

DESCRIPTION DE LA STRUCTURE INTERNE ET DES
PROPRIÉTÉS DE COMBINATOIRE DES LOCUTIONS
POINTÉES DANS LE DiCo
ET CONSTRUCTION DU COMPLÉMENT DU DiCo

Par
BEN HAMOU Sabrina

Juin 2004

Rapport de stage de maîtrise (Université Grenoble 3)

Directeur de stage : Alain Polguère

Fait à l'Observatoire de linguistique Sens-Texte (OLST)
Département de linguistique et de traduction
Université de Montréal

1. INTRODUCTION.....	1
1.1 Contexte et objet de l'étude.....	1
1.2 Méthodologie utilisée et structure du présent rapport.....	2
2. APPROFONDISSEMENT DE LA NOTION DE LOCUTION.....	3
2.1. La locution en tant qu'unité lexicale.....	3
2.2 Locutions étudiées.....	5
2.3 Structures syntaxiques des locutions étudiées.....	7
2.3.1 Locutions nominales.....	7
2.3.2 Locutions adverbiales.....	9
2.3.3 Locutions adjectivales.....	10
2.3.4 Locutions verbales.....	11
2.3.5 Autres parties du discours.....	12
3. CARACTÉRISATION DES LOCUTIONS DU CORPUS.....	12
3.1 Caractérisation de la structure interne.....	12
3.1.1 Tableau récapitulatif.....	12
3.1.2 La particularité des locutions propositionnelles.....	14
3.2 Combinatoire des locutions.....	17
3.2.1 Locutions nominales.....	17
3.2.2 Locutions propositionnelles.....	18
3.3 Corrélation entre la structure interne et la.....	19
combinatoire des locutions.....	19
3.3.1 Locutions nominales.....	20
3.3.1.1 Hypothèses générales.....	20
3.3.1.2 Possible corrélation entre la structure et les.....	21
hypothèses émises.....	21
3.3.2 Locutions propositionnelles.....	22
4. CONSTRUCTION DU COMPLÉMENT DU DiCo.....	23
4.1 Caractérisation de la base construite.....	23
4.1.1 Étiquettes pour les locutions nominales.....	24
4.1.2 Étiquettes pour les locutions propositionnelles.....	25
4.2 Statistiques sur les données.....	27
4.2.1. Statistiques générales.....	27
4.2.1.1 Locutions nominales.....	27
4.2.1.2 Locutions propositionnelles.....	31
4.2.2 Corrélation entre structure et traits.....	35
4.2.2.1 Récapitulation des hypothèses et liaison avec la.....	35
structure syntaxique.....	35
4.2.2.2 Résumé des résultats.....	39
5. CONCLUSION.....	41
6. BIBLIOGRAPHIE.....	45
ANNEXES : Traits relevés pour chaque structure syntaxique.....	46
1. Locutions nominales.....	46
2. Locutions propositionnelles.....	48

1. INTRODUCTION

1.1 Contexte et objet de l'étude

Le present travail à été effectué dans le cadre du DiCo (acronyme pour Dictionnaire de Cooccurrences), projet développé depuis plusieurs années à l'Observatoire de linguistique Sens-Texte (OLST), par I. Mel'cuk et A. Polguère. L'objectif de ce projet est l'élaboration d'une base de données lexicales du français où chaque lexie est décrite selon deux axes : les dérivations sémantiques qu'elle entretient avec les autres lexies de la langue et les collocations qu'elle contrôle. Pour une description plus complète du DiCo, nous invitons à consulter *L'Introduction à la Lexicologie Explicative et Combinatoire* (ILEC) [MEL 95].

Le DiCo « s'inspire » du Dictionnaire Explicatif et Combinatoire (DEC) [MEL 95] où chacune des entrées (lexie) est soit un lexème soit un phrasème. Brièvement, un lexème est un « mot », au sens de mot graphique, et un phrasème est une expression linguistique. Selon l'ILEC, chacune des lexies du DEC :

est prise dans une acception bien déterminée et munie de tous les renseignements qui spécifient totalement son comportement dans un texte.

Notre étude porte sur les locutions. Une locution est une unité fonctionnelle composée de plusieurs mots graphiques, appartenant à la langue et devant être apprise en tant que forme globale non divisible (il s'agit donc d'un phrasème). Elle est soumise aux mêmes règles qui s'appliquent aux lexies simples, de manière à assumer la fonction d'unité apte à être reprise pour être intégrée dans une structure plus importante. Elle se comporte comme une lexie simple mais possède cependant des traits propres et différents d'une locution à une autre. Ces deux caractéristiques constituent les points majeurs de notre étude qui se concentrera plus particulièrement sur la structure syntaxique et les propriétés de combinatoire des locutions.

1.2 Méthodologie utilisée et structure du présent rapport

En premier lieu, il est nécessaire de définir correctement la notion de locution, dans le cadre de la théorie Sens-Texte, et de repérer dans le DiCo l'ensemble des locutions pointées (dans la zone Phraséologique de la lexie). Une fiche du DiCo se compose de onze champs dont un, Ph, qui indique les pointeurs vers les phrasèmes liés à la lexie et qui « contiennent » cette lexie. Prenons l'exemple (schématisé) de la lexie **ASSIETTE a**.

```
<Nom> ASSIETTE
...
<es> Pièce de vaisselle/récipient
...
<ph> _assiette anglaise_
      _pique-assiette_
      ...
```

Chacun des phrasèmes de la zone Ph de la lexie **ASSIETTE a** doit faire l'objet d'une fiche DiCo particulière.

La notion de locution ayant été définie, un regroupement des locutions par parties du discours et par structure syntaxique (sous forme de structure de dépendance) permet d'établir des sous-classes.

Par la suite, nous avons étudié le comportement de chacune des locutions, premièrement, en observant leur combinatoire générale (similaire à celle des lexies simples) et deuxièmement, en tentant de mettre en évidence les caractéristiques propres aux locutions. L'objectif est de mettre en avant une éventuelle corrélation entre les caractéristiques particulières des locutions, notamment au niveau de la structure syntaxique, et leurs propriétés de combinatoire.

Enfin, nous avons élaboré un complément du DiCo pour les locutions traitées.

Le plan du présent rapport sera donc le suivant :

Section 2 : Approfondissement de la notion de locution

Section 3 : caractérisation des locutions

Section 4 : Construction du complément DiCo

Section 5 : Conclusion

Section 6 : Bibliographie

Nous mettrons en annexe l'ensemble des caractéristiques relevées pour chacune des structures syntaxiques des locutions nominales et propositionnelles.

2. APPROFONDISSEMENT DE LA NOTION DE LOCUTION

2.1. La locution en tant qu'unité lexicale

Dans l'Introduction à la Lexicologie Explicative et Combinatoire, la lexie (décrite dans un article de dictionnaire) est définie de la manière suivante :

Une lexie est une expression qui ne peut être décrite complètement par la grammaire c'est-à-dire par des règles générales mais qui nécessite une description individuelle donc une description par un article de dictionnaire. Une lexie est soit un lexème soit un phrasème ».

Ce dernier étant :

Une locution prise dans une seule acception bien déterminée et munie de tous les renseignements qui spécifient totalement son comportement dans un texte.

Il existe trois types de phrasèmes au sein de l'ensemble des expressions idéomatiques. Nous en donnerons la définition succincte et nous nous contenterons, pour plus d'informations, de renvoyer à [MEL 03].

La signification des symboles utilisés est la suivante :

X = lexème X

'X' = sens de X

\subset = opérateur d'inclusion ;

'X' \subset 'Y' : sens de X est inclut dans sens de Y

\oplus = union linguistique

- Phrasème complet ou locution figée, locution idiomatique

Le sens du phrasème ne peut être retrouvé à partir des sens propres aux constituants.

⇒ AB 'S' où 'A' ∩ 'S' et 'B' ∩ 'S'

S n'inclut ni le sens 'A' ni le sens 'B'.

Ex : **Avant-goût** = sensation que procure l'idée d'un mal, d'un bien futur

Avant = Qui précède dans l'espace, le temps ou dans une hiérarchie.

Goût = saveur

⇒ 'Avant-goût' ≠ 'Avant' ⊕ 'Goût'

- Semi-Phrasème ou locution semi-figée

Il s'agit des collocations. Ces structures sont composées de deux éléments dont un (la base) qui garde son sens. Le deuxième élément (le collocatif) ne garde pas son sens originel et est sélectionné en fonction de la base. Plusieurs collocatifs peuvent exister pour une base donnée.

⇒ AB 'S' où 'A' ∩ 'S' et 'B' ∩ 'S'

Ex : **Colère noire / bleue** = une colère violente, terrible.

Une *colère noire / bleue* est une « sous-classe » de *colère*. On note que *noire / bleue* ne garde pas son sens basique mais fonctionne comme un intensifieur de *colère*.

- Quasi-Phrasème ou locution quasi-figée

Le sens du phrasème est formé sur la base des sens des mots qui le composent mais aucun de ces sens n'est dominant dans le tout sémantique. Par ailleurs, le sens de la locution est plus riche que la somme des sens des lexèmes.

⇒ AB 'S' où 'A' ∩ 'S' et 'B' ∩ 'S' ; 'S' ≠ 'A' ⊕ 'B'

S n'est pas une sorte de A ni une sorte de B.

Ex : **point-virgule** = est formé d'une virgule et d'un point mais c'est avant tout un signe de ponctuation particulier (et non un point particulier ou une virgule particulière).

Cette classification des phrasèmes, propre à la théorie Sens-Texte, offre une trichotomie non ambiguë. En passant outre ces définitions, des erreurs sur la nature phraséologique de certaines structures peuvent apparaître. C'est le cas notamment avec l'expression COUPES SOMBRES qui, bien que répertoriée en tant que phrasème complet [REY 97], est un semi-phrasème. On parle en effet de COUPES SOMBRES *dans le budget*, pour désigner la réduction importante du budget, mais aussi de COUPES *dans le budget*. SOMBRES est un collocatif de COUPES, dont il intensifie le sens. COUPES SOMBRES est donc une collocation.

