

Article

« Systèmes énonciatifs et analyse de données textuelles »

Jean-Pierre Desclés

Études littéraires, vol. 10, n° 3, 1977, p. 453-499.

Pour citer cet article, utiliser l'information suivante :

URI: <http://id.erudit.org/iderudit/500447ar>

DOI: 10.7202/500447ar

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : info@erudit.org

SYSTÈMES ÉNONCIATIFS ET ANALYSE DE DONNÉES TEXTUELLES

*jean-pierre desclés*¹

I. Plusieurs techniques d'analyse de textes (analyse de contenu, analyse (automatique) de discours, analyse structurale des récits) se présentent comme neutres, indépendantes de l'utilisateur, transmissibles et conduisant à des résultats reproductibles. Aussi tendent-elles à se substituer aux analyses plus traditionnelles. Sans porter un jugement sur la fécondité des nouvelles techniques utilisées dans l'analyse des données textuelles, nous voudrions en rappeler quelques traits².

I.1. (a) Toutes les méthodes citées plus haut *se réclament de la linguistique générale* : elles valident les techniques employées en élargissant la portée des concepts linguistiques et des procédures analytiques empruntées soit à la phonologie (principe de commutabilité par exemple), soit à la grammaire transformationnelle (structures de surface/profonde; compétence/performance; recherche de cas profonds; génotexte/phénotexte; grammaire générative de textes; recherche d'équivalence inter-phrases . . .), soit à la sémantique structurale (analyses sémique, componentielle . . .), soit à la syntaxe (analyses en termes de noyaux et d'expansions . . .). Dans tous les cas, on déforme des analyses linguistiques conçues pour la dimension de la phrase pour les appliquer à des textes entiers.

(b) Dans la mise en fonctionnement de ces méthodes, *on fait appel à un univers sémantique* défini implicitement ou présenté comme « naturel » parce que relevant du « bon sens » ou d'un « sens commun ». Cet univers est placé comme

¹ Université de Paris VII — 2, Place Jussieu 75005 Paris, Département de Recherches Linguistiques et UER de mathématiques; ERA 642, CMRS.

² Citons, par exemple Le Flem « Linguistique et littérature : le malentendu », *Études littéraires*, vol. 9, n° 2, août 1976; J.C. Gardin : *Les analyses de discours*, Delachaux et Niestlé, 1974, qui adressent des critiques similaires aux nôtres.

une toile de fond sur laquelle se projettent les représentations souvent issues d'un espace économique ou technico-socio-culturel dominant aux intentions qui ne sont pas toujours trop pures et en tout cas pas universelles.

(c) La plupart de ces techniques *se réfèrent à des catégories dont il faudrait argumenter* sinon le statut ontologique, du moins le statut linguistique. On met également en œuvre des symboles qui sont censés jouer le rôle d'opérateurs formels sans que la définition, l'usage et le caractère opérationnel en soient toujours bien délimités.

(d) *On manipule les textes par des procédures qui ne sont pas toujours invariantes* par rapport aux utilisateurs. Parfois ces procédures sont présentées sous forme d'algorithmes, garantissant ainsi la parfaite objectivité; un algorithme transforme des données et si les données ont déjà été manipulées de façon non contrôlées, les résultats obtenus seront, eux, non reproductibles (« laissez moi maître du choix et de la constitution de mes données et je vous prouve ce que vous voulez ! » dit-on).

(e) Il y a rarement une réflexion sur *la nature et la constitution des données* traitées. Par contre, on insiste sur les techniques manipulatoires, sans établir de relations entre traitement et objets traités. On présente ainsi certaines méthodes comme universellement validables quels que soient les textes que l'on soumet.

(f) Enfin, la plupart des méthodes de traitement ne peuvent être guère prises en défaut puisqu'*aucun objectif* n'a été proposé, qu'*aucune problématique* n'a été tissée (mais après tout, en parodiant Wittgenstein, il n'y a peut-être pas de problèmes mais surtout des « embarras » ou des maladies, car des problèmes sans solutions ne sont pas des problèmes du tout !). En effet, comment une méthode peut elle conduire à un échec puisque *l'on ne sait pas reconnaître la présence de l'échec* (la reconnaissance de l'échec est un indice de scientificité) ? On transcrit un texte dans un certain métalangage et l'on a alors, dit-on, constitué un modèle, mais on voit rarement fonctionner ce modèle en sens inverse : chaque formule du métalangage devant se réaliser par un texte acceptable (réversibilité entre représentation et réalisation). On ne sait pas non plus qu'elle est la portée approximative d'une analyse puis-

qu'aucune règle d'ajustement entre modèle et empirique observable n'a été énoncée.

1.2. Ces nouvelles méthodes ont donc pour légitime ambition : le renforcement du caractère objectif et généralisant des démarches intellectuelles. Elles réintroduisent cependant l'arbitraire à un niveau plus discret, laissant ainsi croire aux naïfs (mais y en a-t-il ?) qu'enfin on posséderait une panoplie de techniques auxquelles on soumettrait *tous* les textes qui ainsi livreraient les secrets que les méthodes traditionnelles étaient bien incapables de révéler. Le mythe du scientisme risque de prendre de nouveaux masques (toujours aussi dangereux) !³

Si l'on espère tenir un jour des discours scientifiques sur les discours (est-ce possible pour tous les discours ?) il faudra en payer le prix par une plus grande ascèse nourrie de réflexions méthodologiques et épistémologiques, en particulier sonder la genèse phylo/onto-génétique des concepts, évaluer leur portée théorique et opératoire, délimiter et articuler les niveaux de représentations (non seulement, signifiant/signifié; langage objet/métalangages, mais aussi linguistique/métalinguistique; occurrence d'un représentant/représentant/représenté; systèmes sémiotiques naturels/artificiels, notamment symboliques à vocation significative . . .).

1.3. L'analyse des données textuelles (de quel type de données s'agit-il ? comment ont-elles été constituées ?) doit être restituée, selon nous, dans la problématique des *systèmes de représentations symboliques*, sous-classe des systèmes sémiotiques : d'une part, un texte représente, il est mis pour quelque chose, d'autre part, son analyse nécessite un ou des système(s) de représentations (linguistiques ou non) et le choix du système de représentations dépend des objectifs fixés.

³ Faut-il citer J.C. Gardin (*op. cit.*, p. 56) : « les sciences du discours ne sont sciences que si elles pratiquent le discours de la science, tout le reste est littérature » ou encore D. Le Flem : « le postulat d'homologie entre phrase et discours, la métaphorisation conséquente de la linguistique et l'absence généralisée de réalisme opératoire (. . .) ne sont pas les seules tares qui affectent cette tendance de la sémiotique contemporaine (. . .) Pourtant ce serait naïveté de croire que ces considérations puissent hâter de quelque manière un revirement indispensable (*op. cit.*, p. 282).

Dans *Métalangue, métalangage, métalinguistique*, j'ai (en collaboration avec Z. GUENTCHEVA) proposé une thèse que l'on peut formuler ainsi⁴ :

« Le langage qui sous-tend l'analyse de données textuelles n'est pas homogène mais fait intervenir plusieurs langages qui doivent être distingués et articulés. Si nous partons des traces textuelles, nous devons au moins distinguer (1) *les langues d'analyse linguistique* des (2) langages que j'appelle évaluatifs. Les langages linguistiques traitent des représentations (linguistiques et métalinguistiques) du texte ».

Par *langages évaluatifs*, je vise tous les langages critiques où un analyste formule des jugements du type : *on pense que . . . , il est connu que . . . , X préfère Y à Z, je* (auteur critique) *déclare que . . .* À ces jugements, il conviendrait d'ajouter des propositions issues d'autres domaines (psychanalyse, connaissances littéraires, historiques, sociologiques, réflexions rhétoriques et stylistiques . . .). C'est donc tout un « jeu de langages » évaluatifs qu'il est indispensable de démêler et de séparer, dans une analyse, des langages d'analyse linguistique⁵.

Nous allons nous questionner sur la constitution de ces derniers langages.

II. Peut-on expliquer l'extrême faiblesse de la linguistique actuelle (structurale ou générative) devant des données ayant une dimension supérieure à la phrase ?

II.1. Avant de répondre à cette question, il faut rappeler que si la linguistique n'a pas atteint un degré de scientificité comparable à la physique ou à la biologie, il existe cependant certains indices qui témoignent du souci de la reproductibilité des analyses. En particulier, la majorité des linguistes

⁴ Voir [DES, GUE, 77].

⁵ Dans ses *Investigations philosophiques*, Wittgenstein a introduit la notion de *jeu de langage* montrant par là qu'il y a multiplicité de langages comme il y a une multiplicité des jeux et que la signification des constituants du langage est déterminée par les règles de leur usage. En ce qui concerne l'analyse des données textuelles, la séparation des langages fait mieux apparaître les spécificités respectives plutôt qu'un discours général qui mélangerait tous les points de vue.

reconnaît les limites actuelles de la linguistique contemporaine (quelles que soient les écoles). Celle-ci a mis au point des méthodes extrêmement robustes : méthodes structurales (Troubetzkoy, Jacobson, Hjelmslev, Blomfield, Harris, Tesnière, . . .) puis des méthodes transformationnelles (Harris, Chomsky, Saumjan, Mel'cuck . . .) mais ces méthodes ont un champ d'applicabilité dont on délimite de mieux en mieux les frontières.

Nous pouvons reconnaître que :

(a) les analyses qui réussissent relativement bien sont délimitées à une dimension supérieure, celle de la phrase (et non de l'énoncé); au-delà, les méthodes proposées restent encore très peu reproductibles par tout un chacun;

(b) aucune théorie de la paraphrase n'est actuellement complète, satisfaisante et opératoire; or, toute analyse de discours n'est possible que si l'on a un solide contrôle des équivalences et déformations sémantiques;

(c) Toute analyse de contenu sémantique fait, à un moment ou à un autre, référence à un univers physico-culturel et l'on doute de plus en plus d'une possibilité de constituer une Sémantique Générale des lexèmes (celle de Katz, par exemple) qui, par vocation, se voudrait universelle⁶.

On sait bien analyser les phrases par des méthodes taxinomiques mais l'on ne sait pas analyser et reproduire le processus par lequel l'énonciateur constitue un énoncé muni de valeurs référentielles. De plus, Z.S. Harris a essayé, en étendant les méthodes structurales aux dimensions du discours de fonder une « analyse du discours » mais, très vite, il

⁶ « Ils [Katz et Fodor] supposaient, eux, une analyse entre la phonétique et la sémantique. De même que la représentation phonétique s'appuie sur un système universel de traits phonétiques, de même la représentation sémantique reposerait sur un système universel de traits distinctifs sémantiques. Le système universel de Katz est censé représenter toute la pensée. Katz dit encore aujourd'hui que sa théorie sémantique donne une caractérisation complète des énoncés de la langue de chaque énoncé de toutes les langues, indépendamment de toute considération extralinguistique. À mon avis, il n'existe aucune raison de croire à l'existence d'un tel système sémantique universel (. . .) Je crois qu'on ne peut pas séparer la représentation sémantique de la connaissance du monde » (N. Chomsky, *Dialogues avec Mitsou Ronat*, Flammarion, 1977, p. 146-7).

s'est rendu compte qu'il lui fallait dominer *et* le problème des transformations entre phrases (l'analyse du discours est à l'origine de l'analyse transformationnelle) *et* le problème de la paraphrase (c'est l'objet de ses actuelles recherches)⁷.

II.3. Les limites de la linguistique sont peut-être provisoires mais il faut en rechercher les causes, faire le constat de certaines frontières à l'intelligibilité d'un texte et développer rationnellement les domaines susceptibles d'être approfondis. Je vais donc livrer quelques réflexions qui me sont personnelles en réexaminant certains fondements de la linguistique et en traçant quelques pistes de réflexion.

II.3.1. *Faiblesse des systèmes de représentations*

Il nous semble que les linguistes acceptent trop facilement la séparabilité entre syntaxe, sémantique et pragmatique, introduite par C. Morris. Rappelons ce que la sémiotique entend par ces mots : la syntaxe vise les relations entre les expressions; la sémantique les relations entre expressions et ce qu'elles désignent; la pragmatique les relations entre les expressions et les usagers de ces expressions. Cette tripartition a été féconde pour la logique et surtout pour l'enseignement des systèmes constitués puisque d'un système on peut faire une présentation syntaxique séparée de son interprétation sémantique et de son usage théorique. On sait, par exemple, présenter l'arithmétique élémentaire par un langage où l'on donne le vocabulaire puis les règles de formation des expressions et les règles de transformations sans s'occuper des significations des expressions et des termes. On peut, ensuite, donner une interprétation logico-mathématique de ce langage en assignant des significations aux expressions et termes sous forme de modèles (par des ensembles, des applications, des opérations, des relations).

