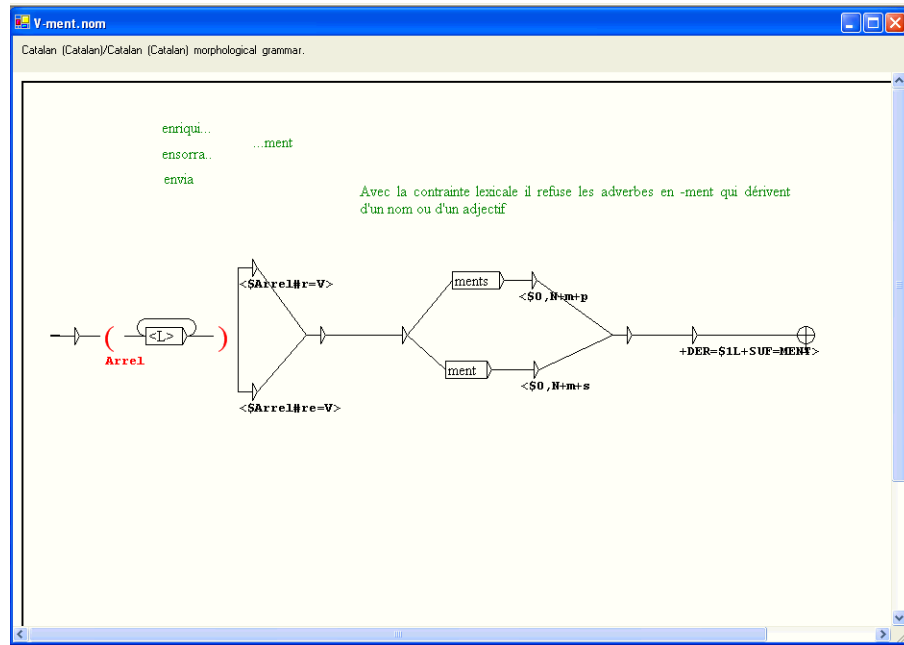


FIG. 11.3: Grammaire des substantifs en -ment

lettres retrouvée avant le suffixe est gardée en mémoire sous forme d'une variable appelée (arbitrairement) *Arrel*. Avant de retrouver la forme suffixale (*-ment*), une contrainte s'impose : il faut que la séquence de lettres de la variable « *Arrel* » corresponde à une entrée du dictionnaire étiquetée *Adjectif* et à la forme féminine. Finalement, nous lui demandons de donner en sortie le mot forme analysé (\$0), la catégorie grammaticale (ADV) et l'adjectif dont il dérive (\$1L) suivi du suffixe (*ment*).

Le graphe de la figure 11.3 est semblable ; dans ce cas, dans la contrainte nous précisons qu'il faut rajouter la lettre *r* à la variable « *Arrel* » pour qu'il puisse retrouver l'entrée dans le dictionnaire. Ici la contrainte impose qu'il s'agisse d'un mot déverbal. Dans les deux nœuds suivants nous signalons la catégorie grammaticale et les informations morphologiques du genre et du nombre. Ainsi, en sortie, il nous donne le mot cible (\$0), la catégorie grammaticale, le genre, le nombre, le verbe dont il dérive (\$1L) suivi du suffixe (*-ment*).

Un autre exemple de grammaire morphologique correspond à la figure 11.4. Cette grammaire, composée de deux graphes, représente deux types d'opérations. D'un côté, la reconnaissance des substantifs déverbaux ayant le suffixe *-ció* (cf. Fig. 11.4). D'un autre côté, l'ajout du préfixe *re-* à la forme précédemment dérivée (cf. Fig. 11.5). Pour rendre compte de la récursivité morphologique, ces grammaires, contrairement aux flexionnelles, admettent des combinaisons de

graphes à l'intérieur d'une même grammaire.

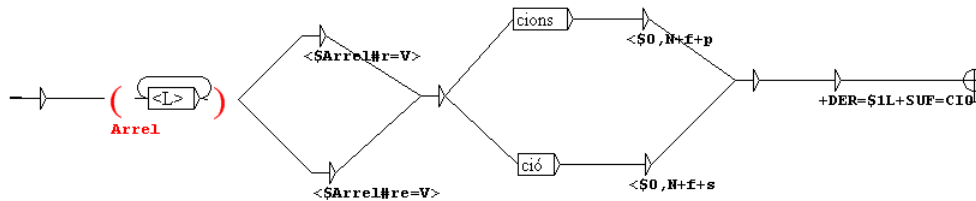


FIG. 11.4: *Sous-graphe : cio*

Dans le module NooJ du catalan nous avons inclus 14 grammaires lexicales, parmi lesquelles 6 représentent des phénomènes dérivationnels divers :

- Les adjectifs déverbaux en *-tiu*: *decoratiu*, *significatiu*...
- Les adjectifs déverbaux en *-ble*: *edificable*, *recuperable*...
- Les substantifs déverbaux en *-ció*: *construcció*, *comparació*...
- Leur variante préfixée avec *re-*: *reconstrucció*, *reinversió*...
- Les substantifs déverbaux avec le suffixe *-ment*: *seguiment*, *pensament*...
- Les verbes avec le préfixe *re-*: *reeditar*, *reescriure*...
- Les adverbes déadjectivaux avec le suffixe *-ment*: *equivocadament*, *desesperadament*...

En vue de l'utilisation de ces grammaires sur des *corpora*, l'une de leurs caractéristiques est la possibilité de donner des priorités aux différentes grammaires. En effet, nous pouvons faire en sorte que ces grammaires (ou seulement quelques unes d'entre elles) ne soient appelées qu'en dernier ressort. Cela présente un très grand intérêt car il est possible de décrire des phénomènes dérivationnels concaténés. Par exemple, dans notre module du catalan nous avons attribué à « la grammaire des adjectifs déverbaux en *-tiu* » une priorité supérieure à celle des « adverbes déadjectivaux en *-ment* ». De cette façon, des mots comme *equitativament*, *aproximativament*... peuvent être reconnus par nos grammaires. En effet, avant d'appliquer la grammaire des adverbes en *-ment* (cf. Fig. 11.2), il faut avoir reconnu le mot *equitatiu*. Si ce mot n'apparaît pas dans le dictionnaire nous aurons besoin de « la grammaire des adjectifs déverbaux en *-tiu* » pour pouvoir le reconnaître. C'est donc sur cette dernière base, c'est à dire du mot reconnu précédemment, que sera appliquée « la grammaire

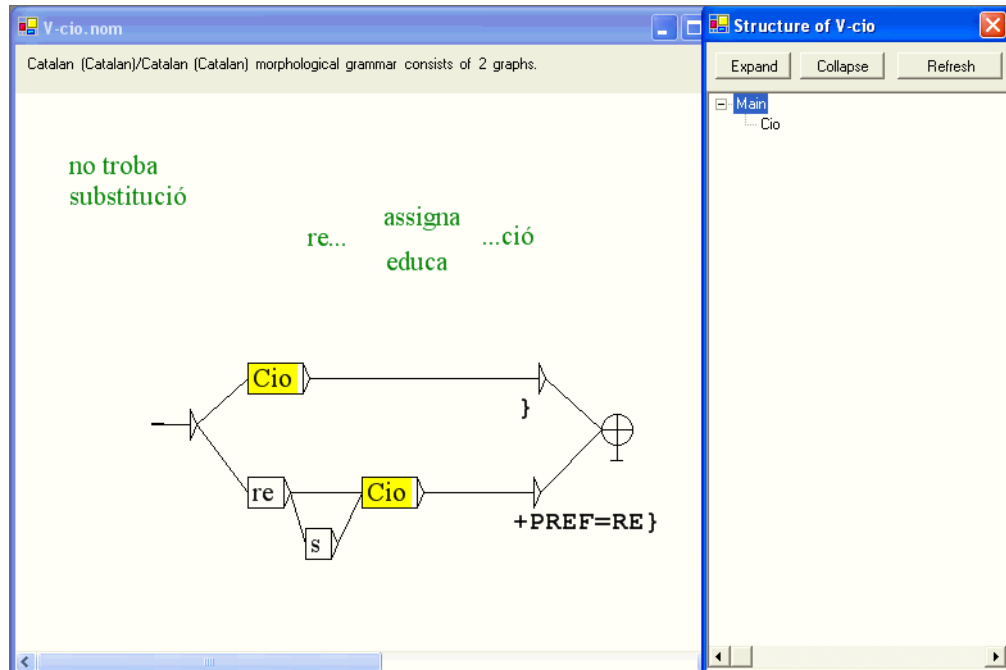


FIG. 11.5: Grammaire des substantifs en -ció

des adverbes déadjectivaux en *-ment* » pour finalement reconnaître les formes équitativement, aproximativement, etc.

Comme nous venons de le voir, les grammaires lexicales jouent un rôle complémentaire et nécessaire lors de l'élaboration de dictionnaires électroniques. Elles nous permettent, d'une part, de séparer les unités lexicales des éléments linguistiques propres à la linéarisation, tels que les pronoms, les articles... D'autre part, ces grammaires nous permettent de représenter aisément les phénomènes de dérivation régulière par composition, sans avoir à augmenter considérablement la taille du dictionnaire. Par ailleurs, ces grammaires nous donnent la possibilité de prévoir certains phénomènes de création lexicale et d'être en mesure, non seulement de reconnaître les nouveaux termes, mais aussi de leur associer les informations correspondantes.

## 11.2 Les grammaires syntaxiques

Le module NooJ du catalan intègre également un autre type de grammaires, les grammaires syntaxiques. Ces grammaires agissent au niveau des suites de mots-formes et non pas au niveau des caractères comme les grammaires

morphologiques vues dans la section §11.1. Les grammaires syntaxiques peuvent donc servir à la désambiguïsation des mots en contexte, à la reconnaissance, à l'indexation d'expressions dans un texte donné ainsi qu'à l'analyse sémantico-syntaxique des textes, à la génération automatique de paraphrases ou à la traduction<sup>60</sup>.

Nous présenterons par la suite une grammaire locale des expressions de « date » ainsi qu'un graphe de désambiguïsation des pronoms et des articles et, finalement, l'application de ces grammaires sur des textes.

### 11.2.1 Un exemple de grammaire locale

En nous basant sur le module NooJ du français, nous avons réalisé pour le module du catalan une grammaire locale des dates. Elle est censée reconnaître toute séquence linguistique comprenant une indication quelconque d'heure et/ou de date.

L'un des grands avantages des grammaires sous forme d'automates est qu'elles permettent de factoriser de façon significative un nombre d'expressions difficile à traiter sous forme de listes. Notre grammaire locale des « dates » est composée de 20 graphes au total.

L'utilisation de sous-graphes nous permet de réutiliser certaines séquences pouvant apparaître à plusieurs reprises, sans avoir à les recopier ou à les réécrire à chaque fois. Ainsi, par exemple, le nom des mois en catalan (*mois*) et les ordinaux (*nums*) et les cardinaux (*cardinals*) en toutes lettres sont appelés à différents endroits de la grammaire.

De la même manière, nous pouvons récupérer les graphes faisant partie des grammaires morphologiques vues précédemment. C'est le cas, par exemple, du graphe des chiffres romains (*num\_rom*) dont nous avons eu besoin pour la description d'expressions telles que : *El segle III a.C.*

Les deux graphes 11.7 et 11.8 nous servent à la reconnaissance d'expressions telles que :

EXEMPLE 107 :

- 107.a : *La setmana passada*
- 107.b : *A principis de la setmana vinent*
- 107.c : *Durant els mesos de gener i febrer*
- 107.d : *Després del mes de juny*
- 107.e : *Cap a principis de segle*
- 107.f : *A mitjans del mes vinent*
- 107.g : *Durant el segle IV a.C.*

---

60. Pour ce qui est de notre module, pour l'instant, ces deux dernières fonctionnalités n'ont pas été appliquées au catalan.

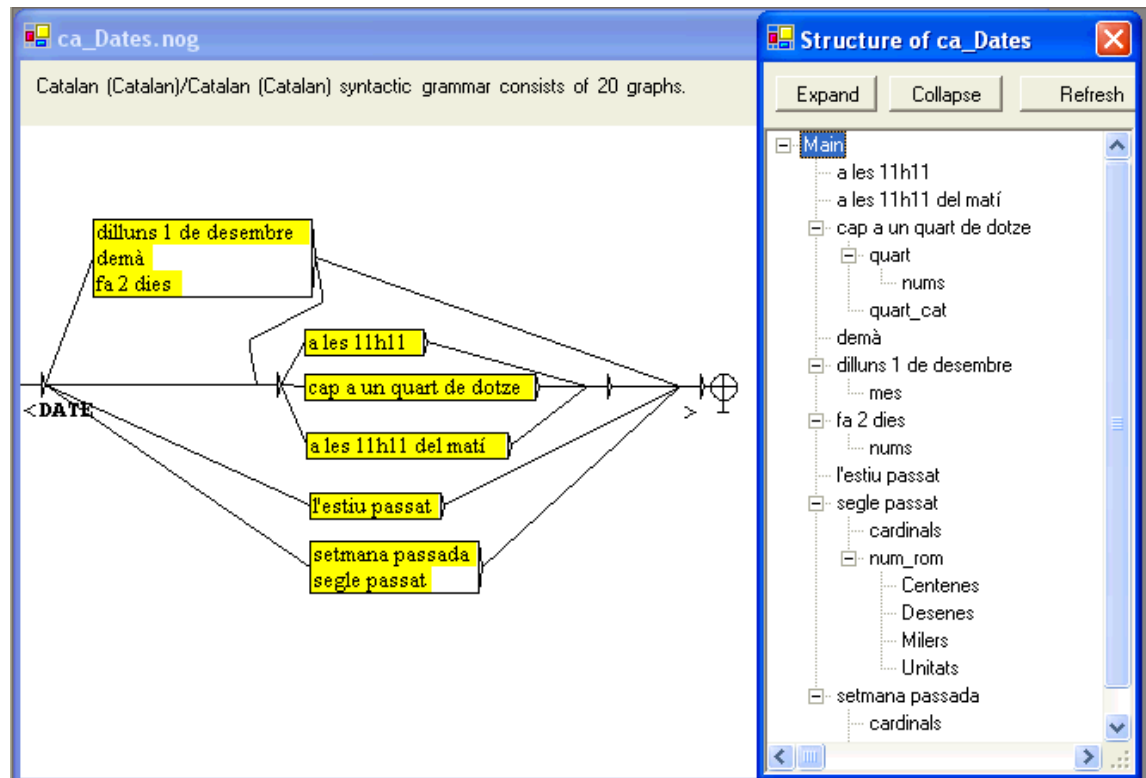


FIG. 11.6: Grammaire locale des dates : MAIN

107.h : *Després de l'any 1714*

EXEMPLE 108 :

108.a : *Durant el mes de setembre de l'any 2006*

108.b : *La darrera setmana del segle XX*

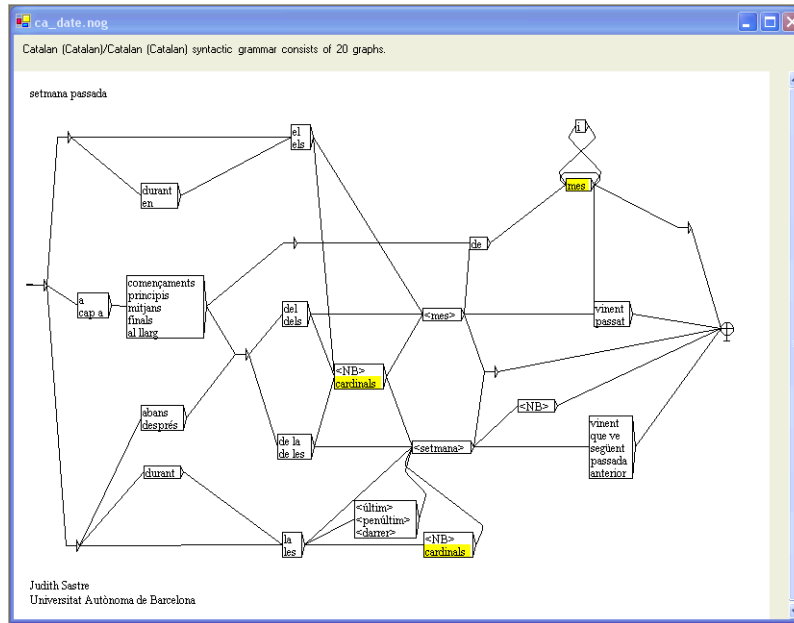


FIG. 11.7: Grammaire locale des dates : la setmana passada

Le graphe de la figure 11.7 est imbriqué dans le graphe de la figure 11.8 afin de pouvoir reconnaître des expressions comme celles des exemples [108] où des séquences appartenant aux deux graphes se succèdent.

