



HAL
open science

Trois essais de formalisation en linguistique : Harris, Saumjan, Chomsky

Mitsou Ronat

► **To cite this version:**

Mitsou Ronat. Trois essais de formalisation en linguistique : Harris, Saumjan, Chomsky. *Action poétique*, 1970, 43, pp.45-73. halshs-04005652

HAL Id: halshs-04005652

<https://shs.hal.science/halshs-04005652>

Submitted on 26 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open licence - etalab

« Si bornée que soit en fait la nature humaine, il y a cependant en elle une très grande part d'infini, et je vais jusqu'à soutenir que si elle n'était pas elle-même infinie sous bien des rapports, on ne saurait expliquer la conviction et la certitude assurées où nous nous savons tous unis, touchant l'être de l'absolu. »

CANTOR.

(Fondement d'une théorie générale des ensembles).

A l'heure actuelle, les recherches en linguistique se poursuivent essentiellement dans deux grandes directions :

- *l'une, qui est l'aboutissement du structuralisme et de la linguistique descriptive, et dont le but est la découverte d'une grammaire d'une langue à partir des données d'un corpus déterminé fini (par exemple : les énoncés recueillis dans les journaux). Les travaux de Harris et de Saumjan sont les plus représentatifs de cette tendance ;*
- *l'autre, élaborée par Chomsky, qui a pour base la théorie des modèles context-free (grammaires indépendantes du contexte), et dont le but est la construction d'un mécanisme qui pourrait engendrer toutes et rien que les phrases d'une langue, c'est-à-dire une infinité.*

Avant de développer de façon quelque peu détaillée le contenu de ces théories, il convient de les replacer dans leurs contextes historiques et géographiques, et de rappeler également que, dans ce domaine comme

dans tous les autres domaines scientifiques, l'évolution ne s'est pas faite de façon isolée, « spontanée », mais en fonction des échanges constants que permettent les publications internationales.

U.R.S.S. 1920

C'est l'époque très féconde et effervescente des formalistes russes. Elèves eux-mêmes des linguistes Fortunatov et Baudoin de Courtenay, les homologues de Saussure en Europe de l'Est, ils mirent au programme de leurs recherches tout ce qui concernait le langage, la poésie, la littérature. Après le départ de Jakobson et de Troubetzkoï pour l'Europe centrale, la linguistique comme science a subi une longue hibernation. L'appui officiel accordait le monopole à la théorie dite « Japhetologie » dont N. Ja Marr fut le promoteur. Cette théorie démente, prétendue marxiste, associait l'histoire de l'évolution des moyens de production à celle du langage, la langue étant considérée comme une superstructure au même titre que les institutions juridiques, la religion, etc.

Les sociétés « primitives » auraient utilisé d'abord un langage gestuel, puis un langage composé seulement de quatre syllabes (sal, ber, jon, rosh), qui se serait compliqué avec l'évolution des techniques et de la division du travail.

L'ont payé cher, les opposants — tel Polivanov — à cette théorie.

En 1950, l'intervention de Staline met un terme à ces fantaisies dangereuses. La linguistique soviétique devra encore réagir contre ce dogmatisme nouveau : la méfiance contre la propagande du capitalisme entraîne la mise à l'index des œuvres scientifiques occidentales.

Le rôle de Saumjan, dès 1952, pour la réhabilitation du structuralisme tel qu'il avait été défini à Prague, est capital. Après la mort de Staline, il met en avant les succès des recherches mathématiques et structu-

rales en linguistique appliquée, et les profits que les capitaux privés et l'armée américaine comptent obtenir des résultats de telles investigations. Par un curieux renversement, le même Saumjan, qui a lutté contre toute confusion entre science et idéologie, est aujourd'hui proposé comme caution d'un autre dogmatisme. Ses travaux se situent dans le cadre des recherches pour la construction des traductrices automatiques.¹

C'est pourquoi il nous a paru plus adéquat de le situer par rapport à Harris (Pennsylvania University) avant de l'opposer directement à Chomsky (M.I.T.).

Un bref rappel des théories de Harris s'impose pour éclairer l'étude ultérieure du modèle proposé par Saumjan.

Z.S. HARRIS

Harris s'est trouvé l'héritier de l'école structuraliste américaine, dirigée par Bloomfield, Bloch, etc., et dont l'analyse principale, en ce qui concerne la syntaxe, était la décomposition des énoncés en Constituants Immédiats, c'est-à-dire une catégorisation hiérarchique des diverses séquences d'éléments, de l'énoncé aux éléments minimaux appelés morphèmes.

Un exemple de ce genre d'analyse est donné par les fameuses « boîtes » de Hockett complétées par les « étiquettes » catégorielles :

<u>Le</u>	<u>détective</u>	<u>suiv</u>	<u>ait</u>	<u>l'</u>	<u>espion</u>
		racine verbale	affixe temporel	article	nom
article	nom	verbe		Syntagme nominal	
Syntagme nominal		Syntagme verbal			
Phrase					

1. Ceci explique l'emploi indifférent des termes « linguistique mathématique », « structurale », « appliquée », dans la « littérature » linguistique soviétique.

Harris a d'abord² entrepris de procéder à cette analyse en parcourant le chemin inverse, établissant ainsi les bases de sa théorie distributionnaliste : chaque élément (ou séquence d'éléments) est défini par l'environnement dans lequel il peut apparaître, chaque élément d'une même classe étant substituable l'un à l'autre dans un même environnement : Ex :

la classe des substantifs (clochard, cheval, vélomoteur, cochonnet, gouvernement...) dans le contexte :/ le — va tout droit/.

Cette analyse devrait être affinée par diverses restrictions, puisque, par exemple, on ne peut pas trouver dans le contexte :/le — chante/des mots tels que : amidon, chapeau,³ qui sont pourtant des substantifs ; on devra donc établir des sous-catégories, et leurs relations de co-occurrence.

