

A. CULIOLI

**À propos d'opérations intervenant dans le traitement
formel des langues naturelles**

Mathématiques et sciences humaines, tome 34 (1971), p. 7-15

http://www.numdam.org/item?id=MSH_1971__34__7_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1971, 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/>

A PROPOS D'OPÉRATIONS INTERVENANT DANS LE TRAITEMENT FORMEL DES LANGUES NATURELLES

par

A. CULIOLI ¹

Il n'est pas question de revenir ici sur les problèmes fondamentaux que pose la formalisation des langues naturelles : j'avoue être toujours d'accord avec le cadre théorique esquissé dans l'article *La formalisation en linguistique (Cahiers pour l'analyse, n° 9, juillet 1968)* et repris dans les *Considérations théoriques à propos du traitement formel du langage* ². En particulier, je me permettrai de citer la fin de l'article qui marque les objectifs modestes que la linguistique devrait chercher à atteindre.

« Construire de tels modèles, c'est refuser de *réduire* le langage, et refuser de ramener la linguistique à n'être qu'une collecte de phénomènes individuels ; c'est permettre de poser les problèmes théoriques, se contraindre à une métalangue commune et à des modes de raisonnement rigoureux. C'est ainsi que l'on pourra axiomatiser la linguistique et peut-être la formaliser. »

Cette exigence théorique, alliée à la rigueur technique et au souci de respecter les phénomènes observés, s'impose d'autant plus qu'à l'heure actuelle, la linguistique, sous son apparence « formalisée », joue un rôle idéologique qui mériterait d'être considéré de plus près : l'emprise de la démarche technologique se lie à l'espoir, une fois de plus, de parcourir un espace homogène, grâce à quoi, à petits pas, nous nous glisserions au cœur de la réalité, à moins que, par nos procédures même, nous ne dissolvions cette dernière, au fur et à mesure que la science progresse. Il est remarquable et inquiétant de voir tant d'efforts récents confondre théorie et observations locales, métalangue et techniques de ré-écriture, pour ne citer que ces problèmes. Le danger est qu'en procédant de la sorte on ne sait pas à quel niveau on travaille, on risque d'articuler de façon triviale des domaines qui supposeraient une analyse plus complexe, on tranche dans le vif, par suite d'une conception naïve de la distinction surface/profondeur, masquant ainsi des problèmes qui ou bien resteront cachés ou bien réapparaîtront au hasard des recherches.

Sans entrer dans le détail, indiquons ici à grands traits quelques directions fondamentales qui devraient permettre au mathématicien et au linguiste de collaborer de façon fructueuse ³ et, pour ce faire, nous décrirons rapidement la façon dont nous avons procédé. A partir d'observations quasi-

1. Département de Recherches Linguistiques, Université Paris VII.

2. Antoine Culioli, Catherine Fuchs, Michel Pêcheux, *Considérations théoriques à propos du traitement formel du langage (TFL)*, Document de Linguistique Quantitative n° 7, 1970.

3. Cette formulation pompeuse signifie très simplement que le mathématicien devra comprendre : 1) que toute théorie formalisée doit être interprétable ; 2) que les langues naturelles ne sont pas de la logique qui aurait mal tourné ; 3) que la logique n'est pas un langage qui a subi une ré-éducation.

exhaustives et souvent ténues en apparence, nous nous sommes construit un ensemble d'hypothèses, d'êtres métalinguistiques et d'opérations. Trois règles de conduite se sont révélées fondamentales.