Les phrasèmes complets et les quasi-phrasèmes forment l'ensemble des locutions et ce sont ces deux types de phrasèmes qui nous intéressent ici dans la mesure où ils nécessitent des entrées distinctes de dictionnaire ; nous laisserons donc de côté le vaste sujet des collocations. On note tout de même que ces dernières sont encodées dans le Dictionnaire Explicatif et Combinatoire à l'aide des fonctions lexicales.

Ce bref aperçu permet de rendre compte des caractéristiques générales des locutions. Nous pouvons, maintenant, nous pencher sur l'élaboration de notre corpus d'étude.

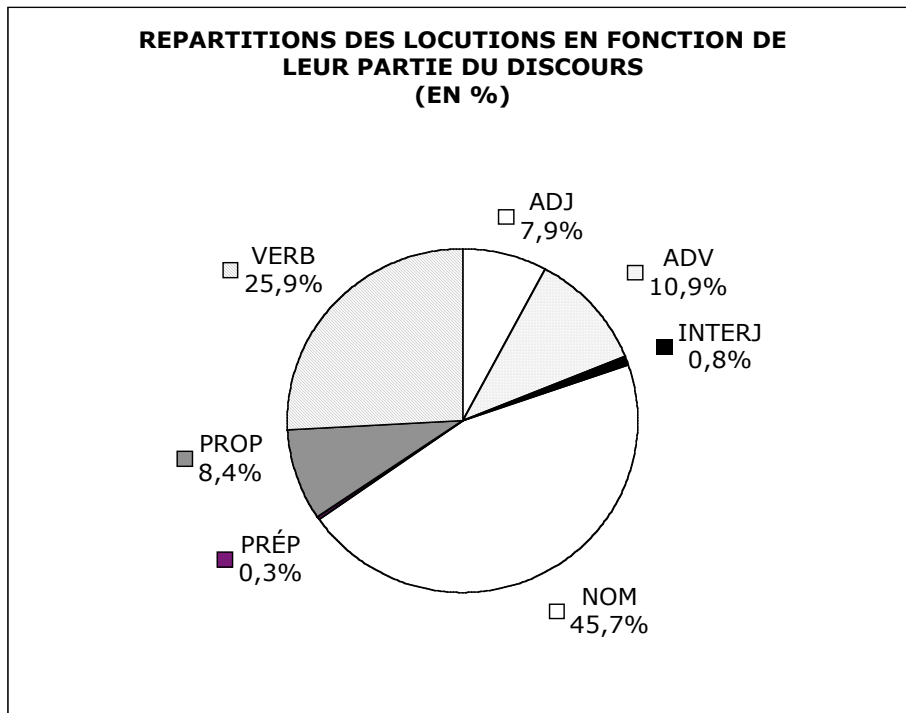
2.2 Locutions étudiées

Le corpus est basé sur l'ensemble des locutions pointées dans le DiCo c'est à dire celles qui n'ont pas leur propre entrée dans le dictionnaire. La locution pointée est simplement notée dans la zone phraséologique des lexies qui apparaissent dans la locution en question. Le corpus est constitué de 878 locutions qui se répartissent inégalement selon leur partie du discours. Les parties du discours relevées sont les suivantes:

- Locutions adjectivales (loc adj)
- Locutions adverbiales (loc adv)

- Locutions Interjectives (loc interj)
- Locutions nominales (loc nom)
- Locutions prépositionnelles (loc prép)
- Locutions Propositionnelles (loc prop)
- Locutions verbales (loc verb)

Le schéma ci dessous permet de rendre compte de la répartition des locutions dans le corpus d'étude:



Pour notre étude nous nous intéressons aux locutions nominales et propositionnelles. Mises ensemble, ces locutions représentent plus de 50 % de l'ensemble. Nous avons choisi les locutions nominales parce qu'elles sont le plus répandues dans le corpus. Des études sur ce type de locutions pourrait s'avérer très pertinentes. Les locutions propositionnelles sont, elles, intéressantes dans la mesure où très peu d'études ont été faites les concernant. Par ailleurs, la description de leur structure interne et de leur combinatoire pose des problèmes particuliers.

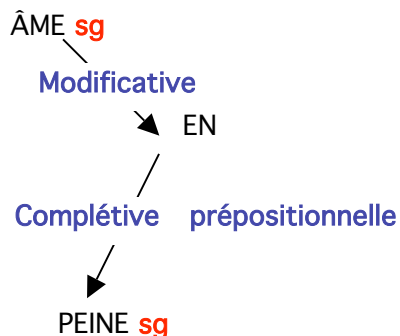
On note, néanmoins que les structures syntaxiques ont été établies pour les 878 locutions. La sélection des locutions nominales et propositionnelles ayant été faite pour la caractérisation des locutions et les propriétés de combinatoire de ces dernières.

2.3 Structures syntaxiques des locutions étudiées

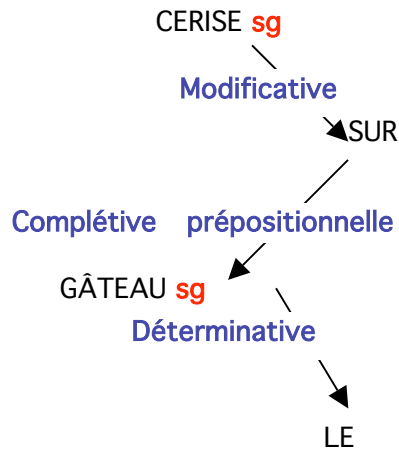
La tête d'une locution appartient (souvent mais pas tout le temps) à la même partie du discours que la locution elle-même. Ainsi, une locution nominale sera souvent gouvernée par un nom, une locution prépositionnelle par une préposition. Les structures syntaxiques peuvent varier au sein d'une même partie du discours. Nous ne citerons que quelques structures (accompagnées d'exemples), sous forme d'arbres de représentation syntaxique de surface, en fonction de certaines parties du discours ; la section 3.1 présente en détail, les structures syntaxiques des locutions du corpus.

2.3.1 Locutions nominales

Structure du type N->prép->N (structure la plus répandue des locutions nominales)



Accablé par la douleur, il errait comme une âme en peine.



Être parmi les trois premiers c'est déjà bien mais être le grand vainqueur, ça c'est la cerise sur le gâteau.

Un cas particulier pour les locutions nominales : Prép->N

La locution nominale CONTRE LA MONTRE a une structure syntaxique particulière dans la mesure où la tête est une préposition. On retrouve de manière éparse ce genre de construction mais la forme graphique de ces noms implique toujours soit un trait d'union, entre la préposition et le nom, soit l'agglutination de ces deux éléments. Il semblerait que la norme admette, certes, ce genre de structure atypique mais elle la « fige » fortement de sorte que la préposition et le nom ne forme qu'une seule entité indivisible. Le cas de CONTRE LA MONTRE, structure syntaxique plus « libre », pourrait être dû au fait que la locution est encore nouvelle. En effet, cette dernière est une forme condensée de la locution COURSE CONTRE LA MONTRE (structure de type **N->prep->N**) qui, ayant un sens différent désormais, est, à l'origine, propre au cyclisme. D'ailleurs on retrouve le nom CONTRE-LA-MONTRE qui désigne, de nos jours, une épreuve cycliste. Ce sens de CONTRE-LA-MONTRE, ayant un sens plus ciblé, est plus ancien que le sens CONTRE LA MONTRE, beaucoup plus général de surcroît.

Ci- dessous des exemple de l'utilisation des locutions COURSE CONTRE LA MONTRE et CONTRE LA MONTRE :

Il me reste un jour pour finir mon rapport ; c'est une véritable course contre la montre.

Trois minutes pour répondre à toutes ces questions c'est pire qu'un contre la montre.

Enfin, quelques exemples de noms ayant une structure calcifiée, dans la diachronie, du type **Prép->N** :

EN-CAS, EN-AVANT, PARDESSUS, APRES-GUERRE, AVANT-GUERRE, ENTRE-DEUX-GUERRES, POURBOIRE, POURPARLER, POURTOUR, SANS-FAUTE, ENTRE-JAMBES, ENTRE-BANDE, ENTRECHAT, SANS-ABRI, SANS-CŒUR etc.

2.3.2 Locutions adverbiales

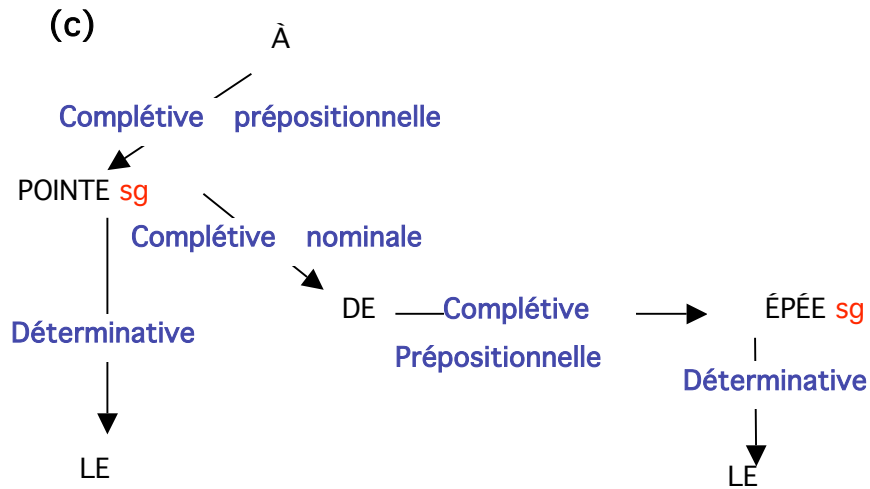
Structure du type adv->N. Cette structure, propre aux locutions adverbiales, est très peu répandue (a). Les locutions adverbiales, admettent plutôt une structure du type **prép->N** gouvernée par une préposition (b).



Le silence régnait en maître, il n'y avait pas un chat dans la ville.

Coiffée à la chien, les yeux limpides, elle ressemblait à un ange.

La structure (c) est un peu plus complexe mais est formée sur la même base que (b). Dans le corpus, elle ne concerne que quelques locutions (A LA POINTE DE L'ÉPÉE, A LA VIE A LA MORT...)

