

## ANALYSEURS SYNTAXIQUES EQUIVOQUES ET PROBLEMES

Jean-Marie MARANDIN\*

### Résumé - Abstract

Le statut épistémologique de l'analyseur syntaxique (dorénavant analyseur) est examiné dans trois programmes de recherche: le traitement du langage naturel (TAL), la syntaxe et la compréhension du langage naturel (CNL). On montre que l'analyseur occupe dans le TAL la place de l'instance de performance telle que la définit le modèle standard de la grammaire générative et transformationnelle; il en hérite les difficultés. On montre ensuite qu'un parseur ne peut pas être une instance d'expérimentation pour la syntaxe. L'analyseur a un statut clair dans la CNL: c'est un dispositif; mais son contenu de connaissance n'est pas assuré tant qu'il ne remplit pas les impératifs liés à ce statut: permettre l'analyse critique des conjectures dont il donne une image concrète.

**Mots clefs:** épistémologie de la linguistique, épistémologie du dispositif, analyse syntaxique automatique, expérimentation en linguistique, performance, compréhension du langage naturel.

The epistemological status of the syntactic parser (henceforth parser) is examined in three different research programs: natural language processing (NLP), syntax and natural language understanding. The parser is reducible to the performance device defined in the standard model of the Transformational Generative Grammar in NLP. The parser cannot be an experimental tool for any theory of syntax. The parser is a concrete image of a conjecture in NLUnderstanding; as such it must allow the critical analysis of the conjecture it embodies.

**Key words:** epistemology of linguistics, syntactic parsing, experiment in linguistics, performance, understanding.

---

\* URA 1028, Tour centrale, 9e étage, Université Paris VII, 2 Place Jussieu, 75251 Paris cedex 05, Jean-Marie Marandin @ linguist.jussieu.fr

## 1. INTRODUCTION

Cet article est le résultat d'une longue période de sueurs et de colères froides face aux confusions et aux dialogues de sourds qui entourent la conception d'analyseurs syntaxiques (abrégé dorénavant en *analyseur*). *Machine, dispositif, modèle, analogue, algorithme, système expérimental*, l'analyseur est un caméléon, et il n'est pas rare de le voir changer de couleur au fil des discussions ou des présentations. A l'évidence, selon qu'il est une machine ou un dispositif, un modèle ou un système expérimental, un analyseur est pris dans des programmes de recherche ou de travail qui ont des logiques et des objectifs distincts. D'où la confusion; d'où la question à laquelle est consacré cet article:

- (1) Quel statut prend, ou peut prendre, un analyseur dans un programme de recherche consacré au langage?

La question peut sembler naïve ou excessivement théorique. Entre naïveté et raffinement théorique, je voudrais donner place au rappel d'un certain nombre de distinctions qui font que deux programmes consacrés à la réalisation d'un analyseur syntaxique peuvent être aussi incommensurables que le sont la confection d'un parapluie et une opération chirurgicale. Il y a, bien sûr, un lieu de rencontre: un horizon de références informatiques identique, mais il ne justifie pas que l'on identifie des programmes qui n'ont en commun que leur nom.

J'examinerai la place que peut occuper un analyseur dans les deux domaines distincts que sont la syntaxe et la compréhension du langage naturel (CLN). Auparavant, j'examinerai la conception, communément partagée dans le champ du traitement automatique du langage (TAL), selon laquelle "un analyseur est un analyseur": un système de description des énoncés (par segmentation et catégorisation) qui est caractérisé par la théorie ou le formalisme qu'il exploite (*un analyseur LFG, un analyseur catégoriel, un analyseur basé sur l'unification*, etc.) et qui est neutre à l'égard des modèles de performance. Je désignerai, par convention, cette conception par l'acronyme 3A (Analyse automatique autonome).

La question (1) prend un sens en dehors du champ du traitement automatique. Les sciences, (ou les arts) du langage ont, de tout temps, eu recours à des dispositifs techniques pour donner forme à leur développement: grammaire, dictionnaire, glossaire, etc<sup>1</sup>. L'analyseur apparaît comme un nouvel outil; il est donc essentiel de bien cerner son statut épistémologique<sup>2</sup>.

---

## 2. L'ANALYSEUR DANS LE TAL

On considère couramment, dans le traitement automatique du langage, que l'analyse syntaxique automatique constitue un domaine

1 . Voir Auroux 1992.

2. J'admets tout au long de ce texte qu'un analyseur analyse des énoncés: les éléments de l'ensemble infini des productions langagières (*E-language* dans Chomsky 1986).

autonome<sup>3</sup>. Cette autonomie, qui n'est pas problématisée dans le TAL, doit être explicitée pour dégager son contenu de connaissance. C'est l'objet de ce premier paragraphe.

## 2.1. Le programme

Le programme de travail adopté dans le TAL peut être défini par la proposition suivante :

- (2) Concevoir, étant donnée une grammaire, l'algorithme ou l'heuristique qui permet de décrire des énoncés dans les termes de cette grammaire; implémenter la procédure.

La définition suivante (3) illustre parfaitement ce programme; elle est tirée d'un texte de synthèse, par ailleurs remarquable, consacré à l'analyse syntaxique en général:

- (3) "Le problème du passage [= analyse syntaxique automatique] consiste à trouver une procédure utilisant un ensemble de connaissances (i.e. une grammaire  $G$  et, éventuellement, d'autres types de connaissances), qui permettent de traduire une représentation  $r_i$  (e.g. une chaîne de mots  $\Omega = m_1, \dots, m_k$ , appartenant à un ensemble  $C$  de chaînes possibles, i. e. un langage) en une (ou plusieurs autre(s) représentation(s)  $r_j$  ( $r_k, r_m, \dots$ ), appartenant à un (ou plusieurs) autre(s) ensemble(s)  $R$  ( $R', R'', \dots$ ) de représentations plus riche(s) et mieux manipulable(s) que la première; représentation(s) sur laquelle (lesquelles), par exemple, divers traitements (vérification d'erreurs, interprétation sémantique, déduction, traduction, compression, etc.) pourront s'effectuer directement et de la façon la plus harmonieuse possible, traitements qu'il serait impossible de faire de façon cohérente sur une simple chaîne de mots (Morin, sd, p.4).

Cette définition est caractéristique de la problématique 3A. L'analyseur est déproblématisé: son existence va de soi comme va de soi le problème qu'il pose. L'analyseur est fonctionnellement évident: il rend les énoncés "mieux manipulables"<sup>4</sup>. La manipulabilité est une propriété à double visage. En tant qu'elle est une condition générale de l'expérimentation (Prigogine & Stengers 1979, entre autres), elle range l'analyseur parmi les outils de la science; en tant qu'elle est la condition de l'exploitation technologique, elle fait de l'analyseur une machine. C'est ce double statut qui est symptomatiquement exhibé dans la liste d'utilisations qui clôt la définition. A première lecture, cette liste est digne d'un inventaire chinois: les objectifs de connaissance (la question de l'interprétation sémantique) côtoient sur le

3. C'est sur cette base qu'ont été organisées les journées de l'ATALA consacrées aux analyseurs syntaxiques du français (Abeillé 1991).

4. On pourrait paraphraser la formule qu'emploie Barthes pour définir la lecture (*la lecture ajoute de l'intelligibilité au texte*) en disant que l'analyseur "ajoute de la manipulabilité aux énoncés".

même plan les objectifs technologiques (vérification d'erreurs, compression, traduction)<sup>5</sup>.

## 2.2. Archéologie de la problématique 3A

La définition (3) énonce clairement un problème à résoudre. Mais d'où viennent les termes de ce problème? Où est-il constitué? Quelle est sa cohérence interne? La réponse paraît s'imposer: la problématique 3A est entièrement déterminée par la problématique de la grammaire générative et transformationnelle (GGT) dans sa version standard. Autrement dit, la GGT est le socle "connu, mais non reconnu" qui donne cohérence et sens au programme défini en (2). Deux faits attestent cette dépendance:

- la reprise de l'analyse grammaticale,
- la forme d'ensemble de la problématique.

### 2.2.1. L'analyse grammaticale

Un analyseur effectue une analyse grammaticale des énoncés. *Analyse grammaticale* est à prendre au sens de la tradition grammaticale:

- (4) L'analyse grammaticale consiste dans l'identification des parties constitutives d'un énoncé, de leur catégorie, de certaines de leurs propriétés (morpho-syntaxiques) et des relations grammaticales qu'elles entretiennent.

La reprise de cette tradition est caractéristique du programme génératif à ses débuts. Chomsky la justifie de la façon suivante:

- (5) L'information, donnée par une analyse grammaticale est "en grande partie exacte (...) essentielle à tout traitement rendant compte de la manière dont la langue est employée ou acquise" (Chomsky 1971, p. 92; je souligne).

Karttunen et Zwicky présentent le passage syntaxique en faisant la même référence à la tradition de l'analyse grammaticale: "comme beaucoup d'autres aspects des cadres intellectuels modernes, l'idée de passage a ses racines dans la tradition Classique" (Karttunen & Zwicky 1985, p. 1). Cet appel à la tradition est loin d'être une pure figure de rhétorique.

L'analyse grammaticale a une double fonction dans le cadre de la GGT. D'une part, elle est le vecteur d'une connaissance de langue: elle stipule les unités et les configurations bien formées caractéristiques des langues. D'autre part, elle prend place dans une hypothèse sur l'usage du langage: le processus de compréhension des énoncés implique nécessairement une information syntaxique. La reconceptualisation subie

<sup>5</sup> . On peut distinguer dans le TAL une autre ligne de raisonnement basée sur la proposition suivante: "étant donné des algorithmes (existants pour les langages de programmation), concevoir les grammaires qui puissent les utiliser". Elle aboutit par une autre voie à la même déproblématisation de l'analyseur. Je remercie A. Abeillé d'avoir attiré mon attention sur ce second programme de travail.

par l'analyse grammaticale dans le cadre contemporain, que notent Karttunen & Zwicky (ibid.), consiste précisément à séparer l'analyse grammaticale de sa visée pédagogique et à lui conférer la fonction définie par l'hypothèse chomskyenne: fournir une information "essentielle à tout traitement rendant compte de la manière dont la langue est employée ou acquise". Ce n'est pas l'analyse grammaticale en tant que telle qui est reprise, c'est l'analyse grammaticale revisitée par l'hypothèse chomskyenne énoncée en (5). C'est ce qui fait de la définition fonctionnelle de Morin en (3) autre chose qu'un inventaire hétéroclite.