Enfin, on peut montrer l'usage théorique que l'on fait de ce langage par une voie axiomatique-déductive et ainsi faire l'étude formelle de l'arithmétique élémentaire (recherche de la non-contradiction, de la complétude, de la décidabilité, de la catégoricité...). Cette séparabilité n'est cependant pas si

⁷ Voir, par exemple, « On the theory of language » — *Journal of Philosophy*, Vol. 73, n° 10, mai 1976, p. 253-76.

nette même pour les langages formalisés car les trois aspects sémiotiques sont imbriqués entre eux dès que l'on constitue un langage formalisé chargé de refléter les propriétés d'un système abstrait (mathématique ou logique) : les règles de formation et de transformations de la syntaxe visent à une certaine signifiante.

En linguistique, la syntaxe semble concerner les procédures de catégorisation et les procédures de construction et de constitution des « expressions bien formées » (appelées phrases (*sentence*) ou énoncés avec diverses nuances). La pratique actuelle de son fonctionnement fait apparaître un recours soit à une norme (donc à une classe socio-culturelle d'usagers), soit au sens (donc aux désignations des expressions), soit aux conditions d'acceptabilité (ce qui implique dans le dispositif théorique de description la prise en considération d'énonciateurs réels ou idéalisés sous forme d'un énonciateur universel omniscient à comprendre et à produire).

Après les succès et limites reconnues de l'analyse distributionnelle, on a tenté de « réintroduire » le sens dans les descriptions, sens qui était proscrit par un censeur comme Bloomfield. La sémantique s'est alors développée en réaction aux méthodes purement formelles qui manipulaient les seuls signifiants, mais de façon spontanée, sans contrôle suffisamment raisonné aboutissant sur la lancée aux projets de Sémantique Universelle : que l'on se rappelle les illusions de Katz et de Fodor ! que l'on se reporte aux difficultés rencontrées par le modèle standard de N. Chomsky sur l'introduction des items lexicaux et sur la place et la pertinence, dans le modèle, des représentations sémantiques, ce problème a abouti aux querelles que l'on sait !

Devant l'échec d'une systématisation trop rapide et finalement décevante de la sémantique, en réaction bon nombre de chercheurs s'en détournent pour explorer la composante pragmatique, décrétant que tel phénomène est de nature pragmatique plutôt que sémantique sans avoir précisé, pour autant, ce qu'ils entendaient exactement par sémantique et pragmatique. Après avoir quitté le terrain rigide et ascétique de la syntaxe, la sémantique est apparue comme le « paradis », mais bien vite ce paradis est devenu le « paradis perdu » et la pragmatique apparaît maintenant comme le lieu, où l'on es-

père traiter tout ce que l'on ne sait pas traiter ailleurs. Les linguistes redécouvrent, après le cheminement qui leur est propre, la nature fondamentalement pragmatique du langage, nature qui avait été oubliée, en tout cas depuis C.S. Pierce⁸.

Les procédures de construction des objets (catégories linguistiques), qualifiées de syntaxiques, sont en linguistique essentiellement taxinomiques. C'est évident pour les méthodes structurales (analyses en constituants ou apparentées) : les objets formels manipulés sont des classes d'équivalence d'objets empiriques reliées entre elles par des relations de préordre (hiérarchie, emboîtements). La classe des grammaires transformationnelles de Z.S. Harris (*Mathematical Structure of Language*, 1968) est taxinomique également puisque les opérateurs transformationnels sont par construction biunivoquement associés à leurs traces. Quant aux modèles (Sémantique interprétative/sémantique générative) issus du modèle standard de N. Chomsky, c'est toute une discussion qui devrait être menée pour montrer que les procédures de constitution des objets manipulés (catégories et chaînes de catégories) restent essentiellement classificatoires (les catégories, même en « structure profonde », sont directement déduites des « parties du discours » et les objets de la « forme logique » (au sens de Chomsky) qui sous-tend la représentation sémantique, sont de nature prédicative et construits par classification⁹).

⁸ Il faut introduire la dimension pragmatique dans les principes d'une théorie du langage (voir les principes d'énonciation et de repérage ci-dessous) et non comme une composante chargée d'ajuster le modèle linguistique aux phénomènes comme déictiques, personnes, temps, aspects, modalités... qui ne sont pas décrits adéquatement par le modèle. Cet ajustement est manifeste dans la « grammaire de Montague » (R. Montague : *Formal Philosophy*, Yale University press, 1974), où le langage pragmatique vient s'ancrer sur le langage syntaxique ou encore conçu comme une généralisation de la sémantique (p. 93 à 118; p. 119-147). Le langage humain a une nature pragmatique puisqu'il existe des *indexical expressions* dans toutes les langues; il faut par conséquent construire cette pragmatique qui servira de fondement aux représentations syntaxiques (voir [DES, 78 a] et ci-dessous : II.3.3; III.1; III.2).

⁹ Si par prédication nous entendons « attribution d'une propriété à un terme », il est clair que pour qu'une telle opération puisse s'effectuer, il faut que les éléments mis en jeu par la prédication soient au préalable catégorisés, par exemple sous forme d'un « sujet » et d'un « prédicat ». On

Les systèmes de représentations utilisés (S.R.) sont des systèmes *classificatoires* qui reposent sur des catégorisations taxinomiques (classes d'équivalence, prédication d'une propriété métalinguistique à une expression), des relations hiérarchiques entre catégories, des relations entre analyses hiérarchisées de phrases (nous pensons aux indicateurs syntagmatiques et aux relations transformationnelles qui les relient).

Ces systèmes de représentations classificatoires ne correspondent pas aux analyses empiriques et descriptions des linguistes qui accompagnent les représentations formelles proposées d'un discours qui, lui-même, n'est pas adéquat aux représentations. En d'autres termes, les systèmes de représentations classificatoires (S.R.c) (arbres de dépendance, indicateurs syntagmatiques, expressions parenthésées,...) ne peuvent pas représenter certaines descriptions linguistiques. J'irai plus loin : la classe des S.R.c, du moins tel qu'ils sont actuellement constitués, apparaît comme un obstacle à la claire formulation de certains concepts linguistiques, à l'exacte représentation formelle de certains phénomènes, au traitement manipulatoire de certains problèmes, à la reproduction (au niveau de la représentation symbolique) de certaines argumentations et déductions effectuées par le linguiste. Ces systèmes sont peut-être, selon moi, un obstacle au développement harmonieux de la linguistique.

Évoquons trois exemples :

(a) La théorie des cas (« *Case for cases* ») de C. Fillmore a été intégrée dans un S.R.c mais les notations ne rendent pas compte des intuitions et exigences exprimées par le concept de « relation casuelle ». Le cas abstrait — et non pas le cas morphologique — n'est pas, au niveau de la « structure profonde », ni attaché au prédicat, ni au terme mais est

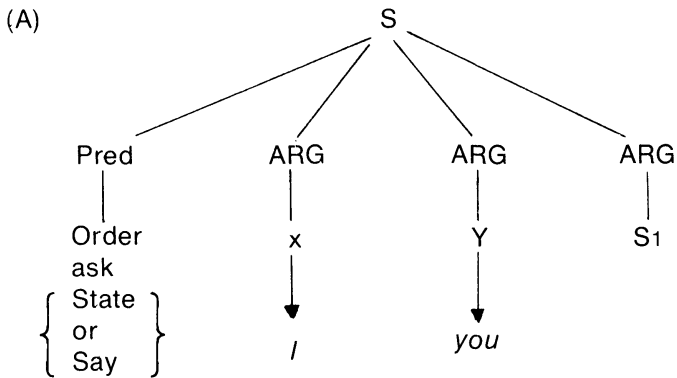
connaît la nature ambiguë du « sujet » toujours mis en question ou en procès ! Souvent, le « sujet » se confond, dans la pratique, avec un élément de la classe nominale « mis en position sujet » et le prédicat avec un verbe. Toute une discussion que l'on pourrait mener, montrerait que les éléments qui sont reliés dans la « forme logique » du Modèle Standard Étendu ne sont ni plus ni moins abstraits que les éléments d'une classe d'équivalence taxinomique. Sur une critique des représentations, voir [CULI, 76].

l'expression d'une relation entre un prédicat et un terme ou plus exactement entre une place non instanciée du prédicat et un argument susceptible d'instancier cette place. Il y a donc une inadéquation fondamentale entre la notation du « cas » dans la représentation arborescente et le concept que sous-tend cette notation (la même difficulté reparaît avec la représentation de la « théorie localiste » de J. Anderson).

(b) La distinction essentielle dans la description des langues comme « topic »/« comment » ou encore « thème/rhème » qui, de plus en plus, envahit la littérature linguistique n'est pas adéquatement représentable dans les S.R.c. En effet, cette distinction n'est pas de nature classificatoire, mais relève d'opérations. Si l'on définit « topic » comme « rappel d'une information connue », il faut représenter ce phénomène de rappel (par qui, pour qui ?) par une opération. Mais le terme qui est topicalisé est lui-même coulé dans une autre structure (prédicative), c'est-à-dire qu'il peut être aussi bien sujet qu'assurer toute autre fonction dans l'énoncé. Prenons : *Il y a Jean, son père, il est ministre* où *Jean* ne peut pas être considéré comme le sujet, mais comme un certain élément servant de base à la prédication; le terme *Jean* est lui-même pris dans une autre relation (relation entre *Jean* et *père* : *Jean a un père*) : on dit que *Jean* est le thème de l'énoncé à propos duquel on énonce une propriété. Comment représenter cette opération dans un système classificatoire (en se refusant, bien entendu, à ramener une représentation symbolique et réglée à une simple notation « ad hoc » sous forme d'un symbole qui serait accolé au terme sans autres justifications)¹⁰ ?

(c) Examinons l'organisation de la catégorie de l'assertion donnée par G. Lakoff dans « Linguistics and Natural Logic » (*Synthèse*, Vol. 22, n° 1-2, 1970); la représentation arborescente (p. 21, dans la traduction française parue chez Klincksieck, 1976) :

¹⁰ Le lecteur pourra se reporter à [CULI, DES; 78] pour voir comment est traité et représenté formellement un phénomène de ce genre qui remet en cause la structure arborescente classificatoire.



ne permet pas de noter le jeu des contraintes entre l'énonciateur, l'asserteur (qui n'est pas nécessairement confondu, comme dans : *tu dis . . .*), le co-asserteur (*tu lui dis . . .*), le sujet de l'énoncé (*tu lui dis : je . . .*) selon que le repérage est direct (énonciation rapportée directe : *tu lui dis : je*) ou indirect (*tu lui dis que tu . . .*). La règle grammaticale d'accord en personne, « qui fait qu'un NP s'accorde en personne avec son antécédent » est soit fautive, soit peu précise¹¹.

II.3.2. Systèmes de représentations métalinguistiques/linguistiques.

Les problèmes précédents (on aurait pu mentionner aussi la représentation des relations anaphoriques, les opérations de focalisation, le problème des réfléchis) font recourir, pour une bonne description, à des systèmes de représentations plus abstraits et plus complexes. Nous avons choisi d'étudier des systèmes de représentations définis non plus par des classes (et une opération de concaténation des classes pour constituer des chaînes) mais par des opérations combinées entre elles par des opérateurs. Ces opérations seront responsables des agencements de signes au sein de l'énoncé. La constitution de l'énoncé ne sera plus expliquée par de simples

¹¹ Dans [DES, 76 a] et [DES, 78 a], je propose une représentation formelle qui permet de noter les contraintes; cette représentation rend mieux compte du fonctionnement de la modalité assertive et ne présuppose pas, comme chez G. Lakoff, qu'une telle modalité soit constituée et qu'il suffise de l'accrocher, sous forme d'une étiquette, à une branche de l'arbre syntaxique.

manipulations élémentaires de la chaîne des signifiants (ajout à gauche, à droite; emboîtement d'une chaîne; permutation; effacement) mais par des *intrications* de signes pris dans un réseau de relations (et pas simplement dans une relation hiérarchique), l'énoncé se constituant par enchevêtrement de chaînes de marqueurs¹².

Nous avons dit que les systèmes devaient être plus abstraits, cela signifie, en particulier, que les opérations et opérateurs ne sont plus, comme dans les systèmes classificatoires, *biunivoquement associés*, et de *façon stable*, à des types de marqueurs syntagmatiques mais reliés aux *occurrences* de ces marqueurs (il s'agit de la distinction : *type/token*), c'est-à-dire à l'occurrence de signifiants textuels — contigus ou non — en fonctionnement. On ne peut donc plus assurer que tel opérateur ou telle opération aura toujours pour trace tel marqueur ou telle classe fixe de marqueurs. Cela ne signifie pas non plus que la correspondance sera aléatoire ou soumise au bon vouloir du descripteur. Au contraire, la correspondance est assurée par un système déterminé de règles ajustant le système des opérateurs et opérations au jeu des occurrences des marqueurs morpho-syntaxiques. Cette correspondance est contrainte par deux exigences complémentaires : (i) chaque combinaison significative d'opérateurs et d'opérations doit avoir une réalisation empirique observable (i.e. doit être le représentant de quelque chose); (ii) chaque occurrence d'une production linguistique interprétable doit être la trace d'une combinaison significative d'opérations et d'opérateurs sur ces opérations, qui ainsi représente cette production¹³.

