



HAL
open science

Les tables de Méthodes en syntaxe: introduction à un mode d'emploi.

Gabriel G. Bès

► **To cite this version:**

Gabriel G. Bès. Les tables de Méthodes en syntaxe: introduction à un mode d'emploi.. Les cahiers de praxématique, 1994, 22 (1), pp.57-80. hal-00990802

HAL Id: hal-00990802

<https://hal.science/hal-00990802>

Submitted on 14 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

cahiers de **praxématique**

Autour du verbe : théorie et implémentation

Rachel PANCKHURST

Frédérique SEGOND

Philippe BLACHE

Gabriel BÈS

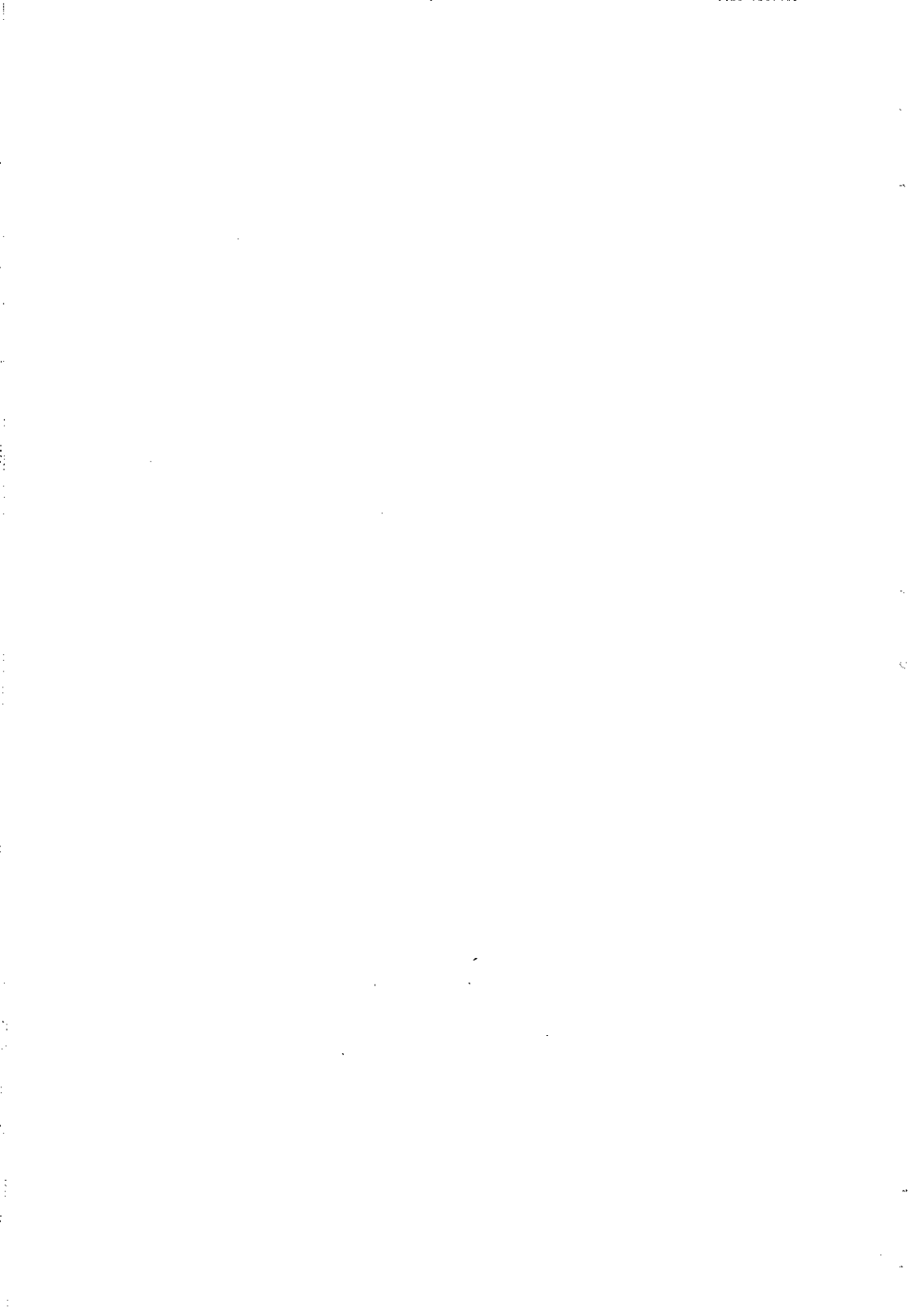
Philippe ALCOUFFE et
Bruno REVELLIN FALCOZ

Pierre OUELLET

Adel EL ZAIM

Hervé BOUCHARD

Jean-Jacques FRANCKEL



Les tables de *Méthodes en syntaxe* : introduction à un mode d'emploi*

Depuis une vingtaine d'années, Maurice Gross et ses collaborateurs ont construit des tables censées enregistrer de manière systématique des données descriptives du français.

Ces tables sont décrites comme étant des matrices où les lignes correspondent à des entrées du lexique et les colonnes à des propriétés syntaxiques. La présence d'un symbole « + » à l'intersection d'une entrée e et d'une propriété P indique que e a la propriété P et la présence d'un symbole « - » indique que e ne l'a pas ; cf. (Gross, 1975 :14), (Leclère, 1991 :112), (Clémenceau, Roche, 1991 :1), (Roche, 1993 :56-57).

On pourrait donc penser :

- une table est une matrice lisible et interprétable.

L'objectif de ce travail est, une vingtaine d'années après sa parution, de caractériser le contenu des tables (désormais T) de *Méthodes en syntaxe* (désormais MS), le livre fondateur de la formalisation par tables de Maurice Gross. La caractérisation de MS peut se résumer dans le constat suivant :

- T est formellement ininterprétable.

Ce constat sera justifié dans la section suivante, §1. Dans §2, on montrera la possibilité d'obtenir T' , objet construit à partir de T et rela-

* Ce travail a été partiellement réalisé dans le cadre du projet LRE 61034 DELIS. Je remercie S. Ait-Mokhtar, K. Baschung, M. Emorine, C. Dafniet, J.L. Rodrigo Mateos et R. Panckhurst de leurs observations à une première version de ce travail ; la responsabilité de toute erreur me revient.

tivement plus explicite et plus interprétable que T , et on y présentera un fragment de T' ¹.