Parmi ces relations, on peut distinguer les constructions endocentriques, où la juxtaposition d'une catégorie « mineure » ne change pas la catégorie du constituant, ex. : $BA \rightarrow A$, ou $\text{Adj.} + N \rightarrow N$ (la flèche signifie : dans tout environnement, le membre droit de la formule peut être substitué au membre gauche). — les constructions exocentriques, où la présence d'un morphème fait changer de catégorie le morphème voisin : $AB \rightarrow C$ ou (anglais) $\text{Adj.} + \text{ly} \rightarrow \text{Adv.}$

Dans une seconde étape de sa théorie, Harris analyse les phrases en y dégagant un *chaînon central*, auquel, pour reconstituer l'énoncé entier, il faut ajouter à gauche ou à droite, et par degré, les divers adjoints définis par les catégories initiales auxquelles ils sont assignés.

Ex. : *Aujourd'hui, la jolie capsule tourne autour du soleil.*

2. *From Morpheme to Utterance*, in *Language* 22 1946, en Français, in *Langage* 9 1968.

3. Il résulte de ceci que, jusqu'à maintenant, on ne peut considérer les textes surréalistes — et poétiques en géné-

Nous avons la séquence élémentaire (centrale) suivante :

— la capsule tourne

[élémentaire

— aujourd'hui : adjoit à gauche de : la phrase

— jolie : » » » : capsule

— autour du soleil : adjoit à droite de : tourne.

La (dé)composition est ainsi figurée par une suite d'énoncés partiels de longueur (dé)croissante.

Une troisième étape est consituée par l'analyse dite *transformationnelle*. Chaque phrase est décomposée en une ou plusieurs séquences élémentaires (avec zéro ou plusieurs adjoints primitifs, accompagnées de zéro ou plusieurs transformations dites « unaires », c'est-à-dire opérant à l'intérieur de cette séquence); les opérations qui relient ces séquences élémentaires entre elles sont appelées transformations « binaires » (de phrase à phrase) : Ex :

Aujourd'hui, des rayons verts sont émis par la jolie capsule dans la galaxie.

	<i>Adj. primitifs</i>	<i>T. unaires</i>	<i>T. binaires</i>
1) <i>la capsule émet des rayons</i>	(de Ph.) aujourd'hui (de V.) dans la galaxie	passif	•
2) <i>la capsule est jolie</i>	•	identité	encastrée dans 1)
3) <i>le rayon est vert</i>	opérateur de Ph. : pluriel	•	• •

Pour Harris, ces trois analyses sont indépendantes, toutes aussi pertinentes — bien que leur puissance aille dans l'ordre croissant — car, de toute façon, le langage les contient toutes les trois.

De telles descriptions conduisaient naturellement à considérer ce qui, dans les mathématiques moder-

ral — qu'en termes de déviation, d'infraction, d'agrammaticalité, d'inacceptabilité.

nes, pouvait servir à la linguistique, en définissant d'autre part ce qui, dans les propriétés du langage, pouvait être formulé mathématiquement.

Il résulte des analyses précédentes que toutes les phrases d'une langue peuvent être réduites à un ensemble fini et restreint de formes élémentaires possibles pour cette langue (la grammaire doit en dresser la liste). On peut obtenir toutes les phrases de cette langue au moyen d'une méthode récursive, à partir de cet ensemble, et avec un ensemble relativement petit d'opérateurs.⁴

Harris a donc été amené à construire un *système abstrait* dont la condition « sine qua non » est d'être adéquat au langage. Ce système doit être le même pour toutes les langues ; il consiste en une famille primitive d'*arguments* et 5 familles finies d'opérateurs, chacune de ces 5 familles étant définie comme opérant soit sur les arguments primitifs, soit sur une autre famille d'opérateurs. La composition interne de ces familles dépend de l'interprétation particulière à chaque langue.

Pour formuler ce système abstrait, — dont les objets sont définis uniquement par les relations dans le système —, il faut :

- 1) un alphabet fini de symboles,
- 2) un ensemble fini de formes de séquences de symboles,
- 3) un processus récursif dépendant.

Ce système abstrait est un sextuple ordonné :

$\langle N, F^*, F_1, F_2, F_3, F_4 \rangle$

dont N (N ens. fini) est l'argument primitif,
(q ens. infini)

4. Nous omettons ici l'étude de l'automate cyclique « d'annulation ». Chaque énoncé y est représenté par des paires de symboles inverses indiquant la relation, le certificat de « bonne construction » étant décerné si toute la phrase peut être effacée cycliquement par l'automate.

et F° , F_1, \dots les symboles fonctionnels.

Chaque F est un ensemble fini d'opérateurs, chaque opérateur introduit un « matériau » qui, soit, 1) est concaténé avec son opérande ; soit, 2) apporte un changement à cette opérande.

On peut faire le produit (: des applications successives) des opérateurs.

Appliqué à l'anglais, l'argument primitif N sera le Nom simple ;

F° : des modifieurs comme : some,...

F_1 : les verbes transitifs et intransitifs,... (Cf. : Math. Str. of Lang.), etc.

Deux exemples d'interprétation :

$F_1 : N \rightarrow N$ F_1 : le *exists* de la phrase « John exists »,

$F_2 : N_h, F_4 \rightarrow N_h$ $F_2 F_4$: le *knows* de la phrase : Jack knows that John came ».

(h = humain)

Ce système est, en définitive, assez lourd, mais il est amélioré par plusieurs procédés, entre autres, un ensemble de transformations unaires (modulation, temps,...) et binaires (relations entre les phrases).

Cet exposé mille fois trop succinct, et par là même peut-être déjà fastidieux, de la théorie de Harris, était cependant indispensable pour éclairer ce qui va suivre.

Mais, il nous faut d'abord voir quelle est la conception de la langue qui est implicite (ou explicite) dans son œuvre.

Pour Harris, et par suite, pour tous les structuralistes, une langue est un ensemble de phrases infini dénombrable. Ce sont les caractères :

- 1) discret — au sens linguistique du terme,
- 2) préorganisé (code sous-jacent connu de l'émetteur et du récepteur, ce qui leur permet de communiquer),
- 3) arbitraire (articulation signifiant/signifié saussurien),

qui sont retenus pour cette définition.