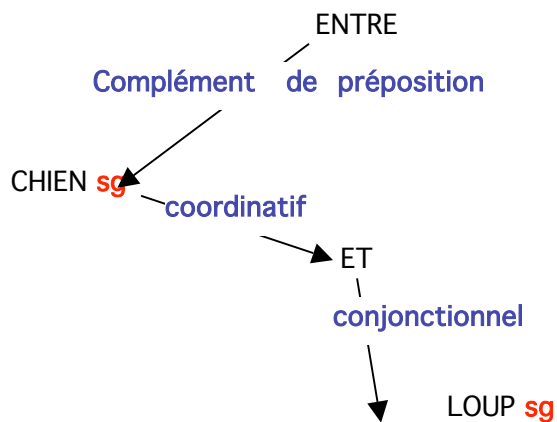


Le courageux guerrier l'emporta à la pointe de l'épée.

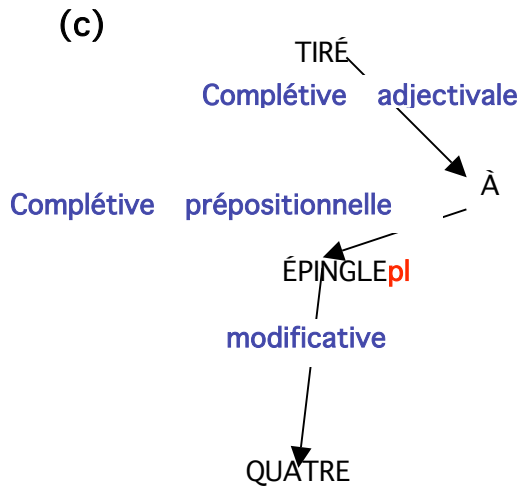
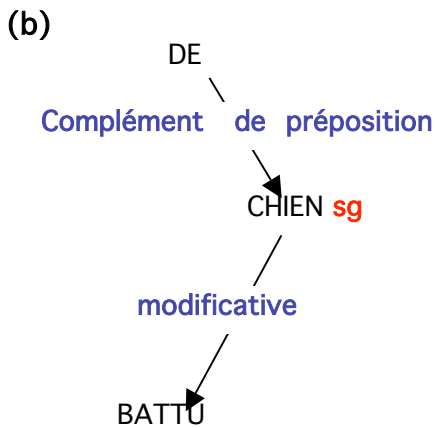
2.3.3 Locutions adjectivales

Structure du type prép->N. La structure gouvernée par une préposition est très répandue pour ce type de locutions (b) et (a). La structure (c), où la tête est un adjectif, est relativement rare au sein du corpus

(a)



A l'heure d'entre chien et loup, il ne distinguait plus ses camarades de guerres de ses ennemis.

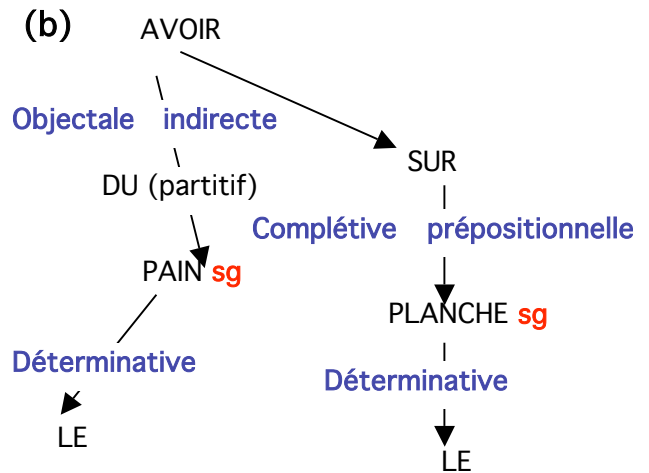
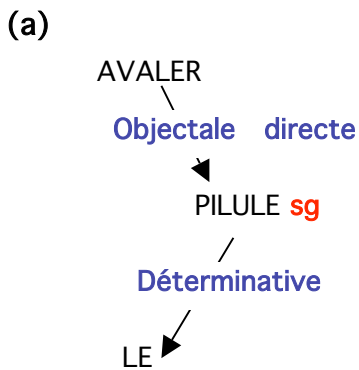


Avec sa mine de chien battu, tout le monde cède à ses caprices.

Le banquier tiré à quatre épingle, la dévisageait, l'écrasant de tout le poids de son regard.

2.3.4 Locutions verbales

La structure (a) correspond au cas où le verbe à un complément et (b) correspond au cas où le verbe à deux (ou plus) compléments.



La mauvaise nouvelle était tombée. Les larmes aux yeux, ils avalèrent la pilule sans protester.

“Onze heures quarante-huit et ils nous restent trois pièces à peindre. Allez les gars ! On a du pain sur la planche !”

2.3.5 Autres parties du discours

Les locutions interjectives ont une structure gouvernée par un nom dans la majeure partie des cas (NOM D’UN CHIEN, CEINTURE !, LA BARBE ! *etc.*). Quant aux locutions prépositionnelles, la tête est majoritairement une préposition.

La structure des locutions propositionnelles sera identifiée sous un autre angle. En effet, nous intégrerons un autre critère, les notions de structures binaires et non binaires (voir section 3.1.2).

3. CARACTÉRISATION DES LOCUTIONS DU CORPUS

3.1 Caractérisation de la structure interne

3.1.1 Tableau récapitulatif

L’étude des locutions a permis de dégager pour chacun des types de locutions, la modélisation de la structure arborescente par une chaîne de parties de discours. Cette modélisation par chaîne de parties du discours est représentée sous forme de sous-classes répertoriées dans le tableau I ci-dessous.

Les sous-classes sont sélectionnées de manière à ce qu’elles discriminent assez bien l’ensemble des locutions tout en évitant les distinctions « abusives », qui n’auraient pas d’utilité. Ainsi, la sous-classe **Prèp N ConjCoord N** admet les structures exclusivement composées d’une préposition suivie d’un nom, d’une conjonction de coordination, d’un nom mais elle admet aussi les mêmes structures où un ou plusieurs articles ont été insérés. Ce que l’on tente de mettre en avant c’est avant tout les éléments les plus intéressants

de la structure syntaxique. Il s'agit le plus souvent de la tête et de quelques « satellites » ; les plus importants étant énoncés dans la structure, les moins importants étant sous-entendus. Une seule contrainte, cependant, est introduite dans le système des sous-classes ; le signe + qui indique que la structure décrite n'admet aucun modifieur ou article quelconque. En d'autres termes, les locutions faisant partie de ces sous-classes ont exactement la même structure que celle décrite par les sous-classes en question.

tableau I : Structures syntaxiques par partie du discours

<i>Partie du discours</i>	Locutions Adjectivales	Locutions Adverbiales	Locutions Nominales	Locutions Interjectives
<i>Sous-classes associées</i>	Prép+N	Prép+N	N Prép N	Prép+N
			Prép N	
			N	
	Prép N	Prép N	N+Prép+N	Prép N
	N	N	N Adj	
	Conj N	Conj N	Adj N	
	Prép V N	Conj Prép N	N+Prép+N+ Prép N	N
	Prép N ConjCoord N	N ConjCoord N	V N	
	Adj Prép N	Prép V	N ConjCoord N	N Prép N
	Adj Prép+N	Conj Prop	N+ConjCoord +N	
Adv Prép N				
	Adv N			

<i>Partie du discours</i>	Locutions Prépositionnelles	Locutions Propositionnelles	Locutions Verbales
<i>Sous-classes associées</i>	Prép+N+Prép	Structure binaire	Un complément
	N+Prép	Structure non binaire	Plusieurs compléments

Adj	=	Adjectif
Adv	=	Adverbe
ConjCoord	=	Conjonction de coordination
Conj	=	Autre type de conjonction (subordination)
N	=	Nom
Prép	=	Préposition
V	=	Verbe

3.1.2 La particularité des locutions propositionnelles

La répartition **Structure Binaire/Non binaire** n'est valable que pour les locutions propositionnelles. Ces dernières sont particulières dans la mesure où elles sont extraites de l'oral, des proverbes véhiculés à travers les générations et propres à chaque culture. Le pouvoir mnémotechnique de la formule énoncée est directement lié à la « symbolique gestuelle et articulatoire » [HAG 85] qui émane de l'ensemble, symbolique dont le procédé général, dérivé de l'intonation, est la répétition. Source de structures symétriques, c'est cette répétition et par extension l'opposition des deux concepts de la structure que nous essayons de mettre au jour ici. La distinction des deux concepts se fait au niveau syntaxique et au niveau sémantique.

Exemples

Structure non binaire : C'EST DU SUICIDE

Structure binaire : peut être une structure phrastique standard (1), ou une juxtaposition de deux phrases (2).

(1) COEUR QUI SOUPIRE N'A PAS CE QU'IL DESIRE

(= *Une attitude plaintive indique un besoin non satisfait*)

L'opposition des deux concepts n'est accessible qu'au niveau sémantique.

(2) LES CHIENS ABOIENT, LA CARAVANE PASSE

(= *les cris, les protestations des envieux n'empêchent pas les grandes entreprises de se faire*)

L'opposition des concepts est aussi visible syntaxiquement.

Le tableau II ci-dessous donne pour chacune des locutions du corpus à structure binaire, les deux concepts qui « s'opposent ». Bin syn (code oui dans le tableau) indique que la structure est binaire syntaxiquement.

Les termes en italique représentent les cas où il y a répétition d'une même forme. En effet, les locutions du corpus concernées par cette répétition sont ambiguës dans la mesure où il est difficile de dégager et de nommer deux concepts opposés. Néanmoins, le sentiment du locuteur est, au niveau perceptif, que la structure semble binaire notamment par la répétition de certains constituants.

Pour certaines locutions, les deux sens sous-jacents n'ont pas été indiqués dans la mesure où il n'y avait pas d'opposition réelle. Il n'y avait pas non plus de répétition de formes. Ces locutions n'avaient, finalement, que la structure syntaxique que nous pouvions qualifier de binaire. Nous avons tenu à les intégrer au sein de ce tableau essentiellement pour cette raison.