La spécification de T' soulève de nouveaux problèmes. Ce sont les difficultés d'interprétation de T' dans un univers linguistique – les énoncés de la langue – et de vérification de l'exhaustivité de T' . Enfin, la question plus générale et importante : quelle est la théorie Th sur laquelle T et MS se fondent ? Ces trois questions seront posées mais à peine développées dans §3. L'évaluation de T est ainsi soigneusement distinguée du reste du contenu de MS où, en 1975, ont été posées des questions essentielles qu'aucune linguistique ne peut ignorer.

1. L'ininterprétabilité formelle de T

Les tables dans T font penser aux horaires SNCF. Ceux-ci sont organisés sous la forme d'un couple <matrice, notes>. P. ex., dans les horaires sur les trains Clermont-Paris pour la période du 27/9/92 au 22/5/93, la matrice présente en colonne un intitulé avec identification des trains, en ligne les gares de Clermont jusqu'à Paris, et dans l'intersection *ligne x colonne*, l'indication d'une heure. Les intitulés des colonnes sont des pointeurs qui renvoient aux notes, celles-ci détaillant les jours durant lesquels les trains sont en service. Ces informations peuvent se traduire par les énoncés dans (1).

(1) Dans la période du 27/9/92 au 22/5/93

Il y a un train au départ de Clermont identifié <X> à Y h qui doit arriver à Paris à Z h et qui circule les jours ouvrables du lundi au vendredi.

1 On résume ici une pratique de l'intérieur des tables du LADL et, tout particulièrement, de celles de MS sur lesquelles ce travail porte exclusivement. Cette pratique a été requise par les thèses préparées au GRIL et par de courtes participations du GRIL au projet GENELEX (du 1/10/90 au 31/12/90 et du 15/5/91 au 30/7/91). Dans (Bès et Emorine, 1991) on trouvera une première version de T' . La méthodologie qui y avait été adoptée est exposée ici. Sur cette même question, cf. (Alcouffe, Falcoz et Zaysser, 1993).

Les tables dans T sont présentées comme des matrices mais elles ne sont pas interprétables directement, intuitivement, comme les horaires SNCF. Une table de T présente la structure générale de (2).

(2)

Le		Le
V	E	V

E est une colonne distinguée où, dans l'intersection avec chaque ligne, sont inscrites les *entrées*. Dans V on trouvera des colonnes qui, à l'intersection d'une ligne, portent l'inscription de *valeurs*. La zone Le est occupée par des légendes placées dans des cartouches

Pour identifier les informations dans T on utilisera des conventions : $T-i$, avec $i = 1$ à 19, identifie la i ème table.

entrée, e , correspond au lemme verbal considéré (lemme simple, p.ex. *finir* dans $T-1$, ou suite de mots ou de symboles, comme *ne manquer* *Neg* toujours dans $T-1$).

N° (*de colonne*) n (on notera parfois N^n) identifie une colonne déterminée. On obtient n en numérotant de gauche à droite de 1 à n les colonnes de chaque table à l'exception de E .

valeur est l'inscription enregistrée dans l'intersection *ligne x colonne*.

Chaque colonne est immédiatement couronnée par un cartouche qui porte une inscription. A partir d'un N° *de colonne* on repère de manière univoque le cartouche qui la couronne immédiatement et la légende que celui-ci porte. On dira que cette légende est un *label*.

Chaque label est une structure de phrase dans laquelle un verbe « peut entrer ou non » ($MS : 152-153$). Or, les structures correspondant aux labels ne sont pas faciles à déterminer. Par ailleurs, tous les labels ne représentent pas des structures (cf. p.ex. $N^{\circ 7}$ de $T-4$). Mais on voit que trois types de questions relativement claires peuvent être posées : (i) quelles sont les structures dans lesquelles une entrée dans une table peut « entrer », avec explicitation des propriétés de ces structures (on dira désormais qu'une entrée *s'intègre* à une structure), (ii) quelle est

l'interprétation dans un univers linguistique – les énoncés de la langue – de ces structures, (iii) quelle est l'adéquation externe de *T* par rapport à l'univers linguistique considéré, en l'occurrence le français ? La deuxième question ne peut être posée qu'après avoir donné une réponse à la première, et la troisième ne peut venir que s'il y a eu réponse à la deuxième. C'est la réponse à (i) sous la forme de *T'* qui doit permettre de poser les deux autres.

Tout langage se décrit en explicitant d'une part sa syntaxe (son vocabulaire et ses règles de bonne formation), et sa sémantique. La spécification des structures relève de la syntaxe de *T*. Or, le langage de *T* est peu ou mal explicité. On trouve dans (*MS* :150-156) quelques indications, mais la syntaxe du langage, d'où l'on devrait pouvoir obtenir les règles de lecture de *T*, n'est pas donnée.

Les tables de *T* ne peuvent donc pas être systématiquement lues comme les horaires SNCF. On est obligé de passer par le processus inverse qui consiste à essayer de comprendre ce qu'elles « veulent dire » et, à la suite de ce processus, d'expliciter les conventions permettant d'aboutir à *T'*. Mais cette technique comporte des risques. Supposons que l'on ait réussi à dégager une explicitation de la notion de « compatibilité entre compléments », évoquée au cours de la présentation du formalisme et de son expression moyennant la disposition des cartouches (*MS* : 150). Il est en général vrai que des compléments « compatibles » sont dominés (cf. ci-dessous) par des cartouches horizontaux distincts qui ne se dominent pas entre eux. Mais cette convention n'est pas valable dans *T-2* : si on se rapporte à la discussion dans (*MS* :165) à propos de *T-2*, on doit accepter que cette notation, dans *T-2*, indique que les compléments notés dans les cartouches placés sous le cartouche *Comp. indirect* ne sont pas compatibles avec le syntagme verbal à l'infinitif. Pour *venir*, p.ex., *T-2* spécifie les structures de (3) mais pas celle de (4) (pour chaque structure on indique entre parenthèses la ou les colonnes avec la valeur + qui permettent de la spécifier, et on en donne un exemple).

