



HAL
open science

Éléments méthodologiques pour la discussion des relations entre sciences cognitives : linguistique, psychologie, informatique

Gabriel G. Bès

► **To cite this version:**

Gabriel G. Bès. Éléments méthodologiques pour la discussion des relations entre sciences cognitives : linguistique, psychologie, informatique. [Rapport de recherche] Université Blaise-Pascal. 1987. hal-01143148

HAL Id: hal-01143148

<https://hal.science/hal-01143148>

Submitted on 16 Apr 2015

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.

Elements méthodologiques pour la discussion des relations entre sciences cognitives : linguistique, psychologie, informatique

Gabriel G. Bès
Formation doctorale Linguistique et informatique,
Université Blaise-Pascal, Clermont-Ferrand

Rapport élaboré pour le document de synthèse de l'ATP « Intelligence artificielle et sciences cognitives - Stratégies d'interprétation de textes », 1987

Résumé

Ce rapport présente des éléments méthodologiques pour la discussion des relations entre linguistique, psychologie et informatique. Le texte présente d'abord une organisation du cadre conceptuel où traiter les problèmes : l'auteur explicite différentes dichotomies entre lesquelles se placent les entités sur lesquelles opère la linguistique, puis il présente la notion de système informatique. Sont ensuite caractérisés trois types d'intégration des modèles linguistiques aux systèmes informatiques : intégration terminologique, observationnelle ou forte. L'auteur défend l'idée que la création récente des modèles de grammaire GPSG, LFG et grammaires catégorielles permet une intégration forte de la linguistique avec l'informatique et qu'il est impossible de traiter informatiquement les langues naturelles sans passer par un modèle linguistique formel.

Voir aussi

Gabriel G. Bès, « Fonctions de représentation et de communication », *in* Piaget J., Bronckart J.-P., Mounoud P. (dir.), *Psychologie, La Pléiade*, Gallimard, Paris, p. 1757-1835, 1987.
<https://hal.archives-ouvertes.fr/hal-01065925>

A.R.I Sciences de la communication
A.T.P Intelligence Artificielle et Sciences Cognitives
Stratégies d'interprétation de textes
Resp : C. Fuchs

**ELEMENTS METHODOLOGIQUES POUR LA DISCUSSION
DES RELATIONS ENTRE SCIENCES COGNITIVES :
LINGUISTIQUE, PSYCHOLOGIE, INFORMATIQUE**

