

A. CULIOLI

J. P. DESCLÉS

**Traitement formel des langues naturelles. Première partie :
mise en place des concepts à partir d'exemples**

Mathématiques et sciences humaines, tome 77 (1982), p. 93-125

http://www.numdam.org/item?id=MSH_1982__77__93_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1982, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

TRAITEMENT FORMEL DES LANGUES NATURELLES

Première partie :

Mise en place des concepts à partir d'exemples .

A. Culioli*, J.P. Desclés**

AVERTISSEMENT

Le texte qui suit n'est pas un article mathématique, Il s'agit de donner un exemple où les concepts théoriques émergent du traitement formel des données linguistiques empruntées à plusieurs langues. Souvent, derrière la diversité parfois déroutante des observations, il apparaît des schémas organisateurs identiques. Seule une analyse abstraite autorise de tels rapprochements et dégage petit à petit quelques opérations invariantes, la diversité - dans le cas des exemples traités ici - provient alors de différentes "mises en ordre linéaire" (permutation, déplacement de marqueurs morpho-prosodiques...) alors que l'agencement fondamental des relations constitutives est assuré par un ensemble stable d'opérations et de macro-règles d'enchaînement, aussi cet agencement abstrait reste-t-il invariant par rapport aux manifestations qui se modifient en des organisations par concaténations différentes selon les langues, d'où l'hétérogénéité apparente.

L'analyse distributionnelle ou toute autre analyse analogue, y compris la plupart des analyses transformationnelles, est incapable de sortir du magma des observations brutes des structurations invariantes et forcément abstraites. En effet, dans ces analyses, les principales manipulations autorisées sont des opérations de substitution d'un élément plus complexe - une chaîne contiguë ou non de morphèmes - à un élément plus simple (par exemple des syntagmes), le critère étant que les occurrences des éléments substitués restent équivalentes lorsqu'elles sont combinées avec les autres termes de la phrase.

* Dpt. de rech. Linguistiques. ERA 642. Univ. Paris 7.

** UER de Mathématiques. ERA 642. Univ. Paris 7.

On comprend très bien que les phénomènes d'imbrication (d'une relative dans un syntagme, par exemple) aient joué un rôle de catalyseur théorique dans la typologie des grammaires syntagmatiques, par essence taxinomiques. Même si les imbrications sont récursives - ce qui resterait à prouver du point de vue d'une adéquation avec les observations empiriques, sauf si l'on fait appel au couple conceptuel chomskyen : compétence / performance qui réconcilliera toujours le modèle théorique avec les observations effectives - les représentations métalinguistiques sont toujours structurables en "arbres" engendrés par des grammaires formelles d'un certain type, en fait les grammaires de type 2 (ou non contextuelles) dans la hiérarchie de Chomsky.

Ce type d'organisation par substitution de chaînes de morphèmes concaténés n'a cependant aucune chance de faire émerger des invariants structurels. Il en résulte une vision classificatoire propre aux observateurs. Cette vision se justifie lorsque une langue est décrite pour elle-même mais la procédure classificatoire reste fondamentalement inadéquate lorsque l'on désire dresser des typologies syntaxiques. Pour s'en convaincre, il suffit de se rappeler les échecs de ces dernières années. C'étaient pour la plupart des projets typologiques et contrastifs. Signalons, entre autres, l'ouvrage édité par Charles N. Li : Subject and Topic, Academic Press, 1975, pour se rendre compte de l'importance attachée à la notion de "thème" et des critiques adressées à celle de sujet.

La plupart des contributions des auteurs au livre signalé ont montré que l'opposition sujet/prédicat n'avait aucune chance d'être universelle, d'autant plus que le terme de "sujet" se prêtait à de multiples interprétations et qualifications (sujet syntaxique, profond, logique, psychologique, théétique etc...). Aussi, la littérature a-t-elle été submergée par des oppositions comme topic/comment, thème/rhème... destinées à mieux expliquer la constitution progressive de l'énoncé alors que l'analyse n'exprimait au mieux que la décomposition classificatoire des termes de la phrase toute constituée. Mais les définitions qui ont été proposées ("information nouvelle opposée à l'information ancienne", "terme ajouté à ce qui est déjà connu", "terme posé à propos duquel on en dit quelque chose") ne présentent pas cette simplicité qui les rendraient transparentes par tout un chacun, elles ne débouchent pas non plus sur des définitions opérationnelles validables par des traitements informatiques finalisés. Pourtant, cette classe de phénomènes linguistiques est pleinement représentative de l'activité langagière observée dans de nombreuses langues orales mais aussi écrites.

Les définitions proposées par la plupart des auteurs ne se coulent pas dans des écritures métalinguistiques qui représenteraient, par les notations employées, le mécanisme de la constitution de l'énoncé effectuée en posant d'abord un terme (thème) pour ensuite lui attribuer des propriétés rhématiques. Chaque acte qui "pose un thème" est le résultat d'une opération mettant en jeu un énonciateur à un instant pris pour origine de sa prédication. En effet, le thème est posé symboliquement par rapport à l'énonciateur, ce qu'attestent des expressions comme :soit, il y a, quant à, en ce qui concerne... Il faut alors inventer des écritures formelles dans lesquelles on puisse formuler des opérations de thématization. Or, l'analyse des agencements par thématization (à divers degrés, du reste) fait apparaître des modes d'organisation qui diffèrent de ceux qui sont obtenus par simple concaténation (observable directement au niveau linguistique), ou par agencement markovien (avec les langages de type 3 ou "à nombre fini d'états") ou par substitution (dans une analyse classificatoire des termes constituants de la phrase).

Un mode plus complexe d'agencement est apparu, au fil des observations et des analyses, qualifié de mode par intrication. En effet, un terme étant "posé" par l'acte énonciatif, ce terme est ensuite inséré dans d'autres relations où il est intriqué. Beaucoup de procédés linguistiques, comme l'anaphore, l'utilisation de pauses prosodiques et d'actualisateurs comme il y a en français, variables avec les langues, sont utilisés pour maintenir les liens de cohésion entre les termes "posés" et leur insertion dans des relations prédictives; ce sont là les traces repérables de cette opération d'agencement par intrication.

Après un examen des différents types de données linguistiques et les traitements qu'elles imposent, l'article vise à montrer comment se présentent au mathématicien intéressé par le traitement formel des langues, les concepts linguistiques qu'il lui faudra épurer sans pour autant les dénaturer. Ce mathématicien est ensuite confronté aux systèmes de représentations métalinguistiques qu'il doit "fabriquer" et enrichir sous le contrôle du linguiste pour leur adéquation. Les systèmes de représentation "classiques", en particulier le "calcul des prédicats" sont inadéquats pour exprimer formellement les diverses thématizations. La nature de cette inadéquation se comprend dès que l'on rappelle les finalités qui les ont motivé, à savoir l'analyse des fondements de l'arithmétique essentiellement; ils ne sont pas destinés à être des instruments d'analyse des langues. On pourrait citer de nombreux auteurs qui, pour remédier à l'inadéquation du

Calcul des Prédicats "inventent" des étiquettes, mais celle-ci ne sont pas constitutives d'un langage représentatif "bien fait" et adéquat à la description de phénomènes comme la thématisation ou les relations casuelles. Que ce soient G. Frege ou B. Russell, H.B. Curry et surtout A. Tarski, tous ces logiciens établissent des barrières entre les langages formalisés logiques et artificiels et les langues naturelles. La mise en place de systèmes métalinguistiques formels s'effectue par tâtonnements successifs. En effet, un "bon" système de représentation doit remplir au moins quatre conditions :

- 1) représenter les phénomènes utilisés ;
- 2) être manipulable par tout un chacun et être transparent ;
- 3) être cohérent ;
- 4) être fécond et doué d'un pouvoir heuristique.