Le graphe de la figure 11.9 reconnaît des expressions telles que :

EXEMPLE 109 :

109.a : *Diumenge vinent*

109.b : *Divendres 19 de juliol de 1996*

109.c : *El proper dilluns 23 d'abril de 2007*

109.d : *El 7 de setembre de 2003*

109.e : *Dimecres, 13 de juny de 1979*

Bien que nous nous placions dans une perspective de reconnaissance plutôt que de génération, nous avons toutefois essayé de tenir compte de l'existence de phrases invraisemblables comme celles de l'exemple [110]. Pour exclure





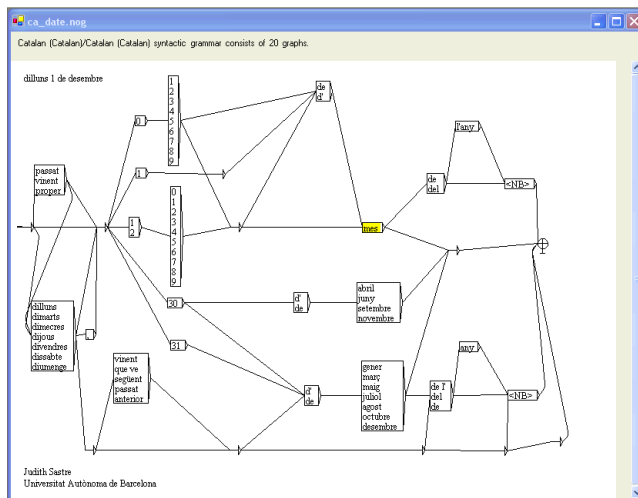
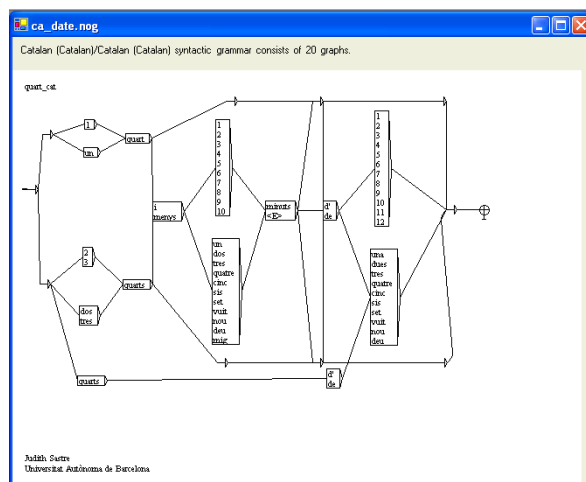


FIG. 11.9: *Grammaire locale des dates : dilluns 1 de desembre*

FIG. 11.10: *Grammaire locale des dates : A quarts de set*

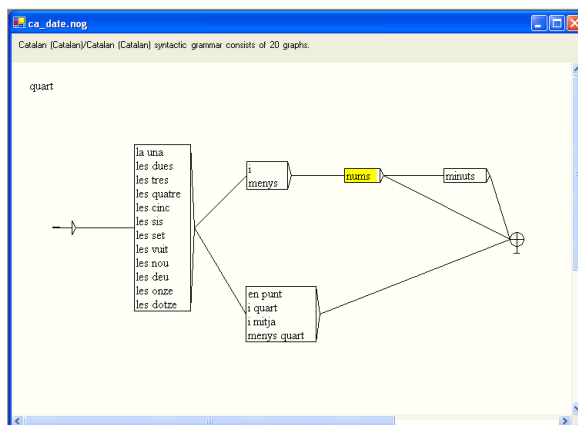


FIG. 11.11: *Grammaire locale des dates : A les dues en punt*

nombreux phénomènes profitant toujours d'un environnement agréable et simple pour le linguiste. Cependant, il existe de grandes difficultés lors de la création de graphes à cause de l'énorme quantité de données à accumuler. Il est indispensable que le linguiste dispose d'une grande rigueur et d'une bonne organisation pour éviter que les données deviennent illisibles. Il est important de pouvoir interpréter rapidement toutes les données de façon à ce qu'elles puissent être modifiées ou complétées à tout moment.

### 11.2.2 Un exemple de grammaire de désambiguïsation

L'un des problèmes majeurs lors de l'analyse de textes et de la description formelle de la langue est sans doute l'ambiguïté des formes. Chaque mot de la langue peut être interprété en moyenne de deux façons différentes. Pour une formalisation de la langue il faut obligatoirement envisager des techniques pour la désambiguïsation. Avec NooJ, la levée d'ambiguïtés peut se faire en créant des règles linguistiques qui font intervenir le ou les mots cibles en contexte. Ces grammaires sont chargées d'associer à chaque mot du texte une étiquette qui déterminera de quelle unité lexicale il s'agit. A titre d'exemple, nous avons réalisé une grammaire de désambiguïsation des pronoms et des articles. Dans les exemples [112] nous pouvons voir de nombreux contextes où la forme de l'article et du pronom à la troisième personne coïncident :

EXEMPLE 112 :

112.a : *Estudia-la bé, la lliçó*

112.b : *Ja l'ompliré jo, l'ampolla*

112.c : *Els diré la veritat*

112.d : *Els nens riuen*

Cependant, la position des pronoms et celle des articles sont assez différentes. Nous pouvons ainsi distinguer une unité de l'autre en décrivant leurs contextes d'apparition. En effet, un article sera toujours antéposé à un nom et s'accorde en genre et en nombre avec celui-ci (cf. Fig. 11.12). Par contre les pronoms à la troisième personne seront antéposés ou postposés à un verbe conjugué (cf. Fig. 11.13).

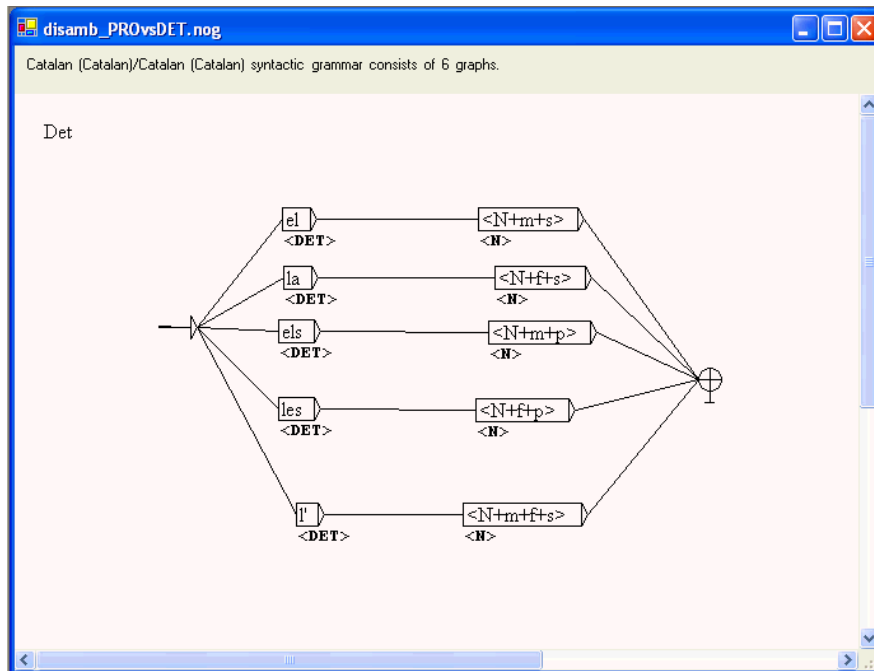


FIG. 11.12: Grammaire de désambiguïsation : les articles

En tenant compte de ces deux différences concernant leur syntaxe, nous pouvons déjà commencer l'élaboration d'une grammaire de désambiguïsation comme celle de la figure 11.13. D'après cette grammaire, les formes *le*, *la*, *els*, *les*, *l'* seront étiquetées *Article* si elles s'accordent en genre et en nombre avec le nom auquel elles précèdent (cf. fig. 11.12). D'autre part, ces mêmes formes seront étiquetées *Pronom* si, par exemple, elles sont suivies d'un verbe conjugué (cf. fig. 11.13). Il est évident que cette grammaire ne décrit que très sommairement les contextes où l'on peut trouver un article ou un pronom et dans de nombreux cas l'ambiguïté demeurera. Par exemple, dès qu'il y aura des insertions entre l'article et le nom les séquences ne seront pas reconnues par cette grammaire :

EXEMPLE 113 : *La desigual distribució de la riquesa i el diferent*

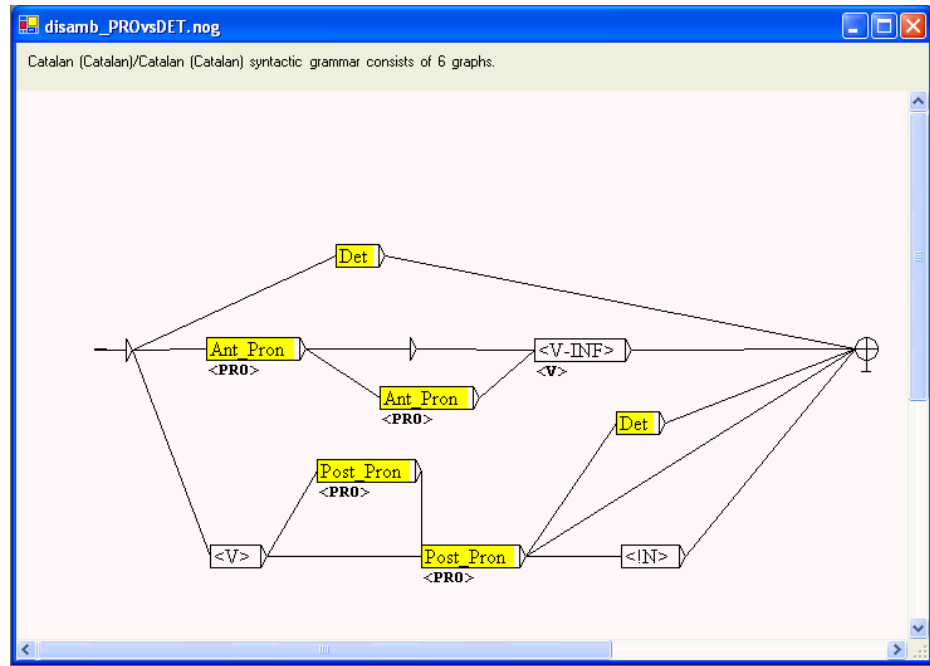


FIG. 11.13: Grammaire de désambiguïsation : Les pronoms vs les articles

*creixement i desenvolupament econòmic.*

Dans l'exemple ci-dessus, l'adjectif placé entre l'article et le nom ne sera pas reconnu et lors de l'analyse de cette phrase l'analyseur nous proposera aussi bien un article qu'un pronom. Pourtant, bien que cette grammaire ne recouvre pas tous les contextes possibles, elle réussit déjà à reconnaître et à désambiguïser certaines séquences, comme celles de [112] et [114].

EXEMPLE 114 :

114.a : *Obre el rebost i treu-ne les postres*

114.b : *Ensenya-li l'os*

114.c : *Els exposaràs la situació?*

114.d : *L'haviem de portar demà*

114.e : *L'àvia no el sap, aquest conte*

Les exemples de 112 et 114 sont reconnus et désambiguïsés par cette grammaire. Il existe toutefois des cas où la grammaire présentée (fig. 11.13) reconnaît la séquence mais ne peut pas désambiguïser. C'est le cas des exemples ci-dessous :

EXEMPLE 115 :

115.a : *I les noies què les veus?*

*115.b : Les hauria de veure, les roses*

Dans ces exemples, non seulement la forme *les* est ambiguë, les mots *veus* et *roses* le sont aussi. En effet *veus* peut être soit un nom (féminin, pluriel) soit l'une des formes conjuguées du verbe *veure*. Il en est de même avec *roses*, qui peut aussi bien être un nom qu'une forme du verbe *rosar*. Notre grammaire de désambiguïsation n'a pas suffisamment de critères dans ces contextes pour lever l'ambiguïté. Si nous avions eu un exemple comme :

EXEMPLE 116 : *Que els veus els nois?*

l'ambiguïté aurait pu être levée puisque la forme *els* ne s'accorde pas en genre avec le nom. C'est pourquoi l'analyseur pourrait désambiguïser les deux formes (*les* et *veus*) en même temps et les étiqueter l'une comme *PRO*, l'autre comme *V*. Mais pour les phrases de l'exemple 115 la grammaire de désambiguïsation s'avère incomplète. Les deux images ci-dessous montrent l'analyse de la phrase [116] avant l'application de la grammaire (cf. Fig. 11.14) et après avoir appliqué la grammaire de désambiguïsation des pronoms (cf. Fig. 11.15).

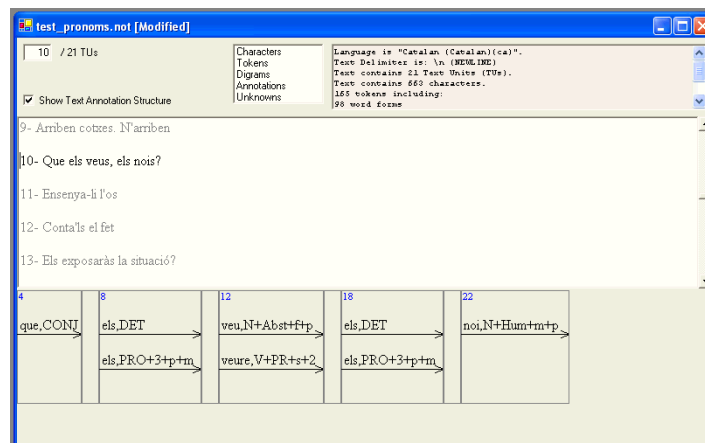


FIG. 11.14: Analyse de la phrase : *Què els veus, els nois?*

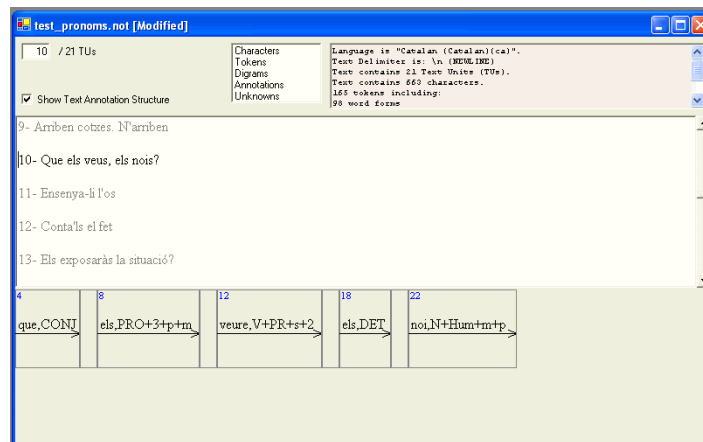


FIG. 11.15: Application du graphe de désambiguïsation

## 12 L'application à des corpora

Le but ultime du module NooJ du catalan est de pouvoir être appliqué sur des *corpora*. Puisqu'il s'agit d'une première version du dictionnaire et des grammaires, une confrontation à des *corpora* nous permettra de réviser, de corriger et de compléter toutes nos ressources. Pour évaluer l'efficacité du dictionnaire ainsi que des grammaires morphologiques, il suffit d'appliquer le dictionnaire sur des *corpora* et de vérifier ensuite la liste de mots reconnus (« Annotations ») et la liste de mots non-reconnus (« Unknowns ») générées automatiquement par NooJ. À partir de ces listes, les dictionnaires et les grammaires morphologiques peuvent être systématiquement soumis à révision perfectionnés petit à petit. C'est donc avec la liste des « Unknowns » que nous avons complété petit à petit notre dictionnaire. La liste des « Annotations » nous permet de réviser aussi bien les entrées de notre dictionnaire que celles générées par la collection de grammaires morphologiques. Les tests quantitatifs constituent l'une des méthodes qui permettent d'évaluer l'efficacité des grammaires. On réalise une analyse manuelle d'un corpus de taille moyenne et ensuite on compare les résultats avec ceux obtenus par l'application des grammaires. Trois critères seront pris en compte : le bruit, le silence et la reconnaissance partielle. Le *bruit* est la quantité d'occurrences incorrectes reconnues par la grammaire. Le *silence* est la quantité d'occurrences pertinentes non reconnues par la grammaire. La reconnaissance partielle est la quantité d'occurrences dont seulement une partie a été reconnue. En appliquant la grammaire des « dates » sur un corpus d'un fichier de texte de près d'1 Mo, nous obtenons les résultats suivants :

TOTAL	RECONNUES	BRUIT	SILENCE	PARTIELLE
320	331	30	19	64

TAB. 12.1: Reconnaissance de la grammaire locale des « dates »

Après avoir réalisé la lecture du texte nous avons relevé un total de 320 occurrences qui expriment, soit l'heure, soit une date du calendrier, soit un moment de la journée. Or, parmi le nombre total d'occurrences reconnues par



notre grammaire, il y a en tout 30 occurrences qui ne sont pas pertinentes. Parmi celles-ci, nous pouvons trouver des expressions comme :

EXEMPLE 117 :

117.a : *una dona d'uns ... 24 anys/<DATE>...*

117.b : *per ... segona/<DATE>... vegada consecutiva*

117.c : *... un ésser/<DATE>... humà*

117.d : *ha renunciat... a les 3/<DATE>... 947 accions*

Plus de la moitié des occurrences « bruit » retrouvées par la grammaire ressemblent au premier exemple de [117], qui fait référence à l'âge. Il s'agit là d'un manque de précision du graphe ; il faudrait lister les éléments susceptibles d'apparaître juste avant ce type de constructions ou, plutôt, les éléments qui ne vont pas donner une expression de temps comme celles que nous recherchons ((tenir) 24 anys; (amb) 24 anys; (de) 24 anys; (d'uns) 24 anys...). Dans le deuxième de ces exemples, il s'agit à nouveau d'un défaut de notre grammaire, car nous n'avons pas précisé que <segon> doit correspondre à un substantif et non pas à un adjectif ; il en va de même pour le troisième exemple, où nous avons décrit trop naïvement le mot <era> sans remarquer que cela correspond à l'une des formes conjuguées du verbe *ser*. Puisque nous l'avons insérée entre les symboles (< >) la grammaire prendra en compte toutes les formes fléchies du verbe *ser*, il faut donc préciser qu'il s'agit uniquement du substantif. En ce qui concerne le dernier exemple, nous croyons que le problème de reconnaissance est dû au caractère « . », plutôt qu'à un manque de précision de la grammaire. Par rapport au silence, voici quelques séquences non reconnues :

EXEMPLE 118 : Silence

118.a : *ahir al matí*

118.b : *setembre del 99*

118.c : *l'estiu de 1964*

118.d : *fins al 2004*

118.e : *durant l'abril i juny*

118.f : *entre abril i juny*

118.g : *al maig*

118.h : *l'endemà*

118.i : *l'any passat*

118.j : *abril del 1994*

118.k : *a les 21.00*

De ces exemples il s'ensuit que toutes les séquences non reconnues correspondent fondamentalement à des oublis lors de l'élaboration de la grammaire. C'est en réalisant ce type d'évaluation que nous avons complété petit à petit les lacunes existantes. Par rapport aux exemples de [118] seulement les

deux derniers ne constituent pas un défaut de la grammaire, car ce sont des expressions peu normatives pour exprimer la date et l'heure. Les séquences les plus habituelles seraient : *abril de 1994* et *a les 21:00*. Malgré tout, nous avons trouvé ce type d'expressions de forme assez récurrente dans nos textes, c'est pourquoi nous croyons qu'il faudrait sûrement les inclure dans la grammaire locale des « dates ». Finalement, nous avons 64 expressions au total qui ont été partiellement reconnues. Il s'agit d'expressions comme celles des exemples [119] (les mots entre crochets sont ceux qui ont été reconnus par la grammaire) :

EXEMPLE 119 : Reconnaissance partielle

- 119.a : Fins al [29 d'abril] a les 20:00*
- 119.b : [24 de febrer] passat*
- 119.c : [dilluns] de la setmana passada*
- 119.d : entre el 1939 i [els primers anys 60]*
- 119.e : els darrers [25 anys]*
- 119.f : des de [fa dos anys]*
- 119.g : avui [dimecres]*
- 119.h : a partir del [28 de maig]*
- 119.i : [a la tardor] del 1928*
- 119.j : [dilluns] a la nit*

La plupart de ces exemples comportent des suites très courantes dans la langue. Pour une bonne reconnaissance, il faudrait donc compléter notre grammaire afin qu'elles puissent être reconnues complètement. Il est évident que, dans le corpus choisi, nous n'avons trouvé qu'une infime partie des séquences réellement représentées dans notre grammaire locale des « dates ». L'application de la grammaire sur un corpus nous permet uniquement d'évaluer et de réviser une petite partie de la grammaire construite. Cependant, et comme nous venons de le voir, c'est en appliquant les grammaires que celles-ci peuvent être corrigées et complétées de telle façon que plus nous utilisons une grammaire, plus puissante elle devient.



