

André VALLI
Equipe DELIC
Université de Provence
Andre.Valli@up.univ-aix.fr

Un dictionnaire de la complémentation verbale en français

1) En TAL, en particulier les recherches sur les NLP, on peut observer un développement considérable des recherches sur les problèmes posés par le rattachement prépositionnel.

Si parmi les principales études consacrées aux compléments prépositionnels on relève des approches strictement sémantiques¹ où intéressées aux problèmes de désambiguïsation², si d'autres se limitent, dans des cadres formels donnés, à mettre au point des patrons de co-occurrences établis sur des bases statistiques, comme le montre notre jeune collègue de Delic dans sa thèse³, on peut relever des études comme celles annoncées sur le site web de l'ERSS - UMR 5610⁴ qui cherchent à exploiter les ressources linguistiques existantes dans le domaine.

2) Devant un tel état de la question des constructions prépositionnelles, en accord avec M. Salkoff, il m'a paru opportun de reprendre, en collaboration avec lui, l'élaboration d'un dictionnaire décrivant les arguments des verbes français, arrêté à la lettre J, il y a plus de vingt-cinq ans, et abandonné depuis. Je rappelle qu'à l'origine, ce lexique de verbes était destiné à servir à l'analyseur syntaxique mis au point par M. Salkoff au LADL dans les années 70–80.

La réalisation de ce dictionnaire dans sa version formalisée originale servira au fonctionnement de l'analyseur ; dans une version en langage naturel qui propose des exemples attestés des constructions admises par les verbes et décrivant par des symboles grammaticaux les arguments acceptés, ce dictionnaire pourra constituer une ressource utile pour les chercheurs en linguistique et en TAL, qui s'ajoutera aux ressources existantes : dictionnaires classiques, GR, TLF, version papier et électronique, Tables du LADL.

3) Une première version de ce dictionnaire de la complémentation verbale en français, limitée aux 975 verbes les plus fréquents listés dans le Frequency Dictionary of French Verbs, est en cours de réalisation et devrait être menée à terme courant 2005-2006.

Les principales caractéristiques de ce dictionnaire sont les suivantes :

¹ Par exemple : Gomez R; Segami C.; Hull R.. "Determining Prepositional Attachment, Prepositional Meaning, Verb Meaning, and Thematic Roles", *Computational Intelligence, February 1997, vol. 13, no. 1.*, : Blackwell Publishing ,pp. 1-31(31)

² Brian Mitchell. 2003. *Prepositional phrase attachment using machine learning algorithms. Ph.D. thesis, University of Sheffield.*

Adwait Ratnaparkhi, Jeff Reynar, and Salim Roukos. 1994. A maximum entropy model for prepositional phrase attachment. In *Proceedings of the ARPA Workshop on Human Language Technology*, pages 250–255.

Jiri Stetina and Makoto Nagao. 1997. Corpus based PP attachment ambiguity resolution with a semantic dictionary. In *Proceedings of WVLC '97*, pages 66–80.

G. Whittemore, K. Ferrara, and H. Brunner. 1990. Empirical study of predictive powers of simple attachmentschemes for post-modifier prepositional phrases. In *Proceedings of ACL '90*, pages 23–30.

Jakub Zavrel, Walter Daelemans, and Jorn Veenstra. 1997. Resolving PP attachment ambiguities with memory-based learning. In *Proceedings of CoNLL'97*, pages 136–144.

³ Nuria Gala Pavia (2003), *Un modèle d'analyseur robuste fondé sur la modularité et la lexicalisation de ses grammaires*, Thèse U Paris XI, CNRS, LMSI

⁴ *Linguistic clues for corpus-based acquisition of lexical dependencies*

Cécile Fabre, Didier Bourigault

ERSS - UMR 5610

Université de Toulouse Le Mirail, 5 allées A. Machado, 31058 Cedex, France.

1 SYNTAX: a tool for the extraction of lexical dependencies

2 Description of the strategy for PPs attachment

- a) Si l'essentiel de l'effort est en faveur de la description des objets prépositionnels, de manière à fournir aux parseurs une aide pour distinguer les arguments verbaux de simples modificateurs, le dictionnaire liste également tous les objets complexes identifiés dans la grammaire en chaîne. (Voir en annexe - 2 la liste des objets retenus)
- b) Comme on peut le voir dans la première annexe, dans l'entrée du verbe *distinguer*, les arguments acceptés par le verbe sont illustrés comme des exemples d'énoncés français attestés et représentés dans des formules qui ne font intervenir que des symboles de catégories grammaticales (N, nom, V, verbe, Adj. Adjectif, D, adverbe, ê pour *être* et de fonctions S, Subj. Pour Sujet, O, Obj. Pour Objet) qui peuvent être récupérés dans les grammaires formelles qui sont utilisées dans le parsing du français.
- c) Les données auxquelles nous avons recours pour décrire les arguments des verbes sont constituées par un large éventail de dictionnaires du français, format papier et électronique (Petit Robert, Grand Robert, TLF), qui donnent, on le sait, une description incomplète du domaine, et une présentation par emplois, en termes sémantiques, ce qui concourt à disperser les renseignements d'ordre strictement grammatical. Les entrées de ces dictionnaires sont particulièrement insatisfaisantes dans la description des arguments prépositionnels..
- L'autre importante source de données est constituée par les tables du LADL, les tables de M. Gross (1975) et en particulier les tables de Boons, Guillet, Leclère, (1976,1992), disponibles dans un classeur Excel, qui fournissent des renseignements précieux pour l'étude des constructions indirectes.
- Enfin, nous avons exploité le considérable corpus de français écrit virtuel que constituent les pages francophones de google.fr⁵ qui offrent de nombreuses occurrences de formes de compléments absentes des dictionnaires et des tables du LADL