Le système de représentation présenté dans l'article répond, du moins dans ses intentions, aux quatre contraintes. Il est destiné à noter les diverses opérations de thématisation. Sa transparence n'est peut-être pas, dans la présentation retenue ici, évidente. Le système métalinguistique présenté ici s'inspire pourtant beaucoup plus des notations frégréennes que des écritures russelliennes, en ne retenant pas notamment l'usage de la variable du Calcul des Prédicats présenté dans les ouvrages de Church ou de Quine. La notation de variable, libre ou liée, est ambiguë. Certes pour traduire les énoncés arithmétiques ou pour écrire des équations et des expressions du calcul différentiel et intégral ou exprimer les identités remarquables, l'ambiguïté est généralement levée par le lecteur sans problème insurmontable mais pour fonder un système représentatif, il faut non seulement dominer mais aussi chasser et éviter ce genre d'ambiguïté. Ce fut la tentative opérée par Schönfinkel et développée pour le programme de la Logique Combinatoire de H.B. Curry et R. Feys : construire un système formel "sans" variables (encore faudrait-il s'entendre sur "sans variable" !)

La Logique Combinatoire et le λ -calcul de A. Church, légèrement différent, sont utilisés maintenant pour une description sémantique des langages de programmation, mais également pour représenter les langues naturelles dans le Modèle Applicationnel de S. K. Shaumyan par exemple (Logique Combinatoire) ou dans la Grammaire Universelle de R. Montague et B. Partee (λ -calcul). La non-transparence immédiate des systèmes esquissés n'est souvent qu'un manque d'habitude, elle ne doit pas non plus surprendre puisqu'il s'agit d'un système en construction et non d'un système figé présenté selon les canons traditionnels des manuels logiques. Le lecteur mathématicien reconnaîtra au fil de l'exposé la trame du système formel sous-jacent au discours.

La recherche de la cohérence - non contradictoire mais aussi l'effectivité opératoire - appartient de plein droit aux mathématiciens et logiciens voire aux informaticiens. Traquer l'incohérence, les paradoxes, les pseudo-définitions, les erreurs de calcul, les imprécisions...mais en même temps déduire les conséquences d'une décision pour mieux en apprécier la valeur ou l'incongruité, ce sont là les tâches quotidiennes du mathématicien confronté à la modélisation des théories linguistiques. Quant à la fécondité des systèmes de représentations métalinguistiques, elle est éprouvée par l'usage, plus ou moins approprié dans certains cas, qu'en font les linguistes. L'outil mis à leur disposition permet-il de mieux décrire les langues et de faire apparaître des faits jusque là ignorés ou du moins passablement sous-estimés ? Nous renvoyons sur ce point à quelques une des thèses de linguistique et aux articles publiés qui utilisent ce type d'écriture représentative pour décrire des faits linguistiques, aussi bien dans les langues indo-européennes que dans le domaine sémitique, chinois, japonais, ou pour les langues africaines.

Nous avons déjà dit que les auteurs avaient préféré montrer, sur un ou deux exemples, comment s'opèrait la conceptualisation et la mise en place progressive d'un système formel local car relatif à un problème précis plutôt que de représenter ce système sous une forme qui serait coupée de ses motivations. Pour traiter ces exemples, il leur a été nécessaire d'introduire un certain nombre de concepts - comme ceux de repérage- et d'évoquer certaines questions épineuses - comme extension /intension -. Il serait bien entendu souhaitable de mieux les cerner par un discours plus serré et mieux argumenté. Ceci montre en tout cas que l'analyse et le traitement formel d'un seul énoncé linguistique impose en fait des ingrédients conceptuels nombreux. Il en serait de même pour traiter d'exemples physiques ou biologiques. Le rapport immédiat que nous entretenons avec notre langage ne doit pas nous laisser croire qu'une appréhension naïve des faits est suffisante pour en découvrir les règles de fonctionnement. L'objection suivante a été émise par des esprits qui ne croient qu'au simple immédiatement compréhensible : "c'est bien difficile ce que vous représentez ! Pourquoi faire si compliqué ? " sous-entendu "quand d'autres collègues ont un discours plus transparent". Ceci est évidemment réfutable : "Pourquoi le

langage sur lequel toutes les civilisations, à toutes les époques, se sont interrogées, se laisserait-il décrire et enfermer dans un discours globalisant parsemé d'assertions qui semblent dépouillées de tout mystère alors que l'étude de la nature ou du vivant a conduit à des modélisations complexes et locales qui imposent une longue formation spécialisée préalable à toutes discussions sur la validité et la pertinence de ces modèles ? "

Le mode d'agencement par intrication conduit à des systèmes formels dont la syntaxe n'engendre pas - nous l'avons déjà dit - des "arbres" comme c'était le cas avec des systèmes agencés par substitution et concaténation ("arbres de dérivation" ou "arbres syntagmatiques" dans le modèle génératif de Chomsky). L'opération d'agencement par intrication engendre des représentations qui, lorsqu'on veut mémoriser "l'histoire constitutive" par une notation appropriée, se laisse décrire par une sous-classe de graphes sans circuits que l'un des auteurs avait avec M.C. Barbault proposé d'appeler "treille" (cf. Transformations formelles et théories linguistiques 1972). Quelle est la définition exacte de cette structure ? Comment est-elle engendrée ? Quelles en sont les propriétés ? Ces questions seront l'objet d'un prochain article.

*

*

*

Dans le prochain numéro, il sera donné une liste de quelques thèses (de 3^{ème} cycle et d'état) portant sur des descriptions de langues (anglais, chinois, français, japonais, malgache, mooré, portugais, tamoul...) et qui utilisent les systèmes de représentation et concepts théoriques développés dans le cadre de l'ERA 642.

La bibliographie qui concerne cet article est identique à celle de l'article précédent.

Ce travail a pu être mené avec l'aide du C.N.R.S., dans le cadre de l'A.T.P. "Informatique et Sciences Humaines" *. Nous remercions cet organisme et ses représentants.

J.P. Desclés

*

*

*

1. LES DONNEES EN LINGUISTIQUE

Les données auxquelles on a affaire en linguistique peuvent être classées, par souci de commodité et de clarté, sous diverses rubriques.

1.1. Etant donné un terme (dans le cas le plus simple, une unité lexicale), on étudie ses propriétés distributionnelles, c'est-à-dire ses latitudes de co-occurrence. On dégage ainsi des cadres de comptabilité contextuelle, et l'on construit des classes d'équivalence à partir de certaines propriétés syntaxiques, ou, pour être plus précis, de certaines contraintes sur les manipulations et les agencements. Il s'agit là d'un travail classique en linguistique descriptive, auquel doit se plier le chercheur, qu'il soit sur le terrain ou dans son bureau, et on a trop souvent tendance à sous-estimer l'importance de cette activité de tri et de classification sans laquelle la construction théorique risque de n'être qu'un mirage. Mais il est également patent que le linguiste est amené à fabriquer une bonne partie de ses exemples hors contexte (alors qu'il faut travailler à contexte explicite), hors situation (ainsi, on ne tiendra pas compte des valeurs aspectuelles, modales

* A.T.P. n° 51 4220. (1979-1981)

ou du jeu des personnes) et hors prosodie (comme si l'oral n'était qu'un avatar de l'écrit). D'où une situation métalinguistique qui, si l'on n'y prend garde, transforme l'objet que l'on s'efforce de caractériser, d'autant que le linguiste n'a à sa disposition que du texte (écrit ou oral) et que, quels que soient ses efforts théoriques pour rendre compte d'une activité à laquelle il n'a pas un accès direct, il devra se fonder au départ sur des suites textuelles, pour, inéluctablement, y retourner.

1.2. Ceci nous fournit un second ensemble de données (parfois produit de façon délibérée, mais souvent, hélas, confondu avec du texte naturel) et que nous appellerons métatexte : phrases de manuels, textes de logiciens ou de spécialistes d'intelligence artificielle. Il n'y a, naturellement rien de choquant à fabriquer et manipuler de tels métatextes. Après tout, l'activité de langage n'est pas la chasse gardée du linguiste et de telles études permettent de dégager certaines propriétés intéressantes, mais il est dangereux de mettre sur le même plan des données hétéroclites, toutes désignées du même qualificatif de linguistiques, alors que, dans le second cas, par exemple, on a supprimé un certain nombre de traits propres aux énoncés produits et interprétés, par des sujets énonciateurs (humains!), dans des situations spécifiques et dans une langue naturelle donnée.