- (3) N° V V° Ω (N°1, N°13)
 Jacques viendra prendre un livre
 N° V ici (N°1, N° 23)

Jacques viendra ici
 N° V là (N°1, N° 24)
 Jacques est entré là

- (4) N° V ici V⁰Ω
 Jacques viendra ici prendre un livre

Les illustrations de (3) et (4) montrent le double aspect de l'ininterprétabilité formelle de T : (i) pour pouvoir dire qu'une occurrence du verbe *venir* s'intègre à des structures données, il faut déterminer celles-ci, c'est la question de la lisibilité de T ; (ii) pour pouvoir associer une structure dans (i) à des énoncés du français, il faut connaître les propriétés de ces structures : propriétés des symboles qui composent ces structures et relations des symboles entre eux dans une même structure et dans des structures différentes. Or, les tables de T sont formellement illisibles et formellement ininterprétables. Formellement illisibles parce qu'il n'existe pas de règles explicites permettant de calculer à partir de la lecture de $T-i$ pour une entrée arbitraire les structures auxquelles cette entrée s'intègre. Formellement ininterprétables, parce qu'il n'existe pas de caractérisation précise des valeurs linguistiques qui doivent être associées aux symboles constitutifs de ces structures et aux propriétés de ces structures. Les structures et les exemples de (3) et (4) ont été présentés suite à une laborieuse herméneutique, le contraire donc de l'explicitation formelle.

2. La spécification de T'

La spécification de T' devrait apporter une réponse à la question de la lisibilité de T : il s'agit d'explicitier les structures auxquelles s'intègre une entrée quelconque dans une table quelconque de T . La spécification de $T'-i$ à partir de $T-i$ doit se fonder sur deux types de connaissances :

- Le déchiffrement des conventions de notation généralisables ;
- La modulation de celles-ci en fonction des commentaires associés à $T-i$ (*MS* :160-213) et de l'exposé introductif de l'ouvrage (*MS* :13-150, *MS* :214-231).

Une structure de T' est notée sous la forme d'une liste de symboles. Dans cette liste il y a un symbole verbal distingué unique (U dans les structures de $T-1$ et V dans toutes les autres) qui peut être conçu comme une variable. Une entrée va instancier cette variable : c'est ainsi que l'on peut caractériser la notion qu'une entrée s'intègre à une liste. Spécifier les structures de e dans $T-i$ c'est spécifier les listes dans lesquelles e instancie le symbole verbal distingué. La spécification des listes doit se faire en fonction des valeurs de e dans $T-i$ et doit porter sur les symboles de la liste, leur ordre et leurs propriétés.

On aura ainsi $L(< e, T-i >)$, c'est-à-dire l'ensemble (fini) de listes où e instancie le symbole verbal distingué selon les valeurs de e dans $T-i$. On souhaite donc se donner une méthode de lecture permettant d'obtenir $T'-i$ à partir de $T-i$. A partir de $T'-i$ on devra pouvoir spécifier $L(< e, T-i >)$. Autrement dit, on souhaite, p.ex., obtenir $T'-2$ de telle manière que **par un procédé mécanique de lecture** de $T'-2$ (et non pas par un processus herméneutique), on sache que pour le verbe *venir*, du fait que les valeurs de *venir* sont + dans les N^o de colonnes 1 et 3, la table spécifie, parmi d'autres, les structures de (3), et que, pour le même verbe, la table ne spécifie pas la structure de (4).

Pour obtenir $T'-1$, on procédera en deux étapes. Dans la première (§2.1) on spécifiera la multiplicité de facteurs dont on doit tenir compte pour aboutir à l'explicitation de T' . Dans la seconde (§2.2) on formulera quelques règles permettant d'aboutir à des fragments de T' , qui seront illustrés dans §2.3.

2.1. Vocabulaires et relations dans T

Pour lire T il est nécessaire de distinguer plusieurs vocabulaires et différents types de relations entre les éléments de ces vocabulaires. Ces facteurs sont explicités dans (5). Ils seront repris un par un dans la suite.

- (5)
- (i) Type de valeurs
 - (ii) Type de cartouche
 - (iii) Type de légende
 - (iv) Disposition des cartouches entre eux dans Le .

- (v) Relation de certaines légendes dans *T-i* avec la structure principale associée à *T-i* dans l'Annexe 1 de (*MS* :232).
- (vi) Expressions dans les entrées

L'ensemble *Va* de valeurs est spécifié dans (6).

- (6) $Va = \{ +, - \} \cup Prép \cup Sax \cup Conj \cup Adv$
Prép = { de, à, par, pour, sur, avec, en, auprès de, dans, parmi, contre, devant }.
Sax = { 0, Loc, Opér } [*Sax* : Symboles auxiliaires]
Conj = { et }
Adv = { comme }

Les cartouches qui enferment les légendes peuvent se caractériser par leur format. On reconnaît les types mentionnés dans (7).

- (7) *Types de cartouche*

Vertical fermé
 Angulaire fermé
 Angulaire ouvert
 Horizontal

Les légendes dans l'ensemble *Leg* ont été groupées en 16 ensembles disjoints. On trouvera le descriptif partiel de cette partition dans (8), où on ne spécifie que les types de légendes qui seront utilisées par la suite ; la spécification complète de la partition est nécessaire pour aboutir à une lecture exhaustive de *T*. Si l'on ajoute la légende *bl* (blanc) on a en tout 139 légendes. Les labels sont les légendes qui couvrent immédiatement une colonne.

(8) Ensemble et mnémotechnique	Eléments de l'ensemble ou propriété caractéristique	Quantité d'éléments
<i>F</i> (fonctions)	{Sujet, Comp.dir., Comp. ind., Comp. dir. ou ind., Comp. dir. ou en <i>pour</i> , Comp. ind. en <i>à</i> , Comp. ind. en <i>de</i> ²	7
<i>SPh</i> (structures phrases)	Existence de N_0 ou de N_1 à gauche de V ou U , et de V ou U Ex. : $N_0 V N_1$	33

Dans les zones *Le* des tables (cf. (2) ci-dessus), les cartouches se trouvent disposés dans des configurations variées. Par la superposition de cartouches, on peut définir la notion de *domination*. Un cartouche vertical ne domine aucun type de cartouche et tous les autres dominent tous les types de cartouche.

Dans (5v), il a été fait état d'une relation de *T-I* avec la ou les structures principales associées à *T-i* dans l'*Annexe 1* de (*MS* :231). Cette relation est cruciale pour l'explicitation d'une méthode de lecture de *T* conduisant à *T'*.

Une structure principale est une liste de symboles. Ces listes de l'*Annexe 1* doivent être découpées en sous-listes. Ce découpage est fait dans l'*Annexe 1'* (cf. (9)). L'*Annexe 1'* est constituée de listes principales (désormais *l-pr*) et sera utilisée pour définir les relations avec *T*.