I. Au fur et à mesure que le champ de la recherche s'élargit et que l'on étudie un nombre croissant de phénomènes dans des langues différentes, on doit vérifier la validité de la théorie et l'adéquation de la métalangue. Dans de nombreux cas, on a affaire à des solutions non uniques et non contradictoires ; si, par contre, on découvre une contradiction, il faudra changer de théorie ou de métalangue, ou des deux. Cette exigence peut paraître banale, mais elle n'a pratiquement jamais joué en linguistique, parce qu'on n'a jamais atteint le stade du *calcul* (rappelons que l'objectif de la linguistique formelle est de pouvoir *calculer*, faute de quoi on remplacera un bavardage par un autre). Or, à l'heure actuelle, nous n'avons pas de description tout bonnement factuelle d'une langue quelconque ; en particulier, les études de la linguistique dite descriptive se révèlent, dans la grande majorité des cas, superficielles et inutilisables. De même, il n'existe pas de procédures de description qui garantissent l'exhaustivité. Être exhaustif, c'est noter, puis expliquer que, dans une famille de paraphrases, certains énoncés soient impossibles (ainsi, on aura : *Jean a cassé la tasse ; la tasse a été cassée par Jean ; c'est Jean qui a cassé la tasse ; c'est la tasse qui a été cassée par Jean ; c'est la tasse que Jean a cassée ; c'est par Jean que la tasse a été cassée ; Jean, il a cassé la tasse ; la tasse, elle a été cassée par Jean ; la tasse a été cassée, c'est Jean ; Jean, la tasse, il l'a cassée*, etc., mais pas * *Jean, la tasse, elle a été cassée* pour ne prendre que ce cas. De même, pour renvoyer à la classe, on peut avoir *le chien aime la viande ; les chiens aiment la viande ; tous les chiens aiment la viande ; un chien, ça aime la viande ; des chiens, ça aime la viande ; les chiens, ça aime la viande*, etc. mais pas * *des chiens aiment la viande*).

Être exhaustif, c'est aussi considérer la traduction comme un cas particulier de paraphrase et, donc, s'appliquer à rendre compte dans la même théorie de phénomènes en apparence hétérogènes, tels que l'effacement obligatoire de l'agent dans la transformation passive (attesté dans de nombreuses langues)¹, le complément d'agent marqué par une préposition fort variable (*par*, équivalent de *pour*, *à*, etc.) ou par une forme copulative (ainsi, en zulu *ushaywé : yitshe* « il a été frappé par une pierre », littéral « il a été frappé, c'est une pierre »).

II. Rien n'autorise le linguiste à délimiter son champ théorique selon de simples critères de convenance, et à confondre sans inquiétude le linguistique, le métalinguistique, le langagier. Une fois de plus, le linguiste devra se donner le moins de facilités possible : il formulera ses hypothèses et construira sa métalangue de telle sorte que le psychologue puisse faire affleurer des phénomènes souvent trop fugaces pour être repérés de façon immédiate. Si les observations psycho-linguistiques confirment les hypothèses, cela signifiera qu'on n'a pas nécessairement tort ; si les hypothèses et les calculs qui s'en déduisent ne sont pas vérifiés par l'expérience, il faudra, là encore, abandonner ce qui n'est pas étayé.

III. Le linguiste n'a pas à singer le mathématicien. Son travail est de construire une théorie pré-formalisée, comportant des expressions primitives et des règles explicites de construction, soit par une découverte graduelle des relations profondes entre unités de surface (les invariants seront découverts par approximations successives), soit en se construisant une métalangue perfectible, mais efficace, à partir d'expérimentations (observations théorisées). Au mathématicien de formaliser cette linguistique que l'on appellera naïve ou axiomatisée, selon l'acception plus ou moins forte que le lecteur donnera à ces qualificatifs.

1. Ainsi, en malais ou en arabe, on ne peut avoir * *l'enfant a été mordu par le chien*, mais seulement *l'enfant a été mordu*.

Problème

On considère une famille de phrases ¹ telle que :

Jean conduit la voiture

Jean conduit

Jean a conduit la voiture

Jean a conduit

La voiture a Jean qui la conduit (acceptable ou inacceptable selon les locuteurs)

La voiture a Jean pour conducteur

a voiture est conduite par Jean

La voiture est conduite

On se demande s'il est possible d'établir une ou des relations entre ces énoncés.

Il nous faut, au préalable, définir un certain nombre de concepts.