1.3. Ce danger est d'autant plus réel que, à première vue, rien ne semble séparer ces métatextes des gloses épilinguistiques, qui forment un troisième groupe de données. Nous appelons ainsi ces textes qu'un sujet produit lorsque, de façon spontanée ou en réponse à une sollicitation, il commente un texte précédent. Nous parlons de glose et non de paraphrase, afin de réserver ce dernier terme à une activité réglée, donc contrôlée par l'observateur (sur ce point, voir plus bas), alors que la glose renvoie à la pratique langagière du sujet énonciateur. Quant au qualificatif épilinguistique, il désigne l'activité métalinguistique non-consciente de tout sujet et se distin-

gue donc de l'activité métalinguistique délibérée. Nous nous voyons forcés de ne mentionner qu'en passant cette question capitale, que masque chez de nombreux chercheurs l'emploi du mot compétence. Or, les gloses épilinguistiques forment une bonne partie de notre discours quotidien et jouent un rôle important dans le discours explicatif d'un informateur qui veut faire saisir le sens d'une phrase dans une langue étrangère ou la signification d'un énoncé mal interprété. Mais on néglige presque toujours ces gloses épilinguistiques alors qu'elles sont une précieuse source de renseignements linguistiques et, à leur manière, constituent un système de représentations interne à la langue, c'est-à-dire une métalangue non totalement contrôlable. Ainsi, la phrase *L'homme qui est venu hier portait bien un imperméable beige* est peut-être glosé par *"Eh bien, il y a quelqu'un qui est venu hier, et ce quelqu'un, c'est un fait qu'il portait effectivement un imper beige"*.

Lorsque le sujet énonciateur se fait (à sa manière) linguiste, il produit des gloses. Lorsque le linguiste se fait sujet énonciateur, il construit des familles paraphrastiques, c'est-à-dire, comme en (1), des classes d'équivalences: on a affaire ici, non plus à des termes simples, mais à des termes complexes ou énoncés, et les propriétés de ces familles sont fort différentes de celles d'une classe distributionnelle. Très vite, on s'aperçoit que l'on a besoin d'un système métalinguistique qui ne soit pas un jeu d'étiquettes, mais qui permette de représenter la dérivation paraphrastique.

Ainsi, si l'on accepte l'équivalence de :

$$\left. \begin{array}{l} \text{il } \left\{ \begin{array}{l} \text{s'est} \left\{ \begin{array}{l} \text{vu} \quad \text{condamné} \\ \quad \text{condamner} \\ \text{fait} \quad \text{condamner} \end{array} \right\} \\ \text{a été condamné} \end{array} \right\} \\ \text{on l'a condamné} \end{array} \right\} \text{ à trois ans de prison ferme}$$

il faudra, grâce à un système de représentations adéquat, marquer formellement l'équivalence, c'est-à-dire dégager les règles qui permettent de passer d'un agencement à l'autre et d'expliquer pourquoi ces agencements particu -

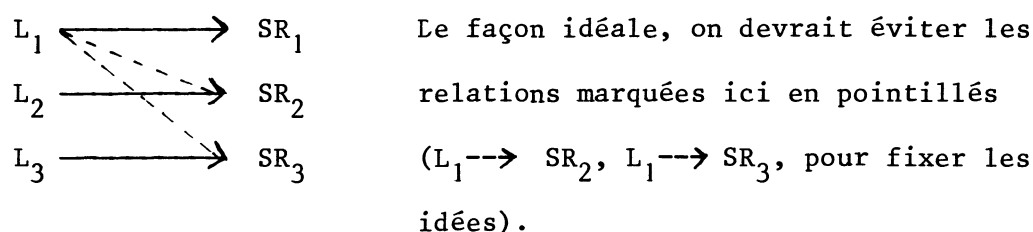
liers ont des valeurs référentielles (significations) équivalentes, à la modulation près dont il faudra rendre compte.

On comprendra l'importance exceptionnelle de cette démarche lorsque l'on s'occupe de plus d'une langue, qu'il s'agisse de description, de traduction, de typologie ou d'analyse dite contrastive. En effet, on ne peut jamais étudier une langue autre en soi (sauf si l'on traite de langues de grande diffusion, comme le français ou l'anglais, où l'on peut avoir la périlleuse illusion de l'immédiateté de l'analyse!): il faudra toujours traduire terme à terme, rendre le sens global en langue courante, mais surtout en passer par un commentaire interprétatif qui oscillera entre la glose, la terminologie héritée ou néologique, sans compter l'astuce technique. Ainsi, de la particule japonaise *wa*, on dit qu'entre autres emplois, elle sert à thématiser -mais la définition de la thématisation, comment se fera-t-elle?- ou alors, qu'elle peut être glosée par les expressions françaises "à propos de", "quant à", "entre autres",..., comme si l'on pouvait utiliser de façon neutre une seconde langue à des fins métalinguistiques (c'est-à-dire pour décrire de façon univoque et stable les phénomènes d'une langue différente). De même, certains croient que l'on peut impunément ré-introduire des termes comme "datif" dans (ou faut-il dire 'sans'?) une théorie générale des cas (on dira, par exemple, que *lui* est "datif" dans *il lui parle*), que cela ne tire pas à conséquence puisqu'il s'agit d'une étiquette conventionnelle, ou qu'en utilisant des majuscules (CAUSE, BE, DO) on élève au rang d'êtres métalinguistiques des termes empiriques. En fait, on peut montrer qu'aucune terminologie composite, ni, au reste, le recours expéditif à des symboles de logiciens (tels le "iota" de Russelle ou le "lambda" de Church) ne permet d'appréhender les phénomènes paraphrastiques de façon contrôlée et fructueuse. Au mieux, elle établira des correspondances étiquetées. Que l'on considère par exemple les phrases suivantes, tirées d'un manuel de japonais* :

* in Modern Japanese for University Students I, Japanese Department, International Christian University, Revised Edition, pp. 117 sq.

Kazuko wa / me ga ookii desu , où *Kazuko* est un nom propre, *me* signifie "oeil, yeux", *ookii* veut dire "grand(s)" et *desu* est une copule (facultative); quant à *wa* et *ga*, ce sont deux particules dont il est dit que *wa* introduit le "thème" et *ga* indique le sujet d'une phrase dans certaines circonstances. Les traductions proposées sont "As for Kazuko, her eyes are large" ("Quant à Kazuko, ses yeux sont grands"); "Kazuko's eyes are large" ("Les yeux de Kazuko sont grands"); "Kazuko has large eyes" ("Kazuko a de grands yeux"), et l'on pourrait ajouter, pour le français, "Il y a Kazuko, ses yeux sont grands / elle a de grands yeux". De même, on aura pour *Tanakasan wa okusan ga byooki desu* (terme à terme: Tanaka Monsieur Wa femme GA malade COPULE), "La femme de monsieur Tanaka est malade"; "Monsieur Tanaka a sa femme (qui est) malade"; "(Il y a) Monsieur Tanaka, sa femme est malade"; etc. Manifestement, le problème posé est d'établir une relation entre la phrase japonaise et ses traductions anglaises et françaises, qui marque l'appartenance des différents membres à une même famille paraphrastique. Un tel problème n'est soluble que si l'on construit les systèmes de représentations pour le japonais, pour le français et pour l'anglais, à la fois en respectant la spécificité des phénomènes propres à chaque langue et en ramenant ces phénomènes à un certain nombre d'opérations élémentaires (prédicatives et énonciatives) qui auront une forme générale; c'est-à-dire une forme que l'on généralisera, par un enchaînement argumenté, à travers les configurations spécifiques de chaque langue. Ainsi, le problème que se pose la linguistique, lorsqu'elle se donne pour objectif l'étude du langage appréhendé à travers la diversité des langues naturelles, est de rechercher comment construire une théorie de la forme générale des opérations prédicatives et énonciatives dans les langues, qui respecte les exigences habituelles de vérifiabilité. Une première condition, on le comprendra aisément, sera l'adéquation empirique, non point au sens vague où il doit être possible de passer du système de représentations (SR) aux phénomènes et des phénomènes au SR, mais de façon plus précise en posant que le SR est le versant méta-

linguistique d'une activité dont la glose représente l'autre aspect, et qu'il doit donc exister des correspondances entre les deux. Une seconde condition est que l'on doit pouvoir passer d'un SR (spécifique à une langue) au SR spécifique à une autre langue. En bref, on a le schéma suivant:



On sait que cela n'est hélas pas le cas, surtout si L_1 est une langue de grande civilisation, avec une tradition grammaticale de surcroît. De même, on comprendra qu'il n'existe pas de relation directe $L_1 \longrightarrow L_2$ ou $L_2 \longrightarrow L_3$, car la traduction pure et simple n'a aucune force explicative. Le problème réside donc dans la relation entre $SR_1 \longrightarrow SR_2$, $SR_2 \longrightarrow SR_3$, ou $SR_1 \longrightarrow SR_3$.