2 Des écritures différentes de la même fonction ont été regroupées pour aboutir à 7 éléments.

(9) Annexe 1'

T	Listes principales
1	[[N ₀], [U], [Prép V ⁰ Ω]]
2	[[N ₀], [V], [V ⁰ Ω]]
3	[[N ₀], [V], [N ₁], [V ¹ Ω]]
4	[[Qu P], [V], [N ₁]]
5.1	[[Qu P], [V], [Prép N ₁]]
5.2	[[II], [V], [Prép N ₁], [Qu P]]
6	[[N ₀], [V], [Qu P]]
7	[[N ₀], [V], [à ce Qu P]]
8	[[N ₀], [V], [de ce Qu P]]
9	[[N ₀], [V], [Qu P], [à N ₂]]
10	[[N ₀], [V], [Qu P], [Prép N ₂]]
11	[[N ₀], [V], [N ₁], [à ce Qu P]]
12.1	[[N ₀], [V], [Qu P]]
12.2	[[N ₀], [V], [N ₁], [de V ¹ Ω]]
13	[[N ₀], [V], [N ₁], [de ce Qu P]]
14	[[N ₀], [V], [à ce Qu P], [Prép N ₂]]
15	[[N ₀], [V], [de ce Qu P], [Prép N ₂]]
16	[[N ₀], [V], [Qu P], [Prép ce Qu P]]
17	[[II], [V], [Prép ce Qu P], [Prép N ₂]]
18	[[N ₀], [V], [Prép N ₁], [Prép N ₂], [Prép Qu P]]
19	[[Qu P], [V], [N ₁], [Prép N ₂]]

Dans (9) on a 21 *l-pr*. A l'exception de *T-5* et *T-12*, *T-i* correspond à une seule *l-pr*. On conserve donc pour *l-pr* la numérotation de la table respective avec un digit supplémentaire pour *T-5* et *T-12*. Dans l'ensemble *Leg*^o on a toute légende appartenant à *F* et les légendes *Exraposition*, et *bl* dans un cartouche blanc horizontal et supérieur. On notera par *leg*^o un élément de *Leg*^o et par *el* une sous-liste particulière dans une liste de (9). Entre les *el* dans une *l-pr-i* et les *leg*^o dans *T-i* on doit définir la fonction *association-f*. L'ordre des *el* dans *l-pr-i* est corrélatif à l'ordre des *les*^o dans *T-i* ; c'est ce qui permet de calculer les *associations-f*. *Association-f* est une fonction bijective partielle entre

les *el* dans *l-pr-i* et les *leg*^o corrélatives dans *T-i*. Selon les *leg*^o utilisées et selon que tous les *el* ou les *leg*^o soient *associées-f*, on a différents types d'associations entre *l-pr-i* et *T-i*. Ces types sont indiqués dans (10), de (i) à (viii). Dans la première ligne de chaque type, on le caractérise ; dans la deuxième, on indique le ou les *l-pr* où il est attesté. *Rien* doit se comprendre aucune *leg*^o \in *T-i* ou aucun *el* \in *l-pr-i*.

- (10) (i) • Tout *el* est *associé-f* à une *leg*^o \in *F* et vice versa.
 • 1,4, 5.1, 7, 8, 9, 10, 12.1, 13, 14, 15, 16, 18, 19.
- (ii) • Le premier *el* est *associé-f* à la première *leg*^o \in *F*.
 • Le deuxième et dernier *el* est *associé-f* à *bl*. La deuxième et dernière *leg*^o \in *F* est associée à *rien*.
 • 2
- (iii) • Les deux premiers *el* sont *associés-f* aux deux premières *leg*^o \in *F*. Le troisième et dernier *bl* domine directement une *leg*^o \in *F* qui est *associée-f* à *rien*.
 • 3
- (iv) • Le premier *el* est *associé-f* à *rien*, le deuxième *el* à la deuxième *leg*^o \in *F* et le troisième à *leg*^o = *Extraposition*.
 • 5.2
- (v) • Les deux premiers et seuls *el* sont *associés-f* aux deux premières *leg*^o \in *F*. La troisième et dernière *leg*^o \in *F* n'est associée à *rien*.
 • 6
- (vi) • Les deux premiers *el* sont *associés-f* aux deux premières et seules *leg*^o \in *F*. Le troisième et dernier *el* est associé à *bl* dans un cartouche blanc supérieur.
 • 11
- (vii) • Les *el* sont [d'après notre interprétation de *MS*] associés à *rien*
 • 12.2
- (viii) • Le premier *el* est associé à *rien* ; les deux *el* restants (deuxième et troisième) sont *associés-f* à la première et deuxième *leg*^o \in *F* ; il n'y a pas de troisième *leg*^o \in *F*.
 • 17

Voici des exemples d'associations-*f*:

1-pr-1	2è el = Prép $V^0\Omega$	T-1 2è leg° $\in F = \text{Comp. ind.}$
1-pr-4	1er el = Qu P	T-4 1ère leg° $\in F = \text{Sujet}$
1-pr-5.2	1er el = il	T-5 rien
1-pr-6	rien	T-6 3è leg° $\in F = \text{Comp. ind.}$

Le dernier élément de (5) est (5.vi), qui porte sur les expressions dans les entrées. On ne les utilisera pas dans la discussion qui suit.

2.2 . L'explicitation de $T'-1$

À partir des vocabulaires et des relations dans T détaillées dans §2.1, on devrait pouvoir spécifier les règles permettant de formuler T' à partir de T . Rappelons que l'objectif ultime est d'obtenir $L(<e, T-i>)$, l'ensemble fini de listes où e instancie le symbole verbal distingué V ou U selon les valeurs de e dans $T-i$.

La complexité des éléments de (5), explicités de (6) à (10) dans §2.1 rend impossible la formulation ici de toutes les règles permettant d'expliciter T' . On se limitera à donner une idée générale de la méthodologie adoptée, des éléments qui concourent à la formulation de ces règles, et dans §2.3 des fragments des résultats obtenus.

La structure générale de $T'-i$ est donnée dans (11.i) et (11.ii).