Tableau II : Sens opposés dans les locutions binaires

locutions propositionnelles binaires (sémantiquement)	bin syn	sens 1	sens 2
À BON CHAT, BON RAT	Oui	attaquant	victime
À CHAQUE JOUR SUFFIT SA PEINE	Non	quotidien	tâche
C'EST LA FAIM QUI POUSSE LE LOUP HORS DU BOIS	Non	besoin	action
CHACUN SON MÉTIER ET LES VACHES SERONT BIEN GARDÉES	Oui	tâche	ce qui doit être
CHAT ÉCHAUDÉ CRAINT L'EAU FROIDE	Non	expérience	comportement
CŒUR QUI SOUPIRE N'A PAS CE QU'IL DÉSIRE	Non	plainte	besoin
DES GOÛTS ET DES COULEURS, ON NE DISCUTE PAS	Oui	jugement de valeur	ce qui doit être
IL FAUT LAISSER LES MORTS ENSEVELIR LES MORTS	Non	<i>morts</i>	<i>morts</i>
L'HOMME EST UN LOUP POUR L'HOMME	Non	<i>homme</i>	<i>homme</i>
LA FAIM FAIT SORTIR LE LOUP DU BOIS	Non	besoin	action
LA FORTUNE SOURIT AUX AUDACIEUX	Non	qqch de bien	comportement
LA FORTUNE VIENT EN DORMANT	Non	qqch de bien	comportement
LA NUIT, TOUS LES CHATS SONT GRIS	Non	situation	perception
LA PAROLE EST D'ARGENT ET LE SILENCE EST D'OR	Oui	comportement	comportement
LE MIEUX EST L'ENNEMI DU BIEN	Non	qqch pas bien	bien
LES AMIS DE MES AMIS SONT MES AMIS	Non	relation	relation
LES CHIENS ABOIENT, LA CARAVANE PASSE	Oui	comportement	comportement
LES MAUVAIS OUVRIERS ONT TOUJOURS DE MAUVAIS OUTILS	Non	type personne	comportement
LES PETITS RUISSEAUX FONT LES GRANDES RIVIÈRES	Non	état	devenir
LES PLAISANTERIES LES PLUS COURTES SONT LES MEILLEURES	Non	comportement	ce qui doit être
MIEUX VAUT UN SAGE ENNEMI QU'UN SOT AMI	Non	relation	relation
NE PAS PARLER DE CORDE DANS LA MAISON D'UN PENDU	Non	instrument du malheur	personne victime
ON NE PREND PAS LES MOUCHES AVEC DU VINAIGRE	Non	personne	action
PLUS Adj QUE N, TU MEURS!	Oui	—	—
QUAND LE CHAT N'EST PAS LÀ, LES SOURIS DANSENT	Oui	situation	comportement
QUAND ON PARLE DU LOUP, ON EN VOIT LA QUEUE	Oui	situation	événement
QUI S'EXCUSE S'ACCUSE	Non	comportement	comportement
QUI VEUT NOYER SON CHIEN L'ACCUSE DE LA RAGE	Non	désir	comportement
SI ON LUI PRESSAIT LE NEZ, IL EN SORTIRAIT DU LAIT	Oui	—	—
SI ON LUI TORDAIT LE NEZ, IL EN SORTIRAIT DU LAIT	Oui	—	—
TANT QU'IL Y A DE LA VIE, IL Y A DE L'ESPOIR	Oui	vie	espoir
UN SOT TROUVE TOUJOURS PLUS SOT QUI L'ADMIRE	non	Type de personne	situation

Cette sous-section a permis de rendre compte de certaines caractéristiques structurales des locutions du corpus. La prochaine étape concerne, désormais, l'étude de la combinatoire des locutions nominales et propositionnelles.

3.2 Combinatoire des locutions

3.2.1 Locutions nominales

La combinatoire des locutions nominales présente deux aspects ; d'une part elle concerne la combinaison avec les grammèmes des lexies nominales en général et d'autre part elle reflète la flexibilité des locutions.

Les noms se combinent avec des grammèmes flexionnels Singulier et Pluriel. Ils possèdent, aussi, des marqueurs lexicaux (articles) du trait indéfini ou défini. Nous avons « testé » ces grammèmes pour les locutions nominales du corpus. Ainsi certaines locutions sont au singulier d'autres sont au pluriel, d'autres encore ne sont qu'au défini ou à l'indéfini. La majeure partie d'entre elles admettent l'ensemble de ces combinaisons. Au niveau de la flexibilité, nous avons testé la possibilité d'insérer des adjectifs ou des adverbes à l'intérieur même de la structure. Nous verrons plus loin que la plupart des locutions nominales n'autorisent pas l'insertion d'éléments. Enfin, nous avons testé des combinaisons avec des modificateurs adjectivaux non plus insérés à l'intérieur de la structure mais antéposés ou postposés à la locution.

Voici quelques exemples de locutions nominales avec leurs traits de combinatoire respectifs :

CERISE SUR LE GATEAU : défini, singulier, sans insertion possible.

BIJOUX DE FAMILLE : pluriel, sans insertion

VIE SAUVE : défini, singulier, sans modification, sans insertion

APPEL AU MEURTRE : aucune particularité

Chacune de ces caractéristiques est testée et attribuée aux locutions nominales, sous formes de traits dans la base du DiCo. Nous verrons dans la section 4 quel genre « d'étiquettes » formalisées nous utiliserons pour rendre compte de ces caractéristiques.

3.2.2 Locutions propositionnelles

Le concept de combinatoire est difficilement appréciable dans le cas des locutions propositionnelles dans la mesure où ces structures sont des phrases autonomes. Néanmoins nous pouvons parler de combinatoire, notamment, avec l'insertion dans la structure de modifieurs ou, pour les cas particuliers, de conjonctions de coordinations ; on parlera plus de substitution avec en particulier le cas (4) et (4') ci-dessous.

Une partie non négligeable des locutions propositionnelles admet l'insertion de modifieurs. Ci-dessous, quelques exemples de locutions propositionnelles accompagnées de leurs caractéristiques :

- (1) LA FORTUNE SOURIT AUX AUDACIEUX : binaire, peut être modifiée
temporellement
- (2) L'HOMME EST UN LOUP POUR L'HOMME : binaire, possibilité d'insertion
- (3) FAITES CHAUFFER LA COLLE : pas de particularité, n'admet aucune
modification ni insertion
- (4) LA PAROLE EST D'ARGENT ET LE SILENCE EST D'OR : binaire, possibilité
d'insertion et de
modification temporelle.

Le trait « modification temporelle » a été rajouté en vue de caractériser un peu plus ces locutions. Nous avons tenté de tester, aussi, la mise au passif des structures propositionnelles mais cela n'est pas discriminant puisqu'aucune des locutions du corpus n'admet le passif.

Ci-dessous, voici les mêmes exemples de locutions pris précédemment avec certaines modifications (autorisées par l'ensemble des caractéristiques accompagnant les locutions) :

Une des modifications temporelles pour (1) serait la suivante :

(1') *La fortune **souriait** aux audacieux*

En cas d'insertion, (2) pourrait prendre la forme suivante :

(2') *L'homme est un **véritable** loup pour l'homme.*

Véritable joue ici le rôle d'intensifieur de la locution dans son ensemble.

On pourrait en effet traduire ce phénomène à l'aide de la fonction lexicale magn :

Magn(L'HOMME EST UN LOUP POUR L'HOMME)=*véritable*.

Véritable se positionneraient syntaxiquement avant *loup* .

Enfin, la locution (4), dans le cas d'une substitution de la conjonction de coordination par une autre, deviendrait :

(4') *la parole est d'argent **mais** le silence est d'or*

Cette simple substitution change le sens profond de la locution. Le contraste entre *parole* et *silence* et surtout au niveau de la valeur respective de chacun de ces termes semble beaucoup plus marqué. La valeur du *silence* n'en est que plus grande, plus précieuse, plus à même d'être le comportement adéquat.

3.3 Corrélation entre la structure interne et la combinatoire des locutions

Une des questions que nous pouvons nous poser, à ce niveau de l'étude, concerne l'existence d'une éventuelle corrélation entre la structure syntaxique des locutions étudiées et leurs propriétés de combinatoire. Par extension, nous cherchons à savoir si une structure syntaxique particulière rend compte d'une locution plus ou moins flexible ou, dans le cas inverse, désigne une locution totalement figée. Nous pouvons émettre certaines hypothèses non pas sur la corrélation directe entre structure et combinatoire mais sur certaines idées plus

générales. En effet, le lien direct entre structure et combinatoire ne peut être relevé qu'en étudiant le comportement des locutions grâce à la manipulation approfondie de la base de données (section 4).

3.3.1 Locutions nominales

3.3.1.1 Hypothèses générales

Les traits attribués aux locutions nominales concernent les grammèmes flexionnels singulier/pluriel et les marqueurs lexicaux indéfini/défini. Ils concernent aussi la possibilité de combiner la locution avec des modifieurs et d'insérer des éléments au sein même de la structure. Le comportement des locutions nominales (elles agissent comme des noms) pousse à penser que le nombre (singulier et pluriel) et la détermination (indéfini/défini) peuvent être utilisés librement l'un comme l'autre. Autrement dit, une locution nominale pourra indifféremment être mise au singulier ou au pluriel, être précédée d'un article défini ou indéfini. Celles qui n'autorisent pas ce genre de « transformations » feraient partie des locutions totalement figées ou plus « métaphoriques ». Les hypothèses **H1** et **H2** résument les suppositions émises ci-dessus :

H1 : *La plupart des locutions nominales peuvent avoir tous les traits indéfini, défini, pluriel, singulier et peuvent se combiner avec des modifieurs.*

H2 : *Les locutions les plus métaphoriques et les plus « riches » sémantiquement possèdent des traits qui ne peuvent être modifiés. Aucune combinaison avec un modifieur n'est autorisée.*

Au niveau de la flexibilité des locutions nominales, l'insertion de modifieurs à l'intérieur de la structure est difficilement réalisable. En effet, la discontinuité dans l'enchaînement des lexèmes casse le sens global. Cependant, si certaines locutions admettent l'insertion sans que le sens global en soit modifié, ces

locutions pourraient être les moins opaques sémantiquement. Ceci permet d'émettre les hypothèses suivantes :

H3 : *La plupart des locutions n'autorisent pas l'insertion d'éléments à l'intérieur de leur structure.*

H4 : *Celles qui admettent l'insertion sont les locutions les moins opaques sémantiquement.*

Désormais nous allons tenter d'établir un lien entre certaines structures syntaxiques relevées dans notre corpus d'étude et les quatre hypothèses émises ci-dessus .