La terminologie relève du structuralisme (Cf. : Phonologie).

Les éléments doivent être ordonnés linéairement en séquences successives.

Les phrases, même très longues sont considérées comme finies (avec possibilité d'éléments récusifs), pour que toutes les conditions de leur « bonne forme » soient respectées.

Cette finitude entraîne diverses propriétés de la grammaire, et permet de traiter les « idioms » et autres exceptions comme des phénomènes marginaux, comme autant d'extensions des opérations régulières.

La contiguïté des objets (n'existant qu'en relation avec leurs voisins) est un impératif fondamental.⁵

Emetteur, discret, arbitraire, linéaire, contigu, pré-organisé, récepteur : ce sont des concepts de la théorie de l'information.⁶

5. Citons, à ce propos, la phrase merveilleuse de Harris, glaçant à blanc la poésie moderne :

« Toute structure de phrase bien formée doit par conséquent exiger une séquence contiguë d'objets, la seule propriété qui fait de cette séquence un schème de la grammaire étant que les objets ne sont pas des mots indifférents mais des mots d'une classe particulière (ou des classes particulières de mots). Mais la séquence doit être contiguë, elle ne peut pas être morcellée, avec des espaces "in between", parce qu'il n'y a aucun moyen d'identifier ou de mesurer les espaces. »
(comme les barres, en musique).

6. L'importation massive des concepts issus de la théorie de l'information telle qu'elle a été établie en 1948-49 par N. Wiener (Cybernetics) et Shannon et Weaver (The Mathematical Theory of Communication) est capitale pour la linguistique. Des notions telles que : information, imprévisibilité, codage, redondance (dans une définition spécifique), bruit, déformation, neg-entropie (issue du vocabulaire physico-mathématique), coût, économie, ou du moins certaines d'entre elles, font très souvent partie du discours linguistique, sans que soit précisée explicitement leur origine.

L'établissement d'un tel système abstrait présuppose la découverte de structures fondamentales communes à toutes les langues, et d'opérations applicables à ces structures, avec pour but final la construction de traductrices automatiques.

SAUMJAN

C'est dans cette problématique que s'inscrit Saumjan, et c'est précisément cette façon de poser les questions que Chomsky rejette.

Dans le bref exposé que nous allons faire du modèle applicatif de Saumjan, nous pourrions constater un

A ce sujet, il est intéressant de consulter le compte rendu du colloque de Royaumont de 1965 sur « le concept d'information dans la science contemporaine ». Le discours inaugural a pour thème la théorie des animaux-machines de Descartes, puis Wiener (titulaire de chaire au M.I.T.) fait un exposé sur les machines à jouer aux dames (et aux échecs) — confirmant ainsi que l'on peut toujours associer une machine à un système de règles —, et sur ses découvertes « humano-mécaniques », tel le bras artificiel mû par l'influx nerveux amplifié par des transistors.

La psychologie behavioriste a aussi sa place, mais de façon renversée : il y a des machines qui apprennent, c'est-à-dire, le système de règles étant donné, qui améliorent leur jeu et leur tactique à force de jouer, se faisant ainsi une « personnalité ». D'autre part, on affirme que le raisonnement déductif est un mécanisme enseigné qui s'insère dans l'organisme, comme une machine à calculer, ceci étant prouvé par le parfait fonctionnement comme correcteur de copies d'un professeur paranoïaque de l'Institut National d'Enseignement par correspondance de Vanves.

Nous sommes donc en plein cœur de la problématique cartésienne, aussi bien en ce qui concerne la Méthode qu'en ce qui concerne l'opposition homme/automate-animal. Témoin encore ces recherches (zoologiques ?) sur le Yéti, l'homme velu, l'hominien, l'homme sauvage chassé par les Tzars (Europe n° 1, samedi 1^{er} novembre 1969, le matin).

parallélisme certain avec le système abstrait de Harris. Et dans la critique que nous ferons de ce modèle, apparaîtront les premiers concepts qui ont décidé Chomsky à repousser ce genre de recherches, et à imaginer les grammaires génératives.

Dans son livre « *Linguistique Structurale* » (1965), Saumjan définit son modèle à partir des structures logiques du modèle abstrait non-linéaire de H.B. Curry, l'« *ob system* » (If X and Y are obs, then XY is an ob).

La seule opération — binaire — est appelée application.

Son hypothèse de départ est la possibilité de construire un modèle dont les éléments ne seraient pas ordonnés comme ils le sont dans le système abstrait de Harris. Ce modèle a d'abord pour tâche d'engendrer une langue idéale appelée « *génotype* », reliée par des règles de correspondance (ou d'interprétation) aux langues naturelles dites langues « *phénotypes* ».

Cette première hypothèse appelle déjà plusieurs critiques.

D'une part, Harris y oppose un argument de détail mais qui a son importance : il est impossible, avec un système non ordonné, de rendre compte des *références croisées*, dans des constructions du type :

« *Bilbao, Bologne, Istamboul sont, respectivement, son oiseau, son cheval, et son puma favoris.* »

D'autre part, et surtout, pour aucune langue, on ne peut affirmer que l'ordre dans lequel apparaissent les mots soit entièrement indifférent. Même dans une langue flexionnelle très riche, où les variantes stylistiques possibles sont en grand nombre, cette indépendance est toujours limitée par le risque de *l'ambiguïté*.⁷

7. Malgré la métaphore de Freud, dans sa quête des analogies avec les indéterminations du rêve, et son étude du rôle de l'auditeur ou du lecteur dans l'inter-

Chomsky donne cet exemple : « Die Mutter sieht die Tochter », où — à moins d'un accent contrastif — « Die Mutter » doit être interprété comme sujet ; la flexion ne suffisant pas à indiquer la fonction grammaticale, une permutation changerait complètement le sens de la phrase.