(11)

	valeurs*	l*
(i)	i...k	Ω
(ii)	i prép-j k	Ω Prép-j Ω

Dans la colonne *valeurs** de (11), i et k notent des N° de colonnes avec valeur +, *prép* j note une variable sur la préposition spécifique qui sera attestée dans le N° de colonne j ; $i < j < k$. Dans la colonne l^* , Ω est une variable sur une liste de symboles utilisés dans les légendes de $T-i$. On trouvera dans (12) l'exemple d'un fragment de $T'-i$, emprunté à $T'-1$.

(12)

valeurs*	l*
5	N_0 est Upp Ω
prép-7 18	N_0 U Prép -7 N_1
22	N_0 U dans N_1

Une ligne $T'-i$ se note $\langle v^*l \dots v^*k ; l^* \rangle$, où v^*_i est soit un numéro de colonne, noté j , soit *prép-j*, et l^* une liste construite à partir de Le dans $T-i$; l^* intègre toujours U ou V . L^*-i est l'ensemble des l^* dans $T'-i$.

La lecture mécanique recherchée de $T'-i$ se fonde sur le principe formulé dans (13).

- (13) $l^* \in L \langle e, T-i \rangle$ ssi l^* est dans une ligne $\langle v^*l \dots v^*k ; l^* \rangle$ de $T'-i$ telle que pour chaque v^*_i
 si $v^*_i = j$, alors valeur ($e x j$) dans $T-i = +$,
 si $v^*_i = \text{prép-j}$, alors valeur ($e x j$) dans $T-i = \text{prép}'$ et *prép'* instance *Prép-j* dans l^* .

Si l'on se rapporte à $T-I$, on voit que *passer* $x N^{\circ 7} = \text{pour}$ et *passer* $x N^{\circ 18} = +$. La 2è ligne de (12) est ainsi satisfaite par *passer* ; la liste N_0 U *pour* N_1 appartient donc à $L \langle \text{passer}, T-I \rangle$.

La définition de L^*-i est cruciale pour la spécification de $T'-i$. En lignes générales, on peut obtenir L^*-i en identifiant des listes de départ et en effectuant des opérations sur ces listes. On a la définition de cet ensemble dans la formule :

Listes de départ de $T-i = \{ \text{liste}(s) \text{ l-pr-i} \} \cup L'-i$

On dispose des liste(s) *l-pr-i* dans l'Annexe I'. $L'-i$ est obtenu à partir de $Le-i$ (légendes de $T-i$) selon (14).

- (14) Toute *lég* $\in SPh$ de $Le-i$ autre que le label $N^{\circ 10}$ de $T-I3$ appartient à $L'-i$ (on note ces légendes $l'-i$).

$L'-i$ est partitionné en $L_1'-i$ et $L_2'-i$ selon les critères définis en (15).

(15.1) $l' \in L'_1$ ssi dans T-i une des trois conditions suivantes est remplie :

- (i) $l' = N_0 U$ (Ex. 1)
- (ii) $l' = N_0 V$ (Ex. 2)
- (iii) $l' = N_0 V \Omega'$ et

l' est dominée par $leg \in F$

$l-pr-i = N_0 V \Omega''$

telle que

$\Omega'' = \Omega^* \Omega$ ou $\Omega'' = \Omega \Omega^*$

$\Omega, \Omega^*, \Omega', \Omega''$ notent des *el*

Ω'' est associé à $leg \in F$

et une des trois conditions suivantes

est remplie :

(a) $\Omega' = \Omega^*$ (Ex. 3)

(b) Ω' porte un numéro de position (N_1, N_2, N_3) correspondant à la position de Ω'' (Ex. 4)

(c) $\Omega' = Prép \Omega^*$ et Ω^* satisfait la condition

(b) immédiatement précédente et $F = Compl. direct$ ou *indirect* (Ex. 5).

(15.2) $l' \in L'_2$ ssi $l' \in SPh$ et $l' \notin L'_1$

L'ensemble L'_1 comprend les sous-structures de (*MS* :55). Elles se définissent par rapport à *l-pr-i*. Les exemples 1, 2, 3, 4 et 5 de (15.1) sont respectivement illustrés par $N^{\circ 8}$ de T-1, par $N^{\circ 7}$ de T-2, par $N^{\circ 25}$ de T-10, par $N^{\circ 15}$ de T-11 et par le $N^{\circ 8}$ de T-16. Un exemple de $l' \in L'_2$ est $N^{\circ 12}$ de T-11.

La spécification de L^*-i se fait par les procédés indiqués dans (16).

- (16) i Instanciation de symboles de *l-pr-i*
- ii Instanciation de symboles de *l'-i*
- iii Adjonction de symboles à *l-pr-i*
- iv Effacement de symboles de *l-pr-i*
- v Remplacement de *el* dans *l-pr-i* par des symboles dans les légendes de T-i

- vi Remplacement de *el* dans *l'-i* par des symboles dans les légendes de *T-i*
- vii Redistribution des *el* dans *l-pr-i*

Dans l'impossibilité d'exposer ici chacun de ces procédés, dont aucun ne l'a été dans *MS*, on se limitera à illustrer la problématique de l'obtention de *T'* pour la question des prépositions. Elles sont concernées par les procédés (16 i, ii, iii et iv).

Pour traiter des prépositions, *T* utilise les symboles *Prép*, *Loc*, *Opér* et *0*, et des prépositions spécifiques telles que *à*, *de*, *par*, *contre* et quelques autres. Le symbole '0' est utilisé pour noter l'absence de préposition. On va considérer *Prép*, *Loc* et *Opér* comme des variables ; *Loc* va être instancié par un sous-ensemble de prépositions qui peuvent également instancier *Prép*. Avec *prép* on note dans (17) une préposition quelconque.

Les différents traitements des prépositions dans *MS* qui concernent la spécification de $L(<e, T-i>)$ sont résumés dans (17).