## **Quatrième partie**

# **Conclusions et Perspectives**



## 13 Conclusions et Perspectives

L'objectif pratique de cette recherche étant la réalisation d'un dictionnaire électronique du catalan, nous avons présenté dans ce travail les principaux mécanismes formels utilisés lors de l'élaboration d'un tel dictionnaire, ainsi que les contraintes auxquelles il faut se soumettre pour, d'une part, rendre compte des entrées d'un dictionnaire et, d'autre part, pour qu'un dictionnaire électronique puisse être intégré dans une application informatique.

Comme nous l'avons vu, notre dictionnaire inclut systématiquement divers champs qui permettent d'introduire des informations morphologiques, syntaxiques, sémantiques, diasystématiques et de traduction. Chaque niveau d'information a été systématiquement décrit sous forme de codes univoques de façon à ce que toutes les données fournies puissent être interprétées et utilisées par un ordinateur. Pour ce qui est de la morphologie, nous avons attribué à chaque entrée la catégorie grammaticale et un code faisant appel à des transducteurs pour générer automatiquement les formes fléchies correspondantes. Au niveau des propriétés distributionnelles, nous avons réalisé un premier classement en traits syntactico-sémantiques (*humain, animal, végétal, concret, abstrait, locatif et temporel*) et, pour certains arguments, nous avons complété cette description à l'aide des classes d'objets. Nous avons également attribué des codes pour représenter les différentes couches de fréquence et, finalement, nous avons donné des équivalents de traduction espagnol et français afin d'intégrer notre dictionnaire dans le système de dictionnaires du Laboratoire fLexSem.

Le résultat de notre travail est ainsi directement utilisable par des systèmes de traitement automatique des langues. Son format et la cohérence interne des données le met à la disposition de différents domaines d'application. Ainsi, par exemple, nous avons vu, dans la troisième partie de cette étude, un exemple d'application du dictionnaire dans le domaine de l'analyse de corpus.

En effet, le module Nooj du catalan permet actuellement de réaliser des recherches et d'analyser de grands corpora textuels. Il est donc désormais possible, à l'aide de notre dictionnaire, d'identifier des suites syntaxiques et/ou sémantiques des textes. D'autre part, notre dictionnaire peut être d'ores et déjà intégré comme correcteur orthographique dans un éditeur de texte.

Naturellement, certains aspects sont restés en suspens. Ainsi, il faudrait en premier lieu, décrire la totalité des substantifs de notre dictionnaire en augmentant tout particulièrement les informations sémantiques concernant les lemmes. Par ailleurs, il conviendrait de terminer la traduction de tous les lemmes du dictionnaire. Nous nous proposons également d'agrandir notre base de données. Néanmoins, comme nous l'avons signalé, notre dictionnaire, en l'état, constitue déjà une ressource utile pour de nombreuses applications.

## **Cinquième partie**

### **Annexes**





## A Tableaux récapitulatifs

TAB. A.1: *Codes de flexion des adjectifs*

<b>CODE</b>	<b>EXEMPLE</b>	
	MASCULIN	FÉMININ
A01	prim/prims	prima/primes
A02	ambigu/ambigus	ambigua/ambigües
A03	oblic/oblics	obliqua/obliqües
A04	bonic/bonics	bonica/boniques
A05	llarg/llargs	llarga/llargues
A06	flonjo/flonjos	flonja/flonges
A07	fals/falsos	falsa/falses
A08	blau/blaus	blava/blaves
A09	anàleg/anàlegs	anàloga/anàlogues
A10	escrupulós/escrupolosos	escrupolosa/escrupoloses
A11	tranquil/tranquils	tranquil·la/tranquil·les
A12	adequat/adequats	adequada/adequades
A13	antic/antics	antiga/antigues
A15	moreno/morenos	morena/morenes
A16	genuí/genuïns	genuïna/genuïnes
A17	succedani/succedània	succedani/succedànies
A19	nimi/nimis	nímia/nímies
A22	dolç/dolços	dolça/dolces
A23	ampli/amplis	àmplia/àmplies
A24	urbà/urbans	urbana/urbanes
A29	carolingi/carolingis	carolíngia/carolíngies
A30	ossi/ossis	òssia/òssies
A33	trist/trists /tristos	trista/tristes
<i>Suite page suivante</i>		

A34	lleig/lleigs/lletjos	lletja/lletges
A35	mig/migs mitjos	mitja/mitges
A36	roig/roigs/rojos	roja/roges
A39	eficaç/eficaços	eficaces
A42	ample/amples	ampla
A43	basc/bascs/bascos	basca/basques
A45	amable/amables	
A46	monoplaça/ monoplaces	
A47	tanoca/ tanoques	
A48	fratricida/fratricides	
A49	<i>invariable</i>	
A52	belga/belgues	
A53	vague/vagues	vaga/vagues
A54	naïf/naïfs	naïve/naïves
A55	maco/macros	maca/maques
A57	espès/espessos	espessa/espesses
A63	jóc/jócs	joca/joques
A64	suís/suïssos	suïssa/suïsses
A65	mus/mussos	mussa/musses
A66	superflu/superflus	supèrflua/supèrflues
A68	ambdós	ambdues
A69	encinta	encintes
A71	afí	afins
A72	aborigen	aborígen
A73	locomotor	locomotriu
A74	qualsevol	qualsevol
A75	cancerigen/cancerígens	cancerígena/cancerígenes
A76	endogen/endògens	endògena/endògenes

TAB. A.2: Codes de flexion des verbes

<b>CODE</b>	<b>EXEMPLE</b>	<b>CODE</b>	<b>EXEMPLE</b>
V001	absoldre	V066	fer
V002	aprendre	V067	tòrcer
V003	atendre	V068	vèncer
V004	carvendre	V069	voler
V005	respondre	V070	beure
V006	confondre	V071	caure
V007	dependre	V072	commoure
V008	doldre	V073	complaure
V009	fondre	V074	coure
V010	moldre	V075	cloure
V011	perdre	V076	creure
V012	prendre	V077	descloure
V013	romandre	V078	deure
V014	vendre	V079	escruiure
V015	admettre	V080	jeure
V016	batre	V081	jaure
V017	metre	V082	moure
V018	ajupir	V083	noure
V019	argüir	V084	raure
V020	bullir	V085	recloure
V021	cobrir	V086	seure
V022	collir	V087	atraure
V023	complir	V088	treure
V024	cosir	V089	veure
V025	cruixir	V090	viure
V026	descosir	V091	cabre
V027	dir	V092	rebre
V028	munyir	V093	rompre
V029	eixir	V094	riure
V030	escopir	V095	heure
V031	fugir	V096	caldre
V032	imprimer	V097	constrènyer
V033	lluir	V098	aparèixer
V034	mentir	V099	malvendre
V035	morir	V100	haver
<i>Suite page suivante</i>			

V036	obrir	V101	anar
V037	oir	V102	estar
V038	omplir	V103	esser (ser)
V039	prevenir	V104	esser (ser)
V040	pruir	V105	cantar
V041	pudir	V106	perdre
V042	reduir	V107	témer
V043	reeixir	V108	servir
V044	sentir	V109	dormir
V045	servir	V111	canviar
V046	sofrir	V112	començar
V047	sortir	V113	torçar
V048	suplir	V114	donar
V049	tenir	V115	dar
V050	tossir	V116	pregar
V051	venir	V117	enaiguar
V052	establir	V119	trencar
V053	atènyer	V120	envejar
V054	conèixer	V121	menysprear
V055	córrer	V122	esglaiar
V056	créixer	V123	valer
V057	desfer	V124	dur
V058	merèixer	V125	obtenir
V060	néixer	V126	pertànyer
V061	nàixer	V127	apendre
V062	péixer	V128	evacuar
V063	poder	v129	donar
V064	saber	v130	oferir
V065	soler	v131	recloure

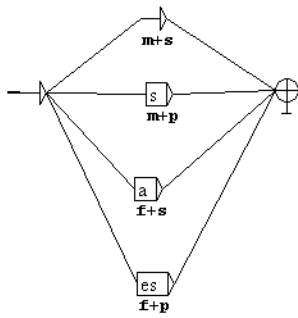
TAB. A.3: *Codes de flexion des substantifs*

<b>CODE</b>	<b>EXEMPLE</b>
N01	peu/peus
N02	dia/dies
N03	braç/braços
N04	disc/discs/discos
N07	desig/desigs/desitjos
N09	bacallà/bacallans
N14	veí/veïns
N15	inca/inques
N16	col·lega/col·legues
N17	paraigua/paraigües
N18	faig/fajos
N19	cabàs/cabassos
N24	ananàs/ananàs
N25	gimnàs gimnasos
N30	país/països
N31	origen orígens
N32	examen exàmens
N33	socors socors socorsos
N40	passeig/passeigs/passejos/passetjos
N41	damassè/domassès
N42	sketch/sketches
N51	font/fonts
N52	casa/cases
N53	plaça/places
N54	oca/oques
N55	pluja/pluges
N56	arruga/arrugues
N57	aigua/aigües
N59	mà/mans
N61	plebs/plebs
N62	aborigen/aborígens

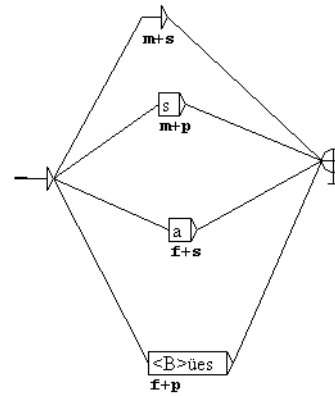


## B Graphes de flexion

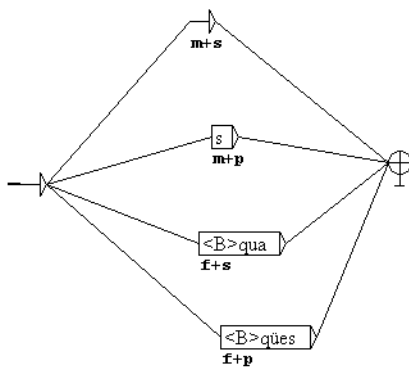
A01



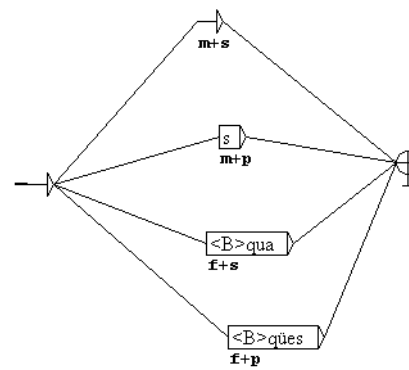
A02



A03

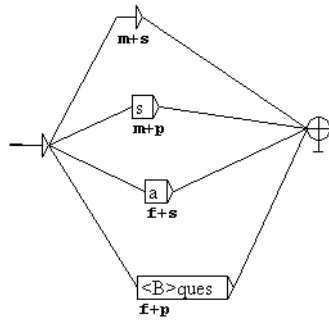


A03

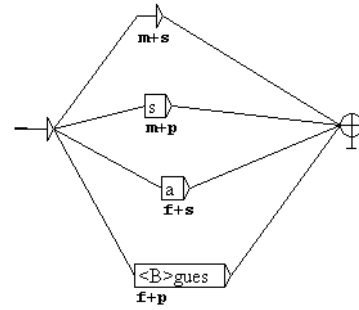




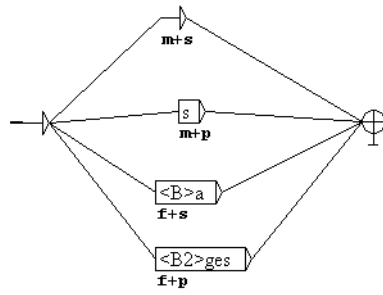
A04



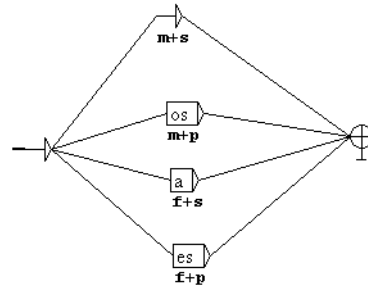
A05



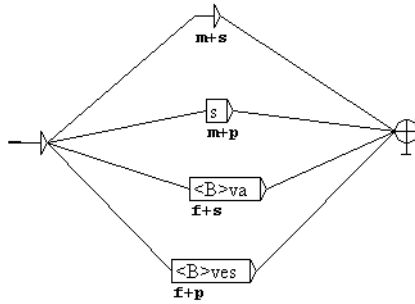
A06



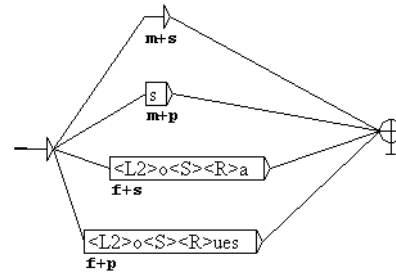
A07

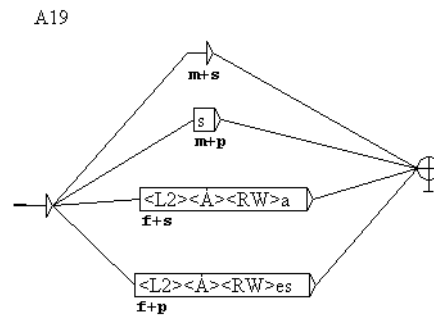
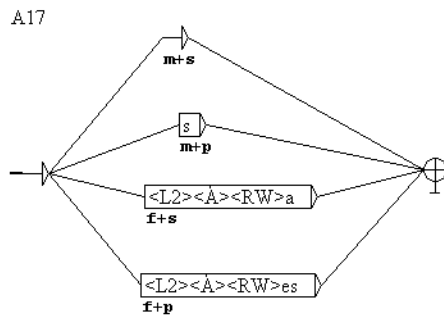
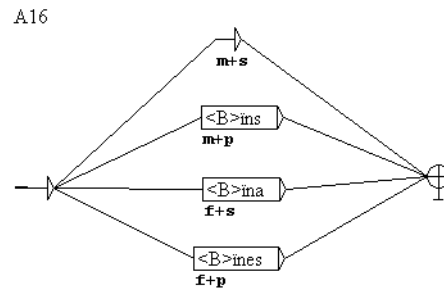
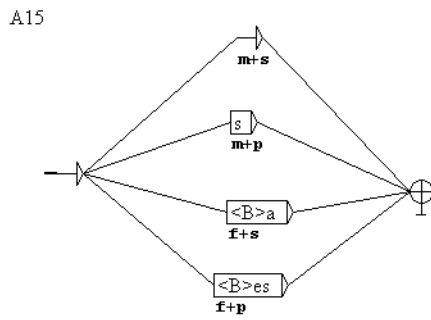
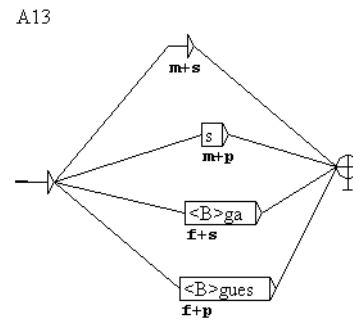
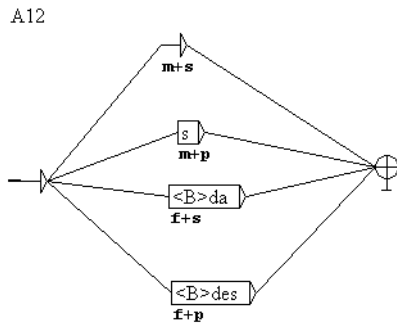
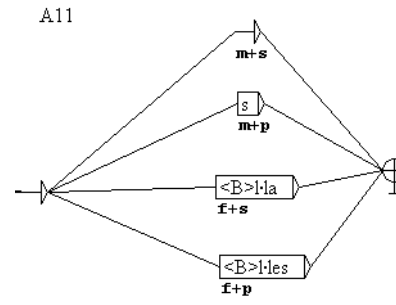
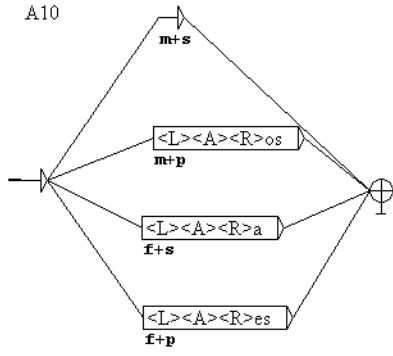