En outre, les diverses recherches ont le plus souvent imposé la détermination précise d'une « structure de base » strictement ordonnée, à laquelle il soit possible d'appliquer toutes les transformations de permutation acceptée par la langue en question. Postuler des règles « ensemblistes » du type $A \rightarrow (B, C)$, où $A \rightarrow BC$ ou $A \rightarrow CB$ indifféremment, revient au contraire à poser un double jeu de règles ; la règle « ensembliste » peut en effet s'appliquer en deux temps :

- 1° concaténation,
- 2° permutation.

(il ne s'agit évidemment pas d'un temps physique).

On peut aussi procéder dans le sens inverse, en ajoutant aux règles de concaténation une règle qui libérerait l'ordre des éléments. Il semble que cela soit une complication inutile.

Cependant, en admettant l'hypothèse, considérons le modèle lui-même. Il se compose de 4 sous-modèles appelés générateurs :

- 1) le générateur abstrait qui engendre la langue idéale, les symboles catégoriels ;
- 2) le générateur de mots ;
- 3) le générateur de phrase ;

prétation, même une langue comme le chinois possède une grammaire, dont la forme est évidemment radicalement différente de celles des langues indo-européennes. Le « pour ainsi dire » (Intr. à la Psych. P.B. Payot, p. 216) marque sa réserve, d'ailleurs, envers son affirmation. Freud projette ses connaissances des catégories grammaticales aristotéliennes sur le chinois, se félicite et s'émerveille de ne pas les retrouver.

4) le générateur du champ transformationnel.

1) le générateur abstrait

L'élément minimal est le sémion, comparable au trait distinctif en phonologie, à la catégorie grammaticale en syntaxe, au *marker* en sémantique.

La combinaison des sémions est dite : faisceau de sémions.

Une relation appelée « représentation » est postulée : un sémion ou un faisceau représente un épisémion déterminé.

Premier travail du générateur abstrait : fournir les deux épisémions initiaux, α et β (Cf. : les arguments primitifs). Les autres épisémions sont engendrés par les règles suivantes :

a) α et β sont des épisémions.

b) Si p et q sont des épisémions, (pq) est un épisémion.

L'opération peut être répétée indéfiniment.

Deuxième travail : engendrer des sémions et des faisceaux de sémions.

Le sémion qui représente l'épisémion x est noté ' x '.

Les faisceaux de sémions sont engendrés par les règles suivantes :

a) Tous les sémions sont des faisceaux de sémions.

b) Si ' x ' est un sémion représentant l'épisémion pq et ' y ' est un faisceau de sémions représentant l'épisémion p , alors ' x ' ' y ' est un faisceau de sémions représentant l'épisémion q .

Dans l'application (' x ' ' y '), ' y ' est l'opérande, ' x ' est l'opérateur.

Le faisceau de sémions (' x ' ' y ') est aussi appelé *image*.

Les faisceaux sont des opérateurs de deux sortes :
— les *identifieurs* : ce sont des opérateurs où les

opérandes et images correspondantes représentent les mêmes épisémions ;

- les *mutateurs* : ce sont des opérateurs où les opérandes et les images correspondantes représentent des épisémions différents. (Cf. supra).

Le système peut être étendu à l'aide d'opérateurs appelés « relateurs » : les adnecteurs (la négation,...), et les connecteurs (et,...).

Quelques exemples d'interprétations de ce modèle pour le russe :

degré 0 α syntagme nominal
' α ' nom au nominatif
 β phrase
' β ' phrase simple

degré 1 $\alpha\alpha$ déterminant de syntagme nominal
 $\alpha\beta$ syntagme verbal

degré 3

$\alpha\beta.\alpha\beta$: $\alpha\beta.\alpha\beta$ déterminants du déterminant du syntagme verbal

' $\alpha\beta.\alpha\beta$: $\alpha\beta.\alpha\beta$ ' adverbess qui sont les déterminants des adverbess préverbiaux, etc.

Les catégories grammaticales sont assez éloignées des catégories traditionnelles. Cependant, il arrive que le plus souvent l'interprétation donnée n'en soit pas très claire, en particulier en ce qui concerne la différence entre sémions et épisémions.⁸

Si la différence entre α (syntagme nominal) et ' α ' (nom) reste claire, il n'en est pas obligatoirement de même pour $\alpha.\alpha\alpha$ (: transformeur de syntagme nominal en déterminant de syntagme nominal), et ' $\alpha.\alpha\alpha$ ' (: affixes formant des adjectifs à partir des noms), à cause de l'imprécision des définitions.

On peut aussi comparer le modèle applicatif avec

8. Nous reprendrons ici, en grande partie, l'analyse faite par Ferenc Kiefer, chef du groupe de recherches pour la traduction automatique de Budapest.

la description de Bar-Hillel (*On categorial and phrase-structure grammar*, 1960, dont Saumjan cite ailleurs les travaux (*La cybernétique et la langue*, Diogène n° 51).

Si nous remplaçons α par n , et β par s (les opérations sont/ou\) :

' α ' $\rightarrow n$

' β ' $\rightarrow s$

' $\alpha\beta$ ' $\rightarrow n \setminus s$

' $\alpha\alpha.\alpha\alpha$ ' $\rightarrow (n/n)/(n/n)$ etc.

Il faudrait justifier la distinction β : phrase et ' β ' : phrase simple. Kiefer va jusqu'à dire que dans la plupart des cas, les *épisémions sont des catégories vides*, c'est-à-dire qu'ils n'ont pas d'interprétation linguistique.

L'engendrement des faisceaux peut se faire selon 4 schémas.

1) itération

' $\alpha\alpha$ ' \times ' α ' = (' $\alpha\alpha$ ' ' α ')

' $\alpha\alpha$ ' \times (' $\alpha\alpha$ ' ' α ') = (' $\alpha\alpha$ ' [' $\alpha\alpha$ ' ' α ']), etc.