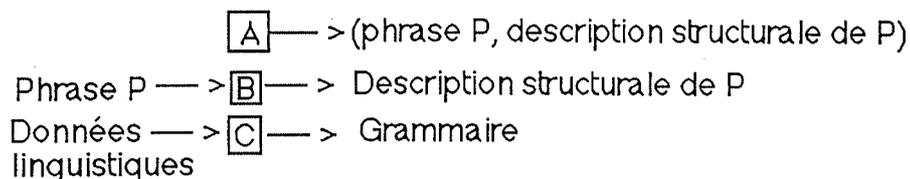
### 2.2.2. La forme d'ensemble de la problématique

Il y a plus important encore: le cadre théorique mis en place par la GGT fournit la structure du programme de travail: 3A admet comme une évidence la répartition suivante:

- (6) L'analyseur est distinct de la grammaire; la grammaire est contenue dans l'analyseur.

La proposition (6) ne va de soi que si on se place dans la configuration que dessine le programme chomskyen et que je rappelle dans sa forme schématique d'origine où les boîtes représentent les unités de traitement et les flèches les entrées et les sorties (Chomsky & Miller 1968, p. 64):

(7)



La boîte B représente l'instance de performance: elle associe à chaque phrase une description structurale. La grammaire (représentée par la boîte A), qui génère les descriptions structurales bien formées, est "contenue à titre de composante fondamentale" dans B (ibid). La relation entre A et B est essentielle: elle caractérise structurellement la problématique de la GGT. Que l'on conçoive la grammaire comme un ensemble de propositions cognisées (*cognized*) ou comme un système de règles générant l'analyse des structures bien formées<sup>6</sup>, elle est contenue dans l'appareil de performance.

C'est dans ce cadre que les notions de *description* et *représentation* employées par 3A prennent sens; c'est précisément celui que leur donne le glossaire de l'ouvrage de référence de l'école cognitive, *Vision* de D. Marr (1982):

6 . La première version correspond à la définition (théorique) néo-cartésienne de la compétence (Fodor 1986; pour la définition de *cogniser*, voir ibid p.15); la seconde correspond à la version qui a été mise en oeuvre dans les modèles syntaxiques qui ont jalonné l'histoire de la GGT.

- (8) a. "A description is the result of applying a representation to a particular entity".  
 b. "A representation of a set of entities S is a formal scheme for describing them, together with rules that specify how the scheme applies to any particular one of the entities" (Marr 1982, p.366).

Ces définitions structurent la division du travail: la syntaxe définit le système formel qui décrit les énoncés, 3A définit les règles qui appliquent les descriptions générées par le système formel. L'analyseur est l'instance d'application qui produit les descriptions d'énoncés. L'analyseur 3A se trouve, de fait, dans la place et dans le rôle de l'instance de performance postulée dans le cadre de la GGT. D'où l'hypothèse:

- (9) L'analyseur 3A occupe la place du dispositif de performance telle qu'elle est définie dans le cadre de la grammaire générative.

La proposition (9) peut surprendre alors que 3A prétend ne pas se commettre avec une théorie de l'usage du langage, que la grammaire générative n'apparaît pas comme la théorie syntaxique la plus fréquemment reprise dans le TAL et qu'il y a une longue tradition critique à l'égard du modèle transformationnel dans la communauté de la linguistique informatique. Mais ce n'est pas une situation inédite: il faut distinguer les hypothèses structurantes qui ont une efficacité implicite et les constructions explicites qui font l'objet des débats. La configuration (7) est structurante, sans doute parce qu'elle fonctionne comme vulgate. Ses éléments essentiels (la distinction entre compétence et performance, l'inhérence de la grammaire dans l'appareil de performance, l'identité du modèle syntaxique construit par la théorie syntaxique et du modèle de compétence du locuteur, etc.) peuvent être conservés indépendamment du modèle grammatical particulier développé par la GGT (grammaire polystratale, usage de règles transformationnelles, etc.). C'est ce qui explique que le cadre de la GGT puisse perdurer dans 3A alors même qu'il a pratiquement perdu, ainsi que le modèle transformationnel, toute validité en théorie de la syntaxe, et ce aux yeux mêmes de ses promoteurs.

J'admets ce point d'archéologie. Si un analyseur occupe une place structurellement définissable comme l'appareil de performance dans la GGT, on s'attend à ce que la conception de cet appareil par la GGT infléchisse le travail investi dans cet analyseur. C'est le cas.

### 2.3. L'appareil de performance

L'appareil de performance est seulement esquissé dans le programme génératif. Par contre, le courant de psycho-linguistique qui se réclame de ce programme, lui a donné un contenu précis; il est au fondement de 3A.

### 2.3.1. L'appareil de performance à partir de la psycholinguistique

La psycholinguistique a donné une forme canonique à l'appareil de performance en reprenant un dispositif très général produit par l'intelligence artificielle: les systèmes de production (SdP). Newell (1973, p. 515) admet, par exemple, que les systèmes de production constituent "un modèle explicatif du comportement humain considéré sous l'angle du traitement de l'information".

Ce qui importe, ce n'est pas le détail du SdP, mais la notion générale de SdP; c'est elle qui donne forme à l'appareil de performance.

Un SdP comprend un ensemble de règles de production, une structure de données et un interpréteur (Rosner 1983). Dans l'approche de la performance, les règles sont celles qui sont données par la grammaire (A). L'application des règles peut entraîner des situations de conflit: lorsque les données valident la condition d'application de plusieurs règles, elles sont applicables simultanément<sup>7</sup>. L'application effective des règles doit donc être ordonnée; c'est la fonction typique d'une instance de contrôle; c'est la fonction que l'on confie à l'interpréteur. Il reste à identifier l'interpréteur et l'appareil de performance: l'appareil de performance applique les règles et gère les situations de conflit. Le mode typique de gestion est la séquentialisation. La fonction précise de l'appareil de performance est décrite par la proposition suivante:

- (10) L'appareil de performance établit l'ordre d'application des règles dans les situations de conflit.

On obtient ainsi la caractérisation des "règles qui spécifient comment le schème formel s'applique à chaque entité particulière" (cf. la définition de *représentation* en (8.b) ci-dessus). Ce sont des règles qui spécifient l'ordre dans lequel appliquer les règles de la grammaire<sup>8</sup>.

### 2.3.2. Situation de 3A

La représentation de l'appareil de performance comme un SdP donne un caractère tangible à la conjecture (7). En acquérant ce caractère tangible, elle perd son caractère spéculatif et/ou conjectural. L'appareil de performance se fige dans les deux propositions suivantes:

- (i) la performance est réductible à l'application des règles de la grammaire,

7. Le recours à la situation de conflit est un lieu commun des recherches psychologiques de l'époque: "This conflict resolution phase of each cycle is where basic cognitive traits like action sequencing, attention focusing, interruptability, and control of instability are realized". (Barr & Feigenbaum 1981, vol 1, p. 197).

8. Les recherches consacrées aux effets de clôture en LFG illustrent parfaitement cette problématique: "ce sont les règles de la grammaire de compétence qui sont utilisées pour construire les représentations internes dans la perception des énoncés (...) l'ordre dans lequel ces règles sont appliquées peuvent donner lieu à des problèmes de clôture ." (Ford et al. 1982, p.741). Cette conception fait le fond de l'hypothèse dite réaliste en GGT (Wasow 1978).

- (ii) l'appareil de performance est une instance finie et "maladroite" d'application: les énoncés peuvent porter les marques de cette finitude et de cette "maladresse"<sup>9</sup>.

On obtient une situation de science normale où les questions de fondation et de légitimité sont forcloses au profit d'un programme de travail aux objectifs bien délimités: c'est le programme qui définit 3A: trouver les solutions algorithmiques ou heuristiques à l'application d'une grammaire.

Or, il est faux de considérer que l'appareil de performance, tel que le propose la GGT, en particulier Chomsky dans *Aspects*, soit réductible à une pure instance d'application des règles syntaxiques. C'est un point capital auquel le courant de psycho-linguistique qui se réclame de la GGT et 3A sont aveugles; c'est la cause d'un certain nombre de difficultés que 3A rencontre dans la mise en oeuvre effective de son programme.

### 2.3.3. La performance dans la GGT

Une proposition, pourtant essentielle à la compréhension de la problématique générative, est souvent oubliée:

(11) "La grammaticalité est seulement l'un des nombreux facteurs qui, par leur interaction, déterminent l'acceptabilité [des énoncés]" (Chomsky 1971, p.23).

Autrement dit, la grammaire ne décrit pas la totalité de l'énoncé ou bien encore, la bonne formation syntaxique (voire sémantique) n'est pas l'unique facteur de l'acceptabilité, ou de l'interprétabilité, des énoncés.

On peut donner contenu à la proposition (11) en recensant les différences entre *grammatical* et *acceptable*. Le prédicat *grammatical* porte sur des structures, et par métonymie, sur les énoncés qui présentent telle structure; le second *acceptable* porte sur les énoncés donnés à l'intuition. Appliqués aux énoncés (directement pour *acceptable* et indirectement pour *grammatical*), les deux prédicats n'ont pas la même extension. Une phrase peut être grammaticale et difficilement acceptable (ou interprétable): c'est le cas classique des phrases auto-enchâssées (Chomsky 1971, p.27) et des culs-de-sac (*garden-paths*)<sup>10</sup>. "A l'inverse", des phrases agrammaticales peuvent être acceptables; ce qu'illustrent les énoncés de (12):

(12) a. Marie a rencontré l'homme que Paul croyait que tu le connaissais.  
b. Je l'ai explosé ce mec.