La relation n'étant plus nécessairement biunivoque et simple entre *le représentant* et *le représenté*, nous qualifierons de

¹² Nous visons à décrire la constitution de l'énoncé et sa production et non pas le produit. Sur le concept d'intrication en linguistique voir [CULI, DES, 78] et [BES, DES, 78] et sur une définition mathématique voir [DES, 76 b].

¹³ Les systèmes d'opérateurs catégoriels de Y. Bar-Hillel, H.B. Curry et J. Lambeck sont des systèmes plus simples que les systèmes d'opérateurs et d'opérations envisagés, puisque chaque opérateur est biunivoquement associé à sa trace linguistique, il est même défini par elle. Ces systèmes appartiennent à la classe des systèmes classificatoires. Dans [DES, 77], je donne un début d'argumentation sur le recours à des systèmes abstraits non classificatoires composés d'opérateurs et d'opérations pour traiter du problème du transfert des catégories d'une langue à une autre.

métalinguistique ces systèmes d'opérateurs et d'opérations chargés de représenter les énoncés. Nous opposons ces systèmes aux systèmes de représentations linguistiques où chaque objet (« représentant ») est associé à un représentant de la classe d'équivalence des objets linguistiques « représentés ».

Les systèmes de représentations métalinguistiques (S.R. μ) sont externes aux langues et leur construction rend nécessaire, sinon indispensable, le recours à la mathématique et à la logique. Ces systèmes entrent dans la classe de la « logique combinatoire pure » (ou un système équivalent) où un objet est une opération fonctionnelle qui opère sur une autre opération donnant pour résultat une opération fonctionnelle.

Du point de vue sémiotique, une langue est un système de représentations. On peut utiliser ce système de représentations pour autoreprésenter trivialement cette langue (usage autonyme des signes) ou bien caractériser, à l'intérieur de la langue, une partie de celle-ci, (chargée de représenter par des énoncés métalinguistiques toute la langue, en constituant une métalangue à l'intérieur de la langue). Une telle autoreprésentation non triviale est construite par Z. Harris qui représente un énoncé de la langue par une coordination d'énoncés de la métalangue (et donc de la langue) chargée de décrire cet énoncé¹⁴. Enfin, on peut représenter une langue par soit une autre langue (problème de traduction), soit par un système autonome — ou tendant à l'être — et fonctionnant en parfaite indépendance par rapport à la langue représentée. Si on constitue un langage formalisé, donc métalangage par rapport à la langue, destiné à jouer ce rôle : refléter les propriétés de la langue, on dispose d'un système symbolique externe à la langue. Y-a-t-il cependant indépendance entre le représenté (la langue à décrire) et le représentant (le langage formalisé jouant le rôle de métalangage) ? L'analyse de la constitution de tout langage formalisé montre que ceux-ci ont *un statut linguistique second*¹⁵ par rapport aux langues naturelles : chaque langage

¹⁴ Voir Z. S. Harris : *Mathematical Structure of Language*, Wiley, 1968. Pour Harris, « la métalangue est dans la langue » est un principe fondateur de sa théorie (voir aussi [DES, GUE, 77]).

¹⁵ Voir sur ce sujet D. Dubarle : *Logos et Formalisation du langage*, Klincksieck, 1977, p. 76-80.

formalisé réclame une métalangue qui assure sa fondation et son mode de fonctionnement (description des symboles, usage des règles de formation et de transformation, etc. . .). Cette métalangue peut être formalisée mais constituant un langage formalisé elle réclamera à son tour une métalangue . . . La métalangue ultime ou fondatrice est alors toujours une langue naturelle qui joue le rôle linguistique primitif.

Étant donné une langue naturelle L , le langage formalisé μL , métalangage de L , ne peut donc être totalement indépendant de L , même s'il est externe et donc séparé de L , puisque $\mu\mu L$ (métalangage de μL) fait appel à une métalangue primitive naturelle. Précisons notre vocabulaire¹⁶ appelons « *A-language* » un langage symbolique artificiel et « *U-language* » une langue naturelle quelconque. Remarquons que chaque « *U-language* » jouit d'une propriété d'ouverture, c'est-à-dire que l'on peut introduire des symboles dans un « *U-language* » et constituer ainsi un « *U+A-language* » qui possède toutes les propriétés du « *A-language* ».

Nous avons donc l'étagement suivant (en indiquant les types sémiotiques : « *A-language* » et « *U+A₁+A_n+ . . .language* » de même type sémiotique que « *U-language* ») :

U-language		A-language	
(0)	L	(1)	μL
(2)	$\mu\mu L = L_1 C L$	(3)	μL_1
(4)	$\mu\mu L_1 = L_2 C L$	(5)	μL_2
	etc.		

Tel que nous venons de le décrire, la description d'une langue naturelle L renvoie, comme dans deux miroirs face à face, à une suite infinie de métalangues et de métalangages

¹⁶ Les deux types sémiotiques *A-language* et *U-language* (artificiel et construit/naturel) sont définis par H.B. Curry (*Combinatory Logic*, p. 25). La propriété d'ouverture a été définie dans [DES, GUE; 77]. Nous distinguons « métalangue » de type sémiotique U-language de « métalangage » de type sémiotique A-language. La métalangue est en général construite par les propriétés qui la définissent et elle est une partie de la langue.

des deux types sémiotiques. Pour rompre cette fuite infinie, nous allons imposer une double contrainte aux métalangages de description :

(iii) chaque formule significative du métalangage μL est glosable dans L par un énoncé métalinguistique (étant une coordination d'énoncés métalinguistiques élémentaires) de la métalangue $\mu\mu L$; cet énoncé indique le fonctionnement des symboles de la formule significative¹⁷.

(iv) la métalangue $\mu\mu L$ comprend une partie finie (comprenant les énoncés de formation et de transformation) et une partie telle que chaque énoncé métalinguistique de la métalangue $\mu\mu L$ soit une glose d'une formule métalinguistique du métalangage μL , indiquant le mode de fonctionnement de cette formule.

Les clauses (iii) et (iv) n'assurent pas nécessairement la clôture de la description mais permettent cependant d'établir des relations entre les différentes métalangues entre elles et les différents métalangages entre eux. Cette question devrait être reprise sous une forme plus technique.

Quoiqu'il en soit, les clauses (i), (ii), (iii) et (iv) assurent l'articulation entre linguistique et métalinguistique.

II.3.3. *Les théories du signe ne sont pas assez opératoires*

Nous vivons trop sous l'influence d'un signe renvoyant à un objet et se dédoublant en un signifié conceptuel et un signifiant matériel. Lorsque nous distinguons un signifiant et un signifié, il faut peut-être s'interroger sur celui qui établit le lien entre les deux et qui désigne quelque chose en utilisant le signe. Cette position réintroduit la dimension pragmatique dès les fondements et non comme une composante surajoutée à la syntaxe et à la sémantique.

Dans la plupart des modèles du signe et de ce qui en découle, le signifieur ou l'énonciateur reste absent ou apparaît de façon constante, ce qui revient à l'effacer. La théorie du signe de Saussure a pour conséquence l'opposition langue/parole et chez Chomsky l'opposition compétence/performance. Puisque la langue n'existe pas (le chinois n'a

¹⁷ On peut donc dire que « le métalangage μL est générateur d'une métalangue » (voir [DES, GUE; 77]).

pas de réalité observable), il faut idéaliser l'objet observable et donc postuler l'existence d'un locuteur idéal, témoin objectif de la langue, omniscient tant pour produire que pour reconnaître les énoncés de la langue. Ce témoin objectif a donc la compétence de la langue puis, à l'aide de la performance, on ajuste sa production aux réalités observables. Il y a donc un énonciateur qui possède en lui toute la compétence de la langue, mais il interviendra, par définition de la compétence, de façon constante et donc il sera toujours représenté par un paramètre constant : f_u (sujet énonciateur universel). Puisqu'il est constant, il est inutile de faire apparaître f_u dans la représentation elle-même. C'est ainsi que partant d'une théorie saussurienne du signe, on arrive à postuler un sujet idéal représentable par un paramètre constant et finalement à éliminer le sujet. Dans le schéma de la communication de R. Jakobson, abondamment repris et commenté, on parle bien d'un auteur ou d'un émetteur et d'un destinataire mais, là-encore, par idéalisation, on suppose qu'ils sont des invariants, il devient donc superflu de leur faire jouer un rôle dans la construction théorique. C'est oublier la relation fondamentale qui existe entre (1) le langage, en tant que système de représentations; (2) ce que représente le langage; (3) celui qui représente quelque chose en utilisant le langage. Il devient alors impossible de « bien » poser¹⁸ le problème de *la constitution d'un énoncé* (comment l'énonciateur constitue-t-il son énoncé, par quelle effectuation d'opérations, qui mettent en jeu une série d'opérateurs et d'opérandes, construit-il étape après étape l'énoncé ?), le véritable problème de *la compréhension d'un énoncé*, le problème de *la transformation d'un énoncé* (recherche d'une paraphrase, déformation grammaticale, modulation stylistique et rhétorique, traduction...) le problème de l'ajustement intersubjectif entre l'énonciateur qui produit un texte et son co-énonciateur qui le comprend, en transformant éventuel-

¹⁸ La bonne constitution d'un énoncé et non la bonne formation d'une phrase ! Il faut en effet observer les productions linguistiques telles qu'elles apparaissent. Ceci m'a amené à articuler : phrase/énoncé/énonçable/chaîne plausible... (voir J.P. Desclés : « Énoncés, énonçables », *Rapport PIT-FALL*, n° 18, 1975, Département de Recherches Linguistiques, Université de Paris VII). Il n'est pas possible non plus, sauf *in abstracto* ou en première approximation de parler de « l'ensemble » des énoncés alors que l'on peut parler de « l'ensemble » des phrases !

lement ce texte pour déclencher un comportement de réponse aux sollicitations impliquées par le texte échangé . . .

L'effacement — conscient ou inconscient — de l'énonciateur repousse très loin toute formulation des transformations d'un système cognitif en un autre qui rendrait compte des processus d'acquisition du langage par un sujet humain; cela conduit au contraire aux solutions extrêmes comme imitation/innéisme négligeant ainsi le rôle dialectique de tout mécanisme cognitif¹⁹.

Le contemporain de F. de Saussure, C.S. Pierce propose une théorie du signe assez originale. Il n'est pas question d'en faire une présentation soumise aux règles de l'exégèse. Nous n'allons même pas tenter de résumer schématiquement cette théorie, souvent présentée de façon trop parcellaire, mais nous allons en faire une « lecture » sans chercher à justifier ici notre interprétation.

« A sign, or *representamen*, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the *interpretant* of the first sign. The sign stands for something, its *object* » (Pierce : *Collected Papers*, t. 2, 228)²⁰.

Un *representamen*, entité matérielle, suite de signifiants phoniques ou graphiques, est un substitut d'un *objet* mais ce substitut ne prend toute sa signification que s'il est interprété par quelqu'un; le résultat de l'acte d'interprétation constitue l'*interprétant*; cet interprétant est lui-même un signe « équivalent » au *representamen* initial et peut donc déclencher, puisqu'il est devenu *representamen*, un nouvel acte d'interprétation conduisant à créer un nouvel

¹⁹ L'énonciateur étant absent, on admet trop vite que l'enfant, en situation d'apprentissage, posséderait un sous-système du système cognitif adulte et le développerait étape par étape pour atteindre l'état optimal. Ceci semble une hypothèse très forte très certainement erronée, et qui ne tient pas compte de l'interaction entre systèmes cognitifs (voir L. Lentin : « Intercompréhension dans le dialogue adulte-enfant », *Rééducation Orthophonique*, vol. 14, 1976, n° 92, pp. 492-513).

²⁰ *Collected papers of Charles Sanders Pierce*, 8 vol. Belnap Press of Harvard Univ. Press, Cambridge, Mass., 1958-60.

interprétant; les deux actes d'interprétations qui, par nécessité, impliquent quelque (*some*) interpréteur, ne sont pas effectués obligatoirement par le même interpréteur.

« A REPRESENTAMEN is a subject of a triadic relation TO a second, called OBJECT, FOR a third, called its INTERPRETANT, this triadic relation being such that the REPRESENTAMEN determines its interpretant to stand in the same triadic relation to the same object for some interpretant » (*op. cit.*, t. 1, 541); « (...) I confine the word *representation* to the operation of a sign or its relation to the object for the interpreter of the representation » (*op. cit.*, t. 1, 541).

Un signe apparaît donc comme le résultat d'une représentation, comme représentant quelque chose d'autre que lui-même (« Every sign stands for an object independent of itself (...) » (*op. cit.*, t. 1, 538). Mais cet acte de représentation suivi de l'acte d'interprétation fait entrer la triade *object, representamen, interprétant* dans la théorie de la Primarité, Secondarité, Tertiarité que nous n'utilisons pas ici.