1) Soit $\vec{P} = (a, b, p)$ un triplet ordonné appelé *relation primitive*, notée aussi $a\vec{p}b$; a est la source et b le but de p . L'orientation primitive (de la source vers le but) dépendra des propriétés de a , b et p (animé/inanimé ; déterminé/indéterminé ; unique/multiple ; intérieur/extérieur ; processus/état, etc.) ou de modulations rhétoriques dépendant de la situation d'énonciation, des présupposés des énonciateurs, pour ne citer que ces cas. L'important est de noter que toute relation est orientée ².

2) A partir de $a\vec{p}b$, on construira un *énoncé primaire*, où l'on a un terme de départ (ou domaine de la relation) et un terme d'arrivée. Le terme de départ sera sélectionné grâce à un opérateur $\underline{\in}$ ³.

Axiome : *L'énoncé primaire est celui qui comporte la même orientation que la relation primitive \vec{P} .*

On aura ainsi :

$$\left\{ \begin{array}{l} \langle \xi_0, \xi_1, \pi \rangle^4 \\ (a\vec{p}b) \end{array} \right. \underline{\in} a$$

que, par convention, nous écrirons $a \underline{\in} a R b$, où R est la valeur que prend p dans le schéma primaire ; (dans la transformée passive, on trouvera $b \underline{\in} b \overset{\cup}{R} a$, où $\overset{\cup}{R}$ marque que l'orientation observée en surface est de b vers a (« $b \rightarrow a$ ») et non pas de a vers b (« $a \rightarrow b$ »). On vérifie que l'on peut écrire l'équivalence :

$$a \underline{\in} a R b \equiv b \underline{\in} b \overset{\cup}{R} a,$$

puisqu'il y a un invariant plus profond : $a\vec{p}b$.

1. Je n'ai pris que le français comme exemple et je me suis restreint à quelques formes, afin de ne pas compliquer la lecture ou alourdir un exposé volontairement schématique.

On pourra étendre l'analyse à l'anglais, de même qu'aux formes plus complexes *la voiture a été conduite par Jean*, par exemple. En fait, on peut généraliser ce qui risque d'apparaître comme une simple astuce, si on le considère séparément.

2. Nous laissons volontairement de côté le problème, au reste aisé à résoudre, des relations réflexives, où l'on a une coïncidence de la source et du but. On n'oubliera pas, en outre, que nous parlons ici de relations profondes (faut-il dire très profondes ?) et non de phénomènes de surface.

3. Sur $\underline{\in}$, voir *TFL passim*.

4. $\langle \xi_0, \xi_1, \pi \rangle$ est un *schéma de lexis*, il comporte trois places, celles des deux arguments ξ_0 , point de départ de la relation, ξ_1 , point d'arrivée de la relation et celle de π , prédicat ou relation entre les deux points. Voir *TFL*, notes VIII et IX, pp. 19 à 25.

3) Dans l'ensemble des relations définies sur l'ensemble structuré à partir du schéma de lexis et du choix du terme de départ, nous considérons une relation distinguée entre a et R qui se caractérise par un ensemble de contraintes, plus ou moins fortes, qui sont trop complexes pour être présentées ici. Disons en bref que l'instanciation $a \in a R b$ ne sera possible que dans certaines conditions précises, variables dans le détail de langue à langue, et plus ou moins apparentes en surface, mais qui comportent nécessairement les opérations suivantes :

I. Le terme de départ sera défini en compréhension et prédiqué existentiellement (« il y a a » « a est là où il est ») ;

II. Il sera repéré par rapport à une situation d'énonciation ;

III. Une opération (liée aux précédentes) restreindra le domaine de la relation du terme de départ (« c'est a et seulement a qui R » / « il y a en tous cas a qui R » / « $a R \dots$ »). Tout ceci définit une *relation lâche* qui sera notée par \cdot^1 ; on aura donc $a \cdot R b$ et $b \cdot R a$