---

9. On notera que la recherche, en psycho-linguistique, a surtout porté sur la simulation du procès d'analyse, en particulier de ses ratés, et non sur le marquage des énoncés par l'appareil de performance. Cette manière de poser le problème est conservée dans le programme Gouvernement et Liage: "The perception problem would be dealt with by construction of a parser that incorporates the rules of the I-language (...) A parser should not map expressions into their structures in the way that these are associated by the I-language. For example, a parser should fail to do so in the case of so-called "garden-path sentences" or sentences that overload memory for left-to-right pass, it should mirror the difficulties experienced ... (Chomsky 1986, p. 25).

10. La littérature consacrée aux culs-de-sac est immense; entre autres Bever 1970, Pritchett 1988 et Milne 1982.

La grammaticalité est une propriété locale des structures syntaxiques alors que l'acceptabilité est "une propriété globale" des énoncés (Chomsky 1971, p. 23). Etc.

Ces différences sont symptomatiques du fait qu'un énoncé est un objet langagier qui n'est que partiellement organisé par la grammaire conçue comme système de règles de bonne formation. Si la grammaire est une théorie partielle de l'énoncé, l'instance de performance ne peut pas être réduite à un dispositif d'application des descriptions de la grammaire. Le dispositif doit, d'une manière ou d'une autre, tenir compte des "facteurs, autres que la grammaticalité, qui, par leurs interactions", déterminent les propriétés, et en particulier la forme, des énoncés. Ces facteurs ne sont pas simplement des contraintes sur les règles qui appliquent les règles; ce qu'illustrent encore les énoncés de (12) qu'on ne peut pas décrire comme le résultat de l'application dans un ordre particulier des règles de la grammaire.

#### 2.4. Le problème de l'énoncé

La distinction entre *grammatical* et *acceptable* est, de fait, terriblement fruste: elle amalgame des distinctions qui se situent sur (au moins) quatre plans différents. Par souci de clarté, et pour interdire les glissements métonymiques, je raisonnerai en termes d'objets et non de propriétés. Le terme *phrase* désigne (par convention) l'objet généré par le dispositif abstrait d'une grammaire et *énoncé* ce qui est appréhendé dans l'échange langagier.

Rappelons tout d'abord les deux premières dimensions qui sont aisément pensables dans le cadre de la GGT:

- (i) la dimension de la production matérielle: l'énoncé porte les marques de sa production psycho-physiologique, alors que la phrase est saisie indépendamment de cet aspect.

- (ii) la dimension des usages: l'énoncé s'inscrit dans un usage sociolectal donné; il instancie de façon particulière les paramètres décrivant la variation caractéristique de certains aspects syntaxiques. La théorie de la phrase fait abstraction des réalisations particulières pour isoler les espaces de variation.

Introduisons les deux oppositions, qui, quand on les radicalise, amènent à quitter le cadre rappelé en (7):

- (iii) la dimension épistémologique: la phrase est un objet construit pour déterminer les propriétés propres du système que la linguistique construit, *le langage* (ou la langue), alors que l'énoncé est ce qui se constate. Rien n'assure, en dehors du cadre particulier et historiquement daté de la GGT, que le dispositif utilisé dans l'étude du langage (par exemple: une grammaire) soit pertinent dans la théorie à constituer des énoncés.

Je ne peux développer ce point ici, mais je rappelle que c'est la conclusion à laquelle on arrive dans le cadre de la théorie de l'usage de Bever (1970) ou bien dans le cadre introduit par Milner (1989). Ce dernier cadre, par exemple, se fonde sur une opposition parallèle entre *possible de langue* et *possible événementiel*. Il pose clairement ce fait que le possible événementiel excède toujours le possible de langue: les énoncés

peuvent toujours échapper aux organisations de langue, en particulier aux organisations syntaxiques. Ce n'est pas parce que les locuteurs se trompent ou maîtrisent mal un savoir, mais parce que le langage est cette organisation singulière qui ne rejette pas ce qui échappe à ses propres lois de formation.

Enfin, (iv), *phrase* et *énoncé* s'opposent sur une quatrième dimension: le type d'organisation qu'ils représentent. Quand on réduit *grammaire* à syntaxe et *phrase* à une organisation particulière, par exemple, à une configuration de positions formant domaine, la phrase est une organisation homogène. L'énoncé, au contraire, est hétérogène: le résultat intriqué de l'interaction de plusieurs principes organisationnels. Or, cela se marque dans la forme même des énoncés: le système des circonstanciels, par exemple, échappe en grande partie au contrôle détaillé de la syntaxe; il en est de même pour les différents systèmes d'anaphore et certains aspects de la sous-catégorisation verbale quand on ne se laisse pas leurrer par le dogme de la projection (ou ses équivalents).

De ce point de vue, rien ne garantit qu'une analyse syntaxique des énoncés puisse être opérée sans tenir compte des autres dimensions organisationnelles, soit qu'on les maintiennent séparées dans une approche modulaire, soit qu'on les conçoive ensemble dans une approche holiste.

## 2.5. Les énoncés dans 3A

3A est aveugle à ces distinctions; si elle peut faire allusion, dans ses attendus, aux dimensions de la production et des usages (cf. (i) et (ii) ci-dessus), elle ignore les distinctions de statut épistémologique et de type organisationnel (cf. (iii) et (iv) ci-dessus). Elle recouvre les écarts entre phrase et énoncé par des solutions continuistes dans les problématiques de la couverture et du relâchement des contraintes. Je n'évoquerai que le problème de la couverture<sup>11</sup>.

La couverture d'un analyseur correspond aux énoncés qu'il peut analyser. La mesure de ce qu'il peut analyser est donnée par la grammaire; c'est, en effet, sa seule source de savoir. La couverture de l'analyseur correspond à "la couverture de la grammaire". Si l'analyseur ne peut pas analyser un énoncé quelconque, "c'est la faute à la grammaire", une déficience de la grammaire. Cette déficience est conçue en termes quantitatifs: la grammaire ne contient pas assez de règles ou d'entrées lexicales pour traiter tous les énoncés<sup>12</sup>. Il en découle un programme quantitatif d'accroissement des règles et des entrées lexicales. Il en découle également la tentation de récuser toute distinction entre grammatical et acceptable: l'objet à traiter est constitué par les énoncés ici et maintenant: le grammatical se résorbe dans l'acceptable et l'acceptable dans l'attesté (ou l'attestable)<sup>13</sup>.

11 . Je renvoie à Chanod et al. 1992, Chanod 1992 pour une illustration de la solution basée sur le relâchement des contraintes.

12 . Tomita (1988) est exemplaire de ce type de raisonnement.

13 . C'est ce que Milner (1989, p. 120) appelle "la tentation documentaire". Cette tentation est caractéristique, par exemple, de l'école des lexiques-grammaires en linguistique.

Plutôt que de problématiser les différences entre phrase et énoncé, 3A reconduit la vulgate tirée de la GGT selon laquelle la grammaire (en particulier la syntaxe) est une théorie globale des énoncés. Les solutions en terme de couverture et de relâchement des contraintes tendent à confondre les énoncés et les phrases; elles occultent le problème de la nature propre des énoncés.

## 2.6. Conclusion

La problématique 3A entretient avec la GGT un rapport de dénégation: tout en se déclarant non commise à une théorie de la performance, elle développe un dispositif<sup>14</sup> qui présente les caractéristiques structurales de l'instance de performance définie dans le cadre rappelé en (7). En un mot, un analyseur 3A est un avatar de l'appareil de performance tel que le définit la GGT et que le façonne l'architecture des systèmes de production.

Si l'acceptation non critique d'un cadre de référence est une situation ordinaire dans le développement scientifique, la dénégation de ce cadre est plus problématique en ce qu'elle entraîne un brouillage des objectifs de connaissance poursuivis et qu'elle ne permet pas une analyse claire des blocages rencontrés. A la question de la place de l'information syntaxique dans le processus de compréhension, l'élaboration des analyseurs 3A n'apporte pas d'élément qui permettrait de travailler en détail l'hypothèse générale de départ qui pose sa nécessité. Elle n'apporte pas grand chose non plus sur la question de l'énoncé qui constitue pourtant "la matière première" des analyseurs. De fait, la problématique 3A se résorbe assez vite dans la problématique technologique dessinée dans la définition (3): rendre les énoncés manipulables, c'est-à-dire exploitables. Les raisons de l'exploitation technologique l'emportent: l'analyseur se justifie d'abord d'être une machine au sens industriel (une instance de transformation avec valeur ajoutée). Il est alors justifiable des seules catégories du langage économique: efficacité, rentabilité, fiabilité.

Il serait, bien sûr, excessif de soutenir que les recherches menées en analyse automatique ne peuvent pas produire de connaissance, mais c'est dans un domaine particulier qui demande de tout autres critères d'évaluation: l'invention et l'étude des langages de représentation, en particulier des formalismes susceptibles de représenter les organisations syntaxiques ou sémantiques et certains processus de compréhension. Se pose alors crucialement le problème de l'adéquation de ces langages: ~~cette adéquation ne peut évidemment pas être évaluée en rejetant le programme de recherche qui lui donne sens.~~

## 3. UN ANALYSEUR DANS UNE THEORIE DE LA SYNTAXE

Il n'y a bien sûr pas de consensus sur ce qu'est une théorie de la syntaxe puisqu'il n'y a pas de consensus sur ce qu'est la syntaxe. Admettons, ici, une théorie qui donne une place centrale à l'analyse grammaticale (cf. (4) ci-dessus), non parce qu'elle est "une information

14 . Au sens introduit dans la proposition (22) ci-dessous.

essentielle dans l'usage du langage " (cf. (5) ci-dessus), mais parce qu'elle ménage un accès aux modes d'organisation des langues<sup>15</sup>. La pertinence d'un système capable d'analyser les énoncés peut ne pas être une question exotique dans un tel cadre. Je reprends, de ce point de vue, la proposition, que l'on avance ici ou là, selon laquelle un analyseur pourrait être un dispositif expérimental pour la théorie syntaxique. Elle est au coeur de ce paragraphe:

(14) Un analyseur peut-il être un dispositif expérimental pour la théorie syntaxique?