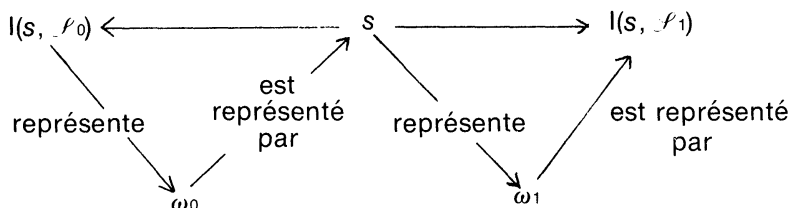
Chaque representamen donne naissance à un interprétant mental (*mental interpretant*, *op. cit.*, t. 2, 274) qui lui-même donne naissance, en tant que *representamen interprétant*, à un autre *interprétant mental* pour un certain interpréteur.

Si nous prenons le schéma de dialogue où un énonciateur (\mathcal{P}_0) est en communication avec un co-énonciateur (\mathcal{P}_1), \mathcal{P}_0 produit un texte matériel, suite de signes phoniques ou graphiques. Puisque l'énonciateur \mathcal{P}_0 est aussi un interpréteur, le signe produit, s est interprété par \mathcal{P}_0 , d'où l'interprétant mental $l(s, \mathcal{P}_0)$; de même, le co-énonciateur \mathcal{P}_1 interprète s et construit l'*interprétant mental* $l(s, \mathcal{P}_1)$. La relation triadique $\langle s, \omega_0, l(s, \mathcal{P}_0) \rangle$ entre *representamen* s , *object* ω_0 et *interprétant* $l(s, \mathcal{P}_0)$ est construite à partir de l'objet ω_0 représenté matériellement par s et interprété par \mathcal{P}_0 en un interprétant $l(s, \mathcal{P}_0)$. Le signe matériel (phonique ou graphique) s entre dans une autre relation triadique $\langle s, \omega_1, l(s, \mathcal{P}_1) \rangle$ où, cette fois-ci, s représente l'objet ω_1 qui est interprété par \mathcal{P}_1 . Alors que c'est \mathcal{P}_0 qui, en produisant s , établit une relation entre s , ω_0 et $l(s, \mathcal{P}_0)$ déterminé par s , c'est \mathcal{P}_1 qui, en comprenant s , détermine $l(s, \mathcal{P}_1)$ et établit une relation entre s , $l(s, \mathcal{P}_1)$ et ω_1 . Pour \mathcal{P}_0 , l'objet ω_0 est dans un rapport premier, alors que

pour \mathcal{F}_1 , ω_1 est dans un rapport second (il s'agit des rapports référentiels).

Cet aspect dynamique de la relation triadique de signification permet de ne pas identifier systématiquement ω_0 et ω_1 car si « the *representamen* determines its *interpretant* to stand in the same triadic relation to the same object for some (SOME) interpretant », cela n'implique pas que le même type de relation tienne pour n'importe quel *interpretant* (résultat d'une interprétation).

Résumons notre position, dans le cas du dialogue, par le diagramme :



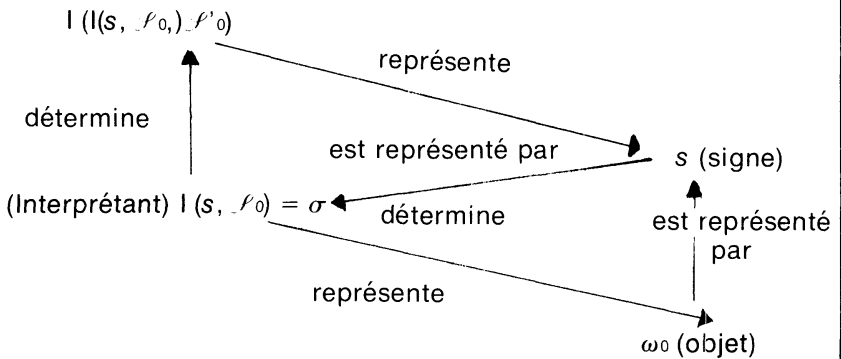
où l'on voit apparaître, d'une part, l'échange du signes dans le dialogue et, d'autre part, la dissymétrie entre la fonction de production de s par \mathcal{F}_0 et la fonction de compréhension de s par \mathcal{F}_1 .

Dans le cas où ω_0 et ω_1 coïncident, il y a intercompréhension, dans le cas contraire, il y a malentendu (ou ambiguïté référentielle).

Nous avons présenté très sommairement cette esquisse de la théorie du signe issue de C.S. Pierce mais si nous l'argumentons mieux nous ferions apparaître les oppositions entre le processus de production et le processus de compréhension. En particulier, ce que nous avons dit, ne doit pas laisser croire que l'objet ω_0 est un « donné (externe à \mathcal{F}_0) réel » d'un monde entièrement extra-linguistique. Dès que l'on aborde les problèmes de production (qui produit quoi ? comment ? par quelles opérations ?) le diagramme précédent devient insuffisant (puisque l'on a, dans le sous-graphe de gauche, un circuit) car on ne sait pas où est l'origine de la production de s (est-ce ω_0 ? $I(s, \mathcal{F}_0)$? ou autre chose ?). Cette esquisse de théorie doit alors être plongée dans une théorie plus vaste de la production/compréhension.

Nous avons dit que chaque *interprétant* constituait lui-même un signe (« It (representamen) addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign »). Comme il n'est pas question de se lancer dans des considérations psychologisantes ou dans des considérations mentalistes sur l'esprit (*mind*) de l'interpréteur (ce que Pierre ne faisait pas, remarquons-le), *l'interprétant mental* ne peut qu'être représenté dans un système de représentations métalinguistique où l'on désigne l'interpréteur sous forme d'un symbole métalinguistique.

$I(s, \mathcal{P}o)$ est donc un signe (métasigne par rapport à s) désigné par σ . Si ce signe σ appartient au métalangage de description μL , de la langue L , à laquelle appartient le signe s , alors le signe σ a pour objets puisque dans μL , s est représenté par σ . À son tour, σ étant un signe, il déclenche dans l'esprit de quelqu'interpréteur, dison $\mathcal{P}'o$, un *interprétant* $I(\sigma, \mathcal{P}'o)$. Etant donné que σ désigne $I(s, \mathcal{P}o)$ on a donc $I(\sigma, \mathcal{P}'o) \approx I(I(s, \mathcal{P}o), \mathcal{P}'o)$, d'où le diagramme :



L'esquisse, très hâtive, de théorie du signe que nous venons de produire, laisse, comme la plupart des théories du signe, dans l'ombre un certain nombre de problèmes et tout particulièrement celui de la dimension du signe (est-ce un mot, une phrase, un énoncé, un nom ?) et ses conditions d'apparition et d'observation. Nous avons fait « comme si » un consensus était bien établi sur cette notion de signe et que l'on connaissait bien ses conditions d'emplois.

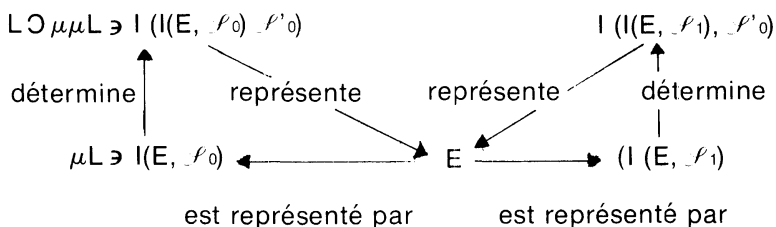
C'est sans doute G. Frege qui a, dans notre époque moderne, le mieux articulé la relation entre le sens (*Sinn*) et la

dénotation (*Bedeutung*) du signe de la dimension du signifiant propositionnel qui se manifeste par un énoncé. Il renouvelait ainsi avec les problèmes posés par Aristote (dans son étude de la prédication et de la vérité ou de la fausseté de ce qui est dit) et poursuivis à l'époque médiévale (théories de la *Suppositio*) rompant avec les recherches sur les seules appellations nominales de la *Logique* de Port Royal ou de J. Stuart Mill.

À quoi réfère un énoncé, conçu comme une unité minimale d'énonciation et ayant un nombre fini d'interprétations (car une suite qui aurait une infinité d'interprétations ne signifie rien et ne peut avoir le statut d'énoncé) ? Alors que l'appellation nominale vise des objets sur lesquels on prédique des propriétés, l'énonciation vise à représenter des « faits » ou des « événements ». C'est l'énonciateur qui rend présent à son co-énonciateur ces faits ou ces événements en échangeant avec lui des suites de signes perceptibles, c'est-à-dire des énoncés (« Dans l'énoncé, la pensée s'exprime de façon perceptible » écrit Wittgenstein).

G. Frege a décidé, en tant que logicien, que la dénotation — *Bedeutung* — de l'énoncé est sa valeur de vérité, constituant par là la classe des langages formels. Cette décision justifiée par les finalités de la logique ne l'est pas en linguistique puisque le problème de la vérité devient secondaire alors qu'il était essentiel en logique. Cela ne suffit pas à évacuer ce problème, la visée linguistique le déplace simplement en construisant ses propres modèles de la référence sans chercher à les importer de la logique.

Reprenons le diagramme précédent. Chaque énoncé E de la langue L est représenté par une formule métalinguistique qui est le résultat (simulé) d'une interprétation de E par \mathcal{I}_0 : c'est l'interprétant $I(E, \mathcal{I}_0)$. Cette formule appartient au métalangage L et détermine un autre interprétant $I(I(E, \mathcal{I}_0), \mathcal{I}'_0)$ qui est le résultat d'une autre interprétation. Si l'on impose (« le métalangage μL est générateur de métalangue », voir ci-dessus) que cet interprétant soit un élément d'une métalangue μL incluse dans L , cet interprétant second constitue bien une représentation à l'intérieur de la langue, de l'énoncé E primitif, la commutativité du diagramme étant assurée par la clause (iii) donnée ci-dessus (cf. II.3.2). Nous obtenons alors, dans un schéma de dialogue, le diagramme :



qui illustre les relations entretenues entre énoncés (E), formules de représentations métalinguistiques $I(E, f_0)$, $I(E, f_1)$ et gloses métalinguistiques.

Nous examinerons plus bas la relation entre l'énoncé et « ce à quoi il renvoie » qui, à ce stade, n'apparaît comme nullement réglée puisque nous avons exclu la « valeur de vérité » comme candidat possible à jouer ce rôle de *denotatum* d'énoncé. Remarquons aussi que l'énoncé apparaît comme l'unité privilégiée d'observation pour le linguiste qui observe « à l'œil nu ». L'énoncé, en effet, apparaît comme l'objet le plus directement observable, plus directement observable que le mot, le phonème, la phrase (qui souvent n'est qu'un énoncé réduit à se couler dans un moule normé) . . . bien que l'énoncé ne soit pas un objet empirique pur qui se présenterait à l'observation du linguiste sous une apparence évidente (« il suffit d'écouter » !) toujours claire et parfaitement reproductible.

II.3.4. *La distinction langage/langues est peu articulée en général*

Si l'on dit que la linguistique est la science du langage, c'est lui imposer une position hégémonique et fautive car d'autres domaines sont eux aussi concernés par le langage. La linguistique ne peut être non plus la science des langues, car une langue n'a pas une homogénéité telle que l'on puisse la cerner par des critères opératoires résistant à l'épreuve des conditions d'observations. Le chinois, c'est la langue parlée par les Chinois mais un chinois du nord ne comprend pas un chinois du sud ! N. Chomsky va jusqu'à dire : « à mon avis, la notion de langue n'est pas une notion linguistique (. . .) Ce sont des raisons politiques qui définissent le chinois (. . .) qu'est-ce qu'une langue ? On dit par plaisanterie qu'une langue, c'est ce

qui a une armée et une marine » (« *Dialogues avec Mitsou Ronat* », Flammarion, 1977).

En admettant que, en toute première approximation, on accepte la notion de langue, nous pouvons dire, à la suite de E. Benveniste, que la linguistique a un double objet, « elle est science du langage et science des langues » ou mieux, en suivant A. Culioli : « la linguistique a pour objet le langage appréhendé à travers la diversité des langues ». Cela signifie deux choses : d'un côté, la linguistique est une science à base empirique, d'un autre côté, en tant que science, elle vise au généralisable donc à la recherche de l'invariance par rapport à la diversité des observations. Le fait de poser l'articulation langage/langues évite, au moins dans les intentions, d'assimiler trop simplement le langage à une langue naturelle particulière ou de voir dans sa propre langue le prototype même du langage, en érigeant les catégories linguistiques d'une langue en catégories langagières. On connaît les erreurs des anciens et des médiévaux qui voyaient dans le grec d'abord, puis le latin ensuite, le modèle par excellence du langage. Pour les premiers, par exemple, les autres langues s'opposaient au grec et formaient un conglomerat de langues « barbares ». Ne voit-on pas les missionnaires jésuites, au XVII^e siècle, s'étonner du caractère « aberrant » du chinois qui ne marque pas morphologiquement la distinction nom/verbe ? On peut, en retour, s'interroger aujourd'hui sur les actuelles descriptions, du japonais par exemple, à partir du modèle anglais en forçant l'énoncé japonais, constitué par diverses opérations de thématization (dont « wa » et « ga » sont, entre autre, les traces) à s'inscrire dans le moule syntaxique « SN/SV » constitué par des opérations de prédication. N'y a-t-il pas, dans la pratique, généralisation hâtive : les structures linguistiques des langues indo-européennes seraient le prototype des structures langagières fondamentales et toute langue, par conséquent, devrait être décrite à l'aide de cette structure ?²¹

²¹ Voir aussi J. Desclés et Z. Guentchéva : « À propos de la distinction langage/langues », *Humanisme et foi chrétienne*, ed. Beauchesne, 1976. Sur le problème de wa et ga et une critique de sa représentation, voir S.Y. Kuroda : « Le jugement catégorique et le jugement thétique, exemples tirés de la syntaxe japonaise », *Langages* 30, 1973, pp. 81-110 et également [CULI, DES; 78].