4) Au lieu de conserver les 2 symboles R et $\overset{\curvearrowright}{R}$, nous calculerons les 2 valeurs de p (resp. $R/\overset{\curvearrowright}{R}$) à partir de \in et des valeurs aspectuelles de p .

a) *Valeurs aspectuelles*: $R/\overset{\curvearrowright}{R}$ (c.-a.-d., \vec{p} ayant reçu une orientation observable en surface) prend l'une des deux valeurs suivantes : soit un procès qui prend ses valeurs dans un intervalle ouvert, noté O , soit un procès qui prend ses valeurs dans un intervalle fermé, noté F (O correspond en gros à un déroulement ; F correspond en gros à un processus stabilisé).

Axiome : *La valeur O est primaire et la valeur F dérivée.*

b) *Opérateur* \in : \in a un dual \ni (epsilon miroir), et l'on aura dans le calcul soit \in soit \ni .

On aura les transformations des dualités suivantes :

Étant donné $a \in a \cdot O b$, quand nous changeons le terme de départ, \in devient \ni et vice versa ; quand nous changeons l'ordre de l'énoncé à droite de l'opérateur \in / \ni , \in devient \ni et vice-versa ; quand O devient F , \in devient \ni , et vice versa ².

Ainsi, étant donné $a \in a \cdot O b$, on a : $b \ni a \cdot O b$, puis : $b \in b \cdot O a$, etc.

5) Nous donnons en outre les transformations (ou règles) d'absorption et de permutation suivantes :

$$\begin{aligned} a \in a \cdot O b &\rightarrow a \cdot \in O b^3 \\ b \in b \cdot O a &\rightarrow b \in O \cdot a \\ a \ni a \cdot F b &\rightarrow a \cdot \ni F b \end{aligned}$$

1. Quant à la relation serrée (notée $+$), elle n'entre pas ici en ligne de compte. A propos de la relation lâche, il est intéressant de noter que, dans de nombreuses langues, le terme de départ doit être déterminé ; s'il est indéterminé, il est nécessaire d'introduire un prédicat d'existence explicite.

Ainsi, en nyanga on ne peut avoir que *la fille chante*, à l'exclusion de * *une fille chante* ; en finnois, on ne peut avoir de partitif sujet (* *des garçons coupent du bois*), il faut un déterminant, par exemple *certains garçons coupent du bois*. En malgache, il est impossible d'avoir * *un chien aboie*, on aura nécessairement *il y a un chien qui aboie* ; la situation est la même en chinois, bien que la contrainte y soit moins forte. En français, *il y a un chien qui aboie* est dans la langue quotidienne plus naturel que *un chien aboie*. En Ijo de Kolokuma, on ne peut passiver une phrase que si le complément (qui deviendra sujet) est déterminé : on peut passiver *un crocodile a mangé la femme*, mais pas * *un crocodile a mangé une femme*. En anglais, on aura *there are boys chopping wood in the yard* et non * *boys are chopping wood in the yard...*

2. Voir note 1 de la page 11.

3. Le déplacement du point (\cdot) pourra surprendre. Il a été placé à gauche de \in pour des raisons techniques (passage ultérieur en surface) qui ne seront pas étudiées ici.

mais $b \underline{\ni} b F . a$ reste $b \underline{\ni} b F . a^1$.

6) *Vidage*: Opération par laquelle on vide une place dans une lexis² dont, par ailleurs, toutes les autres places sont instanciées. La place vidée, qui sera représentée par un parenthésage vide, n'a donc pas de valeur directement assignée, mais possède une valeur liée, qui provient d'une relation d'identification anaphorique.