4) En examinant par exemple, l'entrée du verbe *distinguer* donnée en Annexe – 1, on peut observer que notre dictionnaire retient une liste relativement complète des arguments acceptés par chaque verbe, ce qui en fait un outil important pour les chercheurs. D'autant plus qu'un effort est fait dans la formulation des arguments pour décrire les restrictions lexicales et grammaticales qui leur sont associées. Pour permettre une comparaison, nous avons résumé, ci-dessous, pour ce verbe, les objets recensés par le G R et le T.L.F , dans leur version électronique :

GR

<i>N1</i>	_ un homme, un détail
<i>N1 plu.</i>	_ les hommes
<i>N1 de N2</i>	_ l'homme des animaux
<i>N1 dans N2</i>	_ une chose dans un ensemble
<i>N1 d'avec N2</i>	_ une chose d'avec une autre
<i>N1 entre N2</i>	_ une chose entre plusieurs choses
<i>N1 parmi N2</i>	on le _ parmi ses collègues pour cette fonction
<i>N1 pour N2</i>	on le _ parmi ses collègues pour cette fonction

TLF

<i>N1</i>	_ un chant
<i>N1 plu</i>	_ les couleurs
<i>N1 de N2</i>	cet aveugle _ par le toucher une pièce d'or d'une pièce d'argent
<i>Que P</i>	on _ dans l'obscurité que la table était chargée de grands vases
<i>N1 d'avec N2</i>	_ qqn d'avec un autre
<i>N1 sur N2</i>	_ qqc sur d'autres
<i>entre N1 et N2</i>	_ qqn et un autre
<i>N1 en N2</i>	_ les individus ont été distingués en intellectuels (tu avais signalé cet exemple)
<i>N1 par N2</i>	_ les objets par des noms différents

⁵ Voir la démarche adoptée par Volk. Exploiting the WWW as a corpus to resolve PP attachment. Proceedings of Conference on Corpus Linguistics, p. 601-606, Lancaster, 2001

ANNEXE –1 L'entrée du verbe *Distinguer* dans notre dictionnaire

(Dans une version définitive, les chiffres et les lettres seront remplacés par des formules exprimées en catégories grammaticales; les prépositions seront données en clair)

A60 SE Max s'est distingué (dans ce match)
B2 Max distingue la côte
Max distingue un (chant + visage + bruit)
Max distingue (la pâleur + les paroles + les traits)
?Pour distinguer les paranoïaques constitutionnels
(TLF)
C120 P1 Max a distingué le vrai **du** faux
ce détail distingue Max **des** autres
loi visant à ~ le vin **des** autres boissons alcooliques
Dans la dimension spirituelle, on ne saurait ~ une
seconde **d'**un siècle, une heure **de** toute une ère.
P3 Max distingue Ida à son allure déhanchée (moi)
H| On peut le [chénopode blanc] distinguer à ses
poils
U| en forme d'étoile
M| le [Nh] distinguer à une cicatrice qu'il a au pied.
P7 pour se distinguer **entre** eux, les Tsiganes font..
P13 Les individus ont été ~és **en** intellectuels,
sensitifs et volontaires (TLF)
derrière l'écran où on les [Nh] ~ait **en** ombres
chinoises.
Il les [Nh] distinguait **en** deux classes : ceux qui...
les [genres] ~ **en** genres..marqué et non marqué.
thèses qu'on ~ait **en** mineure, majeure, sabbatine,..
La physique distingue les choses **en** solides et
liquides.
P16 permettre aux clients de vous ~ **par** vos
domaines de spécialisation.
la France...continue de se ~ **par** ses mathématiciens
le [votre nom] ~ **par** une ou deux lettres de
votre
prénom.
distinguer les objets **par** (la couleur + leur
forme)
P17 le [Nh] ~ **pour** ses bons et loyaux services
Le comité..vient de le ~ **pour** la découverte d'une..
P56 Max a distingué le vrai **d'avec** le faux
P96 de le [Nh] ~ **par rapport** à quelqu'un d'autre
le [leurre] ~ **par rapport** à la surface de la mer
le [harcèlement] distinguer **par rapport** à d'autres
problématiques?