3.3.1.2 Possible corrélation entre la structure et les hypothèses émises

La grande majorité des locutions n'a aucune contrainte de Combinatoire (H1). Cependant, rappelons que les structures syntaxiques repérées dans le corpus sont inégalement représentées dans les locutions. On ne peut donc pas affirmer que la structure syntaxique n'influence pas, du fait de l'inexistence de contraintes spécifiques dans la plupart des cas, la combinatoire lexicale des locutions. Il se peut donc qu'une corrélation entre structure et combinatoire soit envisageable. Pour mettre à jour cette corrélation nous nous baserons sur certaines structures syntaxiques (plus « parlantes » que d'autres) en nous fiant à notre connaissance générale de la langue et à des observations superficielles faites à partir base de données. Le tableau ci-dessous permet de rendre compte pour chacune des hypothèses, les structures les plus à même de correspondre à ces hypothèses.

H1	Prep N , NPrép N , Art N , N+Prép+N , N Adj , Adj N , N+Prép+N+Prép N , V N , N ConjCoord N , N+ConjCoord+N
H2	N+Prép+N
H3	Prep N , NPrép N , N , N+Prép+N , N Adj , Adj N , N+Prép+N+Prép N , V N , N ConjCoord N , N+ConjCoord+N
H4	N , N+Prép+N

Pour l'hypothèse H1 et H3 , nous supposons que toutes les structures ont autant de chances de correspondre à ces hypothèses. Pour H4, la structure du type Art N est la structure qui autorise le plus de liberté. Nous l'intégrons dans le tableau ci-dessus tout simplement parce que cette structure est naturellement et logiquement compatible avec H1. Il y a donc une quasi-certitude quant à la liaison entre la structure du type Art N et à la possible insertion intra-structure. Cependant du fait de cette évidence même, la structure du type Art N n'est pas pertinente car l'insertion n'est finalement qu'une des caractéristiques de cette structure. Nous verrons dans la section 4 que cette structure ne fera pas partie des statistiques.

Pour l'hypothèse H2, il est difficile de savoir quelles sont les structures concernées. En effet, il s'agit de notions qui sortent du cadre de la syntaxe et qui s'affirment dans les domaines sémantique et culturel.

Cependant, N+ Prép +N est la structure la plus répandue dans le corpus. Dans le cas de H4 et H2, cette dernière est sélectionnée pour des raisons d'effectif plus important.

3.3.2 Locutions propositionnelles

Nous avons caractérisé la structure des locutions propositionnelles selon le critère de binarité. Ce critère est valable au niveau syntaxique et au niveau sémantique. La combinatoire associée à la locution peut être liée à la structure (binaire ou non) de la locution. En effet, comme nous l'avons précédemment vu (section 3.1), la binarité refète un aspect de symétrie, de

répétition ; élément facilement identifiable au sein des proverbes et autres structures métaphoriques. Ces structures sont celles qui sont les plus opaques sémantiquement et les plus figées syntaxiquement. Elles sont donc celles qui autorisent le moins les modifications ou insertions et donc toute forme de combinaison. L'hypothèse **H5** ci-dessous découle naturellement :

H5 : *Les locutions à structure binaire sont les locutions propositionnelles qui sont les plus figées. Elles admettent moins de combinaisons possibles avec d'autres éléments lexicaux insérés à l'intérieur de la structure.*

Pour l'ensemble des locutions nominales et propositionnelles nous avons pu dégager cinq hypothèses sur le lien potentiel entre structure syntaxique et combinatoire lexicale. Nous allons voir dans la section 4 la validité de ces hypothèses en établissant, préalablement, une caractérisation de la base construite.

4. CONSTRUCTION DU COMPLÉMENT DU DiCo

4.1 Caractérisation de la base construite

Les 878 locutions ont été saisies dans une base de données à l'aide du logiciel FileMaker Pro. Pour chacune de ces locutions, la partie du discours a été indiquée (loc adj, loc adv, loc interj, loc nom, loc prép, loc prop ou loc verb). Des informations supplémentaires sur la structure syntaxique et la combinatoire lexicale ont été rajoutées pour les locutions nominales et propositionnelles. À chacune de ces caractéristiques correspond une « étiquette », rajoutée dans la zone cg (catégorie grammaticale) de la fiche du DiCo. Dans la présente étude, l'étiquette ne prend pas le sens qui lui a donné dans la théorie Sens-Texte. Elle correspond plus à la notion de trait ou de caractéristique.

4.1.1 Étiquettes pour les locutions nominales

En plus de la partie du discours (loc nom), des informations sur les grammèmes flexionnels de type Singulier ou Pluriel et sur les marqueurs lexicaux (articles) indéfini ou défini ont été formalisées. De la même manière, l'insertion potentielle de modificateurs, correspond à une caractéristique notée dans la zone cg de la fiche de la locution traitée. Ainsi si l'on reprend les exemples vu en 3.2.1 les caractéristiques propres à ces locutions seront modélisées de la manière suivante :

Déf/indéf = défini/indéfini

Sing/plur = singulier/pluriel

Sans modif = sans modificateurs (sous entendu antéposé ou postposé)

Sans insert = sans insertion (a l'intérieur de la locution)

Cerise sur le gâteau : défini, singulier, sans insertion possible.

CERISE SUR LE GÂTEAU : loc nom [N1], déf sing, sans insert

Bijoux de famille : pluriel, sans insertion

BIJOUX DE FAMILLE : loc nom [N2], plur, sans insert

Vie sauve : défini, singulier, sans modification, sans insertion

VIE SAUVE : loc nom [N3], déf sing, sans modif, sans insert

Appel au meurtre : aucune particularité

APPEL AU MEURTRE : loc nom [N2]

[Ni] est une étiquette qui « code » la structure syntaxique de la locution. Chacune des structures syntaxiques repérée pour les locutions nominales ont été transcrites en une étiquette de la forme [Ni]. La correspondance entre étiquette et structure syntaxique est exprimée comme suit:

[N0] = Art N

[N5] = V N

[N1] = N prép N

[N6] = N+conj+N

[N2] = N+prép+N

[N7] = N conj N

[N3] = N adj

[N8] = N+prép+N+prép N

[N4] = Adj N

[N9] = Prép N

4.1.2 Étiquettes pour les locutions propositionnelles

La structure des locutions propositionnelles a été identifiée selon le critère de binarité. Celui-ci a été introduit dans la base avec quelques informations complémentaires.

[bin] = binaire (d'une manière générale)

[bin.syn] = binaire syntaxiquement

[bin.syn.permut] = binaire syntaxiquement avec possibilité de permuter les constituants syntaxiques.

La modification temporelle a été modélisée par le trait [tps], tandis que l'insertion de modifieurs à l'intérieur de la locution est représentée par [insertion]. En reprenant les exemples pris dans la section 3.2.2, les traits associés aux locutions seront les suivants :

(1) *La fortune sourit aux audacieux* : binaire, peut être modifiée

Temporellement

LA FORTUNE SOURIT AUX AUDACIEUX : loc prop [bin], [tps]

(2) *L'homme est un loup pour l'homme* : binaire, possibilité d'insertion

L'HOMME EST UN LOUP POUR L'HOMME : loc prop [bin], [insertion]

(3) *Faites chauffer la colle* : pas de particularité, n'admet aucune modification ni insertion

FAITES CHAUFFER LA COLLE ! : loc prop

(4) *la parole est d'argent et le silence est d'or* : binaire, possibilité d'insertion et de modification temporelle.

LA PAROLE EST D'ARGENT ET LE SILENCE EST D'OR : loc prop [bin.syn], [insertion], [tps]

La locution (5) ci-dessous est la seule structure binaire syntaxiquement qui admet la permutation :

(5) *Des goûts et des couleurs, on ne discute pas* : binaire, possibilité de modification temporelle

DES GOÛTS ET DES COULEURS, ON NE DISCUTE PAS : loc prop [bin.syn.permut], [tps]

On note que la locution (5) n'est plus binaire syntaxiquement une fois qu'on la permute ((5') : *On ne discute pas des goûts et des couleurs*). Par ailleurs, il semblerait que la structure initiale de la locution ait été, justement, (5'). En effet, il y a eu antéposition du complément d'objet qui a, d'une part, entraîné la binarité syntaxique, et d'autre part, permis aux constituants syntaxiques d'être permutable. Cette permutation ne permet finalement « que » de retomber sur la version originale de la locution.

L'ensemble des étiquettes que nous venons de voir va nous permettre de manipuler les données dans le but de mettre en évidence certaines corrélations entre structure et combinatoire. Schématiquement, nous allons tenter de voir si

certaines étiquettes sont plus à même d’aller avec d’autres. Nous tenterons aussi de vérifier la validité de nos hypothèses émises en section 3.3.

4.2 Statistiques sur les données

4.2.1. Statistiques générales

4.2.1.1 Locutions nominales

a) Tableaux récapitulatifs

Tableau 1 : Répartition des locutions par structure syntaxique.

Structure	N0	N1	N2	N3	N4	N5	N6	N7	N8	N9	Total
Nb Locutions	4	42	205	107	31	1	2	3	4	1	400
En %	1	10,5	51,25	26,75	7,75	0,25	0,5	0,75	1	0,25	100

Rappel sur les structures syntaxiques :

[N0] = Art N

[N5] = V N

[N1] = N prép N

[N6] = N+conj+N

[N2] = N+prép+N

[N7] = N conj N

[N3] = N adj

[N8] = N+prép+N+prép N

[N4] = Adj N

[N9] = Prép N

Plus de 50% des locutions nominales du corpus ont une structure de la forme N2. Les structures du type N2 et N3 représentent 78% de l’ensemble des locutions.