A08

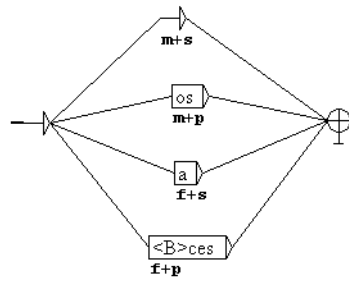


A09

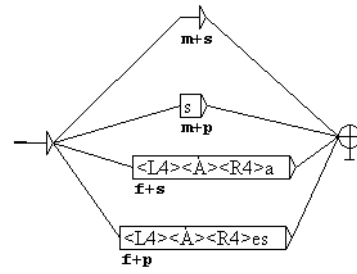




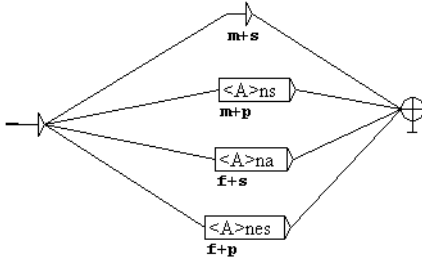
A22



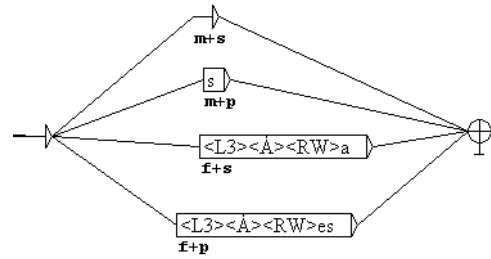
A23



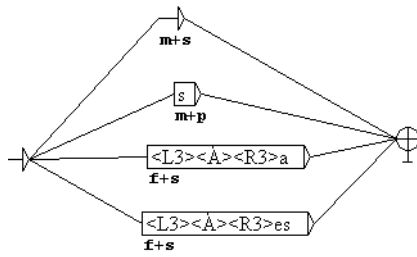
A24



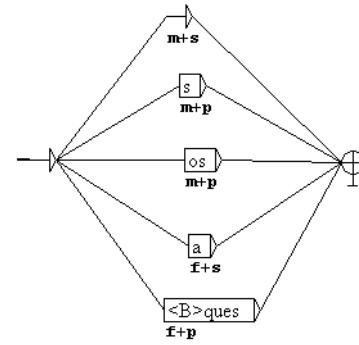
A29



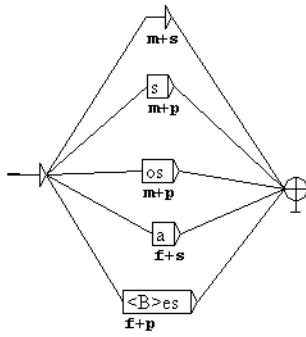
A30



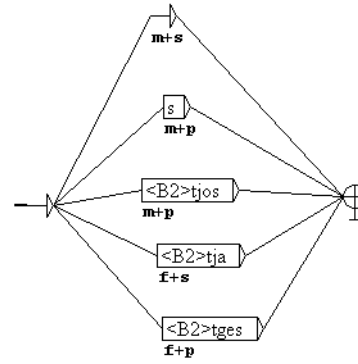
A32



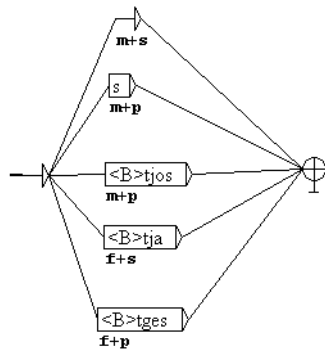
A33



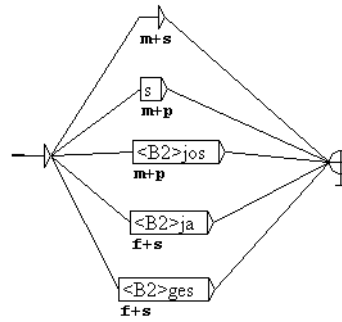
A34



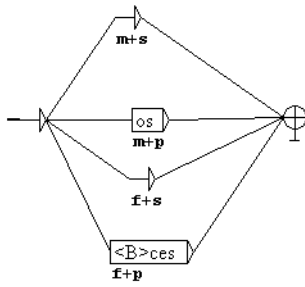
A35



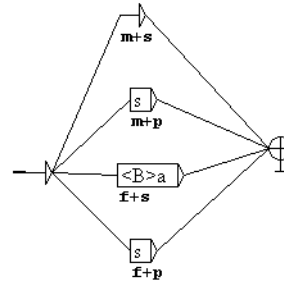
A36



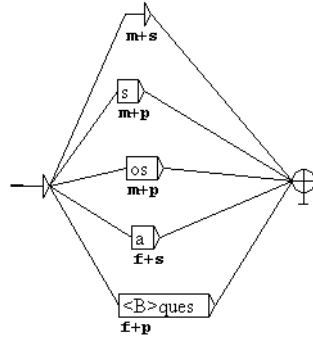
A39



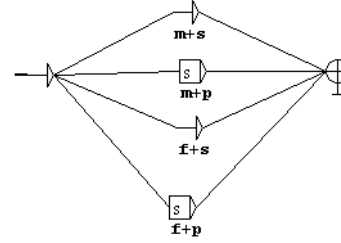
A42



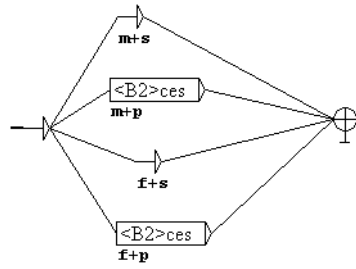
A43



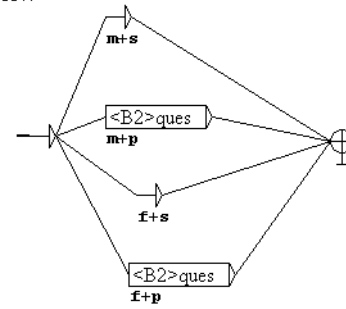
A45



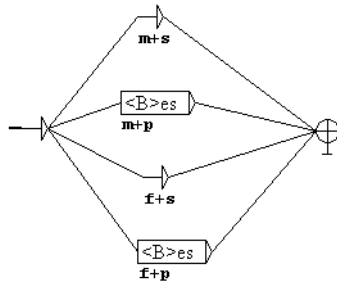
A46



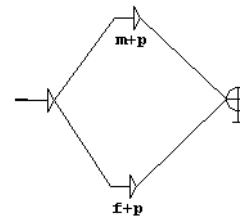
A47



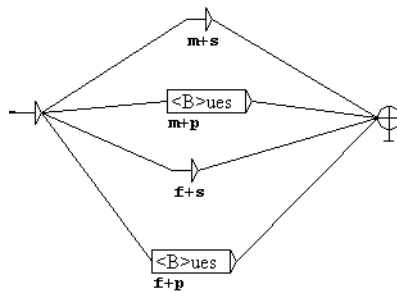
A48



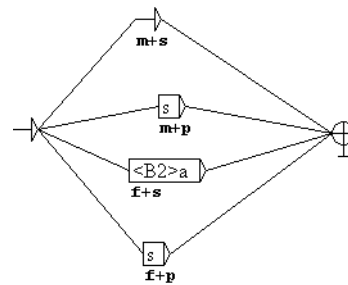
A49



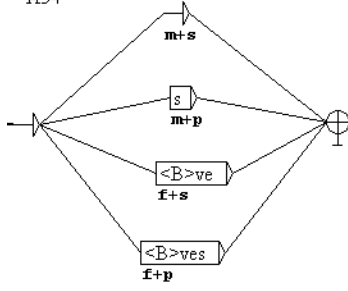
A52



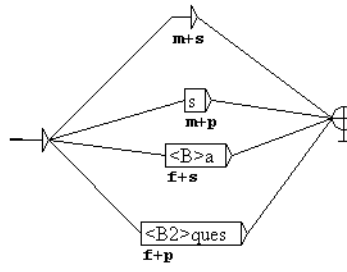
A53



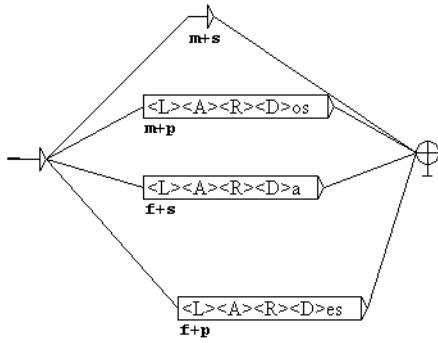
A54



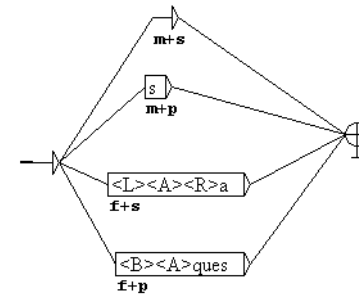
A55



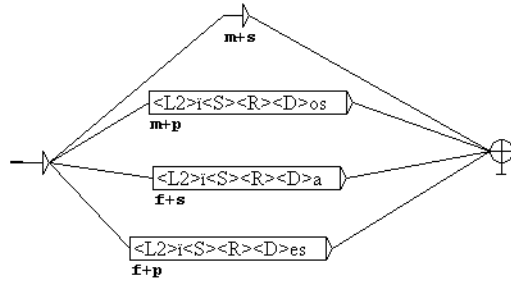
A57



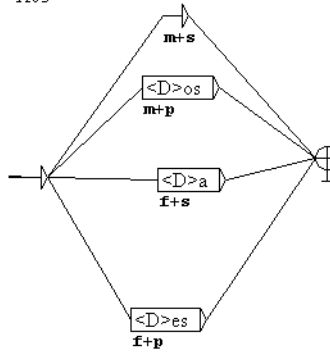
A63



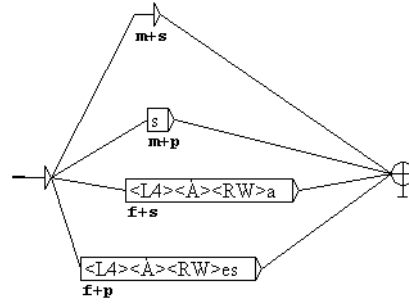
A64



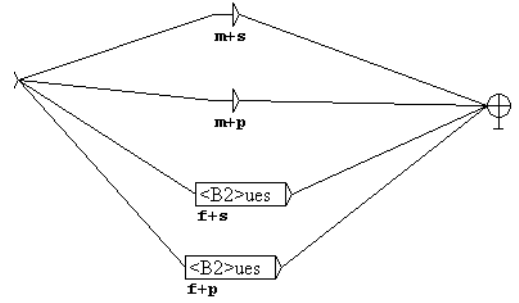
A65



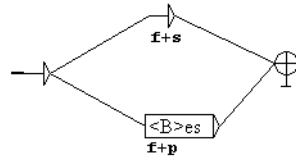
A66



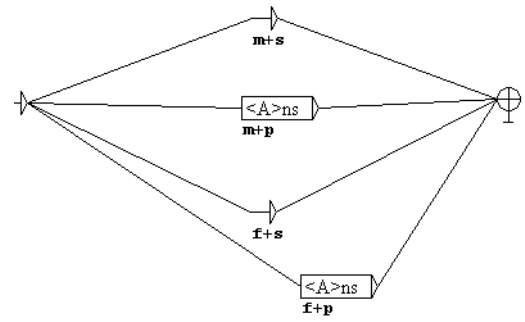
A68



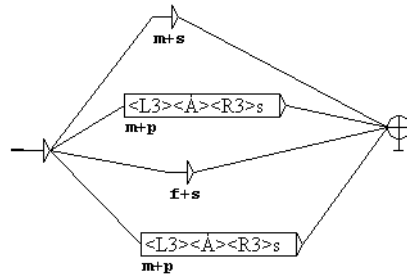
A69



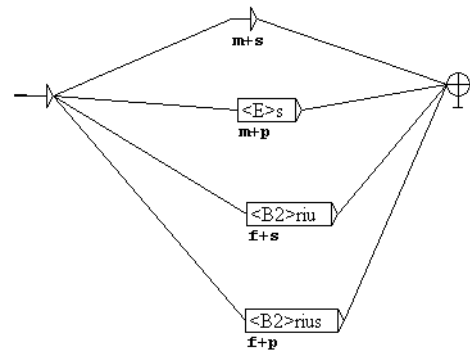
A71



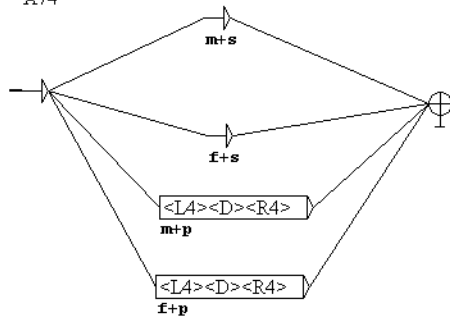
A72



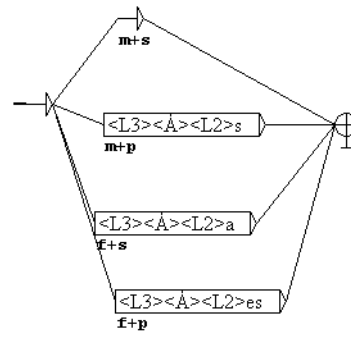
A73



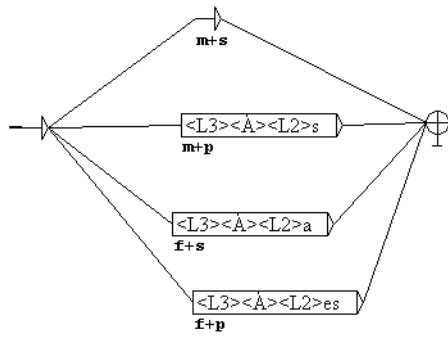
A74



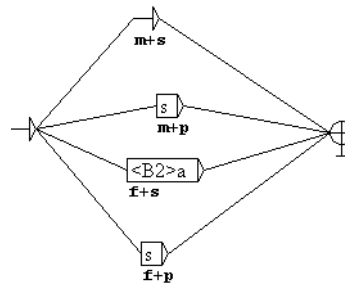
A75



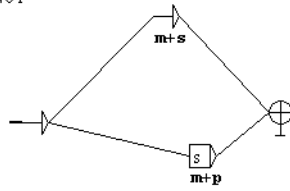
A76



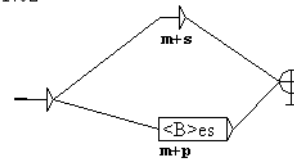
A53



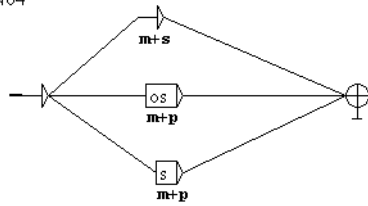
N01



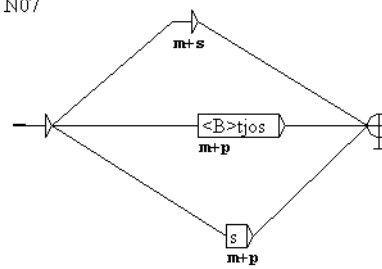
N02



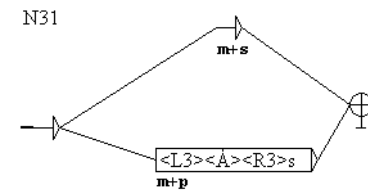
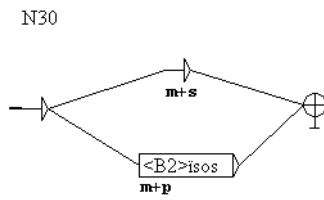
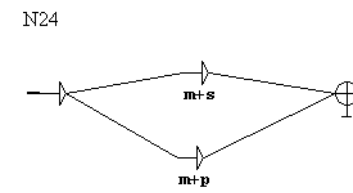
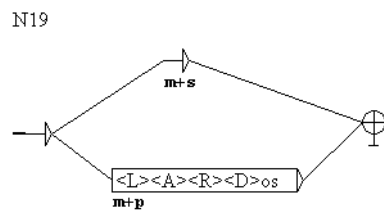
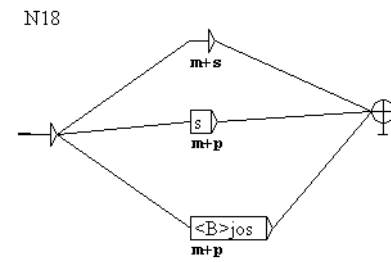
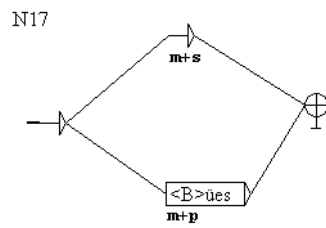
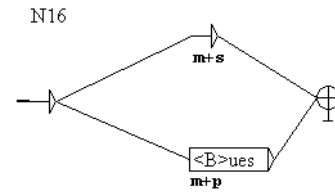
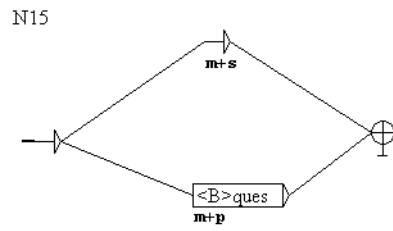
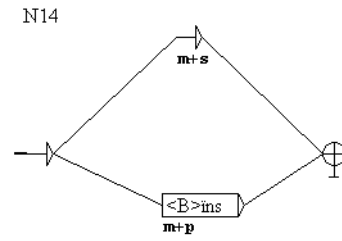
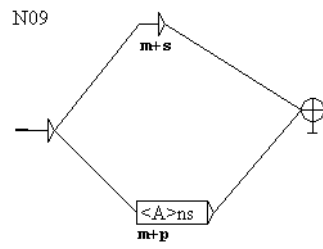
N04