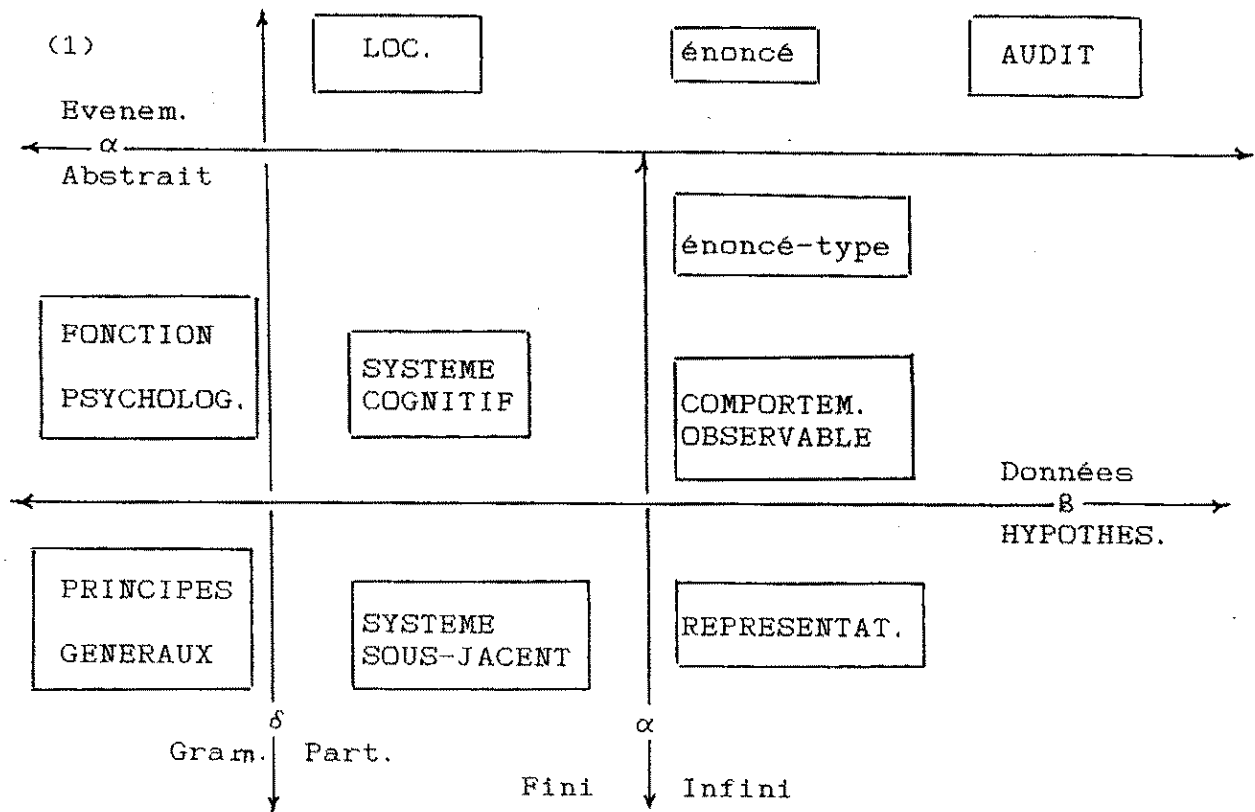
par Gabriel G. Bès

Rapport élaboré pour le document de synthèse de l'ATP

Université de Clermont II Blaise Pascal
Formation Doctorale Linguistique et Informatique
34, Avenue Carnot
F 63037 Clermont Fd

1 - Un cadre conceptuel où traiter les problèmes

En fonction des dichotomies α à δ , (1) schématise le cadre conceptuel où il devrait être possible de situer les objets que l'on doit traiter et les relations qu'ils entretiennent.



La dichotomie α sépare ce qui est événementiel et par conséquent irréproductible, du champ de l'abstrait. Pour parvenir à celui-ci, il a été nécessaire de modéliser (dans le sens d'idéaliser) l'objet événementiel du départ. Dans ce champ abstrait on situe les données, c'est-à-dire les entités sur lesquelles porteront les hypothèses. Ces entités sont par conséquent partiellement déterminées par le type de modélisation choisi. Il est ainsi possible de faire la distinction entre l'énoncé ("token", occurrence) et l'énoncé-type, qui est une classe d'énoncés constituée selon des critères explicites. Les énoncés-types, par le biais des énoncés qui les expriment, sont associés à un comportement observable qu'il faut également définir au départ dans la modélisation adoptée.

La dichotomie α sépare ce qui est infini (les énoncés-type et le comportement observable qui leur est associé) de ce que l'on considère comme fini, car stocké dans un cerveau humain (système cognitif et fonction psychologique correspondante).

La dichotomie β sépare les hypothèses des données, celles-là portant sur celles-ci. La dichotomie δ , enfin, distingue ce qui est général et non déterminé culturellement, de ce qui est particulier.

Les entités sur lesquelles opère la linguistique peuvent se situer sur le schéma de (1). On peut cerner une modélisation A, à deux volets (A-1 et A-2). Dans le cadre de A-1, les énoncés-type sont constitués en épurant les énoncés des conditions énonciatives, le comportement observable est constitué de jugements métalinguistiques, le système cognitif est appelé compétence, le système sous-jacent qui en rend compte, grammaire particulière (= GP), et grammaire universelle (= GU) les principes généraux qui introduisent des restrictions sur les grammaires particulières.

Le volet A-1 est complété par A-2, où le comportement observable devient des opérations concrètes sur le langage : production ou interprétation des énoncés. En parallèle, le système cognitif de A-2 est censé intégrer la GP de A-1. On parle de A-2 comme du modèle de la performance.

La modélisation A n'est pas la seule possible. On peut imaginer une modélisation B qui ne fasse pas la distinction entre les deux volets de la modélisation A. Le système cognitif est alors un système qui détermine des opérations effectives du cerveau, et dans le comportement observable on incorpore le traitement effectif des énoncés. Du système sous-jacent correspondant on dira qu'il est un système langagier plutôt qu'un système linguistique ou un modèle de la compétence. Malgré la possibilité de principe d'une modélisation de type B et malgré des tentatives intéressantes dans ce cadre, le reste de ce rapport ne tiendra compte que des modèles linguistiques qui se situent dans le cadre de la modélisation A.