L'anaphore peut être strictement contextuelle ; ainsi, dans : *C'est Jean qui conduit la voiture*, on peut représenter l'opération à l'aide de quatre opérations élémentaires :

1. $a R b$ (par ex. *Jean conduit la voiture*)

2. Vidage : $() R b$

3. Double anaphore, on aura :

— d'un côté un représentant z de la place gauche dans la lexis $() R b$ (en d'autres termes dans $()$ *conduit la voiture*, l'agent est ...), outre une copule d'identité $=$, d'où $z = a$,

— d'un autre côté, un substitut anaphorique q de *Jean*, afin que la lexis $() R b$ soit bien formée, c'est-à-dire comporte un terme de départ. D'où la notation :

$$\overbrace{z = a} \quad \underbrace{(q) R b}$$

qui donne en surface *c'est Jean qui conduit la voiture*
ou *c'est Jean le conducteur de la voiture*.

On peut aussi avoir une anaphore par inférence ; ainsi, dans *Jean est suivi*, on peut représenter l'agent indéterminé Δ_i (« Jean est suivi par quelqu'un », « Jean est suivi par celui qui le suit ») par les opérations suivantes :

1. $d . R a$ (par ex. *on suit Jean*)

2. $a \overset{\vee}{R} . ()$ (par ex. *Jean est suivi* (« par celui qui le suit »).

3. En juxtaposant le passif $a \overset{\vee}{R} . ()$ et l'actif $d . R a$, où d représente en fait l'agent dans la relation $() R a$, on obtient :

$$a \overset{\vee}{R} \overbrace{()} \quad \underbrace{() R a}$$

D'où, soit *Jean est suivi*, soit *Jean est suivi par celui qui le suit*, où l'on retrouve à quelques règles de surface près :

$$\underbrace{a \overset{\vee}{R} . (z)} \quad \overbrace{(q) R a}$$

1. Nous regrettons de devoir donner une suite de règles sans justifications linguistiques, mais il est impossible d'alourdir davantage cet exposé laborieux. Le lecteur fera confiance ou s'irritera, selon son humeur du moment. Disons que les règles ci-dessus *doivent* se rattacher à d'autres règles et se déduire d'axiomes, sous peine de n'être qu'un petit jeu de décisions *ad hoc*, grâce auxquelles on ne retrouverait à la sortie que ce qu'on aurait bien voulu mettre à l'entrée !

2. Il ne s'agit pas de lexis, au sens strict du terme, puisque la voix est liée à la prédication, c'est-à-dire à l'assertion.

Lorsqu'on a affaire à ce second type de vidage, on assignera à la place vidée la valeur \emptyset . Ainsi, on notera $a \overset{\cup}{R} . \emptyset$ pour l'énoncé français *Jean est suivi* (on parle d'ordinaire d'effacement de l'agent).

On obtient ainsi les transformations :

A. Soit $a . R b, b \overset{\cup}{R} . a$.

1 α .	$a \underline{\in} a . O b \rightarrow a . \underline{\in} O b$	1 β .	$a \underline{\ni} a . F b \rightarrow a . \underline{\ni} F b$
2 α .	$b \underline{\ni} a . O b \rightarrow b \underline{\ni} a . O b$	2 β .	$b \underline{\in} a . F b \rightarrow b \underline{\in} a . F b$
3 α .	$b \underline{\in} b O . a \rightarrow b \underline{\in} O . a$	3 β .	$b \underline{\ni} b F . a \rightarrow b \underline{\ni} b F . a$
4 α .	$a \underline{\ni} b O . a \rightarrow a \underline{\ni} b O . a$	4 β .	$a \underline{\in} b F . a \rightarrow a \underline{\in} b F . a$

B. Soit $a . R \emptyset, b \overset{\cup}{R} . \emptyset^1$.

1 α' .	$a \underline{\in} a . O \emptyset \rightarrow a . \underline{\in} O \emptyset$	1 β' .	$a \underline{\ni} a . F \emptyset \rightarrow a . \underline{\ni} F \emptyset$
2 α' .	$b \underline{\ni} \emptyset . O b \rightarrow b \underline{\ni} \emptyset . O b$	2 β' .	$b \underline{\in} \emptyset . F b \rightarrow b \underline{\in} \emptyset . F b$
3 α' .	$b \underline{\in} b O . \emptyset \rightarrow b \underline{\in} O . \emptyset$	3 β' .	$b \underline{\ni} b F . \emptyset \rightarrow b \underline{\ni} b F . \emptyset$
4 α' .	$a \underline{\ni} \emptyset O . a \rightarrow a \underline{\ni} \emptyset O . a$	4 β' .	$a \underline{\in} \emptyset F . a \rightarrow a \underline{\in} \emptyset F . a$