Une expérience met en place une instance de réfutation/corroboration d'une hypothèse ou d'un système d'hypothèses; elle permet de les tester directement ou indirectement. On envisage donc de demander à un dispositif qui opère des analyses syntaxiques de tester ces analyses. On peut distinguer, chez ceux qui proposent l'analyseur comme dispositif expérimental, deux propositions:

- (15) - a. Réfuter, corroborer une analyse syntaxique (une configuration de constituants catégorisés),  
 - b. Réfuter, corroborer un ensemble d'analyses syntaxiques (une grammaire) "établie sur un petit nombre d'exemples" en l'appliquant au tout-venant des énoncés.

Pour qu'un analyseur soit un dispositif expérimental en syntaxe, il faut que la vérité ou la fausseté des analyses syntaxiques puissent être testées par un dispositif de segmentation/étiquetage des énoncés. Autrement dit, il faut que l'analyse grammaticale soit testable par des opérations sur la chaîne de mots. Cette condition n'est pas remplie dans tout style d'analyse grammaticale. J'opposerai de ce point de vue le style paradigmatique et le style syntagmatique d'analyse. Je raisonne tout d'abord dans le cadre d'une syntaxe qui repose sur le style paradigmatique.

### 3.1. Tester une analyse grammaticale

#### 3.1.1. Un exemple et son analyse

Prenons le groupe nominal français et le constituant en *de* illustré dans l'exemple suivant:

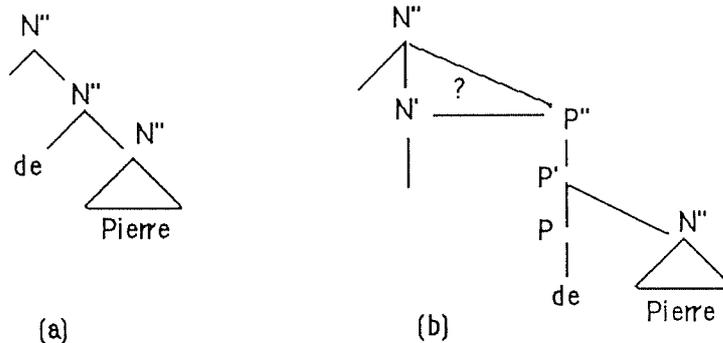
(16) La maison de Pierre.

Le constituant *de Pierre* peut être analysé comme un groupe nominal; *de* est un marqueur casuel du Génitif et le N° *de Pierre* est directement dominé par N°; c'est l'analyse schématisée en (17.a) ci-dessous. L'analyse

15. On peut décrire métaphoriquement cette théorie de la façon suivante: "the statements of a grammar [...] are similar to statements of a physical theory that characterizes certain entities and their properties in abstraction from whatever may turn out to be the mechanisms that account for the properties; say, a nineteenth century theory about valence or properties expressed in the periodic table" (Chomsky 1986, p. 24).

schématisée en (17.b) fait du constituant *de Pierre* un P'' (ou GP, groupe prépositionnel) dont la tête prépositionnelle *de* gouverne un N'' *Pierre* ; P'' peut être rattaché à N' ou à N''.

(17)



On peut, enfin, admettre que ces deux structures existent et que la forme de surface du GN en (16) est structurellement ambiguë; c'est hypothèse de Milner (1982).

Nous sommes bien confrontés à un problème d'analyse grammaticale: si les frontières de constituants ne sont pas en cause, les étiquetages, les fonctions (on a un cas de construction directe en (17.a) de construction indirecte dans (17.b)) et la configuration du GN global sont en litige. Un analyseur peut-il fournir un quelconque élément d'évaluation de ces analyses?

### 3.1.2. Le test

Une première remarque s'impose: un analyseur qui exploite une grammaire qui admet la double géométrie de GN schématisée en (17. a) et (17.b), est incapable, à la seule vue de (16), de choisir entre les deux analyses. La forme de surface de (16) gomme la différence entre (a) et (b).

Il est également immédiatement évident que l'on peut analyser un nombre indéfini de groupes nominaux du français ressemblant à (16) dans autant de contextes que l'on veut et que cette répétition ne nous fera pas avancer d'un iota dans la défense ou la réfutation de l'une ou l'autre analyse. Si c'est le cas, c'est pour deux raisons:

- (i) il n'y a pas de marquage univoque de la structuration syntaxique dans l'énoncé,
- (ii) l'analyse grammaticale est une opération paradigmatique qui ne met pas crucialement en jeu la combinatoire syntagmatique.

### 3.1.3. L'analyse grammaticale est une opération dans l'ordre paradigmatique

Les analyses de la structure de (16), telle qu'elles sont rappelées dans le paragraphe précédent, ne sont pas opérées sur (16) isolément, que ce soit l'énoncé (16), la séquence de "mots" catégorisés ou la classe d'équivalence à laquelle (16) appartient. Elles reposent sur une

paradigmatisation qui rapproche (16) d'autres GN ou d'autres constituants. Entre autres, et pour l'illustration seulement:

- (i) On trace un parallèle entre le constituant en *de* dans GN et le constituant en *de* en position d'attribut. Soit (18):

(18) Cette maison est de moi.

On constate que *de moi* ne peut pas être interprété de façon possessive en (18). Comme il n'y a pas de raison de poser que *être* assigne le cas génitif à son attribut, *de moi* en (18) ne semble analysable que comme un GP (*de* est une préposition); *moi* est normalement possible (Milner 1982, p. 84). On subsume ce comportement sous l'hypothèse suivante: le tour [GP de + GN] est incompatible avec l'interprétation possessive et est normalement indifférent à la forme de GN (le pronom fort est possible) alors que le tour [de-casuel + GN] est compatible avec l'interprétation possessive et bloque une forme marquée casuellement de GN (agrammaticalité du double marquage casuel); autrement dit, [de-casuel + moi] est agrammatical. On conclut que la structure (17.a) est associée à l'interprétation Possessive (entre autres), alors que la structure (17.b) n'est pas ouverte à cette interprétation. On prédit alors qu'un tour comme *la maison de moi* ne peut recevoir d'interprétation possessive et correspond à la structure (17.b).

- (ii) On plonge (16) dans le système d'expansions possibles dans GN; entre autres celui qu'illustre (19):

(19) Ma photo de Pierre.

Le constituant *de Pierre* ne peut pas être interprété de façon possessive (*Pierre* peut être interprété comme "l'agent" ou "le thème" de la photo). C'est un indice: *de Pierre* doit être analysé comme un GP. L'indice est corroboré par une analyse qui admet le principe du diacritique: si l'on admet que *ma* est un GN en position de Spécifieur, il ne peut y avoir deux GN directs dans un même GN: *de Pierre* est nécessairement un P".

L'analyse ne se réduit pas à un découpage-étiquetage de la chaîne; elle rapproche les tours sur des ressemblances et des différences formelles et interprétatives. De plus, elle fait librement appel aux construits de la théorie syntaxique comme le principe du diacritique. La réfutation/corroboration des analyses de (17) passe nécessairement par les questions suivantes: le paradigme est-il bien construit (les modifications apportées au paradigme entraînent-elles des changements dans l'analyse?), le principe du diacritique fait-il des prédictions vraies des domaines syntaxiques? la "catégorisation" de *de* comme marqueur casuel, de *ma* comme GN sont-elles fondées? l'interprétation possessive est-elle suffisamment stable et différenciée pour pouvoir servir d'indice d'une structure? Aucune de ces questions ne met crucialement en jeu le simple découpage de la chaîne ou la mise en classes d'équivalence des éléments découpés.

### 3.1.4. Corollaire

Un analyseur semble parfaitement étranger à une théorie syntaxique fondée sur le style paradigmatique. Je rappelle les deux traits qui caractérisent ce style. Il ne se réduit pas à une analyse en constituants immédiats. Il n'a pas pour fonction première de rendre compte de la bonne formation des énoncés, voire de leur interprétation; il rend compte du paradigme de différences accessibles dans le jugement de bonne formation et dans le jugement qui reconnaît une différence dans l'interprétation<sup>16</sup>.

Pour ces deux raisons, on ne peut rien tirer d'un système qui segmente/étiquette les énoncés pris individuellement; de fait, d'un dispositif qui s'est constitué dans un cadre où l'analyse grammaticale est assimilée à des opérations réitérées de segmentation/étiquetage des énoncés (les analyseurs en chaînes sont typiques de ce cadre). Autrement dit:

(20) Un analyseur ne permet pas de tester des analyses qui impliquent une mise en paradigme qui fonctionne comme espace de contraste pour le jugement de grammaticalité et/ou le jugement interprétatif (ressemblances/différences interprétatives).

### 3.1.5. Elargissement

Le diagnostic (20) n'est nullement surprenant du point de vue de la théorie qui met en oeuvre le style paradigmatique d'analyse; il découle de la nature qu'il assigne aux organisations syntaxiques. Elles ne sont pas données à la perception: "La langue présente ce caractère étrange et frappant de ne pas offrir d'entités perceptibles de prime abord, sans qu'on puisse douter cependant qu'elles existent" (Saussure cité dans Piotrowski 1990, p. 13). Si les entités syntaxiques sont de ce type, si elles ne sont pas données à la perception immédiate, c'est qu'elles ne sont pas entièrement constituées par les relations d'ordre constitutives de la "chaîne parlée". L'identification des éléments syntaxiques ne peut être qu'indirecte (si on admet de qualifier de *directes* les manipulations de la chaîne parlée); c'est ce qui motive la paradigmatisme à la base des analyses de (17)<sup>17</sup>. On ne peut saisir la structuration syntaxique que dans ses effets: les paradigmes sont à l'image des plaques photographiques où l'on repère la trace d'une trajectoire et non la particule elle-même qui se déplace.

## 3.2. Tester une grammaire

Le deuxième contenu expérimental possible est le suivant (je reprends les termes dans lesquels on le formule habituellement): tester "sur un grand nombre de données une analyse forgée sur/pour un petit

16 . Ce style est caractéristique de l'école chomskyenne et, il faut insister sur ce point, tout à fait indépendant de la forme transformationnelle de la grammaire. Il est, à l'évidence, tout à fait distinct de celui que pratique le structuralisme américain: construction de classes d'équivalence dans un contexte identique.