- (17) (i) *prép* s'intègre à un *el* dans les *l-pr*
(ex : *à* dans *l-pr-7*)
- (ii) *Prép* dans *l-pr-i* est instancié par *prép* qui est valeur de *e* dans *T-i* (ex. : les valeurs de $N^{\circ 7}$ dans *T-1*)
- (iii) *prép*, valeur de *e* dans *T-i*, doit être incorporée à $L(<e, T-i>)$
(ex : les valeurs de $N^{\circ 9}$ dans *T-16*)
- (iv) *Prép* dans une légende de *T-i* est instanciée par *prép*, valeur de *e* dans *T-i* (ex : *Prép* dans le label de $N^{\circ 18}$ de *T-1* est instanciée par les valeurs de $N^{\circ 7}$)
- (v) *prép* est indiquée dans des légendes de *T-i* appartenant à *SPh* et à d'autres non explicités dans (8).
- (vi) *prép* dans *el* de *l-pr-i* est effacée par une légende [*pc z*] ou [*prép z*] de *T-i* (ex : le label $N^{\circ 11}$ de *T-7*)
- (vii) L'instanciation de *Prép* dans une légende de $T-i \in SPh$ n'est pas indiquée (ex : *Prép* dans le label $N^{\circ 8}$ de *T-16*)

- (viii) Les instanciations de *Loc*, valeur de *e* dans *T-i* sont indiquées dans *T-i* (ex : les instanciations des valeurs *Loc* dans *N°26* de *T-10* sont indiquées dans *N°29*, *N°30* et *N°31* de *T-10*)
- (ix) Les instanciations de *Opér*, valeur de quelques entrées dans *T-10* ont été impossibles à déterminer ; il se peut que les valeurs de *N°29*, *N°30* et *N°31* y soient pour quelque chose.

On observe qu'il y a neuf manières différentes de traiter les prépositions dans *MS* (certaines y sont mentionnées, cf. (*MS* :64-65)). Si on ajoute que, parfois, un même lemme est inscrit dans deux entrées différentes simplement parce qu'il présente deux instanciations différentes de *Prép* (p. ex. *continuer* dans *T-1*) on arrive à dix traitements différents ; cf. § 3 pour le dédoublement de lemmes.

2.3 . Un fragment de *T'-i*

Les définitions proposées dans §2.2 concourent à la définition de *T'-i* à partir de *T-i*. On présentera ici un exemple comprenant trois parties (*a*, *b*, *c*). La partie *a* illustre un fragment de *T'-i*. On y considère toutes les colonnes de *T-i* à l'exception de celles explicitement indiquées comme telles. Dans la partie *b*, on trouvera quelques entrées de *T-i* avec leurs valeurs. La partie *c* présente pour chacune des entrées dans *b* l'ensemble des listes auxquelles chaque entrée s'intègre. Sur les indices ajoutés à *ppv*, cf §3.1.

Fragment de *T'-8*

Colonnes non considérées : 1,2, 3, 4, 5, 10, 14 à 18, 22, 23, 26

valeurs*	I*
6	N ₀ est V-a
7	N ₀ V
8	N ₀ est V _{pp} Ω
9	N ₀ V de ce que P
11	N ₀ V que P
12	N ₀ V si P ou si P
13	N ₀ V de V ⁰ Ω
19	N ₀ N ₀ V de ce (ci + là)
20	N ₀ V ppv ₁
21	N ₀ V là
24	N ₀ V de le fait Qu P
25	N ₀ V ppv ₂
27	N ₀ de N ₁ V N ₀
28	N ₀ V contre N _{hum}

Entrées dans T-8

N° de colonnes	6	7	8	9	11	12	13	19	20	21	24	25	27	28
s'insurger		+	+	+	+		+	+	+		+	+		+
douter				+	+		+	+	+		+	+		

Ensembles des listes associées aux entrées

L (<s'insurger, T'-8>) =

- { N₀ s'insurger,
- N₀ est s'insurger_{pp} Ω
- N₀ s'insurger de ce que P,
- N₀ s'insurger que P,
- N₀ s'insurger de V⁰ Ω,
- N₀ s'insurger de ce (ci + là),
- N₀ s'insurger ppv₁,

N_0 s'insurger de le fait Qu P,
 N_0 s'insurger ppv₂
 N_0 s'insurger contre N_{hum} }

$L(<douter, T'-8>) =$

{ N_0 douter de ce que P,
 N_0 douter que P,
 N_0 douter de $V^0 \Omega$,
 N_0 douter de ce (ci + là),
 N_0 douter ppv₁,
 N_0 douter de le fait Qu P,
 N_0 douter ppv₂ }

3 Questions soulevées par T'

Supposons que T' ait été formulé et que pour une entrée quelconque il soit possible de calculer $L(<e, T-i>)$.

Pour l'utilisation de T' deux questions se posent maintenant : (i) quelle est la relation d'un l quelconque appartenant à $L(<e, T-i>)$ avec les énoncés du français ? ; (ii) quelle est la couverture de T' ? Si l'on avait des réponses pour (i) et pour (ii) on pourrait d'une part évaluer l'adéquation externe de T' et d'autre part, modifier, adapter ou étendre T' . Ces deux questions seront successivement abordées dans §3.1 et §3.2. Dans la section finale §3.3, on esquissera une caractérisation globale de Th , la théorie sur laquelle T et MS sont fondées. Les trois points soulevés dans cette section 3 posent des questions et esquissent des réponses demandant dans tous les cas à être étayées et confirmées.

3.1. L'interprétation de T'

Les objets appartenant à $L(<e, T-i>)$ sont des listes constituées de symboles. Interpréter une liste de ce type c'est être capable de l'associer à des énoncés du français. Pour cela, on devrait être capable de prendre chacun des symboles d'une liste et pouvoir dire à quelle partie d'un énoncé il doit correspondre et ce qu'il représente comme information

sur cet énoncé. Dans le cadre de cette problématique, on ne s'intéressera ici qu'à la question de savoir quelle est la relation entre listes différentes définies pour un même lemme. Cette relation concerne l'interprétation des symboles qui composent ces listes, quelles que soient ces interprétations. Par exemple, selon l'exemple de §2.3 *s'insurger* s'intègre, parmi d'autres, aux listes de (18).

- (18) (i) N_0 s'insurger
 (ii) N_0 s'insurger de le fait Qu P
 (iii) N_0 s'insurger ppv₁
 (iv) N_0 s'insurger ppv₂

N_0 doit-il être interprété de la même manière dans tous les cas ? Pourquoi a-t-on deux ppv indicés ? Selon les informations que l'on peut glaner dans *MS* (cf. *MS* :55-59), les N_0 de (18) ne doivent pas forcément être interprétés de la même manière ; les indices de ppv ont été rajoutés aux labels dans *T-8* pour que, justement, les ppv ne soient pas non plus interprétés de la même manière. Autrement dit, une identité de notation dans *T* ne doit absolument pas être comprise comme entraînant une identité d'interprétation.

Dans (16) on a décliné les procédés permettant d'obtenir des listes à partir des listes. Comme dans toutes il y a le symbole *V* (ou *U*) on peut le(s) noter à gauche et poser le problème général en (19).