La modélisation A situe les hypothèses linguistiques dans le champ des sciences empiriques : les hypothèses linguistiques sont falsifiables. La modélisation laisse en revanche ouvert le niveau de formalisation dans la présentation des hypothèses. Celles-ci peuvent s'exprimer dans une échelle qui va des formulations en langue naturelle jusqu'à une formalisation forte à l'aide d'entités dont les propriétés mathématiques sont connues. De manière analogue, les hypothèses linguistiques peuvent être plus ou moins générales.

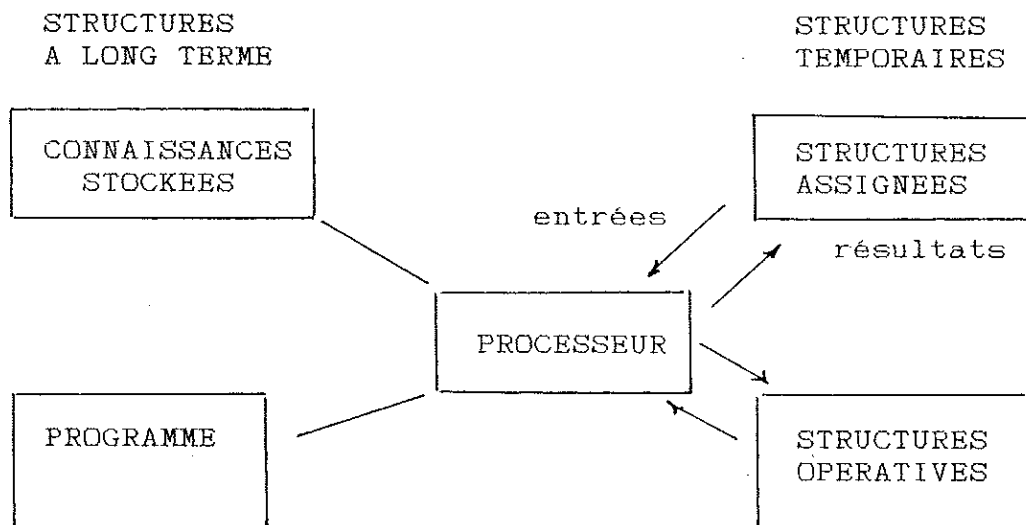
Sur un plan général, on peut penser que le schéma de (1) peut intégrer toute science cognitive : celle-ci, qu'elle soit, vise en effet à proposer des hypothèses sur des systèmes cognitifs sous-jacents, non observables en eux-mêmes et qui se manifestent par des comportements observables spécifiques. Ceux-ci, en tant que tels, ne constituent donc pas l'objet de connaissance visé ; ils ne sont que la manifestation d'un système qu'il s'agit de représenter sans

pouvoir l'observer directement. Sur un plan plus spécifique, le volet A-2 va permettre d'établir la relation entre linguistique et psychologie. On peut y situer les travaux de psycholinguistique qui admettent la nécessité d'intégrer une GP au système cognitif sous-jacent. Mais tout modèle psychologique portant sur la langue naturelle, même s'il se laisse appréhender par le schéma de (1), ne correspond pas au volet A-2. On peut même constater que l'évolution de la psycholinguistique montre une certaine désaffection pour A-2. Dans ce cas, le problème du rapport entre modèle linguistique de A-1 avec système cognitif portant sur le langage se pose dans toute son acuité (cf (2) ci-dessous).

Que ce soit dans le cadre de A-2 ou dans un cadre plus autonome par rapport à la linguistique, les hypothèses sur les systèmes cognitifs -tout comme les hypothèses linguistiques- peuvent s'exprimer à différents niveaux de formalisation et de généralisation.

Le schéma computationnel. Suivant Winograd (p.18), on peut penser qu'un système informatique (= SI) peut être représenté par le schéma de :

(2)



Un système informatique tel qu'il est visualisé dans le schéma de (2) exige une explicitation totale des structures traitées : celles-ci doivent être toujours représentées et manipulées par le biais des informations explicites introduites dans la machine. Contrairement aux modèles cognitifs, un SI exige une explicitation totale, faute de quoi, par définition, il n'est pas un SI. A la différence des modèles précédents, on n'aperçoit pas la possibilité de le falsifier empiriquement. Il est bien probable que les informaticiens opèrent avec des critères d'adéquation, mais

ceux-ci semblent se définir plutôt en fonction de considérations d'efficacité propres au système qu'en fonction d'une relation avec des objets qu'il s'agit de connaître et d'expliquer.