17 . Elle retrouve la démarche structuraliste européenne (Milner 1973, p. 11).

nombre d'exemples" (cf. (15.b) ci-dessus). Il s'agirait en quelque sorte de renverser la démarche caractéristique de la linguistique qui, par la catégorisation et la construction d'exemples, tend à passer de l'infini des productions à un petit nombre d'exemples.

### 3.2.1. Analyse de la proposition

La proposition selon laquelle on pourrait/il faudrait tester une grammaire sur un grand nombre d'exemples présuppose:

- un continuum entre les exemples et l'ensemble des énoncés,
- un jugement négatif sur la pratique de l'exemple.

Le statut de l'exemple, et indissociablement de l'énoncé, est crucialement en jeu dans le raisonnement. La proposition (15.b) revient à considérer qu'un exemple est un énoncé extrait, à la manière d'une citation, de l'ensemble des énoncés.

Un exemple dans le style paradigmatique n'est pas un énoncé. En effet, à la mise entre parenthèses des conditions d'énonciation s'ajoute l'abstraction liée à l'opération de paradigmatisation: un exemple ne vaut que comme élément d'un ensemble de contrastes. Seule la dimension constituant le thème du jugement contrastif est pertinente; les autres dimensions sont suspendues. En ce sens, un exemple est toujours simple alors que l'énoncé dont il est homonyme est essentiellement intriqué<sup>18</sup>. De fait, les analyses grammaticales (comme celle qui est exemplifiée dans les paragraphes précédents) présupposent une décomposition en niveaux organisationnels; cette décomposition est opérée par la paradigmatisation. Là encore, elle ne se réduit pas à une opération sur la chaîne.

De ce point de vue, cela n'a tout simplement pas de sens de vouloir passer de l'ensemble restreint des exemples à l'ensemble des énoncés dont ils seraient extraits, puisque à proprement parler, il n'y a aucune continuité entre ces deux ensembles.

### 3.2.2. Quelques conséquences

Il est une critique, maintes fois renouvelée, selon laquelle la "couverture empirique" de la syntaxe serait insuffisante du fait même qu'elle travaille sur un petit ensemble d'exemples (qui plus est, toujours les mêmes) (entre autres, Tomita 1988). S'il y a quelque vérité dans ce diagnostic (cf. la très grande inégalité dans les structures étudiées), il est mal formulé: si la syntaxe est défaillante, ce n'est pas parce qu'elle travaille sur des exemples, c'est parce qu'elle construit des paradigmes trop étroits ou incomplets. La syntaxe travaille sur des exemples (en quelque sorte *in vitro*) par nécessité et non par paresse ou cécité empirique.

L'analyse grammaticale porte sur des exemples décrits partiellement et en s'autorisant l'ignorance quant aux aspects qu'elle n'étudie pas. La

18. Je reprends dans ce paragraphe l'analyse du statut de l'exemple donnée dans Milner (1989). Un énoncé est une concrétion: "Toute donnée de langue, si minime qu'elle puisse être en extension, combine toujours plusieurs dimensions: pas d'atome lexical qui ne soit aussi une forme phonologique et un emploi syntaxique, pas de phrase qui ne résulte de l'application combinée de plusieurs règles, etc." (Milner 1989, p. 113).

description des énoncés doit gérer tous les aspects de l'énoncé "à la fois", même si la description que l'on vise à donner est partielle. Autrement dit, un analyseur syntaxique peut devoir tenir compte de tous les autres modes d'organisation de l'énoncé même s'il ne se cantonne qu'à la seule description syntaxique.

Ce point n'est pas assez reconnu. Si un analyseur n'ayant accès qu'à des données syntaxiques au sens étroit (configurations bien-formées de catégories) échoue à traiter des énoncés, ce n'est pas un argument contre l'autonomie de la syntaxe (précisément, contre l'abstraction d'un niveau indépendant de description syntaxique), c'est tout simplement la conséquence de l'intrication des énoncés qui est la principale motivation pour adopter une méthodologie de la décomposition, en particulier une méthodologie qui distingue des niveaux autonomes de description.

C'est une autre manière de dire qu'un analyseur syntaxique ne peut pas être une instance expérimentale pour la syntaxe; il faudrait qu'il puisse reconnaître et gérer ce qui ne relève pas de la syntaxe. Ou bien, il doit tenir compte de tous les aspects des énoncés et ce n'est plus un analyseur syntaxique au sens strict; ou bien il ne prend en considération que les aspects syntaxiques et il "se plante" sur le premier énoncé venu. De toute façon, rien ne permet d'assurer que l'échec est lié à la théorie syntaxique qu'il exploite ou à l'interférence de phénomènes externes à la syntaxe.

### 3.3. Raisonnement a contrario

J'ai montré qu'un analyseur ne peut pas valoir comme instance d'expérimentation pour une syntaxe qui pratique le style paradigmatique d'analyse. Etant donné qu'il n'y a pas un seul style d'analyse dans le champ de la syntaxe, une question demeure: y a-t-il un style d'analyse qui soit susceptible de conférer à l'analyseur un statut expérimental? Faisant fond sur ce qui a été introduit au paragraphe précédent, on peut caractériser un tel style: il faudrait qu'il présente les traits suivants:

- 1) appréhender l'énoncé dans sa globalité sans décomposition en plans différents autonomes,

- 2) privilégier la dimension syntagmatique des entités syntaxiques: les unités sont entièrement données par les règles de concaténation d'unités segmentables dans la chaîne,

- 3) donner un sens fort au "dialogue empirique" avec les énoncés<sup>19</sup>.

Ces trois traits caractérisent bien un style d'analyse; c'est celui qui est sous-jacent aux analyses qui se donnent dans le formalisme des grammaires catégorielles (GC): je l'appelle *style syntagmatique*. Un analyseur acquiert-il pour autant un statut expérimental dans ce cadre? C'est la question que j'aborde brièvement dans ce dernier paragraphe; je m'appuierai essentiellement sur Karttunen (1989) et Steedman (1989).

#### 3.3.1. L'analyse grammaticale en GC

19 . Je rappelle que le dialogue empirique n'est pas noué avec les énoncés dans le style paradigmatique, mais avec les jugements sur les énoncés.

Une grammaire GC met en avant la dimension concaténationnelle des unités: le type d'un foncteur est une règle combinatoire, le type d'un argument permet la combinaison (Karttunen 1989, p. 45). Les règles combinatoires respectent un principe d'adjacence: "les règles combinatoires ne peuvent s'appliquer qu'à des entités qui sont linguistiquement réalisées et adjacentes" (Steedman *ibid.*, p. 203). L'assignation des types respecte un principe de transparence: "l'information dans le type syntaxique d'une expression inclut l'information dans son type sémantique" (*ibid.*). Autrement dit, l'approche est unidimensionnelle<sup>20</sup> et prend toujours la forme d'une combinaison entre unités adjacentes. L'accent est mis sur la synthèse sémantique plus que l'analyse syntaxique ou sémantique de l'énoncé. Comme le note Karttunen (*ibid.*, p. 56): "La structure de l'arbre [d'analyse] n'a pas de signification linguistique dans notre système; de ce point de vue, nos arbres sont différents des arbres syntagmatiques. Tout ce qui compte, c'est l'ensemble de traits que l'on obtient". Le jugement essentiel porte sur l'équivalence sémantique produite à partir de découpages qui peuvent être non isomorphes et donc sur l'interprétation globale des énoncés.

On peut admettre donc que les analyses d'une GC présentent les trois traits qui les rendent susceptibles d'être testées sur les énoncés. Il faut maintenant s'attacher au contenu du test.

### 3.3.2. Le test

Dans ce cadre, le test porte essentiellement sur:

- la réussite ou l'échec de la synthèse sémantique,
- la capacité de cette synthèse à donner des interprétations équivalentes à partir de découpages non-isomorphes,
- sa capacité à capter l'interprétation des énoncés.

Le test ne porte que secondairement sur le découpage lui-même. Comme le note avec insistance Steedman, la notion de constituant (fondamentale dans le style paradigmatique et l'analyse grammaticale) est redéfinie: elle recouvre, de fait, tout découpage requis par le calcul sémantique<sup>21</sup>.

Du coup, si on envisage l'analyseur comme une instance expérimentale, il teste non pas une analyse grammaticale mais un calcul sémantique. L'analyse grammaticale est secondaire et ne constitue pas l'enjeu d'une description pour elle-même: l'opposition entre "proper and spurious components of well-formed strings" (Quine 1977, p. 447), qui fait tout l'enjeu du style paradigmatique, est révoquée.

Quant à la proposition (15.b), la situation est confuse étant donné que le statut de l'exemple n'est pas vraiment thématique dans cette approche. Si

20. "A cause de l'hypothèse encapsulée dans le principe de transparence, on peut écrire les règles syntaxiques et les règles sémantiques en une seule, associant dans une règle chaque catégorie syntaxique à l'interprétation sémantique qu'elle reflète de façon transparente" (Steedman, *ibid.*, p. 205).

21. Steedman appelle ces découpages des "constituants de surface": "Si une chaîne comme *I believe that Harry eats* est un constituant pour (*for the purposes*) la conjonction et l'extraction, elle doit tout aussi bien être un constituant dans des phrases canoniques comme *I believe that Harry eats these apples*" (Steedman, *ibid.*, p. 220).

l'exemple est simplement conçu comme un énoncé prélevé de l'ensemble du tout venant des énoncés, il fait sens de tester "à grande échelle" des assignations de type et des règles combinatoires élaborées sur un sous-ensemble. Dans ce cas, l'extrême diversité des formes de surface nécessitera une multiplication des types, et partant il y aura prolifération des analyses. A proprement parler, cette prolifération ne remet pas en cause les hypothèses théoriques du cadre puisque son principe est accepté au départ: "Un des résultats de cette généralisation [de la notion de constituant de surface] est que les analyses syntaxiques de surface tendent à proliférer" (Steedman, *ibid.*, p. 202). Mais du coup, on ne voit pas clairement ce qui est en jeu dans une expérience d'analyse "à grande échelle" qui ne puisse être établi sur un corpus représentatif de petite taille.

### 3.4. Conclusion

L'hypothèse selon laquelle un analyseur pourrait être un dispositif expérimental pour la théorie syntaxique, tout spécifiquement pour l'analyse grammaticale, n'a aucune espèce d'évidence. A coup sûr, elle ne peut pas être transthéorique.