2) réduction

' $\alpha\alpha.\alpha\alpha$: $\alpha\alpha\alpha\alpha$ ' \times ' $\alpha\alpha.\alpha\alpha$ ' = ([' $\alpha\alpha.\alpha\alpha$: $\alpha\alpha.\alpha\alpha$ '] [' $\alpha\alpha.\alpha\alpha$ '])
 $\alpha\alpha \quad \bullet \quad \alpha\alpha$

(Cf. : plus haut : adj. + $N \rightarrow N$)

3) conversion : 2 opérations : permutation, interapplication.

4) connection : engendrement au moyen des connecteurs.

2) *Le générateur de mots.*

Il engendre un nombre infini de mots à partir d'un ensemble fini d'objets initiaux, par le moyen de relateurs

— un ensemble fini de mots élémentaires m_a, m_b, \dots, m_d

— un ensemble fini de relateurs de mots R_a, R_b, \dots, R_d

Chaque mot élémentaire a une certaine structure,

résultant de l'application de l'un des mots relateurs à un objet auxiliaire, la « racine » du mot, dont le rôle est joué par un sémion vide \emptyset : $R_n\emptyset, R_n\emptyset\dots$,

On engendre les mots au moyen des deux règles suivantes :

- a) les mots élémentaires sont des mots,
- b) si X est un mot, alors RX est un mot, et R est un relateur variable ; RX est la *dérivation* R_n du mot X.

Saumjan distingue les mots de 1^o, 2^o..., n^o dérivation ; le générateur travaille par cycles successifs :

1^o cycle : $R_1\emptyset, R_2\emptyset, R_3\emptyset, R_4\emptyset, R_5\emptyset$

2^o cycle : $R_1R_1\emptyset, R_1R_2\emptyset\dots$

$R_2R_1\emptyset, R_2R_2\emptyset\dots$

.....

etc.

Deux exemples d'interprétation pour le russe :

$R_n\emptyset$ = noms au nominatif,

$R_nR_1\emptyset$ = noms dérivés de verbes, infinitifs et gérondifs.

Il ressort de l'étude de ce modèle que des mots de structures différentes ont la même description, et qu'il ne rend pas compte de certains processus de dérivations. Les classes de mots sont définies par leurs propriétés fonctionnelles seulement, c'est-à-dire que sont classés dans une même catégorie, de façon assez sommaire, tous les mots qui jouent le même rôle dans la phrase. (Cf. : analyse distributionnelle).

Le générateur de mots fait partie, en réalité, du générateur de phrases.

Une des premières critiques que la grammaire générative a formulée à l'égard des analyses de type structuraliste fut de ne pas distinguer la structure profonde sous-jacente, et de ne s'attacher qu'à l'étude de la structure superficielle. Cette distinction est absolument nécessaire à l'explication de bien des phénomènes du langage. Barbara Hall (c.r. de Saumjan et Soboleva, 1963, Language XL) donne les exemples sui-

vants, où tous les noms qui suivent le verbe sont à l'accusatif :

1. Ivan čital čas : Ivan lisait une heure,
2. Ivan čital knigu : Ivan lisait un livre,
3. Ivan čital gazetu : Ivan lisait un journal,
4. Ivan čital knigu i gazetu : Ivan lisait un livre et un journal,
5. Ivan čital knigu čas : Ivan lisait un livre une heure.

Dans le système Šaumjan, les phrases (1, 2, 3) ont la même description, et (4, 5) aussi. Il ne rend pas compte de l'agrammaticalité de la phrase : « Ivan čital knigu i čas ».⁹ Ce système permet donc l'engendrement d'une phrase qui n'appartient pas au langage.

Barbara Hall ajoute que si le modèle applicatif développe largement la morphologie aux dépens de la syntaxe, c'est conformément à la tradition russe — mais nous ne pensons pas que ce genre d'arguments soit d'un grand intérêt.

Le système du générateur de mots est aussi très redondant. Šaumjan ne donne que quelques exemples d'interprétation, mais il apparaît que même au 2^e degré de dérivation, il y a des formules qui n'ont pas d'interprétation linguistique.

3) *Le générateur de phrases.*

Il utilise un processus identique au générateur de mots, et appelle les mêmes commentaires.

Une phrase est un énoncé qui se divise en 2 sous-énoncés, représentant les épisémions α et $\alpha\beta$, respectivement : syntagme nominal et syntagme prédicatif.

Les objets initiaux du générateur de phrases sont :

- 1) un ensemble *infini* de mots, M_1, M_2, \dots , appelés *phrases élémentaires*,
- 2) un ensemble *fini* de relateurs R_1, R_2, \dots, R_d ,

9. La G.G. attribuerait à chacune de ces phrases une description structurale différente : (voir *infra*).

- 3) un connecteur C,
- 4) un adnecteur A.

B. Hall remarque encore que si l'on peut, avec ce système, engendrer de longues suites d'adjectifs modifiant un nom, des syntagmes verbaux et nominaux (objet) composés, il n'est pas possible d'engendrer une phrase ayant un sujet composé. C'est là un grave défaut pour un modèle.

4) *Le générateur du champ transformationnel.*

N'est pas justifiée la déclaration de Šaumjan selon laquelle le modèle applicatif est « beaucoup plus général que le modèle des constituants immédiats ou le modèle transformationnel ».

En effet, il définit la transformation comme une règle établissant l'invariance des relations entre les phrases, c'est-à-dire de la même façon que l'avait définie Harris dans son article *Co-occurrence and transformation in linguistic structure*, (Language XXXIII).

Une phrase grammaticale (normale) le reste après transformation, une phrase agrammaticale (anormale) aussi :

OK : le chat mange le dessert \longleftrightarrow OK : le dessert est mangé par le chat.

* : le dessert mange le chat \longleftrightarrow * : le chat est mangé par le dessert.

Cette notion a été affinée par la suite en « différence d'acceptabilité ».

Il formule d'autre part sa définition en termes d'application, qui indique le processus génératif.

Là aussi, la redondance est inévitable, à cause de la croissance très rapide du nombre des transformations à chaque degré de dérivation. Déjà, au second degré, 1.750 transformations sont possibles, dont les trois quarts n'ont pas d'interprétation linguistique.

Šaumjan dit que son modèle peut ainsi « prévoir » toutes les transformations de toutes les langues.