Les tableaux qui vont suivre représentent la répartition des locutions en fonction des traits qui ont été définis dans la base de données. Les combinaisons de traits notées pour chaque trait sont celles qui sont les plus nombreuses avec le trait en question. Ainsi pour le trait Sing, la combinaison DEF+SING+SANS INSERT est la plus courante.

Rappel sur les traits :

Déf/indéf = défini/indéfini

Sing/plur = singulier/pluriel

Sans modif = sans modificateurs (sous entendu antéposé ou postposé)

Sans insert = sans insertion (a l'intérieur de la locution)

Nous mettrons ces traits en majuscule pour plus de visibilité lors de l'interprétation des tableaux ci-dessous.

Le signe + désigne ici la combinaison de traits. Déf+sing veut dire le trait Déf et le trait Sing.

Tableau 2 : Répartition des locutions en fonction du trait DEF/INDEF.

Nb A de locutions	DEF 39 soit 9,75% de l'ensemble des locutions	INDEF 2 soit 0,5% de l'ensemble des locutions	
Nb B de loc et % par rapport à Nb A	DEF+SING 34 soit 87,2 %	DEF+PLUR 5 soit 12,8 %	INDEF+SING 2 —
Nb C loc et % par rapport à Nb B	DEF+SING+SANS INSERT 25 soit 73,5%	DEF+PLUR+SANS INSERT 3 soit 80 %	INDEF+SING+SANS INSERT 2 —

Plus de 87 % des locutions ayant le trait DEF ont aussi le trait SING. Parmi ces locutions, 73,5 % n'admettent pas l'insertion (trait SANS INSERT). Ces dernières représentent finalement la majeure partie des locutions ayant le trait

DEF, soit approximativement 64% (rapport Nb C/Nb A). Le trait INDEF est peu répandu dans le corpus, les deux seules occurrences de ce trait sont combinées aux traits SING et SANS INSERT. De ce fait on ne peut pas faire de statistiques pertinentes pour ce trait.

Tableau 3 : Répartition des locutions en fonction du trait SING/PLUR.

Nb A de locutions	SING 42 soit 10,5% de l'ensemble des locutions	PLUR 13 soit 3,25% de l'ensemble des locutions	
Nb B Loc et % par rapport à Nb A	DEF+SING 34 soit 80,9%	PLUR+SANS INSERT 7 soit 53,8%	DEF+PLUR+SANS INSERT 4 soit 30,7%
Nb C loc et % par rapport à Nb B	DEF+SING+SANS INSERT 25 soit 73,5%		

Sur l'ensemble des locutions admettant le trait SING près de 60% d'entre elles ont aussi les traits DEF et SANS INSERT (rapport Nb C/Nb A). Le trait SING est trois fois plus fréquent que le trait PLUR. Ce dernier est plus souvent combiné avec le trait SANS INSERT.

Tableau 4 : Répartition des locutions en fonction du trait SANS MODIF

Nb A de locutions	SANS MODIF 5 soit 1,25% de l'ensemble des locutions
Nb B Loc et % par rapport à Nb A	DEF+SING+SANS MODIF+SANS INSERT 5 soit 100%

Les cinq locutions n'admettant pas de modifieurs, ont le trait DEF, SING et SANS INSERT. Il n'y a aucune occurrence du trait SANS MODIF seul ou sans au moins un des traits cités ci-dessus.

Tableau 5 : Répartition des locutions en fonction du trait SANS INSERT

Nb A de locutions	SANS INSERT 365 soit 91,25%	
Nb B Loc et % par rapport à Nb A	SANS INSERT (seulement) 316 soit 86,5%	DEF+SING+SANS INSERT 25 soit 6,8%

Le trait SANS INSERT est largement répandu dans le corpus et est une caractéristique de la quasi-totalité des locutions. Il est, par ailleurs, souvent la seule caractéristique des locutions pour près de 80% d'entre elles (rapport Nb B/NB A).

b) Résumé des tableaux

La plupart des locutions ont la seule caractéristique de ne pas autoriser l'insertion (80% des locutions nominales). Parmi ces locutions, la grande majorité de celles qui sont définies sont aussi au singulier et celles qui sont indéfinies sont toutes au singulier.

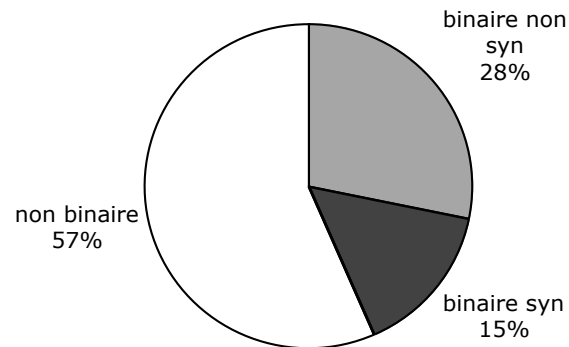
Les locutions qui sont au singulier sont majoritairement définies.

Les locutions au pluriel n'autorisent pas , dans la majeure partie des cas, l'insertion. Aucune locution qui est au pluriel est indéfinie.

Enfin, les locutions qui n'admettent pas de modificateurs sont toutes définies, au singulier et elles n'autorisent pas l'insertion.

4.2.1.2 Locutions propositionnelles

Le schéma ci-dessous permet de visualiser la répartition des locutions propositionnelles en fonction de leur trait BINAIRE.



Sur les 74 locutions propositionnelles du corpus 43% (soit 32 locutions) sont des locutions à structure binaire dont 15% sont binaires syntaxiquement. Dans les structures binaires syntaxiquement une seule locution admet la permutation des deux phrases juxtaposées qui l'a compose (voir section 4.1.2). Ce détail n'apparaît pas sur le graphique ci-dessus dans la mesure où il s'agit d'une structure unique, néanmoins, nous tiendrons compte de cette locution dans les statistiques générales.

Les tableaux ci-dessous représentent la répartition des locutions en fonction des traits BINAIRE, TPS, INSERTION. Le trait TPS désigne la possibilité de modifier temporellement la locution et le trait INSERTION indique que l'insertion de modificateurs, au sein de la structure de la locution, est autorisée.

a) Tableaux récapitulatifs

Tableau 1 : Répartition des locutions en fonction du trait BINAIRE

La combinaison de traits notée dans la troisième colonne correspond à la combinaison (ou à l'absence de combinaison) la plus fréquente pour les locutions ayant le trait de la colonne 2.

Nb A de loc	Nb B de loc et % par rapport au Nb A	Nb C de loc et % par rapport au Nb B
STRUCTURE BINAIRE 32 soit 43% de l'ensemble des locutions propositionnelles	BINAIRE NON SYN 21 soit 65,6%	BIN (seulement) 13 soit 61,9%
	BINAIRE SYN 10 soit 31,2%	BIN SYN (seulement) 7 soit 70%
	BINAIRE SYN PERMUT 1 soit 3,1%	BIN SYN PERMUT+TPS 1 –

La plupart des locutions binaires ne sont pas binaires syntaxiquement. Cependant on note que la binarité syntaxique concerne un tiers des locutions binaires en général. Dans le cas des locutions binaires non syntaxiquement, la plupart de ces locutions n'admet aucune modifications (61,9%). L'étude des locutions binaires syntaxiquement permet le même constat à hauteur de 70% des effectifs. Nous rattacherons le cas de la locution ayant le trait BINAIRE SYN PERMUT admettant la modification temporelle, au fait que cette dernière a subi une transformation (permutation des constituants syntaxiques) qui lui a donné cette structure binaire syntaxiquement avec permutation des constituants (voir section 4.1.2 exemple (5)). Par ailleurs cette locution semble moins opaque que les autres locutions binaires et plus particulièrement celles qui sont binaires syntaxiquement. Il serait, donc, intéressant de voir si le fait qu'une locution est plus intuitive, donc moins opaque sur le sens qu'elle véhicule, lui donne la caractéristique d'être « malléable » comme une structure phrastique quasi-standard. Les mots auraient, en effet, la signification qu'ils portent

originellement et il serait possible, selon les règles de grammaire générale, d'admettre permutation et autres types de transformations. Cependant une étude plus approfondie avec un ensemble plus vaste de locutions de ce type, est nécessaire pour pouvoir conclure, ou tout moins réfléchir sur la question.

Les locutions non binaires admettent « facilement » la modification temporelle et/ou l'insertion (30 locutions au total soit 71,3%). En effet, sur les 42 locutions non binaires, 10 (soit 23,8%) n'autorisent que la modification temporelle, 2 (soit 4,7%) n'autorisent que l'insertion d'un modifieur à l'intérieur de la structure et 18 (soit 42,8%) admettent les deux types de modifications.

Tableau 2 : Répartition des locutions en fonction du trait INSERTION

Nb A de loc	Nb B de loc et % par rapport au Nb A
INSERTION 24 soit 32,4% de l'ensemble des locutions propositionnelles	TPS+INSERTION 18 soit 75%

La combinaison de traits TPS+INSERTION est la plus répandue dans l'ensemble des locutions qui admettent l'insertion. Autrement dit, les trois quarts des locutions ayant le trait INSERTION ont aussi le trait TPS.

Tableau 3 : Répartition des locutions en fonction du trait
TPS

Nb A de loc	Nb B de loc et % par rapport au Nb A
TPS 39 soit 52,7% de l'ensemble des locutions propositionnelles	TPS+INSERTION 18 soit 46,1%
	TPS 10 soit 25,6%

La plupart du temps, le trait TPS est combiné au trait INSERTION. Si l'on regroupe les deux résultats de la colonne 2, on note que 71,8% des locutions ayant le trait TPS ne sont pas binaires. Nous verrons dans la sous-section 4.2.2 les corrélations éventuelles entre structure et traits associés.

b) Résumé des résultats

La plupart des locutions propositionnelles du corpus ne sont pas binaires. Celles qui sont binaires ne le sont pas syntaxiquement dans 70% des cas. Les locutions qui admettent l'insertion et qui sont non binaire admettent dans 75% des cas la modification temporelle. Le trait INSERTION (combiné ou non au trait BINAIRE) est très peu répandu (3 cas soit 12% des locutions qui admettent l'insertion). Parmi ces trois cas, deux ne sont pas binaires. Les locutions qui autorisent la modification temporelle, permettent le plus souvent l'insertion et sont non binaire. Quand les locutions admettent l'insertion il y a de fortes chances qu'elles admettent aussi la modification temporelle, que ces locutions soient binaires ou non.