Un analyseur est un dispositif exogène pour une syntaxe qui pratique le style paradigmatique d'analyse grammaticale: l'instance de réfutation dans ce cadre est à construire dans la dimension paradigmatique. L'hypothèse est plus plausible dans le cadre d'une syntaxe qui pratique le style syntagmatique. Mais, ce n'est pas une analyse grammaticale au sens strict qui est mise à l'épreuve. Dans les deux cadres, la répétition indéfinie de l'analyse ne paraît devoir apporter rien de plus que ce que peut discerner une description sur des paradigmes d'exemples ou sur un corpus dont la représentativité a été construite avec méthode.

Un point est sûr: les analyseurs ne fournissent aucun argument décisif pour choisir entre l'un ou l'autre des deux styles d'analyse grammaticale. C'est dire qu'ils ne fournissent aucun argument décisif dans un des problèmes cruciaux de la théorie syntaxique.

## 4. STATUT D'UN ANALYSEUR DANS UNE THEORIE DE LA COMPREHENSION DES ENONCES

Les théories de l'usage du langage (*language use*), en particulier de la compréhension du langage naturel (CLN), se sont développées en dehors des disciplines traditionnelles de la linguistique: on a vu comment la notion de performance a pris forme dans le giron de l'intelligence artificielle. L'intelligence artificielle abstraite (IA dorénavant) a développé un style épistémologique particulier: l'épistémologie du dispositif<sup>22</sup>. C'est

22. J'adopte la distinction opérée par J. Proust (1987, p.88) entre IA technologique "qui se détermine en fonction d'impératifs pragmatiques et vise moins à poser des problèmes qu'à trouver des solutions" et IA théorique ou abstraite "encline à poser des questions de fondement". Je ne considère que la seconde. Je reprends la formule d'*épistémologie du dispositif* à Milner (1989). J'emploie le néologisme *usage du langage* au lieu de *performance* pour marquer que je ne me place pas dans le cadre de la GGT (rappelé en (7) ci-dessus).

de ce troisième point de vue que j'examine le statut possible d'un analyseur:

- (21) Conçu comme un dispositif, quelle fonction un analyseur remplit-il et à quelle condition ?

En effet, si l'analyseur acquiert d'emblée un statut sur le modèle des implémentations que conçoit et analyse l'IA, ce statut ne garantit pas son contenu de connaissance. L'épistémologie du dispositif met en place des manières de raisonner qui ne sont pas courantes en linguistique<sup>23</sup>; il faut prendre la mesure de cette différence.

#### 4.1. Épistémologie du dispositif

##### 4.1.1. Définition

Un dispositif n'est pas un modèle formel. Sa fonction est de permettre que l'on examine un objet ou un problème "*in concreto*, c'est-à-dire en se donnant une "intuition", soit un système particulier, présentant les caractéristiques générales de la solution" (Proust J. 1987, p. 90)<sup>24</sup>. Son statut est double. C'est d'abord un outil heuristique dans la phase de construction des conjectures et un outil d'analyse conceptuelle. C'est, dans un deuxième temps, un objet qui peut être observé. Alors que les processus intellectuels humains échappent à l'observation, en particulier les processus langagiers de compréhension ou de production, on peut prendre la machine construite comme objet d'observation. Un dispositif n'a donc pas pour mission d'instaurer le dialogue avec l'empirie; d'une certaine manière, il s'y substitue. J'admets la définition suivante de dispositif:

- (22) Un dispositif est une réalisation matérielle particulière de conjecture(s) générale(s) (sous forme d'une machine, en particulier sous forme d'un système de traitement de l'information implémenté).

---

23. "In computer science many fundamental conceptual advances occur by (scientifically) uncontrolled experiments in our own style of computing. [...] Computer science is not unique in having modes of progress that don't fit easily into the standard frames. In the heyday of paleontology, major conceptual advances occurred by stumbling across the bones of immense beasts. Neither controlled experimentation nor theoretical predication played appreciable roles" (Newell, 1982: 93).

24 . Cette définition rapproche, selon J. Proust, les machines de l'IA abstraite de la mathématique selon Kant.

#### 4.1.2. Fonctions d'un dispositif

L'épistémologie du dispositif fait courir un risque bien précis: prendre les vessies de l'imagination pour les lanternes de la réalité. Il faut donc cerner au plus près ses fonctions. Je les reprends.

Un dispositif doit, tout d'abord, introduire de la clarté et de la distinction dans une conjecture globale (Milner 1989, p.159). Prenons un exemple. Raisonçons dans le cadre du cognitivisme: la compréhension des énoncés est vue comme un enchaînement de transformations de représentations<sup>25</sup>. C'est une conjecture globale qui n'a aucun contenu de connaissance particulier.

Dans ce cadre, l'enjeu d'un dispositif particulier est de déterminer les différentes représentations, leurs éléments, les calculs qui les produisent et les calculs qui permettent de passer de l'une à l'autre. La réalisation oblige à l'explicitation maximale de chacun des constituants pris isolément et dans leurs rapports au tout. On comprend dès lors qu'un dispositif est d'autant plus fructueux qu'il est analytique: en morcelant les conceptions spéculatives en un système de propositions de détail se situant sur des plans différents, le dispositif permet d'envisager que l'on pourra, un jour, rencontrer une région empirique, qu'elle soit biologique, linguistique ou comportementale, et, du coup, des éléments qui permettent de conforter ou de rejeter les propositions de détail<sup>26</sup>. Autrement dit, si le dispositif n'instaure pas par lui-même de dialogue avec l'empirie, il est censé le préparer et il ne vaut que dans cette perspective.

D'autre part, on observe les dispositifs. Cette observation doit être critique. Comme le remarque J. Proust à propos de l'IA, "ce en quoi l'IA se distingue des mathématiques vues par Kant, c'est que, faute de disposer encore d'une théorie parfaitement maîtrisée, elle doit se livrer à un travail critique à partir des objets ainsi élaborés afin de découvrir celles de leurs caractéristiques qui sont des traits nécessaires de la cognition" (Proust *ibid.*). Par conséquent, un analyseur ne vaut que soumis au questionnement portant sur les traits nécessaires, ou constitutifs, de la compréhension. Cette attitude est certes difficile à adopter quand on connaît les investissements de nature diverse dont les analyseurs sont l'objet, mais elle est impérative pour que l'analyseur ne soit pas seulement un montage spéculatif.

#### 4.2. L'analyseur comme dispositif

Restons dans le cadre cognitiviste. La notion d'analyseur y prend un sens bien particulier: sa fonction première n'est pas d'analyser les énoncés, mais de les représenter. L'analyseur est un dispositif de traitement qui a pour fonction définitoire de donner une représentation des énoncés dans un processus de compréhension.

25 . Le cadre cognitiviste repose sur "la croyance [*belief*] selon laquelle certains aspects de l'esprit [*mind/brain*] peuvent être utilement construits sur le modèle de systèmes de règles qui forment et transforment des représentations et qui sont utilisées dans l'interprétation et l'action" (Chomsky 1986, p.5) . Voir aussi Marr (1982).

26 . De ce point de vue, la modularité est d'abord un impératif épistémologique.

Si l'analyseur est évidemment un dispositif externe à l'objet étudié quand il est envisagé à partir d'une théorie syntaxique, il constitue une partie de cet objet quand on étudie la compréhension puisqu'il en constitue une partie propre. A ce titre, rien ne l'impose: un dispositif de compréhension peut fort bien traiter des "informations" typiques de celles qui sont données par une analyse grammaticale sans distinguer d'analyseur, de phase autonomisable de représentation syntaxique ou de représentation spécifiquement syntaxique. On ne conçoit d'analyseur que si on pose (i) que la compréhension traite syntaxiquement les énoncés, (ii) que l'on peut isoler ce traitement dans le traitement global de la compréhension et (iii) qu'il y a un niveau de représentation syntaxique isolable (cette dernière proposition n'est pas nécessaire).

De ce point de vue, l'analyse grammaticale (cf.(4) ci-dessus) n'est pas reprise pour elle-même, elle n'est pas reprise parce qu'elle fournit une information "essentielle pour la compréhension"; elle est reprise parce qu'elle constitue une phase ou un aspect du processus de compréhension. Autrement dit, un analyseur n'a de sens que comme un module dans l'image que l'on se fait du processus de compréhension ou de l'esprit. Mais, on ne dit là que la forme d'un programme de travail; il ne prend contenu qu'en spécifiant dans le détail quels sont les modules que l'on distingue, quel est le mode de leur interaction, bref le détail de l'architecture modulaire que l'on explore. Je résume ce point par la proposition suivante:

- (23) Conçu comme partie d'un dispositif réalisant une conjecture portant sur la compréhension des énoncés, un analyseur est un module. Sa fonction est de contribuer à l'explicitation de l'architecture modulaire dont il fait partie.

### 4.3. Questions de modularité

La modularité n'est pas une conjecture homogène; tant s'en faut. Plusieurs conceptions s'affrontent et peuvent se contredire; un analyseur devrait permettre de les approfondir et de les comparer selon la fonction heuristique qui est la sienne. Je vais, dans les paragraphes suivants, prendre comme exemple l'approche réaliste de la modularité; c'est une des conceptions parmi d'autres. Elle me permet surtout ici de marquer les différences qui existent entre les programmes de recherche que je distingue<sup>27</sup>.

L'approche réaliste pose nettement l'existence d'un module de traitement syntaxique, elle est complètement étrangère au cadre résumé en (7) et elle est compatible avec le postulat de base du programme Gouvernement et liage (GB) sinon avec sa forme.

#### 4.3.1. La modularité de la tâche

Admettons l'hypothèse de la conception modulaire proposée par Marr (1982):

27. On peut se reporter à Marandin (1992) pour l'exploration d'autres conceptions de la modularité à l'aide d'un analyseur.