C. Il existe enfin pour des raisons complexes (relation voix-aspect) une règle qui peut se formuler ainsi :

étant donné 3 α' : $b \underline{\in} O . \emptyset$, on a nécessairement 5 β' : $b \underline{\in} F \neq$ et non 3 β' ,

où \neq symbolise la non-référence à a . La relation devient donc réflexive, puisque b est à la fois départ et arrivée (sur ce point, voir *TFL passim*).

Si l'on « traduit » en français ce qui précède, on obtient ² à partir de / Jean /, / conduire /, / voiture / :

1 α . *Jean conduit la voiture* (en anglais : *John is driving the car*) ;

2 α . *La voiture a Jean comme conducteur* ³ (« La voiture a Jean qui la conduit » littéralement « la voiture a Jean conduisant la voiture ») (the car has John driving it / as driver).

3 α . *La voiture est conduite par Jean*.

4 α . * *Jean a la voiture conduite par Jean*.

On pourrait démontrer pourquoi ce schéma à double orientation $a \underline{\ni}$ et $b O . a$ est mal formé ⁴. Qu'il suffise de remarquer que si l'on avait $a \underline{\ni} b O . c$, cela deviendrait acceptable, surtout si l'on établit une relation supplémentaire entre a et b (relation de possession) : *Jean a sa voiture conduite par Paul*.

1 β . *Jean a conduit la voiture*.

2 β . Impossible (mal formé, $b \underline{\in}$ et $a . F b$).

3 β . Impossible (mal formé, « la voiture a la voiture qui ; »).

4 β . Impossible (mal formé, $a \underline{\in}$ et $b F . a$).

1. On s'interdira de partir de \emptyset , car cela ne conduit à aucune interprétation dans ce type d'assertion.

2. Nous laissons naturellement de côté les règles de réécriture qui permettent d'obtenir un énoncé terminal.

3. On peut démontrer que *comme* joue le rôle d'une copule d'identité.

4. Au contraire, (2 α) est possible ; ceci provient de la dissymétrie qui existe entre le schéma primaire $a . R b$ (2 α) et le schéma dérivé $b \overset{\cup}{R} . a$ (4 α).

1 α' . *Jean conduit.*

2 α' . Impossible (mal formé).

3 α' . *La voiture est conduite* (« en train de »).

4 α' . Impossible (mal formé).

1 β' . *Jean a conduit.*

2 β' . Impossible (mal formé).

3 β' . Impossible (mal formé).

4 β' . Impossible (mal formé).

5 β' . *La voiture est conduite* (« résultat »). Ceci est difficilement acceptable avec un prédicat à aspect non terminatif. Avec un terminatif, on a naturellement *la voiture est réparée*.

Le lecteur comblera aisément les lacunes volontaires du raisonnement et la correspondance entre

— participe « présent » et $a \text{ O } b$,

— participe dit passé et $a \text{ F } b$,

— participe (dit passé) passif et $b \text{ O } . a \text{ ou } b \text{ O } . \emptyset$,

— participe (dit passé) de propriété et $b \text{ F } \neq$.