- (19) $V\Omega \rightarrow V\Omega_1$

Chaque procédé devrait être décrit en indiquant tous ses effets sur l'interprétation des symboles dans Ω et dans Ω_1 . Or il n'en est rien. La clef pour comprendre les opérations sur les listes est la notion de transformation. Or, pour résumer schématiquement la situation, on peut dire que *MS* parle partout des transformations et ne les définit nulle part, ou à peine (cf. *MS* :27-30), et, ce qui est étrange, lorsqu'elles sont (à peine) caractérisées, on dit aussitôt que la caractérisation n'est pas respectée, p. ex. à propos de la contrainte de la non-insertion de morphèmes vides qui n'est pas respectée par la transformation passive (*MS* :28-29). Certes (*MS* :29-30) évoque « le principe de conservation du sens par

transformation » et indique que « La solution transformationnelle consistera à décrire P_1 avec toutes ses contraintes, et à y appliquer une opération $[T_1]$ qui transforme P_1 en la phrase P_2 sans affecter les contraintes ». Mais la lecture attentive de MS et des discussions qui y sont faites à propos des transformations utilisées et non utilisées dans T aboutit à un bilan bien différent. En général on peut dire que si l_i et $l_j \in L(<e, T-i>)$ sont associées par une opération de transformation, elles ont plus de chances d'être associées à un même « sens » – qui n'est jamais défini – et de conserver les propriétés entre leurs éléments, que si elles n'étaient pas associées par transformation. Mais comme toujours en MS , un principe général subit de nombreuses entorses selon les besoins particuliers. Dans $T-9$ on trouve l'entrée *dire* avec les complétives indiquées dans (20).

- (20) Paul dit à Marie qu'elle viendra
Paul dit à Marie qu'elle vienne

Les sens différents des deux énoncés de (20) sont bien reconnus dans ($MS :138$), où on ajoute que « nous devrions en fait décrire *dire* au moyen de deux entrées, une pour chacun des deux sens ». Mais cette possibilité est écartée au profit d'une autre solution : une transformation nouvelle, non utilisée dans $T-9$, à peine esquissée dans ($MS :139$) qui consisterait à dériver l'enchassée au subjonctif dans (20) de *Paul dit à Marie qu'il exige qu'elle vienne*. MS est plein de discussions de ce type (on y reviendra au §3.3) dont on ne remet pas en cause l'intérêt profond. Mais la question posée ci-dessus reste entière : on ne connaît pas les propriétés interprétatives communes à l_i et l_j appartenant à $L(<e, T-i>)$ et associées par transformation. Si l'on se réfère à la possibilité ou non d'effacer des morphèmes pleins par les transformations (un des quatre principes proposés en MS pour caractériser les transformations, $MS :27$), excellent exégète serait celui qui au vu de l'exemple de *dire* et de la discussion dans ($MS :28-29$) pourrait dire autre chose que le principe en question est respecté parfois et parfois pas.

La discussion de *dire* conduit à un autre problème, toujours lié à l'interprétation des listes auxquelles s'intègre un lemme déterminé. Tous les lemmes verbaux considérés dans MS n'ont pas été placés dans

une seule ligne d'une seule table. Certains lemmes occupent deux lignes dans une table et d'autres apparaissent dans des tables différentes.

Si les relations entre $l_1 \in L(\langle e_x, T-i \rangle)$ et $l_2 \in L(\langle e_y, T-i \rangle)$ avec un même lemme dans e_x et e_y pour ce qui est de leur interprétation sont difficiles à établir, les relations entre $l_1 \in L(\langle e_x, T-i \rangle)$, $l_2 \in L(\langle e_y, T-i \rangle)$ et $l_3 \in L(\langle e_z, T-j \rangle)$ avec un même lemme dans e_x , e_y et e_z sont définitivement impossibles à détecter. Strictement aucune discussion sérieuse n'existe dans *MS* sur ce point crucial, où l'on se limite à signaler que l'appartenance des listes à une même entrée est signe de l'existence de relations possibles entre ces listes (*MS* :152) et que la distribution des entrées dans les tables est souvent arbitraire (*MS* :152). Si l'on ajoute que l'observation de cas concrets comme celui de *continuer* dans *T-1* (deux entrées pour un même sens) comparé à celui de *dire* dans *T-8* (une entrée pour deux sens) montre que des solutions opposées ont été adoptées pour un même problème, on est amené à conclure que si T' est lisible et, sur certains points, plus interprétable que T , il reste largement formellement ininterprétable.

Si l'on reprend la comparaison avec les horaires SNCF, on peut dire que T est l'analogue d'un horaire dont on ne peut lire les heures indiquées pour chaque gare et que si dans T' on peut lire ces heures, on ne sait pas toujours à quoi elles correspondent.

3.2. La couverture de T'

D'après (*MS* :21-215) on doit comprendre qu'un ensemble d'environ 3 000 verbes ont été étudiés par rapport à « une centaine de propriétés » (ou plutôt 110 propriétés). On sait aussi qu'une colonne représente une propriété syntaxique (*MS* : 150). Or, comme on constate que le nombre de colonnes dans les tables va d'un minimum de 19 dans *T-4* et *T-19* à un maximum de 42 dans *T-15* et *T-16*, on aimerait comprendre comment avec un nombre de colonnes inférieur à 110 on peut décrire le comportement des verbes par rapport à 110 propriétés. On aimerait aussi avoir l'inventaire des 110 propriétés étudiées.

La question évidente qu'un utilisateur de T ou d'un produit dérivé de T (tel que GENELEX) doit se poser est la suivante : comment tel verbe par rapport à tel type de phénomène peut-il être utilisé en français ?

Toute réponse possible suppose de savoir « ce qu'il y a dans T », c'est-à-dire de connaître les propriétés choisies pour construire T . Or ces propriétés choisies n'ont pas été explicitées. P. ex. : peut-on déterminer pour tout élément d'une liste associée par T à un lemme verbal étudié dans MS ses conditions de pronominalisation en clitique ? La réponse est négative. La conclusion de la discussion est que T est structurellement non exhaustif. Un des résultats, non surprenant, de T' est de montrer les limitations sévères de T : T' est largement ininterprétable formellement et structurellement non exhaustif. C'est un peu comme si la SNCF publiait des horaires avec des indications d'heures non lisibles, avec des gares arbitraires et avec des notes fixant des conventions de lecture non respectées, et comme si les voyageurs découvraient qu'il y a des trains annoncés qui ne circulent pas, des trains qui circulent qui ne sont pas annoncés et des heures indiquées qui correspondent parfois à des gares où le train s'arrête et parfois à des gares où il ne s'arrête pas.