Ainsi on voit aisément qu'à côté de l'adéquation empirique se pose le problème formel de la construction de systèmes de représentation compatibles, qui ne soient pas un simple arsenal de consignes pour établir des correspondances entre des suites en L_1 et des suites en L_2 ou L_3 . Deux systèmes ne sont comparables que s'ils sont formulés avec les mêmes catégories généralisables et les mêmes opérations élémentaires dont on cherche à déterminer l'invariance; ce qui varie, de langue à langue, c'est la constitution de familles d'opérations, l'enchaînement et les contraintes sur les règles et leurs combinaisons. Il reste que tout n'est pas représentable, tout n'est pas généralisable, et personne ne saurait avoir la prétention de travailler sur "toutes" les langues et sur "n'importe quelles" données. Aussi bien notre démarche est-elle inverse: délimiter des types de problèmes, construire une linguistique du généralisable, au lieu de spéculer sur des propriétés universelles à partir d'observations non systématiques.

1.4. Mais il existe un autre domaine où l'expérimentation (au sens d'une reproduction) donc au passage réglé d'un système métalinguistique à la suite

textuelle et vice-versa) paraît difficile ou, du moins, d'un autre ordre : Nous parlerons dans ce cas de corpus contraint. En effet, il existe des situations où le corpus ne peut être étendu, transformé, bref utilisé à des fins expérimentales: c'est parfois le cas lorsque le chercheur ne peut qu'enregistrer, sur le terrain, un minimum de textes et d'informations. Ou encore, lorsqu'on a affaire au délire d'un psychotique, que l'on ne saurait modifier et où l'observateur ne peut intervenir. Le langage du très jeune enfant pose des problèmes d'un autre ordre, mais là encore, le corpus est le plus souvent contraint, car la production discursive ne se prête pas aux désirs de l'expérimentateur, fût-il le plus rusé et le plus patient du monde! Il en va de même pour le texte littéraire, où les seules expériences possibles portent sur la lecture, ou sur le pastiche, sans compter l'analyse philologique ou d'ordre stylistique.

1.5 Le lecteur aura compris que notre entreprise n'a pas la prétention de traiter de tous les domaines, comme si l'on pouvait surmonter, grâce à un quelconque outil formel, les différences radicales qui existent entre les observables. Nous appuyant sur le travail effectué en (1), nous ne pouvons considérer que comme un expédient inévitable (mais transitoire et d'une valeur précaire) la fabrication de métatextes, (2), et nous avouons notre incompetence théorique devant certains problèmes que nous pose le traitement d'un corpus contraint. C'est donc sur le domaine (3) - gloses et paraphrases - que nous nous efforçons de mener nos observations en respectant les trois règles suivantes:

(i) construire des métalangages (nécessairement extérieurs aux langues) à partir des systèmes de représentations élaborés à l'intérieur de chaque langue spécifique et n'utiliser que des êtres métalinguistiques qui ont un statut dans la théorie;

(ii) ne pas séparer théorie du langage et théorie de l'analyse des langues, et ne pas se contenter de règles de bonne formation de phrases en éliminant les règles de bonne constitution des énoncés;

(iii) utiliser des formules où chaque occurrence de symboles métalinguistiques a une trace repérable dans un texte et où chaque constituant du texte acquiert un statut métalinguistique dans le système de représentations.

L'aide du mathématicien se justifie dès lors que, pour échapper aux risques des gloses auto-définissantes, on s'impose de recourir à des systèmes de représentations métalinguistiques extérieurs aux langues, ce qui permet de substituer aux manipulations, plus ou moins explicites et soumises aux décisions subjectives des linguistes, des calculs qui puissent être exécutables par des automates. De tels systèmes sont composés de catégories, d'objets primitifs catégorisés, d'opérateurs élémentaires et de règles destinées à constituer des formules métalinguistiques par agencement de ces opérateurs. Ces systèmes ne permettent pas, cependant, de calculer, par exemple, toutes les représentations d'énoncés paraphrastiquement équivalentes à la représentation d'un énoncé donné, ou encore de produire, à partir d'un schéma commun, une famille d'énoncés. Il est alors nécessaire de restreindre l'ensemble des formules (toute formule du système de représentations n'est pas obligatoirement un représentant d'un énoncé) en définissant: (a) des schémas canoniques initiaux (schémas d'axiomes); (b) des schémas de règles (ces schémas se justifient lorsque l'on en donne une interprétation dans un cadre théorique).

L'observation, puis la conceptualisation du fonctionnement des langues et les représentations métalinguistiques font apparaître des régularités structurelles, mais aussi des procédures de traitement, ce qui permet de revenir aux agencements linguistiques.

2. STRUCTURES ASSOCIEES AUX SYSTEMES DE REPRESENTATIONS METALINGUISTIQUES

En formalisant l'analyse des chaînes linguistiques en termes de classes syntaxiques constituées par les procédures empiriques mises au point par l'école structurale américaine, N. Chomsky a proposé, comme on sait, diffé-

rentes classes de grammaires formelles: (1°) "finite state grammars"; (2) "phrase structure grammars"; (3) "transformational grammars". Les deux premières classes de grammaires décrivent, en partie, la plupart des langages de programmation; la théorie des langages formels s'est développée de façon autonome, entretenant des rapports féconds avec l'informatique théorique. La troisième classe de grammaires n'est pas encore entièrement formalisée malgré quelques essais fructueux.

Les structures utilisées en linguistique ne se ramènent cependant pas aux seules structures arborescentes. En effet, la structure arborescente est essentiellement adéquate aux phénomènes classificatoires et formalise adéquatement les relations hiérarchisées et d'emboîtements. Une telle structure ne peut donc pas formaliser les relations d'intrication où un même objet est en relation de dépendance avec plusieurs objets (un objet a plusieurs pères, alors que, dans l'arborescence, un objet a toujours au plus un père). C'est le cas, par exemple, des relations anaphoriques. Diverses tentatives ont été proposées pour traiter, dans le formalisme des arborescences, ces relations au moyen d'indices ou à l'aide d'un système d'adressage où l'on identifie deux objets situés à des adresses différentes (ce qui revient à utiliser des pointeurs). Ainsi, Z. Harris* décrit les relations anaphoriques et les déictiques à l'aide de certains énoncés métalinguistiques (coordonnés à l'énoncé décrit) représentables par des arborescences (ayant pour sommets non-terminaux des opérateurs d'identification et pour terminaux des adresses). Sans discuter ici de l'adéquation de cette représentation aux données linguistiques, on peut s'interroger sur le coût réel (stockage, temps de calcul) des procédures de manipulation; celles-ci nécessitent, au fur et à mesure de la construction des énoncés, (on peut le démontrer) une succession de balayages des arborescences.

Les analyses linguistiques de nombreux phénomènes nous ont amenés à définir et développer un formalisme propre à traiter directement ces relations

*HARRIS, Z.S. - Notes du Cours de Syntaxe - Seuil - 1976.

d'intrication. La structure proposée est "plus complexe" que l'arborescence, elle a reçu le nom de treillage (ou grillage).

Cette structure est présentée dans un autre article du même recueil, nous y renvoyons le lecteur.

3. QUELQUES CONCEPTS THEORIQUES

Nous allons donner, de façon schématique et non formelle, la définition de quelques termes techniques en montrant comment ils s'intègrent dans un développement théorique. Ces termes seront utilisés dans le traitement de l'exemple donné au paragraphe suivant.