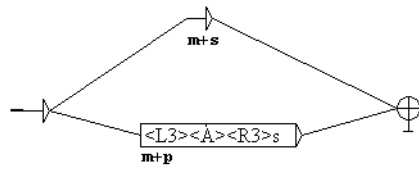
N07



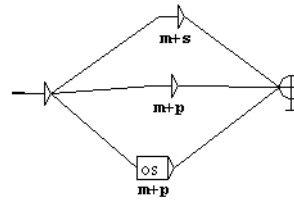




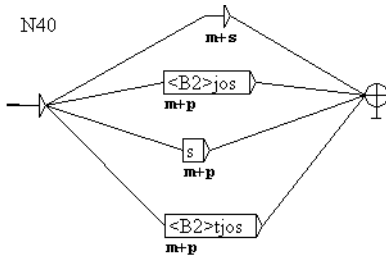
N32



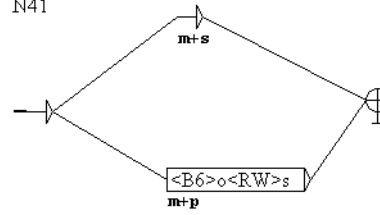
N33



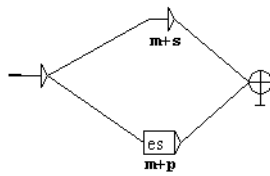
N40



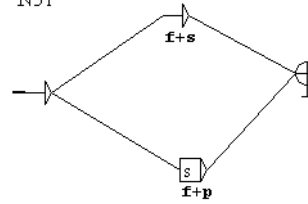
N41



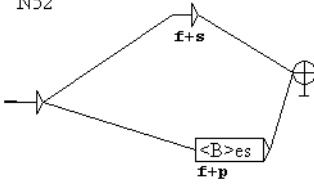
N42



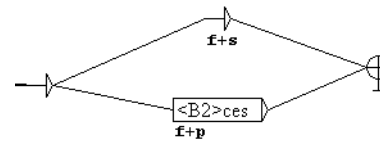
N51



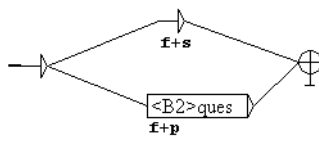
N52



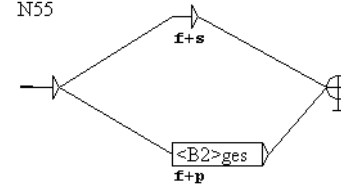
N53

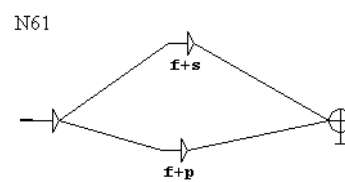
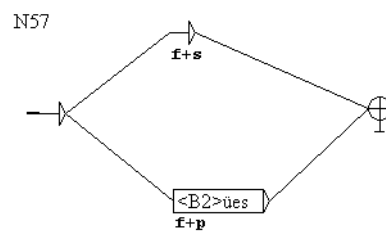
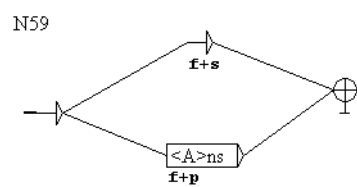
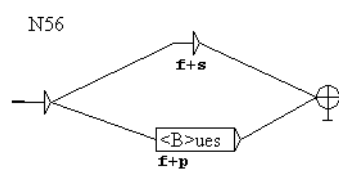


N54



N55







##### VARIABLES ESPRECIFIQUES #####

##### VERBS EN -E #####

```
ePS = (í/PS+s+1 + eres/PS+s+2 + é/PS+s+3 + érem/PS+p+1 + éreu/PS+p+2 + eren/PS+p+3);
eSPR = (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + em/SPR+p+1 + eu/SPR+p+2 + in/SPR+p+3);
eSIMP = (és/SIMP+s+1 + essis/SIMP+s+2 + és/SIMP+s+3 + éssim/SIMP+p+1 + éssiu/SIMP+p+2 + essin/SIMP+p+3);
```

```
Absoldre = (<B3>:Imperfet
            + <B3>gu:ePS
            + <B2>:Futur
            + <B2>:Condicional
            + <B3>gu:eSPR
            + <B3>gu:eSIMP);
```

#####

```
Perdre = (<B2> (o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
          + <B2>:Imperfet
          + <B2>:ePS
          + <B2>:Futur
          + <B2>:Condicional
          + <B2>:eSPR
          + <B2>:eSIMP );
```

#####

```
V012 = :V009;
V016 = :V011;
V028 = :V018;
V044 = :V018;
V045 = :V019;
V047 = :V022;
V049 = :V039;
V050 = :V024;
V058 = :V056;
V061 = :V060;
V068 = :V067;
V078 = :V070;
V081 = :V080;
V082 = :V070;
V083 = :V075;
V084 = :V075;
V086 = :V076;
V099 = :V004;
V104 = :V103;
V106 = :V011;
V108 = :V019;
V113 = :V112;
V121 = :V111;
V127 = :V012;
```

##### VERBS EN -I #####

```
iPS = (í/PS+s+1 + ires/PS+s+2 + í/PS+s+3 + írem/PS+p+1 + íreu/PS+p+2 + iren/PS+p+3);
iSPR = (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + in/SPR+p+3);
iSIMP = (ís/SIMP+s+1 + issis/SIMP+s+2 + ís/SIMP+s+3 + íssim/SIMP+p+1 + íssiu/SIMP+p+2 + issin/SIMP+p+3);
```

```

Ajupir = (<B2>:Imperfet
          + <B2>:iPS
          + <B>:Futur
          + <B>:Condicional
          + <B2>:iSPR
          + <B2>:iSIMP);

##### VERBS EN -A #####

aIMP = (ava/IMP+s+1 + aves/IMP+s+2 + ava/IMP+s+3 + àvem/IMP+p+1 + àveu/IMP+p+2 + aven/IMP+p+3);
aPS  = (í/PS+s+1 + ares/PS+s+2 + à/PS+s+3 + àrem/PS+p+1 + àreu/PS+p+2 + aren/PS+p+3);

##### CODIS #####

#####
#   VERBS EN -DRE
#####

#ABSOLDRE

V001 = <E>/INF # infinitif
      + <B3>ent/G # gerundi
      + <B3> (t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi
      + <B3> (c/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
      + :Absoldre
      + <B3> (<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

#APRENDRE

V002 = <E>/INF # infinitif
      + <B3>ent/G # gerundi
      + <B4> (<À>s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi
      + <B3> (c/PR+s+1 + s/PR+s+2 + <L><À>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
      + :Absoldre
      + <B3> (<L><À>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

#ATENDRE

V003 = <E>/INF # infinitif
      + <B3>ent/G # gerundi
      + <B4> (<À>s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi
      + <B3> (c/PR+s+1 + s/PR+s+2 + <L><Á>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
      + :Absoldre
      + <B3> (<L><Á>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

#CARVENDRE

V004 = <E>/INF # infinitif
      + <B3>ent/G # gerundi
      + <B3> (ut/PP+m+s + uda/PP+f+s + uts/PP+m+p + udes/PP+f+p) # participi
      + <B3> (c/PR+s+1 + s/PR+s+2 + <L><À>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
      + :Absoldre
      + <B3> (<L><À>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

```

## #COMPONDRE

V005 = <E>/INF # infinitif  
 + <B3>ent/G # gerundi  
 + <B4> (st/PP+m+s + sta/PP+f+s + stos/PP+m+p + stes/PP+f+p) # participi  
 + <B3> (c/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + :Absoldre  
 + <B3> (<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

## #CONFONDRE

V006 = <E>/INF # infinitif  
 + <B3>ent/G # gerundi  
 + <B4> (<À>s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi  
 + <B3> (c/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + :Absoldre  
 + <B3> (<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

## #OFENDRE

V007 = <E>/INF # infinitif  
 + <B3>ent/G # gerundi  
 + <B4> (<À>s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi  
 + <B3> (c/PR+s+1 + s/PR+s+2 + <L><À>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + :Absoldre  
 + <B3> (<L><À>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

## #DOLDRE

V008 = <E>/INF # infinitif  
 + <B3>ent/G # gerundi  
 + <B3> (gut/PP+m+s + guda/PP+f+s + guts/PP+m+p + gudes/PP+f+p) # participi  
 + <B3> (c/PR+s+1 + s/PR+s+2 + <L><À>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + :Absoldre  
 + <B3> (<L><À>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

## #FONDRE

V009 = <E>/INF # infinitif  
 + <B3>ent/G # gerundi  
 + <B4> (s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi  
 + <B3> (c/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + :Absoldre  
 + <B3> (<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

## #MOLDRE

V010 = <E>/INF # infinitif  
 + <B3>ent/G # gerundi  
 + <B3><L><À><R> (t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi  
 + <B3> (c/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + :Absoldre  
 + <B3> (<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

## #ROMANDRE

V013 = <E>/INF # infinitif  
 + <B3>ent/G # gerundi

```

+ <B4> (<À>s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi
+ <B3> (c/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ :Absoldre
+ <B3> (<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

#VENDRE

V014 = <E>/INF # infinitif
+ <B3>ent/G # gerundi
+ <B3> (gut/PP+m+s + guda/PP+f+s + guts/PP+m+p + gudes/PP+f+p) # participi
+ <B3> (ut/PP+m+s + uda/PP+f+s + uts/PP+m+p + udes/PP+f+p) # participi
+ <B3> (c/PR+s+1 + s/PR+s+2 + <L><À>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ :Absoldre
+ <B3> (<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

#PERDRE

V011 = <E>/INF # infinitif
+ <B2>ent/G # gerundi
+ <B2> (ut/PP+m+s + uda/PP+f+s + uts/PP+m+p + udes/PP+f+p) # participi
+ :Perdre
+ <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3);

#####
#   VERBS EN -TRE
#####

#ADMETRE

V015 = <E>/INF # infinitif
+ <B2>ent/G # gerundi
+ <B3> (<À>s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi
+ :Perdre
+ <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3);

#METRE

V017 = <E>/INF # infinitif
+ <B2>ent/G # gerundi
+ <B3> (s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi
+ :Perdre
+ <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3);

#####
#   VERBS EN -IR
#####

#AJUPIR

V018 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
+ :Ajupir
+ <B2> (<E>/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3);

```



## #ARGÜIR

```

V019 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2>(eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + eixen/PR+p+3)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>(eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)
+ <B2>:iSIMP
+ <B2>(eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);

```

## #BULLIR

```

V020 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2>(o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2> (eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)
+ <B2>:iSIMP
+ <B2> (eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);

```

## #COBRIR

```

V021 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B3> (ert/PP+m+s + erta/PP+f+s + erts/PP+m+p + ertes/PP+f+p) # participi
+ <B2>(eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + eixen/PR+p+3)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2> (eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)
+ <B2>:iSIMP
+ <B2> (eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);

```

## #COLLIR

```

V022 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> ((<L2><B>u<R2>(o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + en/PR+p+3)) + im/PR+p+1 + iu/PR+p+2)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2> ((<L2><B>u<R2> (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + in/SPR+p+3)) + im/SPR+p+1 + iu/SPR+p+2)
+ <B2>:iSIMP
+ <B2> (<L2><B>u<R2>/I+s+2 + <L2><B>u<R2>i/I+s+3 + im/I+p+1 + iu/I+p+2 + <L2><B>u<R2>in/I+p+3);

```

## #COMPLIR

```

V023= <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B2> (ert/PP+m+s + erta/PP+f+s + erts/PP+m+p + ertes/PP+f+p) # participi
+ <B2> (it/PP+m+s + ida/PP+f+s + its/PP+m+p + ides/PP+f+p) # participi
+ <B2>(eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + eixen/PR+p+3)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2> (eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)
+ <B2>:iSIMP
+ <B2> (eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);

#COSIR

V024 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> ((<L2>u<S><R>(o/PR+s+1 + es/PR+s+2 + <E>/PR+s+3 + en/PR+p+3)) + im/PR+p+1 + iu/PR+p+2)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>((<L2>u<S><R>(i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + in/SPR+p+3)) + im/SPR+p+1 + iu/SPR+p+2)
+ <B2>:iSIMP
+ <B2> (<L2>u<S><R>/I+s+2 + <L2>u<S><R>i/I+s+3 + im/I+p+1 + iu/I+p+2 + <L2>u<S><R>in/I+p+3);

#CRUIXIR
#igual que ajupir excepte la 2 ps PR

V025 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + es/PR+s+2 + <E>/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
+ :Ajupir
+ <B2> (<E>/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3);

#DESCOSIR

V026 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> ((<L2>u<S><R>(o/PR+s+1 + es/PR+s+2 + <L><À>/PR+s+3 + en/PR+p+3)) + im/PR+p+1 + iu/PR+p+2)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>((<L2>u<S><R>(i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + in/SPR+p+3)) + im/SPR+p+1 + iu/SPR+p+2)
+ <B2>:iSIMP
+ <B2> (<L2>ú<S><R>/I+s+2 + <L2>u<S><R>i/I+s+3 + im/I+p+1 + iu/I+p+2 + <L2>u<S><R>in/I+p+3);

#DIR

V027 = <E>/INF # infinitif
+ <B>ent/G # gerundi
+ <B> (t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi
+ <B> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + uen/PR+p+3 )
+ <B2> (eia/IMP+s+1 + eies/IMP+s+2 + eia/IMP+s+3 + èiem/IMP+p+1 + èieu/IMP+p+2 + eien/IMP+p+3)
+ <B>gu:ePS
+ <B>:Futur
+ <B>:Condicional
+ <B>gu:eSPR

```

```
+ <B>gu:eSIMP
+ <B> (gues/I+s+2 + gui/I+s+3 + guem/I+p+1 + gueu/I+p+2 + guin/I+p+3);
```

## #EIXIR

```
V029 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <BW> (ixo/PR+s+1 + ixes/PR+s+2 + ix/PR+s+3 + eixim/PR+p+1 + eixiu/PR+p+2 + ixen/PR+p+3)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <BW> (ixi/SPR+s+1 + ixis/SPR+s+2 + ixi/SPR+s+3 + eixim/SPR+p+1 + eixiu/SPR+p+2 + ixin/SPR+p+3)
+ <B2>:iSIMP
+ <BW> (ix/I+s+2 + ixi/I+s+3 + eixim/I+p+1 + eixiu/I+p+2 + ixin/I+p+3);
```

## #ESCOPIR

```
V030 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> ((<L2>u<S><R>(o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + en/PR+p+3)) + im/PR+p+1 + iu/PR+p+2)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>((<L2>u<S><R>(i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + in/SPR+p+3)) + im/SPR+p+1 + iu/SPR+p+2)
+ <B2>:iSIMP
+ <B2> (<L2>u<S><R>/I+s+2 + <L2>u<S><R>i/I+s+3 + im/I+p+1 + iu/I+p+2 + <L2>u<S><R>in/I+p+3);
```

## #FUGIR

```
V031 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (<L2>j<S><R>o/PR+s+1 + es/PR+s+2 + <L>i<R>/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>:iSPR
+ <B2>:iSIMP
+ <B2> (<L>i<R>/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3);
```

## #IMPRIMIR

```
V032 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B4> (ès/PP+m+s + esa/PP+f+s + esos/PP+m+p + eses/PP+f+p) # participi
+ <B2>(eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + eixen/PR+p+3)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>(eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)
+ <B2>:iSIMP
+ <B2> (eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);
```

## #LLUIR

V033 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B2> (ít/PP+m+s + ida/PP+f+s + its/PP+m+p + ides/PP+f+p) # participi  
 + <B2> (o/PR+s+1 + us/PR+s+2 + u/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)  
 + <B2>/PR+s+3  
 + <B2> (ía/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)  
 + <B2> (í/PS+s+1 + íres/PS+s+2 + í/PS+s+3 + írem/PS+p+1 + íreu/PS+p+2 + íren/PS+p+3)  
 + <B>:Futur  
 + <B>:Condicional  
 + <B2> (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + in/SPR+p+3)  
 + <B2> (ís/SIMP+s+1 + íssis/SIMP+s+2 + ís/SIMP+s+3 + íssim/SIMP+p+1 + íssiu/SIMP+p+2 + íssin/SIMP+p+3)  
 + <B2> (u/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3)  
 + <B2>/I+s+2 ;

#MENTIR

V034 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B2> (o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)  
 + :Ajupir  
 + <B2> (<E>/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3)  
  
 + <B2>(eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + eixen/PR+p+3)  
 + <B2>(eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3)  
 + <B2> (eix/I+s+2 + eixi/I+s+3);

#MORIR

V035 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B> (t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi  
 + <B2> (o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)  
 + :Ajupir  
 + <B2> (<E>/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3);

#OBRIR

V036 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B> (t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi  
 + <B2> (o/PR+s+1 + es/PR+s+2 + e/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)  
 + :Ajupir  
 + <B2> (e/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3);

#OIR

V037 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B2> (ít/PP+m+s + ida/PP+f+s + its/PP+m+p + ides/PP+f+p) # participi  
 + <B2> (eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)  
 + <B2> (u/PR+s+3 + us/PR+s+2)  
 + <B2> (ía/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)  
 + <B2> (í/PS+s+1 + íres/PS+s+2 + í/PS+s+3 + írem/PS+p+1 + íreu/PS+p+2 + íren/PS+p+3)  
 + <B>:Futur  
 + <B>:Condicional  
 + <B2> (eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)  
 + <B2> (ís/SIMP+s+1 + íssis/SIMP+s+2 + ís/SIMP+s+3 + íssim/SIMP+p+1 + íssiu/SIMP+p+2 + íssin/SIMP+p+3)  
 + <B2> (eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);