La linguistique n'est pas « la » science du langage, cela suppose donc qu'elle se refuse à développer une recherche spéculative coupée de toute observation. Elle ne s'impose donc pas la tâche de construire une « grammaire pure » ou une « grammaire spéculative » à partir des *seules* considérations sur les conditions de la communication ou de dialogue entre des êtres vivants. Elle refuse également toute démarche introspective ou purement rationnelle ou formelle « en soi », si ces discours organisés ne sont pas reliés à des observations menées systématiquement.

La linguistique ne peut pas être non plus une théorie cohérente des langues si, voulant atteindre ce qu'il y a de fondamental dans l'activité langagière commune à tous les hommes, elle ne s'impose pas une relation dialectique entre langage et langues, à moins de ramener son propre rôle à celui de simple collecte de « faits » plus ou moins exotiques et déroutants que l'on juxtapose pour mieux montrer l'hétérogénéité irréductible.

La linguistique est une science à base empirique, dont le domaine comprend la diversité des langues. Cela suppose que toute construction théorique doit tenir compte du fonctionnement observable de langues aussi diverses que possible. En effet, nul ne peut prévoir l'extrême richesse de la diversité mais, de plus, tel phénomène qui apparaît comme masqué dans telle langue L_1 devient manifeste dans telle autre langue L_2 , si bien que, lorsqu'on effectuera le retour sur L_1 , le phénomène étudié deviendra plus visible, montrant ainsi le caractère généralisable du phénomène. À titre d'exemple, on pourrait citer le phénomène des classificateurs dans les langues dites « à classificateurs » (chinois, vietnamien, etc.); ce phénomène conduit à « voir » les classificateurs de l'anglais et du français²².

La linguistique est donc à *la quête des invariants* tant par rapport aux langues (c'est le problème de l'équivalence par

²² Les problèmes des aspects des langues slaves et sémitiques invitent à mieux regarder l'organisation grammaticale des temps des langues romanes, par exemple. Le problème de la détermination bien manifesté par certains déterminants des langues « à articles » (grec ancien, bulgare, français, . . .) oblige à reconsidérer ce problème dans d'autres langues sans articles (latin, russe, . . .).

traduction, du transfert des catégories entre langues), que par rapport aux énonciateurs (problèmes de dialectes et niveaux de langue étudiés par la socio-linguistique; problèmes d'acquisition d'une langue première ou seconde...) à l'intérieur d'une même langue (problème de paraphrase et des déformations grammaticales d'un énoncé). Comment se manifestent ces invariants ? Comment les découvrir ? Ce sont, à notre avis, les questions essentielles auxquelles les linguistes doivent apporter des éléments de réponse.

La linguistique doit cependant apprendre à constituer ses données. Nous avons dit que « la » langue n'était pas un « bon » concept linguistique, sauf par approximation. Il semble que la linguistique contemporaine reste trop étroitement enfermée dans le cadre observationnel de la phrase, unité avant tout normative et qu'elle ne sache pas encore bien observer les productions linguistiques. Cela suppose que l'on se donne *une théorie de l'observation* : comment le linguiste, qui lui-même est un générateur de productions linguistiques, observe-t-il en tenant compte des erreurs d'interprétations et de productions, de l'usure bien connue du générateur, par exemple..., de la relativité des degrés d'acceptabilité, du commentaire métalinguistique de l'informateur, des transferts inévitables d'une langue à une autre... Il faut aussi construire *une théorie des observables* : quelles sont les données ? comment sont-elles constituées ? à partir d'où ? par qui ? pour quels objectifs ? quelles sont les unités qui apparaissent le plus directement observables ? quels sont les ajustements approximatifs entre un donné observable et le construit théorique ?²³

En dialectisant langage et langues, les linguistes rejoignent la philosophie du langage mais avec une perspective nouvelle, la psychologie génétique mais en apportant une réponse à la question « quelle est la nature du système langagier ? », la logique qui ne peut être séparée radicalement de l'activité langagière puisque le langage sert à véhiculer nos raisonnements, l'intelligence artificielle mais en lui proposant des fondements peut-être plus assurés, la biologie où l'on peut chercher à établir une relation comparative entre les systèmes

²³ Sur la constitution des données en linguistique, voir [CULI, DES, 78]. Sur l'observation, voir « Énoncés/Énonçables », note 18.

et opérations métalinguistiques et l'organisation neurophysiologique organisée . . .

III. Les divers obstacles qui sont apparus sont-ils infranchissables ? Y a-t-il des indices, dans la linguistique contemporaine, qui laisseraient entrevoir de nouvelles frontières ? En passant en revue quelques-unes des raisons sur les limites de la démarche linguistique, j'ai esquissé certains linéaments programmatiques. Est-il possible d'aller plus loin ?

III.1. Un premier constat s'impose : introduire le sujet énonciateur dans la représentation métalinguistique. Si nous refusons toute approche « mentaliste » et introspective, le sujet ne peut donc qu'avoir un statut métalinguistique, c'est-à-dire être le résultat d'une construction réfléchie et organisée par le linguiste qui cherche ainsi à représenter les productions linguistiques par des langages formalisés manipulables et transmissibles. Cette construction tend alors à se dégager des ambiguïtés inhérentes aux langues, des interprétations instables et donc fuyantes, des pièges terminologiques si pesants dans toute activité épilinguistique, car ces facteurs handicapent fortement celui qui se restreint à une simple représentation linguistique (et non métalinguistique) de la langue analysée. En se conformant aux démarches scientifiques qui visent le général, le linguiste ne représente pas chaque occurrence d'acte d'énonciation où interviendrait un sujet individuel « concret » comme Socrate, Pierre, a, b, c, . . . mais un type d'acte d'énonciation où intervient le sujet énonciateur présent sous forme d'un paramètre, dit métalinguistique.

À la suite de A. Culioli, symbolisons ce sujet par \mathcal{L}_0 . Nous avons repris cette même symbolisations dans nos publications ([CULI, 73] [DES, 76a] [DES, 78a]).

Nous posons donc par principe (*principe de l'énonciation*) qu'il n'existe pas un sujet énonciateur universel \mathcal{L}_u et toute formule métalinguistique significative φ qui représente — ou se réalise par — un énonçable de la langue fait intervenir au moins une occurrence du paramètre \mathcal{L}_0 du sujet énonciateur d'origine, d'où la notation pour une telle formule $\varphi (\mathcal{L}_0)$. L'occurrence \mathcal{L}_0 dans φ a l'avantage de rappeler que cette formule est constituée à partir d'une origine. De même qu'en

physique l'espace et la vitesse ne sont pas des absolus mais relatifs par rapport à un système de référence (principe de la relativité galiléenne), de même toute représentation métalinguistique est relative à une origine. Tout changement par rapport à ce repère devra par conséquent se traduire par des traces textuelles.

Si nous substituons à \mathcal{L}_0 , dans une formule métalinguistique significative $\varphi(\mathcal{L}_0)$, une constante d'individus a, b, c, . . ., nous obtenons la formule « concrète » $\varphi(a)$ ²⁴.

L'introduction de \mathcal{L}_0 , en tant que paramètre dans la représentation métalinguistique et cela dès le principe de toute constitution de formule significative, remet en cause le schéma de Jakobson où le sujet \mathcal{L}_0 a été « effacé ». Les systèmes de représentations métalinguistiques ($SR\mu$) compatibles avec le principe d'énonciation sont qualifiés par nous d'énonciatifs. De tels systèmes fournissent un fondement théorique à la socio-linguistique en harmonisant, tout en les séparant, les recherches sur l'organisation des opérations générales qui sont mises en jeu dans un processus de production ou de compréhension, d'un côté, et les observations sur les variations et réalisations similaires et spécifiques à un groupe social, à une tranche d'âge, voire à un individu unique et isolé d'une époque déterminée, d'un autre côté.

Une formule significative $\varphi(\mathcal{L}_0)$ est le résultat d'une suite d'opérations appliquées dans un certain ordre et soumises à diverses contraintes. Ces opérations engendrent une telle formule à partir d'un axiome. L'ensemble des axiomes, opérations, règles d'application des opérations et contraintes constitue la *syntaxe* du $SR\mu$, chargée de caractériser les formules significatives.

Les $SR\mu$ et leurs syntaxes ne sont pas *a priori* isomorphes aux mécanismes cognitifs. En effet, chaque opération de la syntaxe est une construction du linguiste qui, par là, cherche à organiser — en la décrivant d'une façon cohérente — la pro-

²⁴ Il est clair que \mathcal{L}_0 est une variable. Nous préférons cependant le terme de paramètre pour deux raisons : (1) Le terme de variable est très ambigu comme en témoigne l'analyse de cette notion par H.B. Curry in *Combinatory Logic*, 1958, p. 40 à 57; (2) Nous visons en même temps que la construction d'un formalisme, l'interprétation théorique.

duction d'un énoncé à partir d'un axiome (toujours aux conditions d'ajustement entre formule significative et énonçable observable). L'opération est donc de nature métalinguistique, elle est un construit théorique, elle ne représente donc pas un processus langagier qui serait effectué par un quelconque énonciateur humain.

Supposons un énoncé E observable sous forme d'une suite de signes phoniques ou graphiques. Une représentation adéquate (r) de E dans un SR_μ énonciatif aboutit à une formule significative $\varphi (f_0)$ telle que cette formule se réalise (ρ) par l'énoncé E, c'est-à-dire telle que $\rho (\varphi (f_0)) = E$. Substituons à f_0 la constance d'individu a d'une part, et la constante d'individu b, d'autre part. Nous obtenons deux formules « concrètes » $\varphi (a)$ et $\varphi (b)$. Ces formules ont des organisations isomorphes. Désignons maintenant par $\psi (A)$, respectivement $\psi (B)$, les processus mis en œuvre par le sujet énonciateur humain A, respectivement B, associé à a, respectivement à b. Ces processus sont des enchaînements d'opérations ayant un support neuro-physiologique, c'est du moins une hypothèse raisonnable bien que dans l'état de nos connaissances nous soyons incapables de la valider ou même d'observer sérieusement quoi que ce soit dans ce domaine. On peut imaginer que les deux enchaînements d'opérations $\psi (A)$ et $\psi (B)$ soient distincts alors que les produits observables soient identifiables. Autrement dit, A et B adoptent des stratégies différentes pour produire la même chose. Ceci implique donc que les systèmes cognitifs de A et de B sont organisés différemment, tout en produisant (et comprenant) des classes d'objets identifiables.

On pose alors la relation d'équivalence :

$$\psi (A) \approx_1 \psi (B) \langle \overline{\text{def}} \rangle \varphi (a) \approx_2 \varphi (b) \approx_2 \varphi (f_0)$$

avec $\rho (\varphi (f_0)) = E$

sans que, par nécessité, on ait : $\psi (A) \equiv \psi (B)$ (les relations d'équivalence \approx_1 et \approx_2 sont différentes).

Un système métalinguistique (SR_μ) énonciatif est, par construction, distinct des systèmes cognitifs; il peut cependant entretenir des relations pertinentes avec eux, et servir ainsi d'instrument théorique à leur description. En effet, de la rela-

tion d'équivalence précédente on déduit qu'un système métalinguistique est — à un isomorphisme près — un représentant de systèmes cognitifs « équivalents », c'est-à-dire de systèmes qui semblent avoir les mêmes performances observables (production et reconnaissance d'une même classe d'énoncés).

L'approche précédente est compatible avec le problème des transformations d'un état cognitif vers un autre état cognitif, en interaction avec d'autres systèmes cognitifs (acquisition d'une langue première par un enfant en interaction avec le langage (ou les langages) adulte(s))²⁵.

En ce qui concerne la linguistique, la présence symbolique de l'origine \mathcal{L}_0 dans la représentation oblige le théoricien à formuler ses opérations par rapport à une origine, à les rendre donc *relatives* et à éviter ainsi de les croire absolues ou idéales; ceci permet alors d'étudier l'organisation et la combinaison des opérations entre elles et ensuite de comparer les systèmes métalinguistiques entre eux (voir plus bas l'hypothèse sur l'invariance des opérations).