3.1. Propriétés intensionnelles et notions

A partir d'observations sur le fonctionnement des langues naturelles, il s'agit, rappelons-le, de mettre en évidence des opérateurs et opérations élémentaires qui seront considérés comme les véritables invariants de l'activité langagière. Nous disons "invariants" pour marquer nos distances par rapport aux universaux de la linguistique.

Nous considérons deux classes d'opérations: les opérations énonciatives qui contribuent à fixer les valeurs référentielles (aspectualité, temporalité, détermination...) de l'énoncé et les opérations prédictives constitutives de la relations prédictive. Chaque énoncé est le résultat observable de ces opérations. Les coordonnées énonciatives construites par les opérations énonciatives positionnent la relation prédictive par rapport aux origines de l'énonciation.

Le problème des catégories grammaticales trouve une formulation plus aisée lorsque l'on a recours à des opérateurs et des opérations. Il faut expliquer, en particulier, comment construire la "catégorie nominale", ou la "catégorie verbale", sachant qu'on ne peut se contenter d'un simple jeu d'étiquettes SN (pour syntagme nominal) ou SV (pour syntagme verbal), asso-

ciées à des classes d'équivalence d'objets linguistiques (classes distributionnelles). En effet, l'opposition verbo-nominale n'est vraisemblablement pas universelle et il n'est pas raisonnable de partir d'une telle opposition (ou de son corrélat logique: prédicat / nom) pour fonder les représentations métalinguistiques. Les notions de "prédicat" et de "nom" doivent être construites par des opérations explicites et ne peuvent pas être considérées comme primitives (comme le fait par exemple S. Lesniewski dans son ontologie et dans sa protothétique).

Il nous faut préciser, au moins de façon allusive, comment nous envisageons de "construire" les catégories nominales et prédicatives. Là encore, l'observation des langues nous invite à prendre pour notions primitives des propriétés intensionnelles.

Le concept d'intensionnalité est souvent peu clair car au cours de l'histoire de la logique il a considérablement fluctué. L'idée ancienne était qu'une notion définie intensionnellement est "en dualité" avec celle qui serait définie extensionnellement*. La difficulté revient à donner un sens précis à "en dualité". Alors que l'on a su merveilleusement définir les notions extensionnellement par la théorie des classes, les concepts définis intensionnellement n'ont pas encore reçu de définition mathématique précise qui reflèterait les anciennes exigences. Avec A. Church et R. Carnap, on assiste à une radicalisation du concept d'intension. Cette radicalisation a été reprise par R. Montague qui se ramène tout simplement à la notion de "monde possible".

Le concept d'intensionnalité que nous retenons sera plus large.

Des propriétés sont intensionnelles : lorsqu'elles ne vérifient pas les deux critères d'extensionnalité (EXT_1) et (EXT_2) exposés ci-dessous.

*Note: Ainsi, pour les médiévises: "Quo major extensio, eo minor comprehensio, et quo major comprehensio, eo minor extensio".

Soit deux propriétés unaires p et q . La question de l'intension (et de son corrélat l'extension) se pose dès que l'on cherche à savoir quand ces deux propriétés sont identiques. Notons par ' $p \equiv_e q$ ' l'identité extensionnelle des propriétés p et q . Les deux critères (Ext_1) et (Ext_2) s'énoncent ainsi :

(EXT_1) pour tout objet x ,

SI $p(x)$ est identique (dans un sens à préciser) avec $q(x)$

ALORS $p \equiv_e q$

(EXT_2) pour toute propriété (unaire) u

SI $u(p)$ est identique (dans un sens à préciser) avec $u(q)$

ALORS $p \equiv_e q$

Autrement dit, deux propriétés p et q sont extensionnellement identiques lorsqu'elles s'appliquent aux mêmes objets (EXT_1) et lorsqu'elles vérifient les mêmes propriétés (EXT_2). Certes, ces critères devraient être précisés et intégrés à une théorie générale des objets, des propriétés, de l'égalité, de l'identité, et des classes. Signalons qu'il existe des théories logiques où ces critères de l'extensionnalité ne sont pas posés dès les fondements. Ainsi, la Logique Combinatoire, fondée par Schönfinkel et développée par H.B. Curry essentiellement, ne s'appuie pas fondamentalement sur l'extensionnalité. Aussi n'est-il pas déraisonnable de ne pas imposer a priori aux propriétés de vérifier obligatoirement les critères (EXT_1) et (EXT_2), d'autant plus que de nombreux phénomènes linguistiques nous montrent que les principes d'extensionnalité sont trop exigeants puisque de nombreuses propriétés sont manifestement distinctes (par exemple... *être bouc* et ...*être le mâle de la chèvre* ou encore ...*être un bipède sans plumes* et ...*être un animal raisonnable*) du point de vue de leurs significations (ou intensions) alors que les classes d'objets auxquels elles s'ap-

pliquent sont identiques. Ainsi la classe des bipèdes sans plumes et celle des animaux raisonnables sont identiques puisqu'elles contiennent Platon, Socrate, Alcibiade, Aristote, Napoléon,... bref tous les hommes. Pourtant les énoncés suivants:

Socrate est un bipède sans plumes

Socrate est un animal raisonnable

nous apportent, à propos de Socrate, des informations bien différentes qu'une théorie non extensionnelle devrait essayer de formaliser.

Chaque propriété intensionnelle (par exemple: être-grand , être-bleu , être-chien ...) est insérée dans un faisceau de propriétés constitutives et différentielles, à savoir le champ associé à la propriété. Une même propriété peut appartenir à plusieurs champs.

Chaque champ est muni d'une structure topologique telle que chaque propriété p soit un ouvert. Appelons complémentaire linguistique p' le complémentaire de l'ouvert p par rapport à un des fermés qui contient nécessairement p . La frontière $Fr(p)$ de p fait donc partie de ce complémentaire linguistique. Une notion (prédicative) est définie par la donnée d'un couple $\langle p, p' \rangle$, c'est-à-dire d'un ouvert et de son complémentaire linguistique. Le concept de complémentaire linguistique permet de rendre compte des problèmes de négation et de nombreux problèmes modaux. Ainsi, Jean n'est pas grand ne signifie pas simplement qu'il n'y a aucune relation entre le sujet Jean et le prédicat est grand puisque nous avons aussi Jean n'est pas grand car il est immense, Jean n'est pas grand pour son âge. De même, Jean ne boit pas du café mais du thé; Jean ne boit pas du café mais le déguste. Ceci montre que la négation n'exprime pas toujours cette absence de relation entre le sujet et le prédicat (*Jean et boire du café*) mais plus souvent une relation entre le sujet et un prédicat qui reste "voisin" du prédicat nié, qui se trouve donc dans le complémentaire linguistique.

Etant donné une notion prédicative, diverses opérations, que nous ne donnons pas ici, permettent de construire d'un côté les prédicats et d'un autre côté les objets nominaux qui peuvent être arguments d'un prédicat.

Pour fixer les idées, à la propriété intensionnelle /être - livre/, nous associons le prédicat catégorisé, en français, en verbe: lire, et les expressions qui sont catégorisés, en français, dans la classe nominale: lecteur, livre, lecture... Remarquons qu'un prédicat n'est pas toujours catégorisé en verbe mais peut, selon les contextes, être catégorisé en adjectif, en préposition etc...

Le prédicat ou opérateur prédictif fonctionne tantôt extensionnellement tantôt intensionnellement. Dans le premier cas, une opération de quantifiabilisation permet, entre autres, de construire une classe associée à la notion prédicative (dans Jean est un professeur, est un est la trace de cette opération de quantifiabilisation). Dans le second cas, le prédicat intensionnel ne comporte pas de classe associée, il a un fonctionnement très proche de la notion. Pour construire un objet nominal issu d'une notion prédicative, interviennent diverses opérations dont la désignation, la constitution de la classe extensionnelle associée à la notion, la quantification et la qualification (par une propriété différentielle) (cf. Bibliographie)

Toute prédication s'opère nécessairement à partir d'un réseau de relations primitives liées aux représentations notionnelles filtrées par chaque langue. Ainsi, la notion de /voir - avec - les - yeux/ implique une autre notion "source" (/celui - pour - qui - quelque - chose - est - visible/) et une notion "but" (/ce - qui - est - visible/), d'où trois relations primitives: entre la source et la notion prédicative /voir - avec - les - yeux/, le but et cette même notion, et enfin, la notion source et la notion but.