#OMPLIR

```

V038 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B3> (ert/PP+m+s + erta/PP+f+s + erts/PP+m+p + ertes/PP+f+p) # participi
+ <B2> (o/PR+s+1 + es/PR+s+2 + e/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
+ :Ajupir
+ <B2> (e/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3);

#PREVENIR

V039 = <E>/INF # infinitif
+ <B2><L2>i<S><R>
    (guts/PP+m+s + guda/PP+f+s + guts/PP+m+p + gudes/PP+f+p # participi
    + gu:ePS
    + d:Futur
    + d:Condicional
    + gu:eSPR
    + gu:eSIMP
    + c/PR+s+1
    + gui/I+s+3 + guem/I+p+1 + guin/I+p+3)
+ (<B2>
    (int/G #gerundi
    + (s/PR+s+2 + <B2>é/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
    + :Imperfet
    + <B2>én/I+s+2 + <E>iu/I+p+2);

#PRUIR

V040 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B2> (it/PP+m+s + ida/PP+f+s + its/PP+m+p + ides/PP+f+p) # participi
+ <B2> (o/PR+s+1 + us/PR+s+2 + u/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
+ <B2> (ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)
+ <B2> (í/PS+s+1 + íres/PS+s+2 + í/PS+s+3 + írem/PS+p+1 + íreu/PS+p+2 + íren/PS+p+3)
+ <B>:Futur
+ <B>:Condicional
+ <B2> (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + in/SPR+p+3)
+ <B2> (ís/SIMP+s+1 + íssis/SIMP+s+2 + ís/SIMP+s+3 + íssim/SIMP+p+1 + íssiu/SIMP+p+2 + íssin/SIMP+p+3)
+ <B2> (u/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3);

#PUDIR

V041 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + <B>ts/PR+s+2 + <B>t/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
+ :Ajupir
+ <B2> (<B>t/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3);

#REDUIR

V042 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B2> (it/PP+m+s + ida/PP+f+s + its/PP+m+p + ides/PP+f+p) # participi
+ <B2> (eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + eixen/PR+p+3)
+ <B2> (ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)
+ <B2> (í/PS+s+1 + íres/PS+s+2 + í/PS+s+3 + írem/PS+p+1 + íreu/PS+p+2 + íren/PS+p+3)
+ <B>:Futur
+ <B>:Condicional
+ <B2> (eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)
+ <B2> (ís/SIMP+s+1 + íssis/SIMP+s+2 + ís/SIMP+s+3 + íssim/SIMP+p+1 + íssiu/SIMP+p+2 + íssin/SIMP+p+3)

```

+ <B2> (eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);

#REEIXIR

V043 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B5> (ixo/PR+s+1 + ixes/PR+s+2 + ix/PR+s+3) + <B2>im/PR+p+1 + <B2>iu/PR+p+2 + <B5>ixen/PR+p+3  
 + <B2>:Imperfet  
 + <B2>:iPS  
 + <B>:Futur  
 + <B>:Condicional  
 + <B5> (ixi/SPR+s+1 + ixis/SPR+s+2 + ixi/SPR+s+3) + <B2>im/SPR+p+1 + <B2>iu/SPR+p+2 + <B5>ixin/SPR+p+3  
 + <B2>:iSIMP  
 + <B2> (<B3>ix/I+s+2 + <B3>ixi/I+s+3 + im/I+p+1 + iu/I+p+2 + <B3>ixin/I+p+3);

#SOFRIIR

V046 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B3> (rt/PP+m+s + rta/PP+f+s + rts/PP+m+p + rtes/PP+f+p) # participi  
 + <B2>(eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + eixen/PR+p+3)  
 + <B2>:Imperfet  
 + <B2>:iPS  
 + <B>:Futur  
 + <B>:Condicional  
 + <B2>(eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)  
 + <B2>:iSIMP  
 + <B2> (eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);

#SUPLIR

V048= <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B2> (ert/PP+m+s + erta/PP+f+s + erts/PP+m+p + ertes/PP+f+p) # participi  
 + <B2> ((eix(o/PR+s+1 + es/PR+s+2 + <E>/PR+s+3 + en/PR+p+3)) + im/PR+p+1 + iu/PR+p+2)  
 + <B2>:Imperfet  
 + <B2>:iPS  
 + <B>:Futur  
 + <B>:Condicional  
 + <B2> ((<L2><B>u<R2> (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + in/SPR+p+3)) + im/SPR+p+1 + iu/SPR+p+2)  
 + <B2>:iSIMP  
 + <B2> (<L2><B>u<R2>/I+s+2 + <L2><B>u<R2>i/I+s+3 + im/I+p+1 + iu/I+p+2 + <L2><B>u<R2>in/I+p+3);

#TENIR es igual que prevenir canvia l'imperatiu

V049 = <E>/INF # infinitif  
 + <B2><L2>i<S><R>  
 (gut/PP+m+s + guda/PP+f+s + guts/PP+m+p + gudes/PP+f+p # participi  
 + gu:ePS  
 + d:Futur  
 + d:Condicional  
 + gu:eSPR  
 + gu:eSIMP  
 + c/PR+s+1  
 + (gues/I+s+2 + gui/I+s+3 + guem/I+p+1 + gueu/I+p+s + guin/I+p+3))  
 + <B2>

```

        (int/G #gerundi
+ (s/PR+s+2 + <B2>é/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
+ :Imperfet
+ <B2>é/I+s+2 + <E>/I+s+2 + iu/I+p+2);

#TOSSIR

V050 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (<LW><R>u<S><RW>(o/PR+s+1 + es/PR+s+2 + <B>/PR+s+3 + en/PR+p+3) + im/PR+p+1 + iu/PR+p+2)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>(<LW><R>u<S><RW> (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + in/SPR+p+3) + im/SPR+p+1 + iu/SPR+p+2)
+ <B2>:iSIMP
+ <B2> (<LW><R>u<S><RW>/I+s+2 + <LW><R>u<S><RW>i/I+s+3 + im/I+p+1 + iu/I+p+2 + <LW><R>u<S><RW>/in/I+p+3);

#VENIR viene de v039 canvia el present

V051 = <E>/INF # infinitif
+ <B2><L2>i<S><R>
+ (gut/PP+m+s + guda/PP+f+s + guts/PP+m+p + gudes/PP+f+p) # participi
+ c/PR+s+s
+ gu:ePS
+ d:Futur
+ d:Condicional
+ gu:eSPR
+ gu:eSIMP
+ c/PR+s+1
+ gu/I+s+3 + guem/I+p+1 + guin/I+p+3)
+ <B2>
+ (int/G #gerundi
+ (<B2>és/PR+s+2 + <B>/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + <B2>énen/PR+p+3)
+ :Imperfet
+ <B2>én/I+s+2 + <E>iu/I+p+2);

#ESTABLIR viene de v023 complir

V052 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B2> (ert/PP+m+s + erta/PP+f+s + erts/PP+m+p + ertes/PP+f+p) # participi
+ <B2>(eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + eixen/PR+p+3)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2> (eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)
+ <B2>:iSIMP
+ <B2> (eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);

#####
# VERBS EN -ER
#####

#ATÈNYER viene de v015

V053 = <E>/INF # infinitif
+ <L4><A><R4><B>nt/G # gerundi
+ <B4>s/PP+m+s

```

```

+ <L4><A><R4>
  ((<B4> (sa/PP+f+s + sos/PP+m+p + ses/PP+f+p)) # participi
+ :Perdre
+ <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3));

#CONÈIXER

V054 = <E>/INF # infinitif
+ <L4><A><R4>
  (<B2>ent/G # gerundi
+ <B4>gu (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (<B2>c/PR+s+1 + es/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B4>gu:ePS
+ <B>:Futur
+ <B>:Condicional
+ <B4>gu:eSPR
+ <B4>gu:eSIMP
+ <B2>/I+s+2 + <B4>gui/I+s+3 + <B4>guem/I+p+1 + <B2>eu/I+p+2 + <B4>guin/I+p+3));

#CÓRRER

V055 = <E>/INF # infinitif
+ <L4><A><R4>
  ( <B>ent/G # gerundi
+ <B> (gut/PP+m+s + guda/PP+f+s + ts/PP+m+p + gues/PP+f+p) # participi
+ <B2> (o/PR+s+1 + es/PR+s+2 + e/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B>gu:ePS
+ <B>:Futur
+ <B>:Condicional
+ <B>(<B>(i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3) + guem/SPR+p+1 + gueu/SPR+p+2 + in/SPR+p+3)
+ <B> (em/SPR+p+1 + eu/SPR+p+2)
+ <B>gu:eSIMP
+ <B>/I+s+2 + <B2>i/I+s+3 + <B>guem/I+p+1 + <B2>eu/I+p+2 + <B2>in/I+p+3
+ <B>m/I+p+1 ) ;

#CRÉIXER copiado de perdre PR+s+2

V056 = <E>/INF # infinitif
+ <L4><A><RW>
  ( <B>nt/G # gerundi
+ <B4> (scut/PP+m+s + scuda/PP+f+s + scuts/PP+m+p + scudes/PP+f+p) # participi
+ <B2> (o/PR+s+1 + es/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:ePS
+ <B>:Futur
+ <B>:Condicional
+ <B2>:eSPR
+ <B2>:eSIMP
+ <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3)

+ <B4>squ:ePS
+ <B4>squ:eSIMP
+ <B4>squ(em/SPR+p+1 + eu/SPR+p+2)
+ <B4>squem/I+p+1);

#DESFER

V057 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi

```



```

+ <B2> (aig/PR+s+1 + às/PR+s+2 + à/PR+s+3) + em/PR+p+1 + eu/PR+p+2 + an/PR+p+3 )
+ <B2> (eia/IMP+s+1 + eies/IMP+s+2 + eia/IMP+s+3 + èiem/IMP+p+1 + èieu/IMP+p+2 + eien/IMP+p+3)
+ <B2> (iu/PS+s+1 + eres/PS+s+2 + éu/PS+s+3 + érem/PS+p+1 + éreu/PS+p+2 + eren/PS+p+3)
+ <B2>a:Futur
+ <B2>a:Condicional
+ <B2> (aci/SPR+s+1 + acis/SPR+s+2 + aci/SPR+s+3 + em/SPR+p+1 + eu/SPR+p+2 + acin/SPR+p+3)
+ <B2> (es/SIMP+s+1 + essis/SIMP+s+2 + es/SIMP+s+3 + éssim/SIMP+p+1 + éssiu/SIMP+p+2 + essin/SIMP+p+3)
+ <B2> (es/I+s+2 + aci/I+s+3 + em/I+p+1 + eu/I+p+2 + acin/I+p+3)
+ <B2>acem/I+p+1
+ <B2>(aceu/SPR+p+2 + acem/SPR+p+1) ;

```

## #NÉIXER

```

V060 = <E>/INF # infinitif
+ <L4><A><RW><B2>
    ((o/PR+s+1 + es/PR+s+2 + <E>/PR+s+3 + en/PR+p+3)
    + (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + in/SPR+p+3)
    + (<E>/I+s+2 + i/I+s+3 + in/I+p+3)
    + (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + in/SPR+p+3))
+ <L5>a<S><RW>
    ( <B>nt/G # gerundi
    + <B4> (scut/PP+m+s + scuda/PP+f+s + scuts/PP+m+p + scudes/PP+f+p) # participi
    + <B2> (o/PR+s+1 + es/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
    + <B2>:Imperfet
    + <B2>:ePS
    + <B>:Futur
    + <B>:Condicional
    + <B2>:eSPR
    + <B2>:eSIMP
    + <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3)

    + <B4>squ:ePS
    + <B4>squ:eSIMP
    + <B4>squ(em/SPR+p+1 + eu/SPR+p+2)
    + <B4>squem/I+p+1 );

```

## #PÉIXER

```

V062 = <E>/INF # infinitif
+ <L4><A><RW>
    ( <B>nt/G # gerundi
    + <B5> (ascut/PP+m+s + ascuda/PP+f+s + ascuts/PP+m+p + ascudes/PP+f+p) # participi
    + <B2> (o/PR+s+1 + es/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
    + <B2>:Imperfet
    + <B2>:ePS
    + <B>:Futur
    + <B>:Condicional
    + <B2>:eSPR
    + <B2>:eSIMP
    + <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3);

```

## #PODER

```

V063 = <E>/INF # infinitif
+ <B3>ent/G # gerundi
+ <B3>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (<B2>uc/PR+s+1 + <B>ts/PR+s+2 + <B>t/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B3>gu:ePS

```

```

+ <B2>:Futur
+ <B2>:Condicional
+ <B4>ugu:eSPR
+ <B3>gu:eSIMP
+ <B4>ugu(es/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3);

```

#SABER

```

V064 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B2>u(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (<BW>sé/PR+s+1 + <B>ps/PR+s+2 + <B>p/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <BW>sapig(<L3><À><RW>(a/SPR+s+1 + ues/SPR+s+2 + a/SPR+s+3 + uen/SPR+p+3) + uem/SPR+p+1 + ueu/SPR+p+2 )
+ <B2>:eSIMP
+ <BW>sapig(<L3><À><RW> (ues/I+s+2 + a/I+s+3 + uen/I+p+3) + uem/I+p+1 + ueu/I+p+2 ) ;

```

#SOLER

```

V065 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B2>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (c/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>gu:ePS
+ <B2>d:Futur
+ <B2>d:Condicional
+ <B2>gu:eSPR
+ <B2>gu:eSIMP
+ <B2>(<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

```

#FER cojo model v027 dir

```

V066 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B> (t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi
+ <B> (<B>a (ig/PR+s+1 + s/PR+s+2 + <E>/PR+s+3) + m/PR+p+1 + u/PR+p+2 + <B>an/PR+p+3 )
+ <B2> (eia/IMP+s+1 + eies/IMP+s+2 + eia/IMP+s+3 + èiem/IMP+p+1 + èieu/IMP+p+2 + eien/IMP+p+3)
+ <B2> (iu/PS+s+1 + eres/PS+s+2 + éu/PS+s+3 + érem/PS+p+1 + éreu/PS+p+2 + eren/PS+p+3)
+ <B2>a:Futur
+ <B2>a:Condicional
+ <B2> (aci/SPR+s+1 + acis/SPR+s+2 + aci/SPR+s+3 + em/SPR+p+1 + eu/SPR+p+2 + acin/SPR+p+3)
+ <B2> (es/SIMP+s+1 + essis/SIMP+s+2 + es/SIMP+s+3 + éssim/SIMP+p+1 + éssiu/SIMP+p+2 + essin/SIMP+p+3)
+ <B2> (es/I+s+2 + aci/I+s+3 + em/I+p+1 + eu/I+p+2 + acin/I+p+3)
+ <B2>acem/I+p+1
+ <B2>(aceu/SPR+p+2 + acem/SPR+p+1) ;

```

#TÒRCER

```

V067 = <E>/INF # infinitif
+ <L4><A><RW>
( <B>nt/G # gerundi
+ <B3>ç(ut/PP+m+s + uda/PP+f+s + uts/PP+m+p + udes/PP+f+p) # participi
+ <B2> (<B>ço/PR+s+1 + es/PR+s+2 + <B>ç/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:ePS
+ <B>:Futur

```

```
+ <B>:Condicional
+ <B2>:eSPR
+ <B2>:eSIMP
+ <B2>(<B>ç/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3));
```

#VOLER

```
V069 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B2>g(ut/PP+m+s + uda/PP+f+s + uts/PP+m+p + udes/PP+f+p) # participi
+ <B2> (<B2>ull/PR+s+1 + s/PR+s+2 + <B>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>gui:ePS
+ <B2>d:Futur
+ <B2>d:Condicional
+ <B4>ulgu:eSPR
+ <B2>gu:eSIMP
+ <B4>ulgu(es/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3));
```

```
#####
# VERBS EN -URE
#####
```

#BEURE

```
V070 = <E>/INF # infinitif
+ <B3>vent/G # gerundi
+ <B3>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + vem/PR+p+1 + veu/PR+p+2 + uen/PR+p+3)
+ <B3>v:Imperfet
+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:eSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + veu/I+p+2 + guin/I+p+3));
```

#CAURE

```
V071 = <E>/INF # infinitif
+ <B3>ient/G # gerundi
+ <B3>igu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + iem/PR+p+1 + ieu/PR+p+2 + uen/PR+p+3)
+ <BW>que:Imperfet
+ <B3>igu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>igu:eSPR
+ <B3>igu:eSIMP
+ <B3>(u/I+s+2 + igui/I+s+3 + iguem/I+p+1 + ieu/I+p+2 + iguin/I+p+3));
```

#COMMOURE

```
V072 = <E>/INF # infinitif
+ <B3>vent/G # gerundi
+ <B3>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + vem/PR+p+1 + veu/PR+p+2 + uen/PR+p+3)
+ <B3>v:Imperfet
+ <B3>gu:ePS
```

```

+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:eSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + veu/I+p+2 + guin/I+p+3)

+ <B3>(<À>s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p); # participi

#COMPLAURE

V073 = <E>/INF # infinitif
+ <B3>ent/G # gerundi
+ <B3>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + uen/PR+p+3)
+ <B3>(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)
+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:eSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

#COURE

V074 = <E>/INF # infinitif
+ <B3>ent/G # gerundi
+ <B3>cui(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + uen/PR+p+3)
+ <B3>(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)
+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:eSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

#CLOURE

V075 = <E>/INF # infinitif
+ <B3>ent/G # gerundi
+ <B3> (s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + uen/PR+p+3)
+ <B3>(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)
+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:eSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

#CREURE

V076 = <E>/INF # infinitif
+ <B3>ient/G # gerundi
+ <B3>gu (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + íem/PR+p+1 + íeu/PR+p+2 + uen/PR+p+3)
+ <B3> (ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + <À>íem/IMP+p+1 + <À>íeu/IMP+p+2 + ien/IMP+p+3)
+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional

```

```

+ <B3>gu:eSPR
+ <B3>gu:eSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + ieu/I+p+2 + guin/I+p+3);

```

## #DESCLOURE

```

V077 = <E>/INF # infinitif
+ <B3>ent/G # gerundi
+ <B3> (<À>s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + uen/PR+p+3)
+ <B3>(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)
+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:eSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

```

## #ESCRIURE

```

V079 = <E>/INF # infinitif
+ <B3>vint/G # gerundi
+ <B3>(t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + vim/PR+p+1 + viu/PR+p+2 + uen/PR+p+3)
+ <B3>v:Imperfet
+ <B3>gu:iPS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:iSPR
+ <B3>gu:iSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + viu/I+p+2 + guin/I+p+3)

+ <B3>v:iPS
+ <B3>v:iSIMP ;