III.2. La production d'un énoncé suppose donc un énonciateur. D'autres paramètres sont corrélés à cet énonciateur origine \mathcal{L}_0 .

Tout schéma de dialogue impose que l'échange s'établisse entre deux partenaires, entre l'énonciateur \mathcal{L}_0 et son co-énonciateur \mathcal{L}_1 . Ce co-énonciateur \mathcal{L}_1 est lui-aussi un paramètre qui est différent de \mathcal{L}_0 . Cette relation de différence a des traces textuelles qui apparaissent dans le jeu des personnes (*je/tu*) au fondement même de l'activité langagière²⁶.

La déformation systématique des conditions d'énonciation exhibe la classe des paramètres spatio-temporels dont

²⁵ « Cette période (...) transitoire entre langage enfantin et langage adulte a une durée très variable d'un sujet à un autre. Les tâtonnements de l'enfant peuvent entraîner plusieurs systèmes successifs, ou des systèmes chevauchants (...). L'enfant ne procède pas par *rectification* des formes, par correction des fautes, il s'agit d'une succession de langages (...) qui supposent chez l'enfant tout un travail d'élaboration qui nous est encore très mal connu, mais pour lequel il utilise indubitablement les matériaux fournis par la langue adulte ». (L. Lentin : *Apprendre à parler à l'enfant de moins de 6 ans où ? quand ? comment ?*, t. 1, Éditions E.S.F., p. 61).

²⁶ Voir [DES, 76 a] et [DES, 78 a].

l'origine est symbolisée par ℓ_0 . Le couple (ℓ_0, ℓ_0) constitue le « point de référence » origine. Une variation sur les paramètres spatio-temporels produit d'autres points de référence : $(\ell_0, T_1), \dots (\ell_0, T_i)$ qui seront repérés par rapport à (ℓ_0, ℓ_0) si l'on précise une fonction de repérage entre T_1, \dots, T_i, \dots et ℓ_0 .

Le repère ℓ_0 est l'index origine de l'acte d'énonciation produit par ℓ_0 . Si l'on a la relation d'identification $\langle T = \ell_0 \rangle$, alors tout ce qui aura pour index T sera concomitant à l'acte d'énonciation. Si l'on a la relation de différenciation $\langle T \neq \ell_0 \rangle$, alors tout ce qui aura pour index T sera non concomitant à l'acte d'énonciation. D'autres fonctions de repérage permettent d'engendrer les points de référence repérés par rapport à l'origine (ℓ_0, ℓ_0) . Ces points de référence repérés entre eux sont utilisés pour décrire les conditions d'énonciation, c'est-à-dire les coordonnées énonciatives de l'énoncé produit. Chaque déformation énonciative se manifeste par une trace textuelle repérable. Ainsi, le marqueur morphologique *-ait* de l'imparfait français apparaît comme la trace textuelle d'une non concomitance avec l'acte énonciatif. De même, le marqueur « a + participe passé (d'un verbe) » (dans *Jean a conduit la voiture au garage*, par exemple) apparaît comme la trace d'une non concomitance; la distinction entre l'imparfait et le passé composé (ou « parfait du présent ») décrite par des organisations distinctes de l'opération de non concomitance avec d'autres opérations (dont des opérations topologiques) que nous ne décrivons pas ici.

Les paramètres $\ell_0, T_1, \dots, T_i, \dots$ et les diverses fonctions de repérage sont utilisés pour fonder métalinguistiquement des catégories grammaticales comme le temps et l'aspect. Fonder de telles catégories, c'est montrer comment celles-ci se décrivent et s'organisent à l'aide d'entités jugées plus élémentaires et au nombre très restreint d'une part, ces entités étant généralisables à la description des diverses catégories grammaticales relatives au même phénomène de temps et d'aspect de plusieurs langues même éloignées génétiquement, d'autre part²⁷.

²⁷ Voir [DES, 77]. Nous traiterons de ces problèmes dans une communication présentée avec Z. Guentchéva au colloque international sur l'« Aspect » (Metz, mai 1978).

Si S désigne un sujet énonciateur quelconque, T un index spatio-temporel quelconque, le couple (S, T) désigne un « point de référence » qui servira à décrire les conditions d'énonciation sous forme de coordonnées énonciatives. Donnons quelques exemples très simples. Introduisons auparavant l'identité :

$$\langle S = f_1 \rangle \langle = \rangle \langle S \neq f_0 \rangle$$

c'est-à-dire : « S est identifié à f_1 si et seulement si S est différencié de f_0 ». (On est amené à poser également l'identité

$$\langle S = f_0 \rangle \langle = \rangle \langle S \neq f_1 \rangle.$$

Nous avons alors :

$\left\{ \begin{array}{l} je \\ tu \\ je \\ tu \end{array} \right.$	<i>mang</i>	$\left\{ \begin{array}{l} e \\ es \\ ais \\ ais \end{array} \right.$	↔	$\langle S = f_0 \rangle$	et	$\langle T = t_0 \rangle$
			↔	$\langle S \neq f_0 \rangle$	et	$\langle T = t_0 \rangle$
			↔	$\langle S = f_0 \rangle$	et	$\langle T \neq t_0 \rangle$
			↔	$\langle S \neq f_0 \rangle$	et	$\langle T \neq t_0 \rangle$

où l'on indique la correspondance entre les marqueurs grammaticaux et la combinaison de relations énonciatives engendrées par les opérateurs d'identification et de différenciation.)

Le point de référence (S, T) est l'argument de la situation énonciative symbolisée par Sit (S, T). Ainsi : Sit (f_0, t_0) représente la situation énonciative d'origine. Une situation énonciative peut être substituable à certains termes d'une relation prédicative.

Prenons les deux énoncés *Jean est à Paris* et *Jean est ici* où dans la relation prédicative *ici* et à *Paris* ont la même fonction de localisateur. Représentons les deux énoncés à l'aide du schéma prédicatif :

$$\langle \text{Jean est LOC} \rangle$$

que nous relient à la situation énonciative d'origine $Sit(\mathcal{L}_0, \ell_0)$. À LOC substituons soit *Paris*, soit $Sit(S, T)$. Dans le premier énoncé, le localisateur est déterminé indépendamment des conditions d'énonciation, donc de $Sit(\mathcal{L}_0, \ell_0)$; dans le second énoncé, le localisateur $Sit(S, T)$ est indéterminé et l'on doit introduire une détermination supplémentaire en repérant, par une fonction de repérage, le localisateur $Sit(S, T)$ par rapport à une constante, c'est-à-dire la situation énonciative d'origine, d'où la représentation :

$\langle \text{Jean est } (\text{à}) \text{ } Sit(S, T) \rangle$ et $\langle Sit(S, T) = Sit(\mathcal{L}_0, \ell_0) \rangle$

soit encore, d'une façon plus compacte :

$$\langle \text{Jean est } (\text{à}) \text{ } Sit(S, T) \rangle \quad \overset{=}{\text{---}} \quad \langle Sit(\mathcal{L}_0, \ell_0) \rangle$$

$Sit(S, T)$ est donc, dans ce cas, substituable au localisateur à *Paris* mais cette substitution doit être accompagnée d'une détermination supplémentaire. La formule précédente est glosable par « Jean est à un endroit qui est identifiable à l'endroit défini par l'énonciateur \mathcal{L}_0 en l'instant ℓ_0 de son énonciation ».

L'énoncé *Jean est là-bas* est représentable par

$$\langle \text{Jean est } (\text{à}) \text{ } Sit(S, T) \rangle \quad \overset{\neq}{\text{---}} \quad \langle Sit(\mathcal{L}_0, \ell_0) \rangle$$

glosable par : « -Jean est à un endroit qui est différent de l'endroit défini par l'énonciateur \mathcal{L}_0 en l'instant ℓ_0 de son énonciation ». La représentation et sa glose indiquent immédiatement que celles-ci sont insuffisantes puisque, spontanément, *là-bas* signifie « pas à l'endroit où se trouve \mathcal{L}_0 en ℓ_0 ($Sit(\mathcal{L}_0, \ell_0)$) mais dans un domaine pas trop éloigné (visible, par exemple, ou ayant une certaine détermination contextuelle) ». Cela impose que nous spécifions davantage la sémantique de l'opération de différenciation et le fonctionnement de la catégorie des situations énonciatives, d'autant plus que nous n'avons pas donné les contraintes entre les varia-

tions sur les points de référence et les variations sur les situations énonciatives²⁸.

Une systématisation des opérations d'identification et de différenciation (= et \neq) auxquelles nous ajoutons une troisième opération, dite de rupture (ω), organise les catégories métalinguistiques et énonciatives²⁹.

- des sujets énonciateurs centrés sur l'origine \mathcal{S}_0 ;
- des repères spatio-temporels centrés sur ℓ_0 ;
- des situations énonciatives centrées sur la situation d'origine Sit (\mathcal{S}_0, ℓ_0).

III.3. Chaque occurrence d'un énoncé d'une langue est produite par un sujet énonciateur origine \mathcal{S}_0 , à un « temps » donné ℓ_0 ; \mathcal{S}_0 et ℓ_0 sont arguments d'une situation énonciative Sit (\mathcal{S}_0, ℓ_0) à partir de laquelle la relation prédicative est repérée.

Les paramètres énonciatifs et les fonctions de repérage sont les constituants de base du référentiel énonciatif. Les opérations de repérage agissent sur les divers paramètres énonciatifs, elles engendrent, par itérations éventuelles, les conditions d'énonciation de l'énoncé. C'est l'énonciateur, à chaque énonciation, qui construit son espace énonciatif (qui n'est pas euclidien) et son « temps linguistique » (qui n'est pas absolu, unidimensionnel, linéaire et symétrique comme le temps physique). Le sujet est toujours au centre de sa production. Il ne se repère pas par rapport à un système référentiel qui lui serait externe. Le langage naturel a pourtant des procédures qui ancrent l'espace énonciatif à l'espace de la perception et à une chronologie externe et socialisée (calendrier, par exemple . . .).

²⁸ Voir pour un traitement systématique [DES, 78 a].

²⁹ L'opérateur de repérage ϵ a été introduit par A. Culioli à partir d'observations sur les langues. Soit le schéma de repérage $\langle X \epsilon Y \rangle$. Si la relation induite entre X et Y est réflexive et symétrique, on a une identification; si elle est non symétrique, on a une différenciation; si elle est non réflexive et symétrique, on obtient la valeur de rupture. De la valeur de différenciation on tire, en imposant certaines propriétés supplémentaires, d'autres valeurs (l'inclusion, l'appartenance intuitive, la localisation, le ϵ de l'ontologie de Lesniewski . . .) Cf. [DES, 77] et « Note sur l'opérateur ϵ de repérage » (à paraître).

On inverse le point de vue traditionnel. Alors que la logique s'est centrée, depuis vingt cinq siècles, sur l'étude des propriétés prédicatives avec lesquelles on tentait de décrire les phénomènes énonciatifs, la recherche s'effectue, dans une théorie de l'énonciation, à partir des relations énonciatives et des relations prédicatives, toutes deux jugées fondamentales à l'activité langagière.

Cela conduit au second principe (*principe de repérage*) : « chaque production linguistique est représentée par une formule métalinguistique φ (ℓ_0), centrée sur le sujet énonciateur ℓ_0 , telle que la relation prédicative est repérée — directement ou indirectement — par rapport à la situation énonciative d'origine $\text{Sit}(\ell_0, \ell_0)$ ».

La fonction de repérage est définie à l'aide d'un opérateur primitif (au double sens de premier et de simple) de la théorie symbolisé par ϵ . E désigne un énoncé linguistique; r (resp. ρ) désigne(nt) les procédures de représentation (resp. réalisation); X est une expression contenant la relation prédicative; une relation de déduction (dérivation), sur l'ensemble des formules métalinguistiques d'un $\text{SR}\mu$ énonciatif, est symbolisée par \leftarrow . Le principe précédent s'exprime par :

$$E \begin{matrix} \xrightarrow{r} \\ \xleftarrow{\rho} \end{matrix} \varphi(\ell_0) \leftarrow \langle X\epsilon \text{Sit}(\ell_0, \ell_0) \rangle$$

À titre d'exemple, l'énoncé de III.2. est représenté (en première approximation) par :

$$\varphi(\ell_0) \equiv \text{Jean est } \overbrace{\text{Sit}(S, T)} \text{ } \text{Sit}(\ell_0, \ell_0)$$

en omettant de représenter comment est construit le nom propre *Jean* et la valeur de localisation du marqueur *est*³⁰.

III.4. La théorie du langage et des langues que A. Culioli et son équipe envisagent, doit répondre à quatre questions essentielles :

³⁰ Sur un exposé plus systématique des principes et hypothèses, voir [CULI, 68], [CULI, 73], [DES, 78 b] et [DEs, 77].