C20 P7 il est difficile de ~ **entre** l'anormal et le
morbide
saurais-tu distinguer **entre** les effets et les causes
distinguer **entre** un remblai et une décharge
en refusant de ~ **entre** vin et alcools "durs", ces
dispositions...
distinguer **entre** drogues dures et douces.
il importe de ~ **entre** les personnes qui vivent V8 (Que
Max porte un béret + porter un béret) ~e Max
à domicile
et celles qui sont pensionnaires d'un home.
nous devons distinguer **entre** le temps linéaire..et
les
temps circulaires,
C108 P17 Max distingue Luc **pour** accepter ce poste
C103 Max distingue qu'Ida est capable
C135 Max a distingué le vrai et le faux
C180 P7 Nous est-il si facile de ~ **entre** se croire libre
et être effectivement libre ?
B80 on ~ait qui était encore à cheval et qui gisait à terre,
Il faut bien distinguer à qui on s'adresse ?
Il faut bien distinguer à quoi se rapporte le trait.
Nh...distinguer lequel des deux dirige l'autre,
C81 cet écran plat Hyundai semble se ~ comme offrant
le meilleur rapport qualité prix
Ceci permet de nous ~ comme ayant une expertise
reconnue
que [Nh] l'on souhaite ~ comme étant amies des
Basques,
C136-7 les résultats des législatives qu'il faut ~ comme
approche socio-électorale ;
les quelques entreprises qu' il n' était pas possible..de
~ comme GE ou PME
On peut distinguer comme déchets finals de l'épuration
: Les gadoues de fosses septiques...
nous pouvons ~ comme points principaux de l'œuvre
de G. Monge les deux points suivants :
C100 il distingue être amoureux et aimer.
il faut bien distinguer être maigre et être mince.
C144 P3 ~ que Ph à ce que Ph
C122: P3: ~ à son silence que Max est très fâché
C179 Ce qui m'avait semblé le ~ pour que je puisse
m'en souvenir, c'est..

ANNEXE - 2

Chaînes d'objets

- A60 – objet zéro
B2 -- objet N, R
B10 – objet d'*être*
B16 – adverbe D
B80 – chaînes phrastiques r N
C20 Prep N Mark P
C80 Sbj Vant Obj
C81 Sbj comme Vant Obj
C100 V Obj (l'infinitif)
C101 de V Obj
C102 (jusqu') à V Obj
C103 que C1 (complétive)
C104 que C1-subj
C105 Adj
C108 Sbj de V Obj
C110 Sbj *à, pour, jusqu'à* V Obj
C111 N Prep V Obj
C112 pour, à Sbj de V Obj
C113 P1 N P2 V Obj
C114 P1 Sbj P2 V Obj
C115 Prep N V Obj
C116 de Sbj de V Obj
C117 N que C1
C118 N que C1-subj
C119 Prep N2 N1
C120 N1 Prep N2
C121 N Prep ce que C1
C122 Prep N que C1
C123 Prep N que C1-subj
C124 P1 N P2 N
C125 N Prep ce que C1-subj
C126 P1 N P2 ce que C1
C127 P1 N P2 ce que C1-subj
C128 D Prep N (*augurer*)
(*comporter*)
C129 Prep ce que C1-subj
C130 Prep ce que C1
C131 Vé Obj (*avoir, être*)
C132 Vé Obj-passif
C133 Vant Obj
C134 Pê N
C135 N et N
C136 Sbj-ê Pê Obj-ê
C137 Pê Obj-ê Sbj-ê
C138 Sbj-ê Obj-ê
C139 Obj-ê Sbj-ê
C140 N D
C141 Vê Obj-ê Sbj-ê
C142 V Obj à Sbj/V Sbj Obj (*faire*)
C143 Sbj V Obj
C144 que C1 Prep ce que C1
C145 que C1-subj P ce que C1
C146 que C1 P ce que C1-subj
C147 que C1-s P ce que C1-s
C148 Prep N rN
C149 que C1 à V Obj
C150 que C1-subj à V Obj
C151 V Obj Prep ce que C1
C152 V Obj P ce que C1-subj
C153 de V Obj P ce que C1
C154 de V O P ce que C1-subj
C155 V Obj Prep V Obj
C156 de V Obj Prep V Obj
C157 P N Vê Obj-ê Sbj-ê
C158 de ce que C1 à ce que C1
C159 S V O (C143) P ce que C1
C161 comme C1 (*constater*)
C165 de (FACON) à V Obj
C166 de (FACON) (à ce) que C1subj
C179 N pour que C1-subj
C180 Prep V Obj
C181 pour que C1-subj
C182 Prep rN
C183 N si C1
C184 comme Vant Obj
C185 P N Sbj (*comme/E*) Vant Obj
(*décrire, dépeindre*)
C186 rN Prep ce que C1
C188 si C1
C189 P N Obj-ê Sbj-ê
C190 P N Sbj-ê Obj-ê
C191 P N Sbj V Obj
C192 P N Sbj-ê Pê Obj-ê
C193 N V Obj
C194 de rN à ce que C1
C195 P N C161 (*comporter*)
C196 P N C134 (*comporter*)
C197 Sê Oê P ce que C1 (*croire*)
C198 Oê Sê P ce que C1 (*deviner*)
C199 Vê Oê Sê P ce que C1 (*dép.*)
P D (venir jusqu'ici)
N P D (enlever *N1* de là)