(24) Un calcul important peut être divisé et exécuté sous la forme d'un ensemble de parties qui sont aussi indépendantes les unes des autres que le permet la tâche à accomplir. C'est ce que j'ai appelé *le principe de la conception modulaire*. Son importance vient de ce que lorsqu'il n'est pas appliqué, un petit changement à un endroit a des conséquences dans beaucoup d'autres. Par suite, le traitement pris dans son ensemble est très difficile à corriger ou à améliorer, que ce soit par un concepteur humain ou par l'évolution naturelle, parce qu'un petit changement permettant d'améliorer une partie doit s'accompagner de nombreux changements simultanés compensatoires ailleurs (Marr 1982, p. 102).

Les principes de cette architecture sont bien connus; ils définissent en toute généralité la modularité de la tâche<sup>28</sup>. La confection d'un analyseur effectif doit apporter des réponses aux trois questions qu'elle implique:

- (i) comment individualise-t-on les parties du processus? en particulier, sur quels principes dira-t-on qu'il y a un module de représentation syntaxique (imaginable comme un analyseur)?

- (ii) quelle représentation est donnée par le module d'analyse syntaxique? quel est son grain?

- (iii) le module syntaxique est-il un module au sens fort ou au sens faible? Un module au sens fort est un module fodorien<sup>29</sup>; un module au sens faible est le fruit de n'importe quelle décomposition de la tâche.

On notera d'emblée que les réponses à ces trois questions sont interdépendantes: plus la représentation est "à gros grains", plus l'analyseur peut être informationnellement cloisonné; plus l'analyseur est informationnellement cloisonné, plus il partage les traits d'un système périphérique<sup>30</sup>. Cette interdépendance risque de tourner au cercle vicieux; elle doit être brisée par l'introduction de principes indépendants. C'est l'enjeu de l'introduction de la modularité de l'objet pour fonder la modularité de la tâche.

#### 4.3.2. Un exemple: les relations de dominance immédiate

Marcus et al. (1983) constatent qu'un analyseur déterministe rencontre des difficultés à reconnaître les relations de dominance immédiate dans un domaine syntaxique<sup>31</sup>. Considérons les exemples que je leur emprunte:

28. L'expression anglo-saxonne *modularité de l'objet* est malheureuse; l'expression *hétérogénéité de l'objet* serait plus transparente.

29. Fodor définit *module* à l'aide de neuf traits (ibid, chap III). Je rappelle les deux traits essentiels: (a) un module est un système périphérique; (b) un module est cloisonné informationnellement.

30. L'interdépendance entre grain de la représentation, cloisonnement informationnel et statut périphérique est reconnue par Fodor (Fodor, ibid, p. 100).

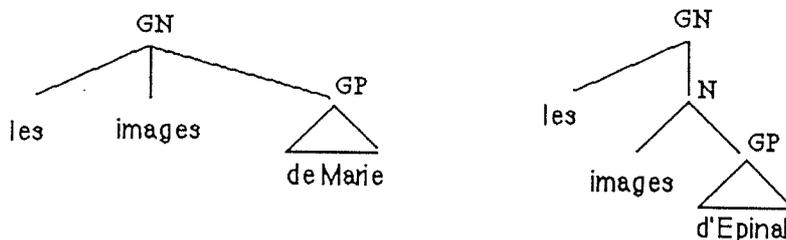
31. *Déterministe* au sens de Marcus: un dispositif déterministe ne fonctionne pas par essai et erreur. Un parseur déterministe s'oppose diamétralement à un dispositif qui applique "à l'aveugle" les règles d'une grammaire. Un dispositif déterministe a accès à l'information pertinente pour décider en toute occasion de la prochaine action. Par conséquent (conséquence qui s'observe), un parseur déterministe dispose

- (25) a. I drove my aunt from Preoria's car.  
 b. He sells ripe apples and cherries.  
 c. He sells ripe apples and bread.

Un analyseur peut aboutir à un cul-de-sac (*garden-path*) quand il analyse *my aunt from Preoria's car* en (25. a) s'il traite *my aunt* comme un GN occupant intégralement la position de l'objet de *drove* ; il ne pourra pas analyser *car*. Un analyseur strictement syntaxique ne peut distinguer sans information supplémentaire les bonnes analyses de (25.b) et (25.c): [ripe [apples and cherries]] vs [[ripe apples] and bread]. Un analyseur du français se trouve dans une situation analogue quand il doit analyser les groupes nominaux (26) quand on les analyse comme (27):

- (26) a. Les images de Marie (sont perdues).  
 b. Les images d'Epinal ( " " ).

(27)



Un analyseur déterministe peut ne pas voir la différence structurale. Résumons l'observation par la proposition suivante:

- (28) Un analyseur déterministe peut analyser les groupes syntaxiques (ce qui se donne en termes de domination), mais ne peut distinguer le détail des groupes (qui se donne en termes de domination immédiate). Autrement dit, il ne peut pas donner une représentation fine des (parties d') énoncés.

L'analyseur est en défaut pour deux raisons : la reconnaissance de la structure dépend d'une propriété lexicale dans un cas (25.b et c) et dans l'autre (25.a) ou (26), il n'y a pas de marquage diacritique dans la forme de surface qui permette de distinguer entre les deux structures. L'analyseur a besoin d'un indice supplémentaire; il ne peut se suffire de la seule donnée de la disposition de surface et des catégories des termes qui s'y succèdent.

---

toujours de l'information qui lui permet de donner la bonne analyse du premier coup. Le point important n'est pas la caractérisation technique à laquelle on réduit souvent le déterminisme: ne pas avoir recours à la marche arrière ou aux analyses en parallèle, mais l'accès à l'information pertinente là où elle est requise. C'est la question de fond posée par Marcus (1980): un parseur syntaxique peut-il être déterministe?

### 4.3.3. Analyse de l'exemple

L'observation est la suivante : un analyseur déterministe ne peut pas reconnaître et, partant représenter, les relations de dominance immédiate. Il faut noter qu'une telle observation ne pourrait être faite en prenant pour objet le comportement d'un locuteur ou en recourant à l'introspection; elle est directement liée à l'observation d'une machine<sup>32</sup>.

Appliquons le raisonnement induit par la définition de la modularité de la tâche rappelée en (24) ci-dessus :

- on distingue plusieurs aspects dans un énoncé ("l'énoncé est modulaire, c'est-à-dire hétérogène"<sup>33</sup>),
- on peut traiter chaque aspect de façon indépendante, de façon "aussi indépendante que le permet la tâche globale",
- on conçoit des unités de traitement spécialisées dans le traitement de chacun des aspects distingués.

Dans notre cas, on distingue la reconnaissance des groupes de la reconnaissance du détail structural interne des groupes qui requiert une information lexicale<sup>34</sup>. On fait l'hypothèse que l'on peut traiter ces deux aspects de façon aussi indépendante que possible. Du coup, le module qui traite des groupes (des relations de dominance) n'effectue qu'une représentation partielle des énoncés: elle est "à gros traits". Elle ne préjuge pas des analyses précises : "elle laisse toutes (et seulement) les possibilités d'analyses syntaxiques correctes ouvertes; [l'analyse] est vague de façon appropriée (*appropriately vague*)" (Marcus et al. 1983, p. 135). Si on trace une analogie avec le dispositif de la vision, c'est une esquisse (sketch) caractéristique des traitements de bas-niveau.

### 4.3.4. L'approche réaliste

Nous venons de suivre un raisonnement analogue à celui que Marr met en oeuvre dans sa modélisation de la vision; il est caractéristique de l'approche réaliste.

(29) Dans le raisonnement réaliste, la modularité de l'objet à traiter détermine la modularité du traitement.

Ceci est clairement explicité par Stillings dans le cas de la vision selon Marr: "the modularity of early vision seems to be possible because of a contingent modularity in the environment. It simply happens (under Marr's view) that reflected light makes available information about the layout of surfaces in the three dimensions in a neatly packaged way. This makes possible the evolution of organisms with modularized early vision. Environmental modularity makes possible fast, hardwired, informationally encapsulated computation" (Stillings 1987b, p. 385).

32. Ce qui illustre la deuxième fonction d'un dispositif: il peut et doit être observé.

33. On reprend, en fait, le contenu de la proposition (10) de Chomsky dont on mesure rétrospectivement l'importance.

34. On exprimerait cette différence de manière beaucoup plus transparente dans le cadre de Milner (1989) en distinguant la reconnaissance des configurations de positions et la reconnaissance des termes dans les positions.

Dans ce cadre, on peut répondre aux trois questions que pose la modularisation de la compréhension (cf §4.3.1) :

- le principe d'individuation des modules est donné par l'objet à traiter,
- chaque module fournit une représentation de l'aspect auquel il est adapté (dans ce cadre, en un sens biologique): la représentation qu'il effectue est constitutivement partielle. Du point de vue du module, le grain de la représentation est maximal; du point de vue du traitement global, le grain peut être extrêmement grossier.
- le module est fodorien si on peut encapsuler l'information requise par chacun des aspects dans les unités de traitement. Autrement dit, il faut pouvoir exprimer les informations dans des sources de savoir indépendantes les unes des autres.

Je ne poursuis pas plus avant : on voit le programme de travail à remplir<sup>35</sup>.

#### 4.3.5. Relations entre CLN et théorie de la syntaxe

Dans le programme de recherche que j'esquisse ici, l'articulation entre la théorie de la compréhension et la théorie des organisations linguistiques (en particulier syntaxiques) s'effectue sur une base profondément différente de celle qui prévaut en (7) pour la GGT et 3A.

La théorie de la compréhension attend de la linguistique qu'elle distingue les aspects des énoncés et qu'elle identifie les ressources nécessaires à leur reconnaissance<sup>36</sup>. Elle n'attend pas que la théorie de la syntaxe lui fournisse une grammaire : elle n'attend pas de système de règles de description des énoncés, ni un dispositif réglant les relations entre niveaux ou type de représentation, ni un ensemble de représentations vraies. Il n'y a aucun sens à maintenir l'image selon laquelle la grammaire est une composante du dispositif de compréhension; de la même façon, il n'y a pas de sens à penser que le dispositif de description des énoncés ait la même architecture que le dispositif de compréhension<sup>37</sup>. Autrement dit, les grammaires que le linguiste construit

35. La représentation globale de l'input est formée par plusieurs modules. Le problème est de savoir comment la représentation globale est formée (si une représentation globale est effectivement produite). On peut poursuivre l'analogie avec la vision: "The various modules contributing information to it [the 2 1/2-D sketch] (stereopsis, motion, etc.) will deliver more or less complete and mutually inconsistent information. Some sort of cooperative computation, which Marr did not describe in detail, is needed to extract a single, optimal, and consistent representation. In addition, the input modules may supply no information at some point or local patches in the image. These areas may be filled in by interpolation computations" (Stillings 1987a, p. 330). De ce point de vue, la question essentielle n'est pas l'interaction entre modules (elle est acquise), mais le contrôle de l'interaction: comment s'opère "la mise ensemble" des contributions des différents modules? y a-t-il "un superviseur"? La réponse de Fodor (1986) est sur ce point purement négative: "le contexte, les attentes d'arrière plan" ne jouent pas le rôle de superviseur.