Le volet A-2 du modèle linguistique s'est en particulier développé dans le cadre de la grammaire générative, transformationnelle ou non, chomskyenne ou non. Suivant toujours Winograd (et d'autres avant lui), c'est en fonction d'une métaphore mathématique qu'il est possible de caractériser la relation entre GP et représentations :

"The concept of abstract competence is closely related to notion of proof in mathematics. We can think of mathematics as a 'language' of formulas, and the job of the mathematician as explaining which combinations of symbols represent true propositions, given a set axioms and rules of inference [...] Mathematics is not the study of how people invent such expressions or what gives on in their minds when they read or try to prove them [...] Generative linguistics views language as a mathematical object and builds theories that are very much like sets of axioms and inference rules in mathematics. A sentence is grammatical if there is some *derivation* that demonstrates that its structure is in accord with the set of rules, much as a proof demonstrates the truth of a mathematical sentence" (p.12).

En fonction de ce qui précède, on peut cerner les difficultés structurelles qui existent pour mettre en rapport la linguistique (au moins telle qu'elle est conçue en A-2), avec l'informatique : dans un modèle linguistique on ne trouvera pas, par principe, l'algorithme permettant d'associer une représentation à un énoncé déterminé. Par ailleurs, la plupart des modèles existants et les GP qui leur sont associées ne sont pas exprimées dans une formalisation forte ; ils sont souvent faiblement explicités. En revanche, l'expérience de généralisation -au moins de principe et souvent au détriment de l'adéquation descriptive- y est grande. Un SI exige, lui, l'explication totale des structures et des opérations qui permettent de les associer, l'exigence de généralisation y étant moins grande et souvent nulle.

2 - L'intégration des modèles linguistiques aux systèmes informatiques.

Un SI est un outil particulièrement adéquat pour exprimer A-2 ; dans le reste de ce rapport on ne fera référence qu'à des systèmes informatiques de ce type, c'est-à-dire, à des analyseurs syntaxiques dont l'entrée est constituée d'une suite de caractères (ou, parfois, une forme vocalique) et dont la sortie est une représentation sémantique. Le problème ainsi posé est parallèle à celui posé depuis longtemps en linguistique, à savoir, l'intégration de la compétence dans la performance.

Après une trentaine d'années de déclarations de principe, plus ou moins oscillantes, il faut tirer un constat d'échec sur l'intégration de la compétence telle qu'elle est conçue dans un modèle chomskyen à un quelconque modèle de la performance qui présente un minimum d'explicitation et de généralisation. Ce n'est donc pas du côté de l'intégration des modèles chomskyens à un éventuel modèle de la performance que l'on trouvera une solution pour le problème évoqué'.

Sur un plan général, on peut distinguer trois "intégrations" des modèles linguistiques aux systèmes informatiques.

"Intégration" terminologique. Le système informatique incorpore un jargon technique propre à un modèle déterminé pour parler de ses entrées et de ses sorties. C'est ce qui se constate dans les systèmes informatiques qui disent employer la ou les grammaires de "cas" : en fait, on utilise les étiquettes véhiculées par les grammaires des cas pour catégoriser des constituants dans les énoncés.

"Intégration" observationnelle. Supposons un analyseur syntaxique du français susceptible d'associer une représentation sémantiquement adéquate à tout énoncé français comportant des clitiques. Ce système aura "intégré" sur le plan observationnel toutes les règles de transition entre clitiques explicités -sans formalisation forte- dans Benveniste et alii. Ce SI va associer une représentation à (3) et à (4), et aucune à (5) :

- (3) Pierre nous y amènera ce soir
- (4)(2) Pierre nous présentera à Jacques
- (5)(3) * Pierre nous lui présentera