```

## #JEURE I JAURE

```

V080 = <E>/INF # infinitif
+ <L4>a<S><RW>
( <B3>ient/G # gerundi
+ <B3>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + iem/PR+p+1 + ieu/PR+p+2 + uen/PR+p+3)
+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:iSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + ieu/I+p+2 + guin/I+p+3))

+ <B3>(c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + uen/PR+p+3)
+ <B3>(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + <À>iem/IMP+p+1 + <À>ieu/IMP+p+2 + ien/IMP+p+3)
+ <B3>(gui/SPR+s+1 + guis/SPR+s+2 + gui/SPR+s+3 + guin/SPR+p+3)
+ <B3>(u/I+s+2 + gui/I+s+3 + guin/I+p+3);

```

## #RECLOURE

```

V085 = <E>/INF # infinitif
+ <B3>ent/G # gerundi
+ <B3> (<À>s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + uen/PR+p+3)
+ <B3>(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)

```

```

+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:eSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

```

#ATREURE

```

V087 = <E>/INF # infinitif
( <B3>ient/G # gerundi
+ <B3>(t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi
+ <B3>(c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + iem/PR+p+1 + ieu/PR+p+2 + uen/PR+p+3)
+ <B3>(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + <À>iem/IMP+p+1 + <À>ieu/IMP+p+2 + ien/IMP+p+3) + <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:iSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + ieu/I+p+2 + guin/I+p+3));

```

#TREURE I TRAURE

```

V088 = <E>/INF # infinitif
+ <L4>a<S><RW>
( <B3>ient/G # gerundi
+ <B3>(t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi
+ <B3>(c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + iem/PR+p+1 + ieu/PR+p+2 + uen/PR+p+3)
+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:iSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + ieu/I+p+2 + guin/I+p+3))

+ <B3>(c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + uen/PR+p+3)
+ <B3>(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + <À>iem/IMP+p+1 + <À>ieu/IMP+p+2 + ien/IMP+p+3)
+ <B3>(gui/SPR+s+1 + guis/SPR+s+2 + gui/SPR+s+3 + guin/SPR+p+3)
+ <B3>(u/I+s+2 + gui/I+s+3 + guin/I+p+3);

```

#VEURE

```

V089 = <E>/INF # infinitif
+ <B3>ient/G # gerundi
+ <B4>is (t/PP+m+s + ta/PP+f+s + tos/PP+m+p + tes/PP+f+p) # participi
+ <B3> (ig/PR+s+1 + us/PR+s+2 + u/PR+s+3 + iem/PR+p+1 + ieu/PR+p+2 + uen/PR+p+3 )
+ <B4> (eia/IMP+s+1 + eies/IMP+s+2 + eia/IMP+s+3 + èiem/IMP+p+1 + èieu/IMP+p+2 + eien/IMP+p+3)
+ <B3> (<B>iu/PS+s+1 + eres/PS+s+2 + éu/PS+s+3 + érem/PS+p+1 + éreu/PS+p+2 + eren/PS+p+3)
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>g (i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + em/SPR+p+1 + eu/SPR+p+2 + in/SPR+p+3)
+ <B3>i (és/SIMP+s+1 + essis/SIMP+s+2 + és/SIMP+s+3 + éssim/SIMP+p+1 + éssiu/SIMP+p+2 + essin/SIMP+p+3)
+ <B3>g (es/I+s+2 + aci/I+s+3 + em/I+p+1 + eu/I+p+2 + acin/I+p+3)

+ <B2>acem/I+p+1 + <BW>(ves/I+s+2 + veieu/I+p+2)
+ <B3>i(eres/PS+s+2 + é/PS+s+3 + érem/PS+p+1 + éreu/PS+p+2 + eren/PS+p+3);

```

#VIURE

```

V090 = <E>/INF # infinitif
+ <B3>vint/G # gerundi
+ <B3>scu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (sc/PR+s+1 + us/PR+s+2 + u/PR+s+3 + vim/PR+p+1 + viu/PR+p+2 + uen/PR+p+3)
+ <B3>v:Imperfet
+ <B3>squ:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>squ:eSPR
+ <B3>squ:eSIMP
+ <B3>(u/I+s+2 + squi/I+s+3 + squem/I+p+1 + viu/I+p+2 + squin/I+p+3);

#####
# VERBS EN -BRE
#####

#CABRE

V091 = <E>/INF # infinitif
+ <B2>nt/G # gerundi
+ <B3>u(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + <B>ps/PR+s+2 + <B>p/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <BW>capig(<L3><À><RW>(a/SPR+s+1 + ues/SPR+s+2 + a/SPR+s+3 + uen/SPR+p+3) + uem/SPR+p+1 + ueu/SPR+p+2 )
+ <B2>:eSIMP
+ <BW>capig(<L3><À><RW> (a/I+s+3 + uen/I+p+3) + <B2>/I+s+2 + uem/I+p+1 + ueu/I+p+2 ) ;

#REBRE

V092 = <E>/INF # infinitif
+ <B2>nt/G # gerundi
+ <B3>u(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + <B>ps/PR+s+2 + <B>p/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B2>:eSPR
+ <B2>:eSIMP
+ <B2> (<B>p/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3) ;

# ROMPRE

V093 = <E>/INF # infinitif
+ <B2>nt/G # gerundi
+ <B3>u(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B2>:eSPR
+ <B2>:eSIMP
+ <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3) ;

# RIURE

```

V094 = <E>/INF # infinitif  
 + <B3>ent/G # gerundi  
 + <B3> gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + uen/PR+p+3)  
 + <B4>e(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + <À>iem/IMP+p+1 + <À>ieu/IMP+p+2 + ien/IMP+p+3)  
 + <B3>gu:ePS  
 + <B2>:Futur  
 + <B2>:Condicional  
 + <B3>gu:eSPR  
 + <B3>gu:eSIMP  
 + <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

# HEURE

V095 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B3>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B4> (ec/PR+s+1 + eus/PR+s+2 + eu/PR+s+3 + avem/PR+p+1 + aveu/PR+p+2 + euen/PR+p+3)  
 + <B2>:Imperfet  
 + <B3>gu:ePS  
 + <B3>u:Futur  
 + <B3>u:Condicional  
 + <B4>egu:eSPR  
 + <B3>gu:eSIMP  
 + <B3>(<B>eu/I+s+2 + <B>egui/I+s+3 + guem/I+p+1 + eu/I+p+2 + <B>eguín/I+p+3);

#CALDRE

V096 = <E>/INF # infinitif  
 + <B3>ent/G # gerundi  
 + <B3>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B3>/PR+s+3 + <B3>en/PR+p+3  
 + <B3>ia/IMP+s+3 + <B3>ien/IMP+p+3  
 + <B3>gué/PS+s+3 + <B3>gueren/PS+p+3  
 + <B> (à/F+s+3 + an/F+p+3)  
 + <B> (ia/C+s+3 + ien/C+p+3)  
 + <B3> (gui/SPR+s+3 + guin/SPR+p+3)  
 + <B3> (gués/SIMP+s+3 + guessin/SIMP+p+3);

#CONSTRÈNYER

V097 = <E>/INF # infinitif  
 + <L4><A><RW>  
 ( <B2>ent/G # gerundi  
 + <B4> (t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi  
 + :Perdre  
 + <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3));

#APARÈIXER

V098 = <E>/INF # infinitif  
 + <L4><A><RW>  
 ( <B>nt/G # gerundi  
 + <B4>gu (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B2> (<B2>c/PR+s+1 + es/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + <B2>:Imperfet  
 + <B4>gu:ePS  
 + <B>:Futur  
 + <B>:Condicional  
 + <B4>gu:eSPR  
 + <B4>gu:eSIMP



+ <B2>(<E>/I+s+2 + <B2>gui/I+s+3 + <B2>guem/I+p+1 + eu/I+p+2 + <B2>guin/I+p+3));

V099 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B3>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B3>(<B>e/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + vem/PR+p+1 + veu/PR+p+2 + n/PR+p+3)  
 + <B2>:Imperfet  
 + <B3>gu:ePS  
 + <B3>u:Futur  
 + <B3>u:Condicional  
 + <B3>g:eSPR  
 + <B3>gu:eSIMP  
  
 + <B3> (guem/SPR+p+1 + gueu/SPR+p+2)  
 + <B3>ig/PR+s+1  
 + <B4> (em/PR+p+1 + eu/PR+p+2);

#HAVER

V100 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B3>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B3>(<B>e/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + vem/PR+p+1 + veu/PR+p+2 + n/PR+p+3)  
 + <B2>:Imperfet  
 + <B3>gu:ePS  
 + <B3>u:Futur  
 + <B3>u:Condicional  
 + <B3>g(i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + <L><À><R>im/SPR+p+1 + <L><À><R>iu/SPR+p+2 + in/SPR+p+3)  
  
 + <B3>gu:eSIMP  
  
 + <B3> (guem/SPR+p+1 + gueu/SPR+p+2)  
 + <B4> (aig/PR+s+1 + em/PR+p+1 + eu/PR+p+2);

#ANAR

V101 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B2> (<BW>vaig/PR+s+1 + <BW>vas/PR+s+2 + <BW>va/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + <BW>van/PR+p+3)  
 + <B>(va/IMP+s+1 + ves/IMP+s+2 + va/IMP+s+3 + vem/IMP+p+1 + veu/IMP+p+2 + ven/IMP+p+3)  
 + <B2>(í/PS+s+1 + ares/PS+s+2 + à/PS+s+3 + àrem/PS+p+1 + àreu/PS+p+2 + aren/PS+p+3)  
 + <B2>i:Futur  
 + <B2>i:Condicional  
 + <B2>(<BW>(vagi/SPR+s+1 + vagis/SPR+s+2 + vagi/SPR+s+3) + em/SPR+p+1 + eu/SPR+p+2 + <BW>vagin/SPR+p+3)  
 + <B2>:eSIMP  
 + <B2>(<BW>vés/I+s+2 + <BW>vagi/I+s+3 + em/I+p+1 + eu/I+p+2 + <BW>vagin/I+p+3)  
  
 + <BW>i:Futur  
 + <BW>i:Condicional  
  
 + <BW>vam/PR+p+1 + <BW>vau/PR+p+1 ;

#ESTAR

V102 = <E>/INF # infinitif  
 + <B>nt/G # gerundi

```

+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (ic/PR+s+1 + às/PR+s+2 + à/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + an/PR+p+3)
+ <B2>:aIMP
+ <B2>igu:ePS
+ <B>:Futur
+ <B>:Condicional
+ <B2>igu:eSPR
+ <B2>igu:eSIMP
+ <B2> (igues/I+s+2 + igui/I+s+3 + iguem/I+p+1 + igueu/I+p+2 + iguin/I+p+3) ;

#ÉSSER o SER

V103 = <E>/INF # infinitif
+ <BW>
+ ( essent/G # gerundi
+ <E>sent/G
+ <E>(estat/PP+m+s + estado/PP+f+s + stats/PP+m+p + estades/PP+f+p) # participi
+ <E>(sóc/PR+s+1 + só/PR+s+1 + ets/PR+s+2 + és/PR+s+3 + som/PR+p+1 + sou/PR+p+2 + són/PR+p+3) #PRESENT
+ <E>(era/IMP+s+1 + eres/IMP+s+2 + era/IMP+s+3 + érem/IMP+p+1 + éreu/IMP+p+2 + eren/IMP+p+3) #IMPERFET
+ <E>(fui/PS+s+1 + fores/PS+s+2 + fou/PS+s+3 + fórem/PS+p+1 + fóreu/PS+p+2 + foren/PS+p+3) #PRET.PERFET
+ <E>se:Futur
+ <E>se:Condicional
+ <E>sigu(i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + em/SPR+p+1 + eu/SPR+p+2 + in/SPR+p+3)
+ <E>(fos/SIMP+s+1 + fossis/SIMP+s+2 + fos/SIMP+s+3 + fòssim/SIMP+p+1 + fòssiu/SIMP+p+2 + fossin/SIMP+p+3)
+ <E>(sigues/I+s+2 + sigui/I+s+3 + siguem/I+p+1 + sigueu/I+p+2 + siguin/I+p+3) ;

# CANTAR

V105 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + es/PR+s+2 + a/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:aIMP
+ <B2>:aPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>:eSPR
+ <B2>:eSIMP
+ <B2> (a/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3) ;

#TÉMER

V107 = <E>/INF # infinitif
+ <L3><A><RW>
+ ( <B>nt/G # gerundi
+ <B2> (ut/PP+m+s + uda/PP+f+s + uts/PP+m+p + udes/PP+f+p) # participi
+ <B2> (o/PR+s+1 + s/PR+s+2 + <L><A><E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:ePS
+ <B>:Futur
+ <B>:Condicional
+ <B2>:eSPR
+ <B2>:eSIMP
+ <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3));

#DORMIR

V109 = <E>/INF # infinitif
+ <B>nt/G # gerundi

```

```

+ <B> (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2>(o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>(i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + in/SPR+p+3)
+ <B2>:iSIMP
+ <B2>(<E>/I+s+2 + i/I+s+3 + im/I+p+1 + iu/I+p+2 + in/I+p+3);

#####
## VERBS EN -A
#####

#CANVIAR

V111 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + es/PR+s+2 + a/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:aIMP
+ <B2>:aPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>(i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + em/SPR+p+1 + eu/SPR+p+2 + in/SPR+p+3)
+ <B2>:eSIMP
+ <B2> (a/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3) ;

#COMENÇAR

V112 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (ço/PR+s+1 + ces/PR+s+2 + ça/PR+s+3 + cem/PR+p+1 + ceu/PR+p+2 + cen/PR+p+3)
+ <B2>:aIMP
+ <B3> (cí/PS+s+1 + çares/PS+s+2 + çà/PS+s+3 + çàrem/PS+p+1 + çàreu/PS+p+2 + çaren/PS+p+3)
+ <B>:Futur
+ <B>:Condicional
+ <B3>c:eSPR
+ <B3>c:eSIMP
+ <B3> (ça/I+s+2 + ci/I+s+3 + cem/I+p+1 + ceu/I+p+2 + cin/I+p+3) ;

#DONAR i dar

V114 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + <L><Â><R>es/PR+s+2 + <L><Â><R>a/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:aIMP
+ <B2>:aPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>:eSPR
+ <B2>:eSIMP
+ <B2> (<L><Â><R>a/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3)

+ <BW>d (em/PR+p+1 + eu/PR+p+2)
+ <BW>d:aIMP

```

```

+ <BW>d(i/PS+s+1 + ares/PS+s+2 + a/PS+s+3 + àrem/PS+p+1 + àreu/PS+p+2 + aren/PS+p+3)
+ <BW>da:Futur
+ <BW>da:Condicional
+ <BW>d (em/SPR+p+1 + eu/SPR+p+2)
+ <BW>d(es/SIMP+s+1 + essis/SIMP+s+2 + es/SIMP+s+3 + éssim/SIMP+p+1 + éssiu/SIMP+p+2 + essin/SIMP+p+3)
+ <BW>d (em/I+p+1 + eu/I+p+2 ) ;

```

# PREGAR

```

V116 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + ues/PR+s+2 + a/PR+s+3 + uem/PR+p+1 + ueu/PR+p+2 + uen/PR+p+3)
+ <B2>:aIMP
+ <B2> (uí/PS+s+1 + ares/PS+s+2 + à/PS+s+3 + àrem/PS+p+1 + àreu/PS+p+2 + aren/PS+p+3)
+ <B>:Futur
+ <B>:Condicional
+ <B2>u:eSPR
+ <B2>u:eSIMP
+ <B2> (a/I+s+2 + ui/I+s+3 + uem/I+p+1 + ueu/I+p+2 + uin/I+p+3) ;

```

#ENAIGUAR

```

V117 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + ües/PR+s+2 + a/PR+s+3 + üem/PR+p+1 + üeu/PR+p+2 + üen/PR+p+3)
+ <B2>:aIMP
+ <B2> (üí/PS+s+1 + ares/PS+s+2 + à/PS+s+3 + àrem/PS+p+1 + àreu/PS+p+2 + aren/PS+p+3)
+ <B>:Futur
+ <B>:Condicional
+ <B2>ü:eSPR
+ <B2>ü:eSIMP
+ <B2> (ua/I+s+2 + üi/I+s+3 + üem/I+p+1 + üeu/I+p+2 + üin/I+p+3) ;

```

#TRENCAR

```

V119 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (co/PR+s+1 + ques/PR+s+2 + ca/PR+s+3 + quem/PR+p+1 + queu/PR+p+2 + quen/PR+p+3)
+ <B2>:aIMP
+ <B2> (<B>quí/PS+s+1 + ares/PS+s+2 + à/PS+s+3 + àrem/PS+p+1 + àreu/PS+p+2 + aren/PS+p+3)
+ <B>:Futur
+ <B>:Condicional
+ <B3>qu:eSPR
+ <B3>qu:eSIMP
+ <B3> (ca/I+s+2 + qui/I+s+3 + quem/I+p+1 + queu/I+p+2 + quin/I+p+3) ;

```

#ENVEJAR

```

V120 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B3> (jo/PR+s+1 + ges/PR+s+2 + ja/PR+s+3 + gem/PR+p+1 + geu/PR+p+2 + gen/PR+p+3)
+ <B2>:aIMP
+ <B2> (<B>gí/PS+s+1 + ares/PS+s+2 + à/PS+s+3 + àrem/PS+p+1 + àreu/PS+p+2 + aren/PS+p+3)
+ <B>:Futur
+ <B>:Condicional
+ <B3>g:eSPR
+ <B3>g:eSIMP

```

+ <B3> (ja/I+s+2 + gi/I+s+3 + gem/I+p+1 + geu/I+p+2 + gin/I+p+3) ;