Il faudrait le vérifier en l'éprouvant dans une typologie structurale.

Cette définition — restrictive — de la transformation implique certaines carences, graves elles aussi.

— Toute opération est conçue comme allant du simple au complexe, ce qui est vrai pour : « *le ciel bleu* → *le bleu du ciel* » mais ne l'est plus pour « *le fou chantant* → *le fou qui chante* ».

— On ne trouve nulle part l'explication des transformations qui entraînent l'effacement ou l'apparition d'un élément. Le modèle applicatif peut relier (1) à (2), mais ne peut pas relier (3) à (4), ni (5) à (6) :

(1) tous les garçons sont venus

(2) les garçons sont tous venus

(3) Mac trouve que Sennet est fou

(4) Mac trouve Sennet fou

(5) que Django tire plus vite effraye Josh

(6) ça effraye Josh que Django tire plus vite

— Les critères manquent pour la détermination de la « correction » des transformations réelles, ainsi que les indications de leurs relations particulières.

En conclusion, si la simplicité est une qualité, il est clair que le modèle applicatif est simple quant au nombre restreint de ses symboles initiaux et à son unique opération fondamentale, mais qu'il ne l'est plus au niveau du nombre total des symboles utilisés pour la description, ni par son degré de redondance générale. Le système proposé ne remplit pas les promesses de son présentateur.

Kiefer rejoint d'ailleurs Harris et Chomsky dans leur conception des rapports entre les mathématiques et la linguistique : l'utilisation des mathématiques n'est possible (et intéressante) que si le modèle mathématique est adéquat au langage, s'il engendre toute et rien que la langue naturelle donnée.

« Il est évident que l'évaluation d'un modèle dépend de ce pourquoi il est construit. S'il veut être un modèle théorique, nous chercherons la base de l'éva-

luation dans le champ théorique. Si, cependant, il a l'intention d'être utilisé, disons, par les calculatrices, son applicabilité pratique et les résultats de son application sont décisifs. » (F. Kiefer, *Mathematical linguistics in Eastern Europe*).

« L'intérêt n'est pas de rechercher un système mathématiquement définissable qui a quelque rapport avec le langage, comme en étant une généralisation ou un sous-ensemble, mais d'ériger en système mathématique toutes les propriétés et les relations nécessaires et suffisantes pour la totalité du langage naturel. » (Z.S. Harris, *Mathematical structures of language*.)

Examinons maintenant si les critiques que Saumjan adresse à Chomsky sur le plan de la théorie générale sont justifiées ou non. Nous avons vu que sa revendication d'une plus grande généralité pour son modèle n'était pas fondée. En ce qui concerne les universaux du langage, Chomsky a toujours affirmé qu'une théorie linguistique adéquate devait fournir :

- 1) des contraintes universelles sur la forme des grammaires (y compris le concept de transformation),
- 2) un critère d'évaluation pour choisir entre 2 (ou +) grammaires qui engendrent le même langage.

Il est inutile de dire que ni l'un ni l'autre n'y sont encore parvenus.

En outre, affirmer que Chomsky n'a su formuler que des théories de grammaires génératives, sans fournir de description de modèles génératifs véritables, est une contre-vérité évidente. C'est oublier le travail de son équipe sur l'anglais, ainsi que les travaux de R. Kayne (*The transformationnel cycle on French Syntax*, M.I.T. Press, à paraître) de R. Lakoff (*Studies in latin transformational grammar*, M.I.T. Press, 1969), de S.Y. Kuroda (*Generative grammatical Studies in the Japanese Language*, M.I.T. Dissertation, 1965), de Postal sur le Mohawk, etc. (ces deux dernières Langues, du moins, ne pouvant être taxées d'indo-européennes).

Il ne reste que la critique de l'arbitraire des listes de transformations, telles qu'elles étaient postulées dans la théorie chomskienne au moment de *Syntactic Structures* (1957) — et dont nous verrons plus loin l'évolution —.

Kiefer démontre que les correspondances multivoques entre les symboles abstraits du modèle applicatif et leurs réalisations concrètes autorisent à organiser les classifications de plusieurs façons, ce qui peut amener à des résultats fâcheux : il est donc nécessaire de dresser une liste des transformations grammaticales pour les distinguer des transformations agrammaticales, ce qui est impossible ici à cause de l'absence de critère formel approprié. Une grande partie des travaux en cours (Ross, Postal, Emonds) traitent justement des contraintes sur les transformations permises.

ETATS-UNIS 1950

Au cours de la discussion qui précède, nous avons vu apparaître çà et là quelques concepts de la grammaire générative. Il est temps maintenant de les situer dans l'ensemble de la théorie.

Avec Jakobson, le « structuralisme de Prague » avait émigré aux Etats-Unis (1941-42).

Lorsque Chomsky faisait ses études à Harvard, Bloomfield était le leader de la linguistique américaine, le représentant dans ce domaine de l'idéologie positiviste, l'homologue de Skinner pour la psychologie du comportement. C'était la grande époque du behaviorisme, de l'étude des stimuli-réponses et des « renforcements », correspondant aux succès remportés par la technologie de l'après-guerre. La « machine à penser » inspirait confiance, on ne cherchait qu'à affiner l'investigation sur les codages (par les émetteurs) et les décodages (par les récepteurs) des messages, abstraction faite du reste.

Toute la période du début des années 50 est placée sous le signe de cette conception de la langue (instrument de communication, véhicule d'information) qui est encore aujourd'hui la plus répandue.