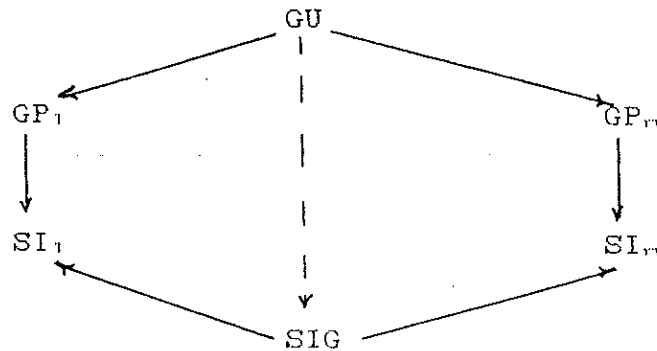
Pour éviter (4), Benveniste et alii proposent des "règles d'éjection". Celles-ci expriment en fait des contraintes de transition entre catégories de clitiques. Elles sont adéquates sur le plan de l'observation mais, elles ne s'insèrent pas dans un modèle formalisé, et le SI ne peut pas les intégrer dans le sens fort du terme : il produit des résultats qui, dans son cadre propre, sont équivalents à une des règles mais il n'y a pas de fonction permettant d'associer les règles d'éjection au SI.

Intégration forte. L'intégration forte des GP à un système informatique peut s'inscrire dans la formule suivante :

$$(6) \text{ fonction } \langle GP_i \rangle = \text{Syst. Inf}_i$$

c'est-à-dire, on cherche une fonction telle qu'elle va associer une GP adéquate pour le langage_i à un système informatique susceptible d'associer des représentations sémantiques adéquates aux suites du langage_i. A un niveau plus général, on souhaite trouver la relation exprimée dans le schéma de (7) :

(7)



On souhaite que la relation soit satisfaisante par toute GP particulière et tout système informatique respectivement spécifiés par une GU et ce qu'on pourrait appeler un système informatique général (= SIG).

Jusqu'en 1982, la fonction de (6) et les modèles linguistiques susceptibles de spécifier des GP qui s'y inscrivent, n'existaient pas ou n'existaient que de manière limitée. A partir notamment des travaux de Bresnan et de Gazdar, des modèles ont vu le jour, et sans prétendre nullement que "tout est réglé" ils montrent l'existence d'une voie de recherche où l'intégration de la linguistique à l'informatique se fait de manière formelle, et, sans doute, avec des retombées technologiques profitables.

Soit la question des clitiques évoqués à l'alinéa précédent. Le modèle GPSG (cf. Gazdar et alii) n'incorpore que des règles syntagmatiques indépendantes du contexte. Dans celles-ci, seule la relation de dominance est exprimée, les relations d'ordre étant exprimées dans les stipulations de précédence linéaire. On a montré qu'au moins les transitions entre clitiques argumentaux en français peuvent s'exprimer dans une grammaire GPSG du français (cf. note 2). Pour obtenir les résultats des règles d'éjection mentionnées dans le paragraphe précédent, on peut avoir recours à une stipulation de précédence linéaire du type :

(8) $C_i < C_i$

qui va éviter que deux catégories identiques se suivent. Comme dans GPSG toute catégorie est une fonction et comme GPSG incorpore des fonctions partielles, il est possible de définir les clitiques de telle manière que la formulation (6) soit susceptible de s'appliquer avec des résultats adéquats. GPSG définit des GP qui vont pouvoir s'intégrer à (6). Un analyseur qui serait le résultat de (6) et qui incorporerait une grammaire GPSG du français rendant compte des clitiques, serait équivalent sur le plan extensionnel, d'un analyseur syntaxique qui "intègre" sur le plan observationnel les effets

des règles d'éjection. Mais l'intégration forte est systématique et formelle, et donc contrôlable, avec les bénéfices de toute formalisation.

GPSG n'est pas le seul modèle existant aujourd'hui : LFG et des modèles issus des grammaires catégorielles existent aussi; qui s'inscrivent tous dans la même ligne de recherche, de même qu'il existe des systèmes informatiques généraux²⁰. Le schéma de recherche de (9) commence à être ainsi mis en place.

Sur ce schéma s'inscrit une intégration forte de la linguistique avec l'informatique. On peut y observer une interaction constitutive de la recherche exprimée moyennant les entrées et sorties notées par les flèches. Ce sont les échecs au niveau de l'analyseur qui vont être utilisés pour corriger les modèles linguistiques qui, par ailleurs, devraient être révisés en permanence en fonction des observations linguistiques. Sur ce schéma est explicitement absente toute flèche allant des données à l'analyseur syntaxique, malgré le fait que cette flèche existe pour la plupart des analyseurs syntaxiques développés en France. Le schéma (9) explicite ainsi une idée claire, qui résume ce rapport :

(10) Il est impossible -et la recherche des années qui vont suivre le montrera- de traiter informatiquement les langues naturelles sans passer par un modèle linguistique formel.