3.3. La théorie de MS

MS est un gisement sinon inépuisable au moins inépuisé d'observations linguistiques sur la sous-catégorisation verbale. On ne peut qu'adhérer à certaines manières de poser les problèmes, face auxquelles toute linguistique devrait prendre position :

« Ces remarques concernant la combinatoire finie dans des suites finies de mots nous ont conduit à penser que divers faits de langue fondamentaux pouvaient ne pas se situer dans les mécanismes récurrents, mais au niveau de phrases simples » (MS :18).

Au moins trois aspects doivent être distingués dans MS :

- (i) Le formalisme de représentation
- (ii) La description de phénomènes
- (iii) L'explicitation de phénomènes

Par rapport à (i), le bilan que l'on peut tirer c'est que la méthode de représentation par tables mène à l'échec. On voit mal comment avec le système de relations que les matrices des tables peuvent formaliser il est possible d'exprimer de manière intuitivement claire et menant à la

possibilité de calculs, les relations complexes entre les formes dans les listes de constituants linguistiques et les sens associés à ces listes. Par rapport à (ii) et à (iii), il ne semble pas vrai que l'objectif premier de *MS* soit d'emmagasiner des phénomènes dans un répertoire réutilisable. *T* se veut – et est – le lexique d'une grammaire qui a des règles qui, par principe ne doivent pas être incorporées dans le lexique. Cette grammaire se veut explicative. Le groupement des entrées dans les tables a une visée explicative. Ce groupement ne se fait pas au hasard et les entorses notationnelles et à certains aspects du formalisme choisi ne sont pas neutres. Est-ce que *MS* a réussi à expliquer ?

En opérant à grands traits, on peut diviser ceux qui se sont occupés de la sémantique de la langue naturelle en deux groupes : d'un côté, ceux qui raisonnent en termes d'un objet dual, le signe, association d'une forme avec un sens. C'est à partir du sens qu'on interprète un énoncé dans un univers. On inclura dans ce groupe Saussure, Chomsky, Sowa et Kamp. De l'autre côté, il y a ceux pour qui le sens ou bien n'existe pas, ou bien est une commodité d'écriture dont en principe on peut se passer. On rangera ici Quine, Harris et Montague. C'est à ce deuxième groupe que *MS* doit être rattaché.

Si l'on suit dans le détail les multiples discussions de *MS* à propos de « l'origine » de telle ou telle construction, (c'est-à-dire, des transformations à appliquer à une liste primitive) on peut entrevoir ce que *MS* recherche : il s'agit de faire se répercuter sur des listes de mots (et non sur des structures sémantiques dans les entrées, structures qui ne sont jamais définies) des variations de sens. Celui-ci peut disparaître ; l'interprétation dans un univers peut se faire à partir des listes de mots. Les variations de sens peuvent être dévoilées par les variations des listes. Le prix terriblement lourd à payer est la non formalisation de ces manipulations.

Peut-on utiliser les tables de *MS* pour construire un réservoir lexical comme celui visé par GENELEX, DELIS ou d'autres projets sur le lexique ? Tout dépend ici de ce qu'on entend par *utiliser*. L'information dans *T* est grande, mais elle est cachée et confuse. Les tables de *T* doivent être déstructurées pour la découvrir. Cette illisibilité est un des échecs de *T*. Peut-on décrire la sémantique des entrées lexicales par des opérations sur des listes de mots où apparaît cette entrée ? Peut-on

décrire la différence entre deux *dire* en disant que l'un est *N₀ dire que N₀ exige* alors que l'autre n'a pas cette exigence sur l'enchâssée ? *MS* montre que cette voie est une impasse, c'est son deuxième échec. Que faire donc des listes dans les tables de *MS* pour les associer en tant que telles à une sémantique ? Plus généralement, est-il possible de modéliser les entrées lexicales sous la forme d'une couche syntaxique, sans tenir compte de l'association des éléments de cette couche avec la sémantique ? Voici les vraies questions qui devraient être mieux posées et pour lesquelles il devrait y avoir des éléments de réponses³.

RÉFÉRENCES

- Alcouffe, Ph., Revellin Falcoz, B., Zaysser, L.
(1993) « Azote : des tables du LADL au format GENELEX ». Colloque « Informatique & langue naturelle », Nantes, décembre 1993 : 387-409.
- Bès, G.G., Baschung, K., Lecomte, A.
(1991) *Une modélisation des entrées lexicales*. Rapport projet Eureka GENELEX.
- Bès, G.G., Emorine, M
(1991) *Une lecture des tables de LADL en vue de la définition de la couche syntaxique de GENELEX*. Rapport projet Eureka GENELEX.
- Bès, G.G., Lecomte, A.
(sous-presse) « Semantics features in a generic lexicon ». Dans Saint-Dizier, P., Viegas, E. *Computational lexical semantics*, Cambridge University Press.

3 La nécessité d'adopter un objet dual pour traiter la sémantique lexicale a été exposée et argumentée dans (Bès, Baschung et Lecomte, 1991), ce qui n'exige nullement d'avoir à incorporer des définitions. Les raisons pour lesquelles il semble bien qu'il soit impossible de faire un calcul de représentations sémantiques à partir d'objets exclusivement syntaxiques ont été partiellement exposées dans (Bès et Lecomte, sous presse). La couche syntaxique du modèle GENELEX venant tout juste d'être publiée, un examen attentif montrera les possibilités et les limites du modèle.

Consortium GENELEX

(1993) *Rapport sur la couche syntaxique*. Projet Eureka GENELEX.

Clemenceau, D., Roche, E.

(1991) « Lexicon grammar » dans Hellwig, P., Minkwitz, T., Koch, H-D « Standards for syntactic descriptions ». Eurotra-7 Study : 41-52.

Gross, M.

(1975) *Méthodes en syntaxe*, Paris, Herman.

Leclère, Ch.

(1990) « Organisation du lexique-grammaire des verbes français ». *Langue française*, 87 : 112-127.

Roche, E.

(1993). *Analyse syntaxique transformationnelle du français par transducteurs et lexique-grammaire*. Thèse de doctorat. Université Paris 7.