(1) Comment sont organisées les catégories linguistiques de chaque langue ? Existe-t-il des catégories langagières plus fondamentales qui seraient constitutives des catégories linguistiques ?

(2) Par rapport à la diversité des langues, y a-t-il des invariants grammaticaux qui permettraient de transférer un système de catégories linguistiques d'une langue dans un autre système d'une autre langue ? Quelles sont les procédures qui exhibent ces invariants (s'ils existent) ? Comment les représenter et les manipuler ?

(3) Étant donné un énoncé, quelles sont les opérations qui constituent la famille paradigmatique des énoncés « équivalents » reliés entre eux par une relation d'équivalence (paraphrase) ? Quelles sont les opérations qui déforment grammaticalement un énoncé d'une famille paraphrastique en une autre famille paradigmatique d'énoncés reliés entre eux par des relations de préordre (déformations grammaticales) ? Quelles sont les valeurs d'un énoncé (ambiguïté), comment les décrire et les relier à la paraphrase ?

(4) Comment un énonciateur, en situation significative, constitue-t-il son énoncé en le produisant de telle façon que son co-énonciateur puisse le comprendre en l'interprétant ?

L'explication des termes utilisés dans la formulation de ces quatre questions nécessiterait quelques développements. Nous ne le ferons pas ici.

Donnons une définition utile pour la suite :

Une catégorie grammaticale est, selon nous, une relation entre d'une part, un système d'opérateurs et d'opérations engendrant les notions grammaticales, et, d'autre part, un système classificatoire composé de classes d'équivalence de marqueurs morpho-syntaxiques. La discussion sur la recherche des invariants grammaticaux débouche sur l'hypothèse dite de *l'invariance des opérations* formulée ainsi :

« ce qui varie de langue à langue, ce ne sont pas les opérations et opérateurs élémentaires, véritables invariants constitutifs du langage, mais l'agencement précis de ces opérations — et les contraintes sur cet agencement. Ces opérations, agencées selon un ordre précis, engendrent les

catégories linguistiques et les formules métalinguistiques d'où l'on dérive les énoncés d'une langue »³¹.

De cette hypothèse sur l'invariance, on en déduit que ce ne sont pas les systèmes de représentations métalinguistiques (SR_{μ}) qui sont invariants mais les opérateurs, les opérations élémentaires et les catégories métalinguistiques. Les SR_{μ} varient avec les langues et même avec les sujets énonciateurs. Il n'existe donc pas un système de représentation universel, obtenu par généralisation et induction/déduction, qui serait utilisé pour représenter n'importe quelle langue. Chaque langue L est représentable dans un système de représentation ($SR_{\mu}(L)$) constitué des mêmes constituants invariants.

IV. Quelles sont les conséquences de cette introduction du sujet énonciateur dans le système de représentations ? Elles sont nombreuses car il s'agit en quelque sorte d'un « renversement de perspective » qui, lorsqu'on poursuit la réflexion, oblige à réexaminer certaines des distinctions les mieux établies. Nous allons en citer quelques-unes.

IV.1. La distinction sémiotique syntaxe/sémantique/pragmatique a du mal à tenir sous la forme « classique ». En effet, nous nous appuyons sur certaines relations qualifiées de pragmatiques — les relations intersujets — pour construire le système des représentations. Les aspects pragmatiques sont inclus dans le système et cela dès sa fondation.

Le langage est un système de représentation pragmatique qui assume plusieurs fonctions, la fonction de communication n'étant qu'une fonction parmi d'autres, les systèmes métalinguistiques doivent représenter ces fonctions pragmatiques. Chaque SR énonciatif est un formalisme destiné à représenter adéquatement la signification des catégories grammaticales (personnes, temps, aspects, modalités, déictiques, détermination, voix, ...). Ce formalisme est ajusté à un système de représentations et de catégorisation

³¹ Cette hypothèse a été introduite dans [CULI, DES; 78]; elle est précisée dans [DES, 77]. Pour prendre une image, les opérateurs et opérations élémentaires sont les analogues des acides aminés et bases nucléiques communs à tous les êtres vivants. L'assemblage de ces matériaux en macromolécules spécifiques, protéines ou acides nucléiques, propres à chaque espèce est essentiellement une organisation de séquences spécifiques complexes.

sémantique des propriétés physico-culturelles où chaque objet de ce système est une notion prédicative (prédicable), non encore catégorisée (ni en « nom », ni en « verbe »), qui selon les opérations qui porteront sur elle, sera catégorisée en « nom », en « verbe » ou en tout autre élément d'une classe linguistique. Chaque SR_{μ} énonciatif vise surtout à une représentation des valeurs, des catégories grammaticales et, de ce fait, peut être considéré comme une sémantique du système classificatoire qui organise les catégories linguistiques mises en évidence par des procédures taxinomiques. Étant donné une langue L, le système classificatoire constituerait la syntaxe de L (position traditionnelle et « en gros » classique); le système métalinguistique énonciatif SR_{μ} constituant la sémantique du système précédent serait plongé dans un système plus vaste qui serait le système sémantique de la langue L.

Les systèmes de représentations SR_{μ} énonciatifs sont aussi des systèmes syntaxiques puisqu'ils sont définis et organisés par des règles qui opèrent sur des symboles sans avoir recours — en principe — à la signification de ces symboles. Les règles et contraintes sur l'application et l'enchaînement des règles ont pour but d'engendrer toutes les formules significatives et rien d'autre. De plus, chaque formule significative peut être reliée, au moyen de règles de correspondance explicites, directement à un énonçable sans passer nécessairement par un système de classes taxinomiques.

On éprouve donc un certain embarras à classer les SR_{μ} énonciatifs. Sont-ils pragmatiques ? sémantiques ? syntaxiques ?

J'ai proposé le qualificatif de « pragmatico-syntaxique ». Ces SR_{μ} énonciatifs cherchent — nous l'avons dit — à représenter la signification des catégories grammaticales, à l'exclusion des propriétés des termes lexicaux, bien qu'on ne puisse pas, dans une analyse qui s'affine, séparer les catégories grammaticales des catégorisations sémantiques du lexique (appelées « relations primitives » par A. Culioli)³². Cette représentation fait d'autre part intervenir les utilisateurs du système représenté (l'énonciateur S_0 et son

³² Voir [CULI, 68] et [CULI, 71].

co-énonciateur (f_1), elle permet donc de représenter certaines des relations intersujets et constitue ainsi une tentative de formalisation de la pragmatique. Ce qualificatif de « pragmatico-syntaxique » ne convient pas car, en fait, c'est toute la tripartition syntaxe/sémantique/pragmatique qui est en cause.

IV.2. Une très longue argumentation montrerait qu'il est absolument nécessaire de distinguer *valeur référentielle* de *référence*. Nous n'allons pas en donner une définition mais poser un problème : peut-on distinguer, et comment, le nom propre qui renvoie à un individu singulier des termes *je* et *tu* qui eux aussi renvoient à un individu singulier ? Comment différencier le nom propre du pronom *il* et ce dernier de *je* et *tu* ?

Prenons les énoncés (1) *Jean vient*; (2) *Je viens*, (3) *Il vient*. Ces trois chaînes, au niveau distributionnel, sont équivalentes : ce sont trois phrases *je* est substituable à *Jean* (en travaillant sur les représentations phonétiques qui ne distinguent pas *viens* de *vient*). Une analyse plus fine montre des différences essentielles³³.

(1) est un énoncé bien formé, résultant d'une prédication. *Jean*, le prédiqué, est un *nom propre*, c'est-à-dire un objet nominal qui désigne un individu unique. Cet individu désigné nominalement n'a pas une existence extralinguistique attestable par la seule énonciation de (1), il n'a qu'une existence nominale. On peut, en effet, énoncer (1') *Ulysse arrive* ou (1'') *Pégasse arrive* sans que l'énonciation de (1') ou (1'') suffise à prouver l'existence perceptible qui entraînerait un consensus sur la « réalité existentielle » d'Ulysse ou de Pégasse. Par un nom propre, on peut désigner un individu qui n'a pas d'autre existence que l'existence nominale.

On dira que l'énoncé (1) a une *valeur référentielle* sans avoir nécessairement une référence. En interprétant (1), l'énonciateur identifie le nom propre et sait que celui-ci désigne un unique individu et donc que l'énoncé a une valeur bien définie.

(2) est un énoncé bien formé, résultat d'une prédication où le sujet syntaxique *je* est déterminé par les conditions

³³ Voir aussi [DES, 76 a].

d'énonciation. L'énonciation de cette occurrence de *je* renvoie à une opération d'identification entre le sujet de l'énoncé (2) et le sujet de l'énonciation ℓ_o . *Je* est la trace textuelle de cette opération et prouve l'existence de l'énonciateur.

On dira que l'énoncé (2) a une *valeur référentielle* et une *référence* puisque *je* est le témoin de l'existence d'un énonciateur concret qui énonce (2).

(3) est une phrase mais n'est pas un énoncé. (3) est bien formé grammaticalement mais, sans autres informations, (3) a une infinité d'interprétations et donc (3) ne signifie rien. L'occurrence de *il* n'est pas déterminée et renvoie à une classe infinie d'individus susceptibles d'être le sujet de la prédication³⁴.

Nous dirons que (3) est un énonçable qui deviendra énoncé dès que l'on spécifie les conditions de son énonciation soit par une détermination anaphorique ((3') *Pierre, il vient*), soit par une détermination énonciative ((3'') [celui qui est désigné par un geste ou par une intercompréhension mutuelle à l'intérieur du dialogue] *il vient*). La chaîne n'a donc ni *valeur référentielle*, ni *référence*.

Comment, dans la représentation métalinguistique, faire apparaître ces différences ? On représente respectivement (1), (2) et (3) par :

$$(I) \quad \langle\langle J \epsilon () v \rangle \epsilon \text{Sit} (\ell_o, \ell_o) \rangle$$

$$(II) \quad \langle\langle S \epsilon () v \rangle \epsilon \text{Sit} (\ell_o, \ell_o) \rangle$$

$$(III) \quad \langle\langle () \epsilon () v \rangle \epsilon \text{Sit} (\ell_o, \ell_o) \rangle$$

où *J* symbolise *Jean*, $\epsilon () v$ la propriété attributive « être en train de venir », $\text{Sit} (\ell_o, \ell_o)$ la situation énonciative d'origine.

Dans (I), la relation prédicative est saturée et déterminée indépendamment des conditions d'énonciation.

Dans (II), la relation prédicative est saturée mais indéterminée; une détermination supplémentaire est établie par la relation $\langle S = \ell_o \rangle$, représentée par le lien d'identification

$$\langle S \quad \ell_o \rangle$$

³⁴ C'est pourquoi *il vient* est un énonçable et non un énoncé.

Dans (III), la parenthèse vide en position sujet syntaxique (ou sujet de l'énoncé) montre que la relation prédicative est non saturée, c'est une forme propositionnelle (au sens de G. Frege), elle est donc indéterminée. L'énoncé (3') serait en revanche représenté par :

$$(III') \langle \langle \overbrace{P \ \epsilon \ \langle () \ \epsilon \ () \ v \rangle}^{\text{=}} \rangle \ \epsilon \text{ Sit } (\ell_o, \ell_o) \rangle$$

où le lien anaphorique introduit une détermination supplémentaire qui permet de saturer la relation prédicative $\langle () \ \epsilon \ () \ v \rangle$.

Le système des valeurs référentielles constitue un *modèle* du système des représentations métalinguistiques. Le problème de la référence (et des valeurs de vérités) consiste à établir une relation entre, d'un côté, les systèmes des représentations et des valeurs référentielles et, de l'autre, un système de référence externe et distinct des deux premiers. Alors qu'un nom a pour valeur référentielle une classe (distributive ou collective), un énoncé a pour valeur référentielle un « événement » ou un « état de choses » au sens de Wittgenstein, constitué à partir d'une relation entre « l'événement élémentaire » auquel renvoie la relation prédicative et « l'acte énonciatif ». Le système des valeurs référentielles est donc un espace référentiel construit par l'énonciateur : chaque « événement élémentaire » associé à la relation prédicative est repéré par rapport aux coordonnées énonciatives définies par les conditions d'énonciation³⁵.