36. Encore une fois, Stillings tire clairement la conclusion: "Carrying out a naturalistic research program requires that the needed analysis of the world be within reach of the cognitive scientist. The immediate value of the strong-modularity thesis is that it carves out a limited domain of the world and demands only that the output of an input system be reasonably reliable and quickly delivered. (...) Thus the modularity thesis allows us to adopt a modular approach to the world as well" (Stillings, 1987b: 384).

37. Contrairement à ce que continuent à poser les chercheurs travaillant la notion de performance à partir de GB, par exemple Pritchett (1988).

dans son propre programme de recherche, peuvent présenter une architecture et manipuler des représentations complètement différentes de celles que manipulent les locuteurs (selon le modèle de compréhension que l'on construit).

- (30) Un programme de recherche consacré à la compréhension des énoncés, qui développe une conception réaliste de la modularité, n'attend de la syntaxe ni grammaire ni analyse grammaticale; il en attend une décomposition des énoncés qui puisse être le principe de décomposition de la tâche en modules.

Les théories syntaxiques actuelles offrent des principes de décomposition qui ne sont pas équivalents. La décomposition en théories du programme GB ne recouvre pas la décomposition en modules de positions vs module de termes de Milner (1989). Ces deux décompositions ne recouvrent pas la décomposition en domaines syntaxiques qui est à la base de Parsifal (Marcus, 1980) ou la décomposition en modules strictement autonomes des approches auto-structurales (Sadock 1991). Selon que l'on choisit l'une ou l'autre, on obtient des architectures d'analyseur très différentes.

#### 4.3.6. Elargissement

Le raisonnement réaliste propose un principe d'individuation des modules: individuation par l'objet à traiter. D'autres principes d'individuation ont été proposés: l'individuation par le type de fonctionnement (théorie de l'auto-organisation) ou encore l'individuation par la source de savoir (théorie du cloisonnement informationnel) Là aussi, opter pour l'un ou l'autre n'a pas les mêmes implications architecturales. La confection d'un analyseur devrait permettre d'explicitier et d'explorer en détail ces implications.

## 5. CONCLUSION

La question (1), posée en introduction, oblige à la naïveté et au raffinement théorique en ce qu'elle engage les fondements des programmes de recherche. C'est, à mes yeux, une des erreurs majeures de 3A que de croire qu'il est possible de s'en tenir à un état indiscutable, ~~car hors de discussion, des conjectures sur le langage et son usage.~~

Je résume les réponses auxquelles nous sommes parvenus. L'analyseur prend le statut d'une machine industrielle dans le TAL; il est soumis aux impératifs de la technologie, ce ne sont pas toujours ceux d'un programme de recherche. Il est probable que l'analyseur n'a aucun statut en théorie de la syntaxe; il n'en a assurément pas dans le cadre d'une conception où la syntaxe est un principe organisationnel des énoncés qui ne se réduit pas aux règles de combinaison des mots. Que l'on puisse concevoir des analyseurs dans le cadre de théories qui "diluent" la syntaxe dans la combinatoire lexicale, n'invalide pas les analyses dans le style paradigmatique. L'analyseur a un statut de dispositif dans l'intelligence

artificielle abstraite: il devrait, dans ce dernier champ, être inlassablement soumis à l'observation critique.

Bref, un analyseur, considéré en dehors du cadre théorique qui préside à sa conception, est une machine célibataire. Un analyseur, considéré à partir du cadre qui préside à sa conception, est le reflet de ses options fondamentales.

## REFERENCES

ABEILLE, A. (ed.) (1991): *Analyseurs syntaxiques du français*, *TA Informations*, vol 32, n°2, Paris, Klincksiek.

ANDLER, D. (ed.) (1992): *Introduction aux sciences cognitives*, Paris, Gallimard.

AUROUX, S. (1992): *Histoire des idées linguistiques*, vol 1 & 2, Liège, Mardaga.

BALTIN, M. et KROCH A. (eds.) (1989): *Alternative Conceptions of Phrase Structure*, Chicago, The University of Chicago Press.

BARR, A. & FEIGENBAUM, E. (eds.) (1981): *Handbook of Artificial Intelligence*, vol. I, Los Altos, W. Kaufmann Inc.

BEVER, T. (1970): "The Cognitive Basis For Linguistic Structures", in HAYES, J. (ed.) *Cognition and the Development of Language*, New-York, John Wiley & sons, 279-352.

CHANOD, J.-P., HARRIEHAUSEN, B. & MONTEMAGNI, S. (1992): "A two-stage algorithm to parse multi-lingual argument structures", in JENSEN, K. HEIDORN, G. & RICHARDSON, S. (eds.), *Natural Language Processing: the PNL approach*, Hingham, Kluwer.

CHANOD, J.-P. (1992): "Dérèglement du langage et parasitisme computationnel: problèmes de robustesse en analyse syntaxique", communication Journée de l'ATALA, mars 1992

CHOMSKY, N. (1965): *Aspects de la théorie syntaxique*, Paris, Le Seuil (trad. fr.).

CHOMSKY, N. (1986): *Knowledge of language. Its Nature, Origin and Use*, New-York, Praeger.

CHOMSKY, N. (1987): *La nouvelle syntaxe. Présentation et commentaire d'Alain Rouveret*, Paris, Le Seuil.

CHOMSKY, N. , MILLER, G. (1968): *L'analyse formelle des langues naturelles*, Paris, Gauthier-Villars (trad. fr.).

DOWTY, D., KARTTUNEN L., ZWICKY, A. (1985): *Natural Language Parsing*, Cambridge, Cambridge UP.

FODOR, J. A. (1986): *La modularité de l'esprit*, Paris, Minuit (trad. fr.).

FORD, M. et al. (1982): "A Competence-based Theory of Syntactic Closure", in BRESNAN, J. (ed.), *The Mental Representation of Grammatical Relations*, Cambridge, MIT Press, 727-796..

GARFIELD, J. L. (ed.) (1987): *Modularity in Knowledge Representation and Natural-Language Understanding*, Cambridge, MIT Press.

KARTTUNEN, L. (1989): "Radical Lexicalism", in BALTIM, M. & KROCH, A. (eds.) , 43-65.

KARTTUNEN, L. & ZWICKY, A. (1985): "Introduction", in DOWTY et al. (eds.), 1-23.

MARANDIN, J.-M. (1992): "La perception syntaxique", *Le gré des langues*, n° 3, 64-91.

MARCUS, M. (1980): *A Theory of syntactic recognition for natural language*, Cambridge, MIT Press.

MARCUS, M., HINDLE, D. & FLECK, M. (1983): "D-theory: Talking about Talking about Trees", *Actes 21st Meeting of the Association for Computational Linguistics*, Chicago, 129-136.

MARR, D. (1982): *Vision: a computational investigation into the human representation and processing of visual information*, New-York, W.H. Freeman & Co.

MILNE, R. (1982): "Predicting Garden-Path Sentences", *Cognitive Science*, n° 6, 349-373.

MILNER, J.-C (1973): *Arguments linguistiques*, Tours, Editions Mame.

MILNER, J.-C. (1982): *Ordres et Raisons de langue*, Paris, Le Seuil.

MILNER, J.-C (1989): *Introduction à une science du langage*, Paris, Le Seuil.

MORIN, J.-Y, sd, *Théorie syntaxique et théorie du passage*, Université de Montréal, non publié.

NEWELL, A. (1973): "Production systems, models of control structure", in CHASE, W. (ed.), *Visual Information Processing*, pp. 463-526, New-York, Academic Press.

NEWELL, A. (1982): "The Knowledge level", *Artificial Intelligence*, vol. 18, 87-127.

PIOTROWSKI, D. (1990): *Structures applicatives et Langage Naturel. Recherches sur les fondements du modèle : "Grammaire Applicative et Cognitive"*, thèse de doctorat, Paris, EHESS, non publié.

PRIGOGINE, I. & STENGERS, I.(1986): *La nouvelle alliance*, Gallimard.

PRITCHETT, B. (1988): "Garden path phenomena and the grammatical basis of language processing", *Language*, vol. 64, n° 3, 539-576.

PROUST, J. (1987): "L'intelligence artificielle comme philosophie", *Le Débat*, n° 47, Gallimard, 88-102.

QUINE, W. (1977): "Methodological Reflections", in DAVIDSON, D. et al., (eds.) *Semantics of Natural Language*, Dordrecht, Reidel, 442-454.

ROSNER, M. (1983): "Production Systems", in KING, M. (ed.) *Parsing Natural Language*, Londres, Academic Press, 35-58.

SADOCK, J. (1991): *Autolexical Syntax*, Chicago, The University of Chicago Press.

STEEDMAN, M. (1989): "Constituency and Coordination in a Combinatory Grammar", in BALTIN, M. & KROCH, A. (eds) , 201-231.

STILLINGS, N.(1987a): "Introduction", in GARFIELD, J. (ed.), 325-332.

STILLINGS, N. (1987b): "Theories of vision", in GARFIELD, J. (ed.), 383-401.

TOMITA, M. (1988): "«Linguistic» Sentences and «Real» Sentences", *Actes 12ème COLING*, Budapest, 453.

VARELA, F. (1989): *Autonomie et connaissance*, Paris, Le Seuil.

WASOW, T. (1978): "On constraining the class of transformational languages", *Synthese*, n° 39, 81-104.