3.2. Prédication et lexis.

Chaque prédication est effectuée en itérant des opérations élémentai-

res qui consistent à: 1°) construire des opérateurs prédicatifs; 2°) construire une suite ordonnée d'objets nominaux, les opérandes des opérateurs prédicatifs; 3°) constituer une lexis ou relation prédicative fondamentale qui est obtenue en agencant entre eux opérateurs prédicatifs et opérandes de façon que si l'on effectue les opérations impliquées par l'agencement formel (c'est-à-dire en faisant agir les opérateurs sur les opérandes) alors la signification attachée à la lexis soit construite et signifiée par ce biais.

Chaque opérateur prédicatif est unaire. L'agencement des divers opérateurs avec leurs opérandes conduit à une lexis élémentaire décomposable en un prédicat et ses divers arguments. Les prédicats binaires (comme *écrire*, *manger*, *voir*...), ternaires (comme *donner*, *recevoir*, *dire*...) sont donc construits par compositions d'opérateurs prédicatifs unaires. Par contre, des prédicats unaires comme *se blesser*, *se casser* sont construits à partir de prédicats binaires (*blesser*, *casser*...) eux-mêmes construits à partir d'opérateurs prédicatifs unaires.

Une lexis (élémentaire) n'est donc pas le résultat d'une simple assignation qui consisterait à remplacer les variables d'un schéma de lexis par des termes catégorisés en prédicat et en objets nominaux (dans la majorité des cas en verbe et nom). Chaque lexis est une expression formelle déjà constituée de façon à présenter le programme opératoire à effectuer pour sortir la signification qui lui est attachée et qui dépend non seulement des notions constituantes mais aussi des agencements constitutifs.

D'après ce qui précède, une lexis est certes proche du concept de contenu propositionnel (Gedanke chez G. Frege) ou du dictum de la tradition médiévale ou encore du lekton stoïcien, mais elle en diffère par les trois caractéristiques suivantes: (1) elle n'est ni assertée, ni non assertée, elle est pré-assertée, c'est-à-dire pas encore assertée, mais devant être nécessairement située dans un espace énonciatif; (2) elle est comparable à une notion et en possède, en tout cas, les propriétés topologiques essentielles (à sa-

voir qu'elle est représentable par un ouvert); (3) elle est agencée formellement et, par là, est une forme génératrice d'autres formes dérivées auxquelles on associera, à l'aide d'opérations ultérieures, une famille d'énoncés.

Une lexis, constituée à partir d'un réseau de relations primitives, apparaît comme la relation prédicative la plus fondamentale qui se trouve être à l'origine de toutes les familles de relations prédictives dérivées.

Une lexis non élémentaire est obtenue en composant, à l'aide d'opérateurs de composition des lexis élémentaires.

Dans l'article "A propos d'opérations intervenant dans le traitement formel des langues naturelles", Mathématiques et Sciences Humaines, N° 34, ou dans l'analyse formelle de la constitution de l'énoncé japonais *Kazuko wa me ga ooki i (desu)* (cf. plus loin) le lecteur verra des exemples de lexis à partir desquelles sont dérivées les relations prédictives. Les lexis et relations prédictives qui s'en déduisent donnent naissance aux familles (structurées par les relations de dérivations entre lexis et relations prédictives) d'énoncés. Toute une famille peut avoir pour invariant commun un même lexis. Par exemple, la famille:

Jean a un trou dans sa poche

La poche de Jean est trouée

Jean, sa poche, elle est trouée

Il y a un trou à la poche de Jean

est issue des deux lexis élémentaires initiales :

$$\lambda_1 \equiv \langle \text{poche, avoir, trou} \rangle \simeq \langle \text{trou-est-à-poche} \rangle$$

$$\lambda_2 \equiv \langle \text{Jean, avoir, poche} \rangle \simeq \langle \text{poche-est-à-Jean} \rangle$$

elles-mêmes agencées dans une lexis non élémentaire: λ où la lexis λ_1 est repérée par rapport à la lexis λ_2 .

3.3. Opérateurs transformationnels et transformations.

De même que toute lexis doit être située dans un espace énonciatif, ce qui revient à affecter la lexis de coordonnées énonciatives, les relations prédicatives seront elles aussi affectées de coordonnées énonciatives (construites, rappelons-le, par les opérations énonciatives); chaque lexis, chaque relation prédicative est alors repérée, dans l'espace énonciatif, par rapport à l'origine de cet espace, le triplet:

$$\langle \mathcal{L}_0, \mathcal{C}_0, \text{Sit}(\mathcal{L}_0, \mathcal{C}_0) \rangle$$

Nous pouvons maintenant concevoir deux types d'opérations formelles qui transforment un énoncé E déterminé en un autre. En effet, la relation prédicative ϕ qui est sous-jacente à E peut être transformée en une autre relation prédicative ψ (dite dérivée de ϕ) de la même famille (c'est-à-dire que ϕ et ψ sont dérivables d'une lexis λ commune) mais les coordonnées énonciatives restent inchangées; on transforme alors, dans ce cas, "à énonciation constante", et le "contenu propositionnel" primitif (par rapport à ϕ et à ψ) reste invariant; on pourra parler ici de transformations paraphrastiques. Si, par contre, on change les coordonnées énonciatives d'une même relation prédicative, on transforme alors les valeurs référentielles des énoncés qui en résultent et, dans ce cas, on parlera de déformations énonciatives.

La famille citée précédemment est produite "à énonciation constante"; il s'agit bien d'une famille paraphrastique ayant pour source la même lexis λ .

Par contre, si nous introduisons quelques déformation énonciatives comme :

Jean a déjà eu un trou à sa poche

Jean aura bien un trou à sa poche

Jean a peut-être un trou à sa poche

Hier, Jean avait un trou à sa poche

La poche de Jean a bel et bien été trouée

etc...

les valeurs référentielles ne sont plus identiques (bien que le contenu propositionnel soit toujours le même) puisque les coordonnées énonciatives ont changé.

Soit λ une lexis. La famille $F_1(\lambda)$ comprend les relations prédicatives dérivées de λ "à énonciation constante". Lorsqu'on déforme (au sens technique du terme) les conditions d'énonciation, on engendre une seconde famille $F_2(\lambda)$ qui étend $F_1(\lambda)$. Chacune des familles $F_1(\lambda)$ et $F_2(\lambda)$ est munie d'un préordre de déduction, noté $\longrightarrow$. Plus précisément, on se donne un ensemble A d'opérateurs dits transformationnels. Un opérateur α de A est transformationnel s'il existe :

- (1) un domaine D non vide et un codomaine D' non vide ;
- (2) au moins une relation prédicative ϕ dans D et une unique relation prédicative ψ (associée à α) dans D' telle que ψ soit déduite formellement de ϕ par α .

Autrement dit, si α est un opérateur transformationnel, alors on associe canoniquement à α , une opération (partielle) - ou transformation - α' qui à chaque élément ϕ fera correspondre, quand il est défini, un unique élément ψ . Nous posons donc :

$$\phi \xrightarrow{\alpha} \psi \text{ si et seulement si } \psi = \alpha'(\phi).$$

L'opérateur transformationnel α sera fonctionnel lorsque l'opération α' canonique est partout définie sur le domaine de α .

On reconnaîtra, dans cette formulation, la problématique des opérateurs transformationnels et des transformations du linguiste Z.S. Harris, du moins telle qu'il l'a présentée dans "Structures Mathématiques du Langage" (*).

(*) Traduction française, chez Dunod, 1971
Voir aussi J.P. Desclés : [DES-75]

Il existe une différence cependant. Les opérateurs chez Harris agissent, théoriquement du moins, uniquement sur des phrases attestables alors que les opérateurs transformationnels agissent ici sur des lexis ou des relations prédicatives auxquelles sont associées ensuite des énoncés.

Cette façon de voir rejoint aussi celle de la grammaire applicative de S.K. Shaumyan avec cependant une différence: la composante énonciative n'apparaît absolument pas dans le modèle applicatif.