Que l'on n'en conclue surtout pas que \in « est » la copule et \ni le verbe *avoir* ! Ceci n'aurait strictement aucun sens. Mais ce calcul fait apparaître une *relation* entre les auxiliaires *être* / *avoir* et les formes participiales du système verbal français. On remarquera en particulier l'ambiguïté de la forme dite de participe passé qui comporte deux valeurs O/F et est compatible avec 2 orientations « $a \rightarrow b$ » / « $b \rightarrow a$ ». La situation serait fort différente dans une langue où cette ambiguïté n'existerait pas (ainsi dans les langues où l'on distingue un participe actif et un participe passif). Tel qu'il est présenté ici, ce calcul permet de comprendre pourquoi l'on ne peut avoir en anglais * *he has driving*, alors que l'on trouve *he is driving* et *he had somebody driving*...

Ce que nous avons présenté ci-dessus n'est qu'un point minuscule de linguistique, à considérer l'étendue du problème, et nous nous sommes volontairement cantonné dans une rigueur technique qui risque d'accroître le caractère apparemment dérisoire d'un effort démesuré pour un résultat minime ¹.

Que le lecteur, cependant, ne s'y trompe pas ; à chaque ligne surgit un problème qui demanderait un traitement approfondi ; chacune des règles, chacune des opérations, chacun des axiomes devrait être explicité et rattaché, par un enchaînement ordonné, à d'autres ; nous pourrions alors traiter de proche en proche de la passivation, des aspects, des formes de potentiels en *-able* (pour ne prendre que le français) qui unissent de façon si remarquable la diathèse à une modalité (*visible* « qui peut être vu »). Nous montrerions alors que de nombreuses questions longtemps séparées se rejoignent dans une théorie plus puissante. Mais nous ne voudrions ici qu'insister sur un point : tout ce qui a été présenté plus haut se fonde sur l'hypothèse qu'il existe des relations primitives et des schémas primaires, des relations orientées et des systèmes dissymétriques, en particulier qu'étant donné $a \xrightarrow{p} b$ c'est $a \text{ O } b$ qui sera posé comme l'énoncé primaire. Cette formule, à première vue fort simple, permet de rendre compte de phénomènes aussi variés que ceux que l'on trouve dans la syntaxe particulière du complément d'agent dans les transformées passives (effacement ; copule d'identification ; relation lâche ; circonstant de localisation

1. Outre que cette rigueur est manifestement insuffisante d'un point de vue formel.

ou de manière ; etc.) ou dans la syntaxe de « l'agi » (ou patient) ¹. En fait, tout se ramène à un schéma d'analyse où source (relations primitives), agent, départ (en surface) jouent un rôle essentiel. Si l'on inverse la relation, afin de prendre comme départ ce qui était l'arrivée, il se produit une dissymétrie : sur le nouveau terme de départ, on prédique une propriété, puis, éventuellement, on adjoint l'agent qui a provoqué ce nouvel état ².

Or, voici ce qu'écrivait Mme Sinclair, à propos d'expériences sur la passivation chez les enfants de 3 à 6 ans ³ :

« A un premier stade, nous observons des descriptions libres où seule une action indifférenciée est exprimée quant à l'agent et au patient : 'Y roule', 'Ça bouge', etc. Pour les sujets de ce stade notre consigne de commencer la phrase par le nom de l'élément qui subit l'action n'a pas beaucoup de sens — exception faite pour le verbe *casser*, seule situation où l'action résulte en un changement d'état irréversible, exprimable par un adjectif-participe.

« A un stade suivant, les enfants différencient dans leurs descriptions libres l'agent et le patient, et notre consigne centre leur attention sur le patient. A ce moment, ils n'emploient pas une autre forme du verbe qu'ils avaient utilisé dans leur phrase active achevée, mais ils décrivent uniquement le résultat de l'action sur le patient : 'Il est propre' ; 'Il est tombé par terre' ; 'Il avance', etc. Le premier verbe qui est maintenu de l'actif au passif est *casser*, suivi de *renverser*.

« A une étape suivante, l'enfant devient capable de se centrer sur le patient sans pour autant exclure l'agent de sa description ; mais maintenant il décrit l'action selon deux points de vue, au moyen de deux verbes différents, dont l'un peut être identique à celui qu'il a utilisé au mode actif dans la description libre. Nous voyons également une tendance à ne pas remplacer les noms par des pronoms, ce qui donne des phrases comme 'Marianne renverse la tasse', 'La tasse tombe parce que Marianne a renversé la tasse'.