La non distinction entre « valeur référentielle » et « référence » conduit la linguistique à poser des problèmes qui paraissent insolubles, mais nous entrons ici dans une discussion qui ne peut se régler en quelques lignes. Disons qu'il y a peut-être deux points de vue duaux et complémentaires, celui de la linguistique qui s'intéresse à la bonne constitution de l'énoncé et celui de la logique qui vise à la maîtrise consciente de la « valeur de vérité ». « Avant l'entrée en scène de la logique formalisée des énoncés,

³⁵ Un énoncé « renvoie » donc non à une « valeur de vérité » mais à un « événement » construit par l'énonciateur. Cet « événement » est le résultat d'une mise en relation de deux « événements » plus élémentaires.

l'énoncé est vrai ou il est faux; à partir de Frege la logique mathématisée dira de plus en plus coutumièrement que l'énoncé a la valeur « vrai » ou qu'il a la valeur « faux ». Le déplacement du langage atteste le déplacement de l'optique de la pensée logicienne — passant du point de vue « linguiste » au point de vue « logicien réfléchissant » à propos des énoncés, tout comme la chose achève de se faire avec la génération de Boole à propos des termes nominaux ou classes ». (D. Dubarle : *Logos et Formalisation du langage*, Klincksieck, 1977, p. 160). Séparer radicalement les deux points de vue, en affirmant que la linguistique n'a rien à voir avec la « valeur de vérité » c'est mutiler toute étude de l'énonciation; confondre les points de vue en suivant jusqu'au bout le logicien Frege pour qui l'énoncé est une nomination d'une valeur de vérité, c'est fausser le problème linguistique de la constitution de l'énoncé par l'énonciateur. Ces problèmes sont délicats. Aussi allons-nous les quitter.

V. Nous avons proposé deux principes (principe de l'énonciation et principe du repérage) qui, par définition, ne sont pas validables directement. Seules les conséquences justifient le bien fondé de ces décisions. Nous avons amorcé (en II) une argumentation en discutant du signe puis de la construction d'un espace référentiel par l'énonciateur. D'autres arguments pourraient s'ajouter mais jamais on ne « prouvera » ou « infirmera » ces principes par une simple confrontation avec les données linguistiques. L'hypothèse sur l'invariance des opérations n'est déjà plus un principe sans lequel aucune construction théorique n'est envisageable. Sa formulation résulte d'une convergence de : (1°) une série d'observations sur le fonctionnement des catégories linguistiques de diverses langues et sur l'articulation et l'organisation des discours des linguistes qui tentent de décrire les langues; (2°) une argumentation déductible tirant les conséquences nécessaires des principes théoriques mentionnés, conformément aux exigences méthodologiques que l'on s'impose en vue de répondre aux questions essentielles sur « le langage appréhendé à travers la diversité des langues ». D'autres hypothèses sont possibles *a priori*. Par exemple, l'irréductibilité des langues : chaque langue est unique, les seuls systèmes de représentations qui conviennent sont des systèmes classificatoires; mais alors comment expliquer le

phénomène de traduction ? Comment expliquer raisonnablement « la faculté de langage » : tout homme placé dans un environnement socialisé apprend à parler et sait manipuler son langage ? Ou encore : existence postulée d'un système de représentations métalinguistique universel assurant le rôle d'un génotype reconstruit qui se manifeste, en tenant compte de l'environnement dans les diverses langues phénotypes, chaque langue reflétant ainsi la structure fondamentale du génotype; mais alors comment expliquer la diversité des langues qui présentent dans l'organisation des énoncés et des catégorisations grammaticales des différences si difficiles à décrire ?³⁶

On peut raffiner l'hypothèse sur l'invariance en précisant davantage mais devant le paradoxe du langage, faculté humaine/diversité des langues, il faut choisir : y a-t-il des universaux ? comment les reconstruire ? comment les découvrir ? y a-t-il des invariants ? comment les représenter ?

Formuler une hypothèse sur l'invariance, c'est non seulement montrer qu'elle est compatible avec l'échantillon plus ou moins vaste des observations recueillies (il n'existe pas d'observations exhaustives en linguistique), non seulement veiller à la cohérence, mais aussi prévoir une procédure de validation qui permettra ou d'intégrer cette hypothèse à la construction théorique ou de lui enlever son caractère de plausibilité. Rappelons qu'une théorie à base empirique n'est jamais vraie ou fausse « en soi », mais que de toute façon elle doit être réfutable sinon elle n'a aucun intérêt.

Comment valider une hypothèse et plus généralement une théorie ?

La linguistique n'a peut-être pas assez réfléchi aux conditions d'observations qui lui sont propres, mais n'a pratiquement pas, ou très peu, envisagé le problème de l'expérimentation. Celle-ci oblige à non seulement représenter un texte dans un langage formalisé mais aussi à « revenir » aux domaines d'entités observables (réversibilité représentation/-

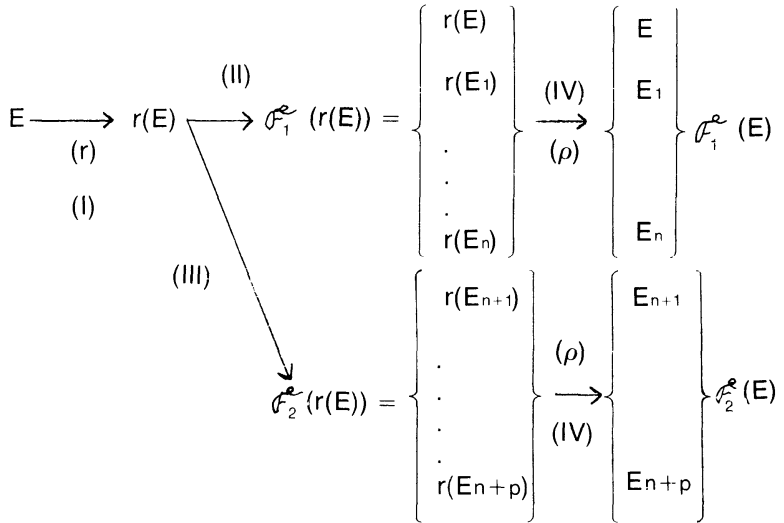
³⁶ Le langage génotype de S.K. Šaumjan est un exemple de métalangage de description « universel ». Sa pertinence théorique et descriptive est discutée dans Z. Guentchéva : « *Présentation critique du Modèle applicatif de S.K. Šaumjan*, Dunod, 1976.

réalisation). Il faut donc se donner à la fois une procédure de représentation (r) et une procédure de réalisation (ρ).

Afin de renforcer la cohérence de la théorie et de mieux répondre au critère de substitution du manipulateur (ou critère de reproductibilité), nous concevons une expérimentation en linguistique comme une *production automatique d'énoncés artificiels* comparables à des énoncés naturels produits en situation signifiante. L'ordinateur peut jouer ce rôle de synthétiseur. De même qu'en chimie on sait synthétiser certaines macro-molécules (on fabrique également des diamants artificiels ayant des propriétés comparables à celles du diamant), il est naturel de vouloir, à partir des constituants théoriques — opérateurs et opérations élémentaires —, produire en « simulant », la production, par un sujet énonciateur paramétré ρ , d'un énoncé « in vitro ». La production artificielle d'énoncés isolés les uns des autres (on parle dans ce cas de génération) risque de ne pas apprendre grand-chose. Aussi cherche-t-on non seulement à produire chaque occurrence d'un énoncé artificiel mais à produire toute la famille paraphrastique associée à un énoncé et puis à déformer grammaticalement l'énoncé en jouant sur certains paramètres de façon à obtenir toute une famille d'énoncés plus ou moins équivalents ou déformés³⁷.

³⁷ Voir [BES, DES, 76], [BES, DES; 78], [DES, 78 b]. Nous avons conduit l'expérimentation sur un problème particulier : J.P. Desclés : « Quelques opérations de localisation » (PITFALL, n° 15, 16, 17, 18, 1975, *Laboratoire de Linguistique Formelle*, juin 1977).

Le schéma complet de l'expérimentation est alors le suivant :



Soit E, un énoncé d'une langue naturelle L; une procédure algorithmique (I) permet de représenter E dans un langage formalisé L sous forme d'une formule significative. Un programme (II) issu d'une grammaire de paraphrase engendre une famille $\mathcal{P}_1^e(r(E))$ de formules significatives « équivalentes » à r(E). Un autre programme (III) engendre une autre famille $\mathcal{P}_2^e(r(E))$ obtenue en déformant certaines des relations entre les paramètres qui occurred dans r(E). Une procédure algorithmique (IV) permet de réaliser explicitement chacune des formules engendrées par un « énoncé artificiellement produit ». Ce qui est commun aux familles $\mathcal{P}_1^e(r(E))$ et $\mathcal{P}_2^e(r(E))$ (« paraphrases » de E et « déformations » de E) c'est un même contenu prédictif λ qui reste invariant sous toutes les transformations. Ce contenu prédictif peut être pris comme « entrée » des programmes (II) ou (III). On peut également partir d'un énoncé E d'une langue L₁, le représenter dans un système SR_μ(L₁) par r₁(E) puis faire un *transfert*, quand c'est possible, de r₁(E) de SR_μ(L₁) vers SR_μ(L₂) d'une autre langue L₂, pour produire des familles d'énoncés artificiels dans une langue L₂. On voit ainsi comment la paraphrase ne se ramène

pas à un cas particulier de la traduction mais aussi comment la paraphrase et la traduction sont étroitement liées³⁸.

Les programmes sont donc les véritables instruments d'expérimentation construits à partir de la théorie. Réaliser un protocole expérimental de ce genre, c'est écrire les programmes informatiques qui permettent à l'ordinateur de jouer son rôle de synthétiseur et de substitut d'énonciateur. C'est l'objectif de notre actuel « Programme Interdisciplinaire de Traitement Formel et Automatique des langues et du Langage (PITFALL)³⁹.

« Rien de plus anticartésien que la lente modification spirituelle des approximations successives de l'expérience, surtout quand les approximations plus poussées révèlent des richesses organiques méconnues par l'information première » (G. Bachelard, *Le nouvel esprit scientifique*).

BIBLIOGRAPHIE

- [BES, DES, 76] H. BESTOUGEFF, J.P. DESCLES. « L'informatique comme moyen d'expérimentation en linguistique » — *Méthodes de validation et d'analyse des données textuelles*, éd. M. Borillo, éditions du CNRS, 1976.
- [BES, DES, 78] H. BESTOUGEFF, J.P. DESCLES. « Programme de recherche de Pitfall » — *Recueil d'articles sur le thème utilisation de l'informatique pour le traitement des langues naturelles* ed. A. Nariniani — Centre de calcul de Novosibirsk — 1978.
- [CULI, 68] A. CULIOLI. « La formalisation en linguistique ». *Cahiers pour l'analyse*, n° 9, Seuil, Paris 1968.
- [CULI, 71] A. CULIOLI. « À propos d'opérations intervenant dans le traitement des langues naturelles » — *Mathématiques et Sciences Humaines*, n° 34, 1971.

³⁸ Le contenu prédicatif ou « lexis » a une forme canonique « à trois places » ([CULI, 68] et [CULI, 71]). Il correspond au *GEDANKE* de Frege. Sur le transfert, voir [DES, 77].

³⁹ Voir [BES, DES, 78], [DES, 78 b], [DGR, 77], [CNR, 77]. Le projet PITFALL est mené dans le cadre de l'ERA 642 du CNRS.

- [CULI, 73] A. CULIOLI. « Sur quelques contradictions en linguistique » — *Communications*, n° 20, Seuil, Paris, 1973.
- [CULI, 76] A. CULIOLI. « Comment tenter de construire un modèle logique adéquat à la description des langues naturelles ». *Modèles logiques et niveaux d'analyse linguistique*, Klincksieck, 1976.
- [CULI, 77] A. CULIOLI. « Note sur la « détermination » et « quantification » définition des opérations d'extraction et de fléchage » — Rapport ERA 642 au CNRS; Juin 1977.
- [CULI, DES, 78] A. CULIOLI, J.P. DESCLES. « Considérations sur un programme de traitement automatique du langage » — *Informatique et Sciences humaines* — U.G.E. (à paraître en 1978).
- [DES, 76 a] J.P. DESCLES. « Description de quelques opérations énonciatives » — *Modèles logiques et niveaux d'analyse linguistique*, Klincksieck, 1976.
- [DES, 76 b] J.P. DESCLES. « Opérations sur des opérateurs typés ». *C.R.Acad.Sci. Paris*, t. 283 — série A, 987; « Transformations d'opérations et de multiopérations » *C.R.Acad.Sci. Paris*, t. 283, Série A, 1045.
- [DES, 77] J.P. DESCLES. « Problème du transfert des catégories », Colloque franco-soviétique sur les aspects pratiques de la traduction automatique, Moscou, décembre 1977, (à paraître).
- [DES, 78 a] J.P. DESCLES. « Représentation formelle de quelques déictiques français » — *Convegno S.L.I. : Logishe, Calcoli, Formalizzazioni e lingua Storico-Naturali*, Catania, septembre 1976, à paraître en 1978.
- [DES, 78 b] J.P. DESCLES. « Présentation du programme PITFALL », *Convegno S.L.I. : Logishe, Calcoli, Formalizzazioni e lingua Storico-Naturali*, Catania, septembre 1976, à paraître en 1978.
- [DES, GUE, 77] J.P. DESCLES, Z. GUENTCHEVA. « Métalangue, métalangage, métalinguistique » Publication Centro Internazionale di Semiotica e Linguistica — Università di Urbino — Urbino, 1977.

[DGR, 77] Rapport final DGRST (PITFALL) contrat 74.7.0341),
Janvier 1977.

[CNR, 77] Rapport ERA 642 au CNRS, Juin 1977.