Donnons un exemple de transformation .

EXEMPLE : *Transformation de dualité.*

Soit la lexis de localisation suivante :

$$\lambda \equiv < x \underline{\epsilon} y >$$

qui se lit "l'objet x est repéré par rapport à l'objet y". Le symbole ' $\underline{\epsilon}$ ' désigne, pour l'instant, un relateur de repérage binaire.(cf.3.4 pour son statut)

L'opérateur transformationnel, désigné par $[\exists]$, construit la converse de la relation précédente, soit :

$$\overset{u}{\lambda} \equiv < y \underline{\exists} x >$$

qui se lit "l'objet y sert de repère à l'objet x". Le symbole ' $\underline{\exists}$ ' désigne, pour l'instant, le relateur converse du relateur de repérage $\underline{\epsilon}$.

Nous avons alors la transformation (ou mieux le schéma de transformation puisque x et y sont des objets quelconques mais de même sorte) :

$$[\exists] \\ < x \underline{\epsilon} y > \longrightarrow < y \underline{\exists} x >$$

3.4. Opérateurs de repérage: esquisse d'une théorie.

Les opérateurs transformationnels de A sont, pour la plupart, complexes

Ils font intervenir, le plus souvent, un opérateur dit de repérage qui joue un rôle fondamental. Nous allons le présenter ici de façon informelle.

Nous nous plaçons dans un Univers (au sens technique du terme); désignons par $x, y, z, \dots$ les objets de cet Univers.

Deux objets quelconques x et y de l'Univers entretiennent a priori trois types de relations: soit ils sont identifiés l'un à l'autre; soit ils sont différenciés; soit ils n'ont rien à voir l'un avec l'autre, on dira, dans ce dernier cas, qu'ils sont disjoints, ou en rupture .

Désignons par $\underline{\epsilon}$ le relateur de repérage non spécifié qui lie x et y entre eux; $\langle x \underline{\epsilon} y \rangle$ se lit alors : " x est repéré par rapport à y "; x est l'objet repéré, y est l'objet repère.

Le symbole $\underline{\epsilon}$ est multivalent puisqu'il peut désigner les trois types de repérages abstraits: l'identification, la différenciation, la ruption.

Le concept de repérage abstrait, tel du moins que nous l'esquissons ici, ne peut se ramener à une simple relation de repérage dans un Univers, cette relation étant définie par son graphe. En effet, repérer un objet x , c'est trouver un objet y qui soit un repère pour l'objet x . Le repérage s'effectue donc au moyen d'une opération qui sera constitutive d'une relation ou plutôt d'une occurrence de relation, en désignant par occurrence de relation la donnée d'un couple (plus généralement n -uple) d'objets pour lesquels la relation est vraie.

Il faut donc considérer non pas la relation constituée mais les opérations constitutives des différentes occurrences de la relation de repérage.

Pour cela, nous introduisons l'opérateur unaire dit de repérage, désigné par $\underline{\epsilon}_1$.

Nous supposons que cet opérateur est déterministe au sens suivant:

" pour tout objet x de son domaine, il existe un et un seul objet y de l'Univers tel que y soit le repère de x ".

L'action de l'opérateur $\underline{\epsilon}_1$ sur une opérande x de son domaine donne pour résultat un objet y .

L'action de $\underline{\epsilon}_1$ sur x , donnant y pour résultat, est constitutive d'une occurrence de relation de repérage désignée par $\langle x \underline{\epsilon} y \rangle$.

Par définition nous avons (en représentant l'opération par une flèche) :

$$\text{SI } x \xrightarrow{\underline{\epsilon}_1} y \quad \text{ALORS } \langle x \underline{\epsilon} y \rangle \text{ est vrai.}$$

Considérons, par exemple, le cas de l'identification :

Identifier x , c'est trouver un identificateur y tel que x soit identifiable à y . Il en résulte que x et y sont identifiés entre eux, sans que, pour cela, x soit, à son tour, nécessairement l'identificateur de y . Nous donnerons, plus bas, des propriétés abstraites de l'identification.

D'une façon générale, $\underline{\epsilon}_1$ est dit avoir la propriété $\mathcal{P}$ si l'occurrence de la relation de repérage constituée possède la propriété $\mathcal{P}$. Plus particulièrement, on dira que :

(1) $\underline{\epsilon}_1$ a la propriété $[\rho]$ (réflexivité) lorsque :
pour tout objet x , on a : $\langle x \underline{\epsilon} x \rangle$ est vrai.

(2) $\underline{\epsilon}_1$ a la propriété $[\sigma]$ (symétrie) lorsque :
pour tout objet x , tout objet y , on a :

$$\text{SI } x \xrightarrow{\underline{\epsilon}_1} y \quad \text{ALORS } \langle y \underline{\epsilon} x \rangle \text{ est vrai.}$$

(3) $\underline{\epsilon}_1$ a la propriété $[\tau]$ (transitivité) lorsque :
pour tout objet x , tout objet y , tout objet z , on a :

$$\text{SI } x \xrightarrow{\underline{\epsilon}_1} y \text{ et } y \xrightarrow{\underline{\epsilon}_1} z \quad \text{ALORS } \langle x \underline{\epsilon} z \rangle \text{ est vrai.}$$

Désignons par $\sim$ le symbole de négation logique.

(4) $\underline{\epsilon}_1$ a la propriété $[\neg\rho]$ (irrégularité) lorsque :
pour tout objet x , on a : $\sim \langle x \underline{\epsilon} x \rangle$ est vrai

(5) $\underline{E}_1$ a la propriété $[A\sigma]$ (asymétrie) lorsque :

pour tout objet x, tout objet y, on a :

SI $x \xrightarrow{\underline{E}_1} y$ ALORS $\sim \langle y \underline{E} x \rangle$ est vrai

(6) $\underline{E}_1$ a la propriété $[D\sigma]$ (dissymétrie) lorsque :

il existe un objet y tel que :

SI $x \xrightarrow{\underline{E}_1} y$ ALORS $\sim \langle y \underline{E} x \rangle$ est vrai

Nous pouvons, à l'aide de ces propriétés, tenter d'établir une typologie rudimentaire des divers types de repérage les plus abstraits selon les propriétés assignées à l'opérateur de repérage.

Un opérateur de repérage qui aura les propriétés $[\rho]$ et $[\sigma]$ sera qualifié : opérateur d'identification ; un opérateur qui aura la propriété $[A\sigma]$ sera qualifié : opérateur de différenciation ; un opérateur qui aura les propriétés $[I\rho]$ et $[\sigma]$ sera qualifié : opérateur de ruption. Nous résumons ceci par le tableau :

Repérages	Propriétés	Symboles
identification	$[\rho], [\sigma]$	=
différenciation	$[A\sigma]$	$\neq$
ruption	$[I\rho], [\sigma]$	ω

L'identification est constitutive d'une relation réflexive et symétrisable.

La différenciation est constitutive d'une relation essentiellement asymétrique.

La ruption est constitutive d'une relation essentiellement irréflexive et symétrisable.

Commentons ces trois types de repérage.

(a) Identifier x, c'est trouver un objet unique y qui lui sert d'identificateur. Il est clair que x peut être considéré comme son propre identi-

ficateur; si x a pour identificateur y , x et y sont identifiables l'un avec l'autre sans que x soit pour autant nécessairement l'identificateur de y .

Autrement dit, y étant l'identificateur de x ($x \xrightarrow{=1} y$) il en résulte que l'on a l'occurrence $\langle x = y \rangle$ sans que pour autant x soit l'identificateur de y . L'identification constitue donc une relation symétrisable.

L'identification n'est ni l'identité, ni l'égalité, ni même une équivalence. L'identification est une opération; si l'identification possède la propriété $[\tau]$ alors celle-ci est constitutive d'une relation d'équivalence.

INTERPRETATION : Dans le langage courant, est comme est la trace d'une identification qui n'a pas la propriété $[\tau]$. De même, la copule est prend souvent la valeur d'identification sans que les deux termes identifiés aient nécessairement des rôles identifiables: de "a est A", on peut déduire que a et A sont identifiables (dans certains cas) sans que l'on ait "A est a" (on retrouve les conditions imposées par S. Leśniewski à l'ontologie, où ce dernier distingue les noms individuels et les noms génériques; voir, par exemple, J.B. Grize, Logique Moderne III, Gauthier-Villars, Paris, 1973; p.79).