4.2.2 Corrélation entre structure et traits

Cette sous-section a pour but de mettre en relief le lien entre structure et traits associés à cette structure. Pour ce faire nous utiliserons les hypothèses émises en section 3.3.

4.2.2.1 Récapitulation des hypothèses et liaison avec la structure syntaxique

Rappel sur les hypothèses émises :

Les hypothèses H1, H2, H3 et H4 concernent les locutions nominales tandis que l'hypothèse H5 est valable pour les locutions propositionnelles.

H1 : *La plupart des locutions nominales peuvent avoir tous les traits indéfini, défini, pluriel, singulier et peuvent se combiner avec des modificateurs.*

H2 : *Les locutions les plus métaphoriques et les plus « riches » sémantiquement possèdent des traits qui ne peuvent être modifiés. Aucune combinaison avec un modifieur n'est autorisé ; ces locutions se suffisent à elles-mêmes.*

H3 : *La plupart des locutions n'autorisent pas l'insertion d'éléments à l'intérieur de leur structure.*

H4 : *Celles qui admettent l'insertion sont les locutions les moins opaques sémantiquement.*

H5 : *Les locutions à structure binaire sont les locutions propositionnelles qui sont les plus figées. Elles admettent moins de combinaisons possibles avec d'autres éléments lexicaux qu'ils*

soient insérés à l'intérieur de la structure, antéposés ou postposé à cette dernière.

Hypothèses H1 et H3

Le regroupement des hypothèses H1 et H3, dans la mesure où toutes deux concernent la plupart des locutions, nous amène à repérer les locutions ayant pour seul trait SANS INSERT. Ces locutions n'ont donc aucune particularité sur le caractère défini ou non (DEF/INDEF), sur le nombre (SING/PLUR) ainsi que sur l'interdiction de combiner la locution avec des modifieurs (SANS MODIF) . Comme nous l'avons vu dans la sous-section 4.2.1, ces locutions représentent 79% de l'ensemble des locutions du corpus ; nous pouvons donc affirmer que les hypothèses H1 et H3 sont validées.

Hypothèse H2

L'hypothèse H2 sous-entend que la locution qui vérifierait cette hypothèse aurait au moins un trait du type DEF ou INDEF, SING ou PLUR. Le trait SANS MODIF est dans tous les cas combiné aux traits DEF, SING ; il n'est donc pas pertinent de le prendre en compte ici. Il nous faut donc rechercher les locutions admettant l'un de ces traits au moins (ou les deux traits simultanément). Ces locutions représentent 13,75% de l'ensemble (soit 55 locutions au total). Pour établir un lien entre la structure syntaxique des locutions et H2, il nous faut non pas fonctionner en terme de valeurs absolues mais en terme de valeurs relatives et ce par rapport au nombre de locutions par sous-classe syntaxique. Reprenons le tableau 1 de la section 4.2.1 qui représente la répartition des locutions par structure syntaxique.

Tableau 1 : répartition des locutions par structure syntaxique

Structure	N0	N1	N2	N3	N4	N5	N6	N7	N8	N9	Total
Nb Locutions	4	42	205	107	31	1	2	3	4	1	400
En %	1	10,5	51,25	26,75	7,75	0,25	0,5	0,75	1	0,25	100

Le tableau 9 ci-dessous rend compte de la distribution des locutions ayant les traits imposés par H2 en fonction de leur structure syntaxique (deuxième ligne du tableau 1).

Tableau 9 Distribution en fonction de la structure syntaxique selon H2

STRUCTURE SYNTAXIQUE	EFFECTIF BRUT	EFFECTIF RELATIF AU NOMBRE DE LOCUTIONS DE LA STRUCTURE	COMBINAISON DE TRAITS
[N0]	2	50%	DEF+PLUR DEF+SING
[N1]	15	35,7%	DEF+PLUR+SANS INSERT DEF+SING DEF+SING+SANS INSERT DEF+SING+SANS MODIF+SANS INSERT
[N2]	24	11,7%	DEF+SING DEF+SING+SANS INSERT PLUR PLUR+SANS INSERT SING+SANS INSERT
[N3]	6	5,6%	DEF+PLUR+SANS INSERT DEF+SING+SANS INSERT DEF+SING+SANS MODIF+SANS INSERT PLUR+SANS INSERT
[N4]	6	19,35%	DEF+SING+SANS INSERT INDEF+SING+SANS INSERT
[N6]	2	100%	DEF+PLUR+SANS INSERT SING+SANS INSERT

La troisième colonne du tableau est obtenue en divisant la colonne 2 par le nombre de locutions par structure syntaxique (deuxième ligne du tableau 1). On note que malgré le nombre élevé de locutions de type N2, seules 11,7% de ces locutions « valident » H2. Par ailleurs, toutes les structures du type N6

implique un trait sur la définition ou le nombre (voir annexe pour la correspondance entre structure et traits associés). La structure qui semble la plus intéressante à étudier est la structure N1. En effet, plus de 35% des locutions appartenant à cette structure admettent un trait DEF/INDEF ou SING/PLUR. Rappelons que cette structure est du type **N prép N** et voici quelques exemple de locution de ce type :

LA CERISE SUR LE GÂTEAU : *Ce qui termine en beauté.*

MALADIE DE LA VACHE FOLLE : *Encéphalopathie Spongiforme Bovine ; variante humaine : Maladie de Creutzfeld-Jakob.*

LA ROUE DE LA FORTUNE : *Symbole de la destinée humaine*

LA BAGUE AU DOIGT : *Avec promesse de mariage / avantage assuré*

On note l'aspect métaphorique de ces locutions qui implique une structure « complète ». En effet, l'article est souvent explicitement indiqué en début de locution et est tout le temps mis à l'intérieur de la structure. D'où une structure syntaxique du type N1.

Hypothèse H4

Les structures syntaxiques des locutions qui admettent l'insertion sont N1, N2, N3. Nous ne tiendrons pas compte de la structure N0 dans la mesure où il ne s'agit que d'un nom (accompagné d'un article) et non pas d'une structure plus complexe où il est possible d'insérer. L'insertion est obligatoirement admise.

Le nombre de locutions (qui admettent l'insertion) pour chacune des structures syntaxiques est le suivant :

N1 = 9 N2 = 16 N3 = 6

Rapporté au nombre de locutions par structure syntaxique, cela nous donne :

N1 = 21,4% N2 = 7,8% N3 = 5,6%

Hypothèse H5

L'hypothèse H5 implique qu'une locution propositionnelle binaire ne peut, dans la majeure partie des cas, accepter des modifications temporelles ou des insertions à l'intérieur de sa structure. Dans ce cas, ces locutions n'auraient que le trait BINAIRE (sous-entendu BIN / BIN SYN / BIN SYN PERMUT). En étudiant le tableau 6, nous avons pu voir que sur les trente-deux locutions binaires les effectifs des locutions ayant pour seul trait BIN ou BIN SYN sont respectivement de 13 et 7. Il y a donc 20 locutions binaires n'ayant pour seul trait que leur caractéristique binaire. Ce qui représente 62,5% de l'ensemble des locutions binaires . L'hypothèse H5 semble être validée.

4.2.2.2 Résumé des résultats

Il semblerait que les hypothèses (pour les locutions nominales) soient, grosso modo, vérifiées. Nous pouvons donc reprendre le tableau (ci-dessous) mettant en évidence la relation entre les hypothèses émises et les structures syntaxiques des locutions du corpus (Section 3.3.1) pour pouvoir réajuster cette relation.

H1	Prep N , NPrép N , N , N+Prép+N , N Adj , Adj N , N+Prép+N+Prép N , V N , N ConjCoord N , N+ConjCoord+N
H2	N+ Prép +N
H3	Prep N , NPrép N , N , N+Prép+N , N Adj , Adj N , N+Prép+N+Prép N , V N , N ConjCoord N , N+ConjCoord+N
H4	N , N+ Prép +N

Le tableau 10 ci-dessous nous donne en fonction des résultats obtenus le lien entre structure syntaxique et les hypothèses. Les pourcentages les plus forts sont en gras.

Rappel sur les structures syntaxiques

[N0] = N

[N5] = V N

[N1] = N prép N

[N6] = N+conj+N

[N2] = N+prép+N

[N7] = N conj N

[N3] = N adj

[N8] = N+prép+N+prép N

[N4] = Adj N

[N9] = Prép N

Tableau 10 : Corrélation entre structure syntaxique et hypothèses émises

H1	N0 N1 N2 N3 N4 N5 N7 N8 N9
H2	N0 N6 N1 N2 N3 N4
H3	N1 N2 N3 N4 N5 N7 N8 N9
H4	N1 N2 N3

La structure la plus hétérogène semble être N1. On la retrouve autant dans les cas de locutions très figées (H2 : n'admettant pas de modifications) que dans les cas de locutions les plus « libres » admettant l'insertion (H4). Dans le premier cas, 35,7 % des locutions du type N1 ont des traits d'ores et déjà définis et ne tolèrent aucune modification. Dans le second cas, 21,4 % de ces mêmes locutions admettent l'insertion.

La structure la plus « standard » est N2. Il s'agit de la structure la plus répandue dans le corpus. Elle constitue aussi une structure de « prédilection » pour les mots composés qui rapelons-le sont des locutions.

La seule différence entre les structures N3 et N4 est la position de l'adjectif. Ce « seul » aspect fait que les structures du type N4 ont plus tendance à être figées en ayant des traits du type DEF, INDEF, SING, PLUR

qui leurs sont propres (tableau 9 section 4.2.2). En revanche, les structures du type N3 peuvent admettre l'insertion.

Les autres structures sont très peu représentées dans le corpus ; il est donc difficile de pouvoir faire des corrélations fiables et avec certitude.