# ESGLAIR

V122 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B2> (o/PR+s+1 + es/PR+s+2 + a/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + <B2>:aIMP  
 + <B2> (<B>i/PS+s+1 + ares/PS+s+2 + à/PS+s+3 + àrem/PS+p+1 + àreu/PS+p+2 + aren/PS+p+3)  
 + <B>:Futur  
 + <B>:Condicional  
 + <B3>(i/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + iem/SPR+p+1 + ieu/SPR+p+2 + in/SPR+p+3)  
 + <B2>:eSIMP  
 + <B3> (ia/I+s+2 + i/I+s+3 + iem/I+p+1 + ieu/I+p+2 + in/I+p+3) ;

#VALER

V123 = <E>/INF # infinitif  
 + <B>nt/G # gerundi  
 + <B2>gu(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi  
 + <B2> (c/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + <B2>:Imperfet  
 + <B2>gu:ePS  
 + <B2>d:Futur  
 + <B2>d:Condicional  
 + <B2>gu:eSPR  
 + <B2>gu:eSIMP  
 + <B2> (<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3) ;

#DUR

V124 = <E>/INF # infinitif  
 + <B>ent/G # gerundi  
 + <B>(t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi  
 + <B> (c/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)  
 + <B> (ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + <Á>iem/IMP+p+1 + <Á>ieu/IMP+p+2 + ien/IMP+p+3)  
 + <B>gu:ePS  
 + <B>:Futur  
 + <B>:Condicional  
 + <B>gu:eSPR  
 + <B>gu:eSIMP  
 + <B> (<E>/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3)  
  
 + us/PR+s+2 + u/PR+s+3  
 + u/I+s+2 ;

#OBTENIR

V125 = <E>/INF # infinitif  
 + <B2><L2>i<S><RW>  
 (gut/PP+m+s + guda/PP+f+s + guts/PP+m+p + gudes/PP+f+p # participi  
 + gu:ePS  
 + d:Futur  
 + d:Condicional  
 + gu:eSPR  
 + gu:eSIMP  
 + c/PR+s+1  
 + (gues/I+s+2 + gui/I+s+3 + guem/I+p+1 + gueu/I+p+2 + guin/I+p+3))

```

+ <B2>
  (int/G #gerundi
  + (s/PR+s+2 + <B2>é/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + en/PR+p+3)
  + :Imperfet
  + <B2>én/I+s+2 + <E>iu/I+p+2);

#PERTÀNYER

V126 = <E>/INF # infinitif
+ <L4><A><RW>
( <B>nt/G # gerundi
+ <B3>gu (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + s/PR+s+2 + <E>/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:Imperfet
+ <B2>:ePS
+ <B>:Futur
+ <B>:Condicional
+ <B2>:eSPR
+ <B2>:eSIMP
+ <B2> (<E>/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3)

+ <B2>u (t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p)); # participi

#EVACUAR

V128 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + es/PR+s+2 + a/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:aIMP
+ <B2> (í/PS+s+1 + ares/PS+s+2 + à/PS+s+3 + àrem/PS+p+1 + àreu/PS+p+2 + aren/PS+p+3)
+ <B>:Futur
+ <B>:Condicional
+ <B2> (ï/SPR+s+1 + is/SPR+s+2 + i/SPR+s+3 + em/SPR+p+1 + eu/SPR+p+2 + n/SPR+p+3)
+ <B2>:eSIMP
+ <B2> (a/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3) ;

#DONAR

V129 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B>(t/PP+m+s + da/PP+f+s + ts/PP+m+p + des/PP+f+p) # participi
+ <B2> (o/PR+s+1 + <L2><A><R2>es/PR+s+2 + <L2><A><R2>a/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + en/PR+p+3)
+ <B2>:aIMP
+ <B2>:aPS
+ <B>:Futur
+ <B>:Condicional
+ <B2>:eSPR
+ <B2>:eSIMP
+ <B2> (a/I+s+2 + i/I+s+3 + em/I+p+1 + eu/I+p+2 + in/I+p+3) ;

#OFERIR

#viene de ESTABLIR

V130 = <E>/INF # infinitif
+ <B>nt/G # gerundi
+ <B2> (t/PP+m+s + ta/PP+f+s + ts/PP+m+p + tes/PP+f+p) # participi
+ <B2>(eixo/PR+s+1 + eixes/PR+s+2 + eix/PR+s+3 + im/PR+p+1 + iu/PR+p+2 + eixen/PR+p+3)
+ <B2>:Imperfet

```

```

+ <B2>:iPS
+ <B>:Futur
+ <B>:Condicional
+ <B2> (eixi/SPR+s+1 + eixis/SPR+s+2 + eixi/SPR+s+3 + im/SPR+p+1 + iu/SPR+p+2 + eixin/SPR+p+3)
+ <B2>:iSIMP
+ <B2> (eix/I+s+2 + eixi/I+s+3 + im/I+p+1 + iu/I+p+2 + eixin/I+p+3);

# RECLOURE

V131 = <E>/INF # infinitif
+ <B3>ent/G # gerundi
+ <B3> (s/PP+m+s + sa/PP+f+s + sos/PP+m+p + ses/PP+f+p) # participi
+ <B3> (c/PR+s+1 + us/PR+s+2 + u/PR+s+3 + em/PR+p+1 + eu/PR+p+2 + uen/PR+p+3)
+ <B3>(ia/IMP+s+1 + ies/IMP+s+2 + ia/IMP+s+3 + íem/IMP+p+1 + íeu/IMP+p+2 + ien/IMP+p+3)
+ <B3>gu:ePS
+ <B2>:Futur
+ <B2>:Condicional
+ <B3>gu:eSPR
+ <B3>gu:eSIMP
+ <B3>(u/I+s+2 + gui/I+s+3 + guem/I+p+1 + eu/I+p+2 + guin/I+p+3);

```

# Remerciements

Voldria agrair al Dr. Xavier Blanco haver acceptat de dirigir el meu treball de recerca.

Gràcies, A. C., S. F. i M. E. per deixar-me fer aquest camí al vostre costat, gràcies pel vostre suport i per tots els vostres comentaris.

A L. B., la rosa d'aquest conte.

A J. Z. per tots els tornejos compartits, per tots els petit moments.

A mis chicas del 6 de mayo, pels 1000 sopars parlant de les 100 pàgines, gràcies.

A L. F. per no deixar que el temps ens aturi.

A G. G., nere kopazco zazpi.

A B. V. por mostrarme que la distancia también es relativa.

A Natalia y a Santos por guiar mis pasos hasta aquí. Y a Víctor por desafiar continuamente mi imaginación.

Y por supuesto, mamá a ti también, pero a ti, al ser la más importante de esta historia, déjame que te lo agradezca mientras tomamos un café.

Gràcies R. V. I. per fer real allò que només podria ser un somni.





## Bibliographie

- [Hip, 1993] (1993). *Hiperdiccionari català-castellà-anglès*. Enciclopèdia Catalana, Barcelona.
- [GDL, 1998] (1998). *Gran Diccionari de la llengua Catalana (GDLC)*. Enciclopèdia Catalana, Barcelona.
- [Blanco, 1999] Blanco, X. (1999). *Lexicographie bilingue français-espagnol et classes d'objets*. Col·lecció Materials (73). Servei de publicacions de l'UAB, Barcelona.
- [Blanco, 2000] Blanco, X. (2000). Les dictionnaires électroniques de l'espagnol (delass et delacs) . *Lingvisticae Investigationes. Revue internationale de linguistique française et générale*, II(XXIII).
- [Bosque, 1999] Bosque, I. (1999). El nombre común. In Demonte, V., editor, *Gramática descriptiva de la lengua española*, volume 1 of *Espasa, Calpe*, pages 3–75. Madrid.
- [Bouillon, 1998] Bouillon, P. (1998). *Traitement automatique des langues naturelles*. Champs Linguistiques. Duculot, Belgique.
- [Català, 2004] Català, D. (2004). Formalización Ingüística dos adverbios compostos do catalán. *Cadernos de fraseoloxia Galega*, pages 11–26.
- [Catena, 2006] Catena, A. (2006). *Contribución a la formalización del adjetivo para la traducción español-francés*. PhD thesis, Universitat Autònoma de Barcelona, Barcelona.
- [Courtois, 1990] Courtois, B. (1990). Un système de dictionnaires électroniques pour les mots simples du français. *Langue Française*, (87):11–23.
- [Flaux and Van de Velde, 2000] Flaux, N. and Van de Velde, D. (2000). *Les noms en français: esquisse de classement*. collection l'essentiel français. Ophrys, Paris.
- [Fuentes, 2006a] Fuentes, S. (2006a). "la notion de classes d'objets adaptés aux humains" seminari de tercer cicle.
- [Fuentes, 2006b] Fuentes, S. (2006b). *Les dictionnaires monolingues coordonnés DicPro*.

- [Fuentes, 2006c] Fuentes, S. (2006c). Étude lexicographique des noms de professions. In *Actes du VII Congrès International de Linguistique Française*, Valencia. Universitat de Valencia.
- [Garrigues, 1992] Garrigues, M. (1992). Dictionnaires hiérarchiques du français. *Langue Française*, (96):88–100.
- [Geiss, 2006] Geiss, C. (2006). Dictionnaire électronique de verbes catalans pour la traduction automatique catalan-français. Master's thesis, Ludwig-Maximilians-Universität München, Centrum für Informations und Sprachverarbeitung.
- [Gross, 1992] Gross, G. (1992). Forme d'un dictionnaire électronique. In *L'environnement traductionnel actes du colloque de Mons*, pages 255–271, Mons. AUPELF-UREF.
- [Gross, 1994] Gross, G. (1994). A quoi sert la notion de *partie de discours*? In de Lyon, P. U., editor, *Les classes de mots: Traditions et perspectives*, pages 217–231.
- [Gross, 1996] Gross, G. (1996). Prédicats nominaux et compatibilité aspectuelle. *Langages*, (121):54–72.
- [Gross and Vivès, 1986] Gross, G. and Vivès, R. (1986). Les constructions nominales et l'élaboration d'un lexique grammairal. *Langue Française*, (69):5–27.
- [Gross, 1975] Gross, M. (1975). *Méthodes en Syntaxe*. Hermann, Paris.
- [Gross, 1998] Gross, M. (1998). La fonction sémantique des verbes supports. *Travaux de Linguistique*, (37):25–46.
- [Harris, 1976] Harris, Z. S. (1976). *Notes du cours de syntaxe*. Travaux linguistiques. Seuil, Paris.
- [Lamiroy, 1998] Lamiroy, B. (1998). Le lexique-grammaire. essai de synthèse. *Travaux de Linguistique*, (37):7–24.
- [Le Pesant, 1994] Le Pesant, D. (1994). Les compléments nominaux du verbe lire. Une illustration de la notion de classes d'objets. *Langages*, (115):31–46.
- [Le Pesant, 2000] Le Pesant, D. (2000). Six études de sémantique lexicale sur les noms communs de lieux. Technical report, Villetaneuse : Université Paris 13. Ouvrage rédigé en vue d'obtenir l'Habilitation à Diriger des Recherches.
- [Mathieu-Colas, 1994] Mathieu-Colas, M. (1994). *Les mots à traits-d'union*. Didier Erudition, Paris.
- [Mathieu-Colas, 1998] Mathieu-Colas, M. (1998). Illustration d'une classe d'objets: les voies de communication. *Langages*, (131):77–90.
- [Mel'čuk, 1984] Mel'čuk, I. (1984). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico sémantiques I*. Les Presses de l'Université de Montréal, Montréal.

- [Mel'čuk, 1993] Mel'čuk, I. (1993). *Cours de morphologie générale*, volume 1. Les Presses de l'Université de Montréal CNRS, Montréal.
- [Mel'čuk, 1994] Mel'čuk, I. (1994). *Cours de morphologie générale*, volume 2. Les Presses de l'Université de Montréal CNRS, Montréal.
- [Mel'čuk, 1997] Mel'čuk, I. (1997). Vers une linguistique sens-texte. leçon inaugurale. page 68, Paris. Collège de France.
- [Mel'čuk, 2000] Mel'čuk, I. (2000). Un fou/une folle: un lexème ou deux? *Lexique, syntaxe et sémantique: Mélanges offerts à Gaston Gross à l'occasion de son 60 anniversaire*, (Numéro Hors série):95–106.
- [Mel'čuk et al., 1995] Mel'čuk, I., Clas, A., and Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Champs linguistiques. Duculot.
- [Polguère, 1998] Polguère, A. (1998). La théorie Sens-Texte. *Dialangue*, (8-9):9–30.
- [Polguère, 2003a] Polguère, A. (2003a). *Lexicologie et sémantique lexicale. Notions fondamentales*. Collection Paramètre. Presses de l'Université de Montréal.
- [Polguère, 2003b] Polguère, A. (2003b). Étiquetage sémantique des lexies dans la base de données dico. *Traitement automatique des langues (TAL)*, 2(44):39–68.
- [Sabah, 1990] Sabah, G. (1990). *L'intelligence artificielle et le langage*, volume 1. Hermes, Paris, 2 édition.
- [Sastre, 2005] Sastre, J. (2005). Vers un dictionnaire électronique du catalan. Université de Franche-Comté (Besançon). Journées Intex/NooJ.
- [Sastre, 2006] Sastre, J. (2006). Paradigmes de flexion et traits syntactico-sémantiques dans le dictionnaire électronique de catalan. Universitat de Valencia. Pour une linguistique plurielle VII Congreso Internacional de Lingüística francesa.
- [Silberztein, 1990] Silberztein, M. (1990). Le dictionnaire électronique des mots composés. *Langue Française*, (87):71–83.
- [Silberztein, 1993] Silberztein, M. (1993). *Dictionnaires électroniques et analyse automatique de corpus: Le système INTEX*. Masson, Paris.
- [Silberztein, 2006] Silberztein, M. (2006). *NooJ's Manual*.
- [Solà and Rigau, 2002] Solà, J. and Rigau, G. (2002). *Gramàtica del català contemporani*. Empúries, Barcelona.
- [Solé, 2002] Solé, E. (2002). *Els noms col·lectius catalans: descripció i reconeixement*. PhD thesis, Universitat Pompeu Fabra.



# Index

- actant, 14, 54  
    actant sémantique, 13, 14, 49  
    actant syntaxique, 14  
action, 49, 53, 57  
adjectif, 7, 9, 27, 28, 35–37, 41, 42, 52, 57, 66, 96, 97, 106, 110  
adverbe, 9, 23, 26, 97  
argument, 9, 11, 14, 47, 49, 52, 75, 76  
argument élémentaire, 48  
article, 104  
aspect, 51, 52, 55, 58  
automate, 6–8, 29, 99  
  
catégorie grammaticale, 7, 25, 27, 28, 96  
classes d'objets, 61  
classes d'objets, 11, 12, 34, 66, 76, 78–80, 82  
conjonction, 26  
  
dérivatème, 35, 72, 73  
dérivation, 36, 38, 53, 70, 94, 97, 98  
diasystématique, 10, 39, 40  
dictionnaire électronique, 5, 10, 12, 29, 30, 49, 91  
  
entité, 48, 66, 72  
état, 49, 57  
évènement, 49, 64, 65  
  
féminin, 37, 40, 51, 60, 62, 67, 96  
flexion, 29, 30, 32–36, 41, 42, 44, 96  
fonction lexicale, 14, 18, 26, 35, 38, 39, 78  
  
genre, 32, 36, 37, 40  
grammème, 35  
graphe, 6, 7, 10, 17, 29, 33, 36, 37, 41, 42, 93, 94, 96, 97, 99, 102  
  
holonyme, 67  
homonymie, 61, 84  
hyperonyme, 54  
  
interjection, 26  
Intex/NooJ, 6, 23, 29, 30, 36, 84, 91, 94, 99  
  
lexème, 13, 14, 34, 37  
lexie, 13, 14, 16–18, 25, 39, 49, 58, 69, 76  
  
méronyme, 67  
macrostructure, 12, 15, 23, 24, 94  
masculin, 37, 40, 51, 60, 62  
microstructure, 13, 15, 23, 24, 28, 94  
morphologie, 33, 35, 70, 73, 94, 96, 99  
  
nom, 26, 27, 32, 36  
nom collectif, 65–70, 73  
nombre, 29  
  
opérateur, 48  
  
paraphrase, 16–18  
partie du discours, 25, 32  
phrasème, 13  
    phrasème complet, 15  
    quasi-phrasème, 15, 38  
    semi-phrasème, 38  
phrase, 9, 48, 50, 53

- plausibilité, 24, 83, 84
- pluriel, 32–35, 66, 67
- polysémie, 11, 35, 40, 53, 61, 65, 69, 84
- prédictat, 9–12, 47–52, 57–63, 65–67, 69, 76, 78, 82
- prédictat approprié, 11, 65, 76, 78, 79, 82
- prédicats, 78
- préposition, 66, 68
- pronom, 35, 62, 94, 98, 99, 104, 107
  
- sémantème, 17
- sémantique, 17, 33, 35, 36, 40, 47, 74
- signifié, 33
- signifiant, 33, 64
- signification dérivationnelle, 72
- signification flexionnelle, 35, 36
- signification grammaticale, 35, 42
- singulier, 33, 34, 66, 67
- substantif, 9, 26, 32, 75
- syntaxe, 62, 98
- syntaxique, 44, 68
  
- trait syntactico-sémantique, 10, 11, 34, 47, 53–55, 57, 58, 61, 62, 70, 75, 76
  - abstrait, 34, 35, 48–51, 54–60, 64, 65
  - animal, 36, 40, 62
  - AnlColl, 69, 70, 73
  - concret, 35, 53, 54, 58–60, 82
  - ConcColl, 72
  - humain, 34, 36, 40, 61
  - HumColl, 69, 73
  - locatif, 63–66, 69, 72
  - végétal, 60, 61
  - VegColl, 69, 70, 72, 73
- transducteur, 41, 93, 94
  
- unité lexicale, 17, 24, 36, 37, 40, 61, 69, 76, 79, 98, 104
- verbe, 7, 23, 35, 38, 39, 42, 44, 45, 52, 59, 94, 105
- verbe support, 9, 12, 50, 58, 60
- vocable, 12, 15, 16