L'égalité et l'identité impliquent la propriété $[\tau]$, ainsi que d'autres lois plus spécifiques (notamment les lois de Leibniz).

(b) Différencier x , c'est trouver un terme y par rapport auquel x se différencie. Toute localisation de x par rapport à un objet y spécifie un type de différenciation, tel que si x est localisé par rapport à y et non pas y par rapport à x , alors x n'est pas un localisateur de y .

Donnons-nous une propriété plus faible que l'asymétrie : la non-symétrie, d'où la différenciation faible. Considérons la propriété suivante d'une différenciation $\neq$ faible (propriété $[\text{Ant}\sigma]$) :

$$\begin{array}{ll} \text{SI} & x \overset{\neq}{\rightarrow} y \text{ et } y \overset{\neq}{\rightarrow} x \\ \text{ALORS} & x \overset{=}{\rightarrow} y \text{ et } y \overset{=}{\rightarrow} x \end{array}$$

(Si x se différencie de y et y de x , alors x s'identifie à y et y à x).

INTERPRETATION : L'opérateur de différenciation faible qui vérifie la propriété $[\text{Ant}\sigma]$ engendre une relation d'inclusion sur les classes de l'Univers.

La propriété $[\text{Ip}]$ imposée à un opérateur de différenciation le rend compatible avec les axiomes ensemblistes (de Zermelo-Fraenkel), y compris l'axiome de régularité dont l'effet est, comme on sait, d'impliquer l'irréflexivité et l'asymétrie de la relation binaire d'appartenance engendrée (théorèmes 12 et 13 de W.S. Hatcher: Foundations of Mathematics, Saunders Company, 1968; p.167).

On sait que la copule est en français, peut avoir pour valeur tantôt l'appartenance (Cette figure est un rectangle), tantôt l'inclusion (Les loups sont des mammifères, tantôt la valeur d'identification dont nous avons parlé (Paris est la capitale de la France).

On peut montrer que la relation ontologique ε de S. Leśniewski est une spécification d'une relation de repérage engendré par un opérateur de différenciation faible ayant la propriété $[\tau]$ (conséquence de l'axiome de l'ontologie; voir J.B. Grize, o.c., p.80-82). Nous rappelons que l'objectif de Leśniewski consistait à caractériser directement la copule est en tant qu'opérateur formateur de "noms" et de "propositions".

(c) Ruption. Un objet x ne peut pas être en rupture avec lui-même. Si un objet x est en rupture avec y , alors x et y sont en rupture et, bien entendu, y et x le sont aussi.

INTERPRETATION : Cette valeur intervient dans la constitution des catégories grammaticales, notamment des personnes (IL est en rupture avec JE

et avec TU), de l'aspectualité (l'aoristique est en rupture avec le système des repères énonciatifs), de l'organisation de l'espace (Ailleurs désigne un espace qui exclue l'espace circonscrit par l'énonciateur et son co-énonciateur...). (Cf. Biblio. [CULI-80] [DES-76a] [DES-79a] [DES-80a])

Les trois valeurs identification, différenciation, ruption sont fondamentales dans l'organisation des repérages sous-jacents aux phénomènes du langage.

3.5. Opérateur dual du repérage.

Considérons le cas de l'identification. Plusieurs objets $x_1, \dots, x_n$ peuvent avoir le même identificateur y .

Alors que l'opérateur d'identification est supposé déterministe (au sens suivant: chaque objet x_i a un unique identificateur), si nous associons à cet opérateur d'identification un opérateur dual, celui-ci ne sera pas déterministe : l'identificateur y situe ou détermine (par cet opérateur dual de l'identification) un des objets x_i identifiable à y . Nous avons :

SI l'identificateur y situe ou détermine un objet x_i par identification
ALORS cet objet x_i est identifié à l'identificateur y .

D'une façon générale, nous désignons par $\underline{\exists}_1$ l'opérateur dual associé à l'opérateur de repérage $\underline{\epsilon}_1$ et nous posons par définition $\underline{\exists}$ en notant par :

$$y \xrightarrow{\underline{\exists}_1} x$$

l'action de l'opérateur $\underline{\exists}_1$ sur y donnant pour résultat x :

$$\text{SI } y \xrightarrow{\underline{\exists}_1} x \quad \text{ALORS } x \xrightarrow{\underline{\epsilon}_1} y$$

c'est-à-dire "SI y situe ou détermine un objet x ALORS cet objet x est repéré par rapport à y ".

De même que l'opérateur $\underline{\epsilon}_1$ était constitutif d'une occurrence de relation de repérage, à savoir :

SI $x \xrightarrow{\underline{\epsilon}_1} y$ ALORS $\langle x \underline{\epsilon} y \rangle$ est vrai

de même, l'opérateur $\underline{\exists}_1$ est aussi constitutif d'une occurrence de relation dite de mise en situation ou de détermination :

SI $y \xleftarrow{\underline{\exists}_1} x$ ALORS $\langle x \underline{\exists} y \rangle$ est vrai

L'occurrence $\langle y \underline{\exists} x \rangle$ se lit :

"l'objet y sert de repère à un objet x " ou encore " l'objet y met en situation ou détermine un objet x ".

INTERPRETATION : Nous avons vu que l'opérateur $\underline{\epsilon}_1$ est un opérateur multivalent (dans [DES-80c], cet $\underline{\epsilon}_1$ est appelé 'archi-opérateur' pour signifier la multivaleur qui lui est attachée). On reconnaît là les problèmes qui sont relatifs à l'ambiguïté de la copule est signifiant tantôt l'identification (*Paris est la capitale de la France*), tantôt l'attribution qui se dédouble, depuis G. Frege, en appartenance (*Jean est médecin*), inclusion (*Les hommes sont mortels*), tantôt la localisation (*Jean est à Paris*). Les considérations évoquées ci-dessus laissent entrevoir que l'appartenance et l'inclusion (jamais distinguées par des expressions linguistiques explicites) sont des spécifications d'un repérage par différenciation.

L'introduction de l'opérateur dual $\underline{\exists}_1$ est liée à la dualité avoir / être.

Jean a un chapeau / ? un chapeau est à Jean

Il y a un livre sur la table / Le livre est sur la table

La poche de Jean a un trou / Jean a un trou à sa poche /

La poche de Jean est trouée

La dualité être / avoir a été étudiée par de nombreux auteurs, dont le linguiste E. Benveniste* qui a montré que les observations de nombreuses langues permettaient d'analyser *avoir* comme la relation converse de *être* à. Cependant diverses considérations montrent que des conditions de détermination

*Problèmes de Linguistique Générale Gallimard (1966)

interviennent dans l'analyse de la constitution des relations. Ainsi, on ne dit pas en français :

un livre est sur la table

mais il y a un livre sur la table

ou le livre est sur la table

De même, on ne dit pas en français :

? un livre est à Jean *

mais Jean a un livre.

L'étude de la dualité être / avoir permet de relier entre eux des phénomènes linguistiques comme la localisation, la possession, la mise en présence. La notation $\in/\ni$ du repérage a été introduite par A. Culioli pour noter et décrire ce genre de phénomènes. Une théorie abstraite du repérage, telle que nous l'avons esquissée ici, permettra d'expliquer, du moins nous l'espérons, la nature profonde de la dualité liée aux conditions de détermination des termes reliés entre eux par prédication.

Une théorie mathématique du repérage (présentée en partie dans la thèse de J.P. Desclés [DES-CCc] montrerait que d'une part les valeurs se composent entre elles constituant ainsi une "algèbre d'opérateurs" et que, d'autre part, les valeurs qui peuvent être attribuées à l'opérateur dual ne sont pas libres mais contraintes par les propriétés générales du repérage.

*

* *

Ceci constitue la première partie de l'article. La deuxième partie qui donne un exemple de dérivation formelle d'un énoncé, traité avec les concepts introduits précédemment, sera présentée dans le prochain numéro.

(à suivre)

*Sauf dans l'emploi suivant : Un des livres est à Jean.