L'hypothèse pour les locutions propositionnelles semble, elle aussi vérifiée. L'aspect de binarité qu'inspire une structure propositionnelle, et notamment les sens que véhicule cette structure, font qu'il est difficile de pouvoir modifier la locution à sa guise, et particulièrement au niveau de l'aspect temporel et de l'insertion « intra-structure » de modificateurs. Nous avons vu que cela est vrai dans 62,5% des cas (section 4.2.2). Cependant nous avons aussi vu que la variation temporelle est la modification la plus courante dans le corpus et ce que les locutions soient binaires ou non. Au niveau des locutions binaires, dans presque tous les cas, la modification n'est admise que si au moins elle est temporelle. Sauf dans le cas de L'HOMME EST UN LOUP POUR L'HOMME. Or, rapelons que cette locution n'a pas, à proprement parler, une opposition de concepts (section 3.1.2). Il n'y a juste qu'une répétition de formes.

5. CONCLUSION

À travers cette étude, nous avons pu mettre à jour certaines caractéristiques, plus ou moins générales, d'une part en définissant la notion de locution à l'aide de la théorie Sens-Texte, et d'autre part en identifiant les structures internes des locutions du corpus d'étude.

Par la suite, notre étude s'est « spécialisée » lorsque nous nous sommes penchés sur la combinatoire lexicale et grammaticale des locutions nominales et propositionnelles, et sur la corrélation, éventuelle, entre structure interne et propriétés de combinatoire.

Grâce à de rapides statistiques sur la base de données, et à des hypothèses préalablement formulées, nous avons obtenu confirmation sur l'existence même d'une corrélation entre structure et combinatoire et nous

avons recueilli de précieuses informations sur les conditions nécessaires pour que cette corrélation soit observable.

Au niveau des locutions nominales nous nous sommes aperçu de plusieurs phénomènes qui se répartissent selon deux aspects d'étude ; l'un se basant sur les données brutes (valeurs absolues) et l'autre se basant sur les données relatives à d'autres données.

Selon le premier aspect, les informations essentielles à retenir sont premièrement, que les locutions nominales ont une structure de type **N+Prép+N** qui est représentée à plus de 50 %. Deuxièmement, la plupart des locutions nominales n'autorisent pas l'insertion et n'ont aucune autre contrainte.

Le second aspect, qui repose sur les données relatives à d'autres, nous informe que la structure la plus hétérogène semble être du type **N Prép N** dans la mesure où elle regroupe des locutions ayant un comportement plutôt figé et d'autres étant plus libres. Par ailleurs, la différence liée à la position de l'adjectif dans les structures de type **N Adj et Adj N**, fait que cette dernière tend à être plus figées (très peu de modifications sont permises). D'autant plus qu'il y a pratiquement quatre fois plus de locutions dans cet ensemble (comparativement à l'autre type de structure), qui ont des traits prédéfinis (Défini, singulier etc.). Enfin, la structure de type **N+Prép+N** est la structure standard dans le cas des locutions nominales. C'est aussi la structure que l'on retrouve assez fréquemment pour les mots composés.

Au niveau des locutions propositionnelles, et toujours selon l'aspect de valeurs absolues, l'insertion implique le plus souvent une possible modification temporelle. Il semblerait en effet, que l'insertion reste une modification plus difficilement observable que la modification temporelle, cette dernière étant plus autorisée, et ce mise à part le critère de binarité. Ce dernier intervient, cependant, lorsque l'on observe les données relativement au trait BINAIRE. Les structures binaires sont plus figées que celles qui ne sont pas binaires. Si modification il peut y avoir, celle ci est avant tout temporelle. Et si il y a

possibilité d'insertion, on note aussi que la structure peut être modifiable temporellement.

Au terme de cette étude les questions qui restent en suspens concernent notamment les structures les moins représentées dans le corpus d'étude (voir tableau 1 section 4.2.1.1). Quels sont leurs comportements ? Y a-t-il une corrélation entre leur structure et leurs propriétés de combinatoire ? Il est fort probable que l'on retrouve le même constat obtenu pour les autres structures d'autant plus que les structures peu représentées sont atypiques voire complexes. Mais pour ce faire, il est nécessaire de travailler sur un corpus plus vaste, permettant de dégager des données pertinentes tant en terme d'effectif qu'en terme de « qualité ».

REMERCIEMENTS

Je tiens à remercier mon directeur de stage, Alain Polguère, pour toute son attention, sa gentillesse et sa bonne humeur. Merci pour vos précieux conseils qui ont su si bien me guider. Merci de m'avoir donné l'opportunité d'approfondir mes connaissances. Au plaisir de re-travailler avec vous...

Un grand merci à l'OLST, laboratoire où règne la bonne humeur, et à tous ses membres pour leur gentillesse et pour cet accueil fort chaleureux. Merci à tous, chercheurs et étudiants, je garderai un souvenir agréable de ce passage dans les couloirs de l'Université de Montréal.

Enfin, je remercie les enseignants de l'université sthendam de Grenoble sans qui cet échange n'aurait pu être fait.

6. BIBLIOGRAPHIE

- [ALT 03] ALTMAN J., POLGUÈRE A., “*La Bdéf : base de définitions dérivée du dictionnaire explicatif et combinatoire*”, *Proceedings of the First International Conference on meaning-Text theory*, Paris, 2003.
- [GRU 03] DE GRUYTER W. “Levels of Dependency in Linguistic Description : Concepts and problems.” In : AGEL V., Eichinger L., EROMS H.-W., Hellwig P., Herninger H.J., Lobin H *Dependency and valency. An International Handbook of Contemporary research*, Vol 1, p 188-229, Berlin-New York, 2003.
- [HAG 85] HAGÈGE C. *L’homme de paroles : contribution linguistique aux sciences humaines*, Paris : Fayard, C, 1985.
- [MEL 95] MELC’UK I., CLAS A., POLGUÈRE A. *Introduction à la lexicologie explicative et combinatoire*, AUPELF-UREF/Duculot, Louvain-la-Neuve, 1995.
- [MEL 97] MEL’CUK I. *Vers une linguistique sens-texte*, Leçon inaugurale, Paris : Collège de France, 1997.
- [MEL 03] “Les collocations dans le dictionnaire” In : SZENDE T. *Les écarts culturels dans les dictionnaires bilingues*, p 19-64, Paris : Honoré Champion, 2003.
- [POL 00] POLGUÈRE A. “Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French”, EURALEX, p.517-527, 2000.
- [POL 03] POLGUÈRE A., “Étiquetage sémantique des lexies dans la base de données DiCo”, *Traitement Automatique des Langues*, volume 44, no 2, p.39-68, 2003.
- [REY 97] REY A., CHANTREAU S. *Le Robert dictionnaire des expressions et locutions*, Le Robert, Paris, 1997.

ANNEXES : Traits relevés pour chaque structure syntaxique

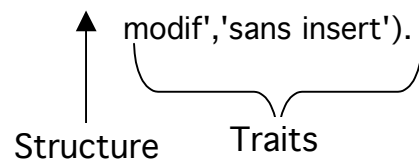
Nous introduisons ici la correspondance entre structure syntaxique, pour les locutions nominales, et structure binaire, pour les locutions propositionnelles, et les traits associés à ces structures.

Cet ensemble a été obtenu grâce à un programme prolog utilisant un fichier de données en format texte (fichenom pour les locutions nominales et ficheprop pour les locutions propositionnelles). La valeur nil indique que la valeur du « champ » du fichier correspondant est nulle. Pour chaque type de locution ci-dessous, la structure du fichier de données est indiquée avant l'ensemble de traits relevés.

1. Locutions nominales

Structure du fichier « fichenom »

fichenom('LA BAGUE AU DOIGT ', 'loc nom', '[N1]', 'def', 'sing', 'sans



Structure : ensemble de traits relevée

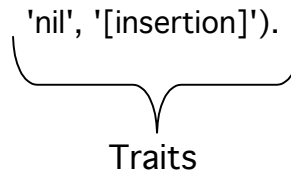
[N0] : def,sing,nil,nil
[N0] : nil,nil,nil,nil
[N1] : def,plur,nil,sans insert
[N1] : def,sing,nil,nil
[N1] : def,sing,nil,sans insert
[N1] : def,sing,sans modif,sans insert
[N1] : nil,nil,nil,nil
[N1] : nil,nil,nil,sans insert

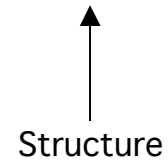
[N2] : def,sing,nil,nil
[N2] : def,sing,nil,sans insert
[N2] : nil,nil,nil,nil
[N2] : nil,nil,nil,sans insert
[N2] : nil,plur,nil,nil
[N2] : nil,plur,nil,sans insert
[N2] : nil,sing,nil,sans insert
[N3] : def,plur,nil,sans insert
[N3] : def,sing,nil,sans insert
[N3] : def,sing,sans modif,sans insert
[N3] : nil,nil,nil,nil
[N3] : nil,nil,nil,sans insert
[N3] : nil,plur,nil,sans insert
[N4] : def,sing,nil,sans insert
[N4] : indef,sing,nil,sans insert
[N4] : nil,nil,nil,sans insert
[N5] : nil,nil,nil,sans insert
[N6] : def,plur,nil,sans insert
[N6] : nil,sing,nil,sans insert
[N7] : nil,nil,nil,sans insert
[N8] : nil,nil,nil,sans insert
[N9] : nil,nil,nil,sans insert

2. Locutions propositionnelles

Structure du fichier « ficheprop »

ficheprop('L'HOMME EST UN LOUP POUR L'HOMME','loc prop', '[bin]',

'nil', '[insertion]').

Traits


Structure

Structure : Ensemble de traits relevés

nil	: nil,nil
nil	: nil,[insertion]
nil	: [tps],nil
nil	: [tps],[insertion]
[bin.syn.permut]	: [tps],nil
[bin.syn]	: nil,nil
[bin.syn]	: [tps],nil
[bin.syn]	: [tps],[insertion]
[bin]	: nil,nil
[bin]	: nil,[insertion]
[bin]	: [tps],nil
[bin]	: [tps],[insertion]