N. CHOMSKY

Chomsky a produit ses premiers travaux en tant qu'élève et collaborateur de Harris, s'occupant, lui aussi, pendant un temps assez bref, de traductrices automatiques. Il s'est très tôt senti mal à l'aise devant l'importance des problèmes écartés ou tenus pour résolus : il fallait se résoudre à la constatation suivante : aussi loin qu'on puisse pousser l'analyse en termes behavioristes, elle ne peut pas rendre compte de la spécificité du langage humain. Le quasi-abandon, à l'Ouest, des centres de traduction automatique est d'ailleurs un symptôme de la désillusion succédant à l'euphorie mécaniste. Malgré l'ampleur et le niveau, jamais atteints auparavant, des recherches de la cybernétique, il est apparu à Chomsky que les défauts de ces analyses n'étaient pas d'ordre quantitatif, mais d'ordre qualitatif : c'était donc les bases mêmes de ces théories qu'il fallait réexaminer. Après dix ans de collaboration avec Harris, Chomsky a été amené à reconsidérer les jugements que les linguistes positivistes avaient portés sur les grammairiens philosophes du XVII^e, XVIII^e, et du XIX^e européens, et par là, à relire leur référent philosophique, c'est-à-dire le Descartes du discours de la Méthode.

La première constatation qui s'imposait était celle du parallélisme entre les climats intellectuels du XVII^e siècle européen et du XX^e américain : Descartes s'est caractérisé par son opposition à la scolastique et à la conception mécaniste de l'« entendement » comme les grammairiens de Port-Royal se sont opposés à Ramus. C'est principalement l'accent mis sur la créativité du langage humain qui sépare radicale-

ment l'homme de l'automate, de la machine, de l'animal.¹⁰

Dans la terminologie moderne, nous dirions que le langage est en majeure partie indépendant de tout stimulus interne ou externe, contrairement au comportement animal entièrement conditionné par le moment présent. Sinon comment expliquer tous ces jeux sur le signifiant (poésie, jeux de mots), et sur le signifié (mensonge, « expression de soi »). Au lieu de considérer la langue comme un ensemble de modèles « emmagasinés », appris et réappris, où l'innovation ne serait qu'une affaire d'analogie, il s'agit de définir le langage dans son aspect créateur : c'est-à-dire de la possibilité d'énoncer et de comprendre des phrases jamais dites ni entendues auparavant. Le nombre de ces phrases peut être augmenté indéfiniment. Ceci implique d'autre part une théorie adéquate de l'acquisition du langage, et conduit à postuler une « Structure mentale » spécifique, préexistante à tout stimulus, qui permet aux enfants d'un même pays de « construire » une même grammaire pour leur langue, indépendamment des différences de climat familial, social, géographique, grammaire qui leur permettra de jouir de cette aptitude « créatrice ». Ce décalage entre le savoir et l'expérience est défini dans les termes de « compétence » et de « performance » introduits par la grammaire générative.

C'est donc essentiellement pour s'opposer au behaviorisme américain que Chomsky a réuni sous le nom de Linguistique Cartésienne une série de concepts et de réflexions qui lui servent à l'élaboration de la théorie de sa grammaire. Cela ne veut pas dire qu'il est contraint d'adhérer au dualisme de Descartes, et aux démonstrations de celui-ci sur l'existence de Dieu.

10. Voir aussi : le vieux fantasme de Pygmalion, Coppélia, Frankenstein, le singe parlant du Diderot à St. Pétersbourg, l'ordinateur qui souffre, du film 2001, Odysée de l'espace.

Ses référents sont plutôt la physique, la biologie, la psychologie modernes (Lenneberg, Mehler et Bever). Il croit qu'il est possible, sans doute, d'établir une explication physique à ces phénomènes, comme on en a fourni une « pour les forces magnétiques et gravitationnelles, les particules sans masse, etc., concepts qui auraient offensé le bon sens des générations antérieures ». (Language and Mind, 1968).¹¹

La tâche de la grammaire sera donc d'étudier la « compétence », neutre par rapport au locuteur et à

11. On peut rapprocher ceci des conclusions de Kopnine (le concept de pensée et la cybernétique - Voprossy Filosofii, 1961) et d'Oukraitsev (Rapport au Congrès de cybernétique à Moscou, 1962) que Luce Langevin cite dans son article « *Les machines à penser* » et la *pensée* (La pensée, n° 147, octobre 69), sur les interprétations idéalistes de la cybernétique, et la réponse des marxistes :

« Le fait de souligner le lien de la pensée avec les autres propriétés de la matière n'est seulement qu'un des aspects de la question, dont l'absolutisme peut effectivement mener au *mécanisme*.

L'autre aspect n'est pas moins important, c'est *l'originalité qualitative de la pensée* en tant que propriété de la matière qu'il ne faut en aucun cas oublier quand on compare le cerveau humain avec les machines électroniques. » (Kopnine)

...« En conséquence l'information cybernétique possède dans les différentes machines, ou automates, les lois du mouvement de la matière inorganique. Elle peut être exprimée par des rapports *quantitatifs*, et revêtir l'aspect de fonctions mathématiques. Une telle information est tout à fait suffisante pour le travail des automates. Les processus de reflet dans la matière inorganique peuvent se réaliser avec une grande vitesse. C'est pourquoi la quantité d'information qui rentre en action dans les automates peut être très grande.

...Les formes les plus élevées du « reflet », sensation et conscience, se réalisent d'après les lois de la matière vivante (nous soulignons) *qualitativement* autres. Elles ne peuvent être décrites et expliquées par les seules lois de la matière inorganique. » (Oukraitsev)

*l'auditeur. Pour rendre compte de cette compétence, on devra construire un certain modèle.*¹²

En définissant une structure profonde sous-jacente à la structure superficielle la grammaire générative met l'accent sur ce que cette dernière a de trompeur, et sur le peu de renseignements qu'elle donne. Les *ambiguïtés* syntaxiques en sont les exemples les plus fameux :

(A) « *Martial a trouvé ce texte illisible* » peut être interprété en :

(1)

Martial a trouvé ce texte

Ce texte est illisible

(2) Martial a trouvé que ce texte

est

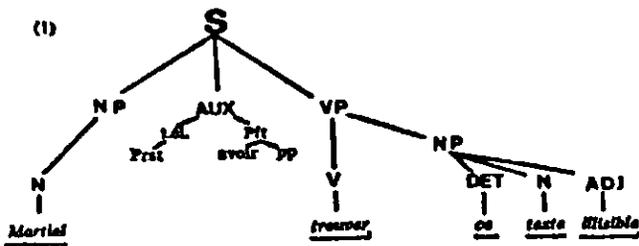
était illisible

La grammaire générative engendre les deux indications syntagmatiques suivants :