« Finalement, la solution simple et économique devient possible et l'inversion se fait au moyen du même verbe. Toutefois, il y a encore un stade intermédiaire où nous observons des phrases telles que 'Le bâton est cassé et c'est le garçon qui l'a cassé dans ses mains' ; 'La tasse elle est tombée parce que la fille a fait tomber la tasse'. Le premier procédé qui permet de réitérer le verbe sous une autre forme et de donner une description complète avec inversion, semble être 'se faire + infinitif + par'. Ceci, nous semble-t-il, rappelle certain stade dans la notion de causalité, où pendant longtemps l'enfant attribue aux objets les actions d'un sujet vivant. L'expression 'se faire + infinitif + par' n'est pas du tout réservée aux êtres animés, mais elle s'applique à tout : la tasse se fait renverser, le camion se fait laver, la bille se fait pousser, etc. Il nous semble que la préférence pour ce procédé s'explique en partie par la tendance à attribuer une action ou au moins une intention au sujet de la phrase. Il est à noter que toutes les phrases à inversion correcte reprennent le verbe utilisé dans la phrase active donnée en description libre. ⁴ »

1. Il n'existe pas de formule magique, et la formule en question demanderait à être développée et commentée. Une fois de plus, que le lecteur se résigne ou s'irrite, à son gré.

2. Nous ne donnerons pas la représentation formelle qui soulèverait des problèmes complexes d'écriture. On a, très approximativement :

$$b \text{ O } \cdot \overbrace{(\quad)}^{\quad} a \cdot \text{O } b \quad \text{où} \quad b \text{ O } \cdot (\quad) \text{ est très voisin de } b \text{ F } \neq.$$

3. H. Sinclair, « L'acquisition des structures syntaxiques », *Psychologie française*, 13 (2), 1968, pp. 167-174. Que le lecteur pardonne la longueur de cette citation, mais ces très belles observations se prêtaient mal aux coupures.

4. On pourra rapprocher ceci du passage suivant tiré de *l'Esquisse de la langue hongroise* d'A. Sauvageot, Klincksieck, Paris, 1951, « le hongrois ne possède pas de voix passive du verbe et ne se sert, pour exprimer la notion d'action subie, que d'expédients divers dans le genre de ceux que nous venons d'analyser. Ce que les grammairiens hongrois appellent le passif n'est qu'un factitif rendu inchoactif par l'adjonction de la désinence *-ik* » (cette dernière a une valeur complexe réfléchie — intransitive, pour ne citer que ces deux traits, A.C.).

On peut démontrer qu'un causatif et un réfléchi donnent une relation équivalent à un passif. Ce dernier se présente en effet comme un schéma où l'on a une propriété prédiquée (départ et arrivée confondus), donc formellement de type réflexif (voir *TFL*). Quant au causatif avec réfléchi, il peut se représenter comme suit : $b \text{ R } \emptyset \text{ R } b$, c'est-à-dire un schéma où départ et arrivée coïncident.

Cette convergence des recherches linguistiques et des travaux psycho-linguistiques paraît être à l'heure actuelle l'une des meilleures « démonstrations » des « théorèmes » de toute théorie pré-formalisée du langage appréhendé à travers la diversité des langues naturelles. Il faudra encore beaucoup d'observations et de vérifications avant que l'on puisse bâtir un modèle mathématique, si l'on veut bien ne pas donner un sens trop dérisoire au terme. Entre temps, le mathématicien et le logicien auront sans nul doute aidé le linguiste à traquer les inconsistances, à renforcer la rigueur sans laquelle on travaille à l'aveuglette, à simplifier une écriture trop lourde. Que le mathématicien, chemin faisant, ne cède pas à la tentation de se transformer en métaphysicien, démiurge ou, à l'inverse, fabricant de trivialités.