Notes

1- Pour la justification de ces observations, cf. Bès 1987 a

2- Cf. Gazdar (1982) et Bresnan (1982) Pour aboutir à ce stade, il a été déterminant de montrer qu'il est possible de traiter les langues naturelles moyennant des grammaires de contexte libre ; cf. Pullum et Gazdar (1982)

3- Cf. une analyse des clitiques selon GPSG dans Bès (1987 b) et un examen critique de GPSG dans Bès et Baschung (1986) Comme exemple de grammaire catégorielle, cf. Zeevat 1986 b. Comme SI, cf. Shieber et alii (1984) et ses aspects théoriques dans Shieber (1986), Calder (1986) est une implémentation PROLOG de PATR-II. Le modèle UCG a été utilisé pour le français dans Baschung et alii (1987 a) et Baschung et alii (1987 b). Cette dernière GP du français portant notamment sur les structures interrogatives est maintenant implémentée sur SUN 3/50 aux Laboratoires de Marcoussis et à l'Université de Clermont II Blaise Pascal.

Dans Zeevat et alii (1986 a) et Panckhurst (1986) on trouvera la présentation d'un langage descriptif permettant d'obtenir les "Observations systématisées" du schéma de (9).

Références

- Baschung, K, G. G. Bès, R. Panckhurst, A. Corluy et Thierry Guillotin (1987- b) *Syntactic French Parser for Dialogue*. Rapport PROJET ESPRIT 393 ACORD.
- Baschung, K, G.G. Bès, A. Corluy, A. et T. Guillotin (1987) "Auxiliaries and clitics in French UCG grammar". Communication présentée à *The Third Conference of the European Chapter of the Association for Computational Linguistics*, Copenhagen.
- Blanche-Benveniste, Cl, J. Denlofen, J. Stefanini et K. Van Den Eynde (1984) *Pronom et syntaxe : l'approche pronominale et son application au français*. Paris, Selafl.
- Bès, G.G. (1987 a) "Fonctions de communication et de représentation" Dans J. Piapet et alii [eds] *Psychologie*. Paris, Gallimard, La Pleiade.
- Bès, G.G. (1987 b) "Clitiques et constructions topicalisées dans une grammaire GPSG du français" Dans *Lexique*, N°5, à paraître.
- Bès, G.G. et Baschung, K (1986). "Feasibility of a GPSG French Grammar" Rapport PROJET ESPRIT 393 ACORD.
- Bresnan, J. (ed) (1982). *The mental representation of grammatical relations*. Cambridge, Mass. The M.I.T. Press.
- Calder, J. (1986) PIMPLE. Edinburgh, Center for Cognitive Science.
- Gazdar, G. (1982). "Phrase structure grammar", dans P. Jacobson et G.K. Pullum (eds). *The nature of syntactic representation*, Dordrecht, Reidel, p. 131-186.
- Gazdar, G., E. Klein, G. Pullum et I. Sag (1985). *Generalized phrase structure grammar*. Cambridge, Mass., Harvard University Press.

Panckhurst, R. (1986). Annex to Zeevat et alii (1986). Rapport PROJET ESPRIT 393 ACORD.

Pullum, G. et G. Gazdar (1982). "Natural languages and context-free languages", *Linguistics and Philosophy*, 4, p. 471-504.

Shieber, S.M. (1986) *An introduction to unification based approaches to grammar*. Stanford, CSLI.

Shieber, S.M., H. Uszkoreit, F.C.N. Pereira, J. Robinson and Tyson (1984). "The formalism and implementation of PATR-II". Dans B. Grosz et M. Stickel (eds). "Research on the interactive acquisition and use of knowledge" SRI final report.

Winograd, T. *Language as a cognitive process : Syntax* (1983). Massachusetts, Addison-Wesley.

Zeevat, H. G.G. Bès, R. Panckhurst, A. Corluy et Th. Guillotin (1986 a) *Contextual phenomena in dialogue*. Rapport PROJET ESPRIT 393 ACORD.

Zeevat, H, E. Klein et J. Calder (1986 b). *Unification Categorical Grammar*, Edinburgh, Center for Cognitive Science.

(7)