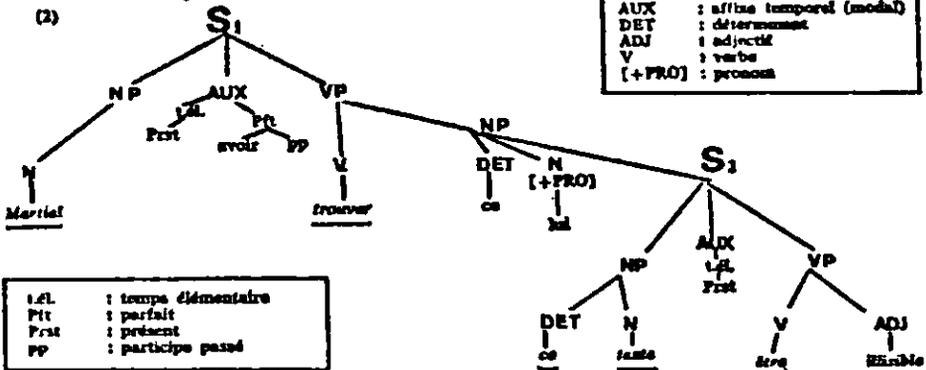
12. Son modèle est construit à partir de l'étude des grammaires formelles, particulièrement des grammaires dites « indépendantes du contexte ». Celles-ci sont constituées par un ensemble fini de règles dont la forme est : $\varphi \rightarrow \psi$, où $\psi \geq \varphi$ et φ est un symbole unique. Ces règles sont dites « de réécritures », elles doivent engendrer toutes et rien que les « phrases » (ou séquences de symboles) du langage donné, au moyen de processus récursifs, en leur attribuant une description structurale (possibilité de leur associer un « arbre »), à partir d'un axiome choisi dans le vocabulaire auxiliaire.

Pour le langage naturel, l'axiome est $S = \text{phrase}$; le vocabulaire auxiliaire : les symboles catégoriels (syntagme nominal, syntagme verbal, adjectif...) ; le vocabulaire terminal : les mots de la langue donnée dans leur transcription phonologique. Les arbres correspondants sont appelés indicateurs syntagmiques.

La base (la structure profonde) est engendrée par les règles de réécriture et les premières transformations obligatoires (Ex. : nombre, affixes temporels). Cette base est reliée à la structure superficielle par la série des transformations facultatives ou non (Ex. : passif, etc.).



S	: Phrase
NP	: syntagme nominal
VP	: syntagme verbal
N	: nom
AUX	: affixe temporel (modal)
DET	: déterminant
ADJ	: adjectif
V	: verbe
[+PRO]	: pronom



l.e.l.	: temps élémentaire
Pt	: parfait
Prst	: présent
pp	: participe passé

Dans les deux cas, les transformations différentes peuvent engendrer la phrase (A).¹³

Depuis *Syntactic Structures* (1957), la théorie chomskienne s'est de plus en plus libérée de son héritage

13. Cette conception de la structure profonde diffère de celle inaugurée par les grammairiens de Port-Royal. En effet, ces derniers décomposent de la façon suivante l'exemple souvent cité :

- « Dieu invisible a créé le monde visible » en trois autres :
- Dieu est invisible,
 - Dieu a créé le monde,
 - Le monde est visible, [analyse assez proche de celle de Harris (3^e forme) et qui correspond à leur logique ainsi qu'à leur confiance à l'introspection].

Dans sa structure profonde au contraire, Chomsky est souvent amené à supposer des éléments sous-jacents qui n'apparaissent pas dans la structure superficielle, du type

structuraliste. Dans *Aspects of the Theory of Syntax* (1965), la grammaire est redéfinie ainsi : elle contient, en premier lieu, une *base* comprenant elle-même une *composante catégorielle* et un *lexique* (ensemble d'entrées lexicales, chaque entrée lexicale étant un système de traits spécifiés). Cette base, du type context-free, engendre des indicateurs syntagmatiques ayant pour seul élément terminal le symbole vide Δ . L'*insertion lexicale* consiste à remplacer le symbole vide par les entrées lexicales dans la mesure de leur compatibilité : l'objet formel ainsi construit est la *structure profonde*.

La grammaire contient, en second lieu, un système de transformations qui transforme les indicateurs syntagmatiques en d'autres indicateurs syntagmatiques : l'application des séquences de transformations à la structure profonde, selon certains principes universaux et selon les contraintes particulières à la grammaire, détermine l'indicateur syntagmatique ultime : la structure superficielle.

Depuis, les collaborateurs de Chomsky se sont divisés en deux grands camps :

- 1) les partisans d'une théorie radicalement lexicale (sémantique),
- 2) les partisans d'une théorie transformationnaliste (syntaxique et formaliste).

Chomsky mentionne cette évolution dans *Remarks on nominalisation* et dans *Deep Structure, Surface Structure, and Semantic Interpretation* (à paraître) ; il remet en question sa définition des catégories grammaticales, au profit de l'hypothèse d'une seule catégorie plus abstraite, considérant les symboles gram-

[celui] précédant les complétives (Cf. : arbre 2) afin de donner une seule et même explication à des phénomènes apparemment indépendants. (Pour l'argumentation, Cf. : M. GROSS, *Grammaire transformationnelle du français, syntaxe du verbe*, Larousse, 1968.)

maticaux comme des *ensembles de traits*, en précisant bien leur condition de formation.

Chaque argument nouveau, plus subtil que les précédents, oblige à reconsidérer l'ensemble du modèle et d'en « déconstruire » un autre plus adéquat empiriquement, en faisant sienne la formule de Bachelard : « L'essence même de la réflexion, c'est de comprendre qu'on avait pas compris ». Si l'horizon s'éloigne à mesure qu'on s'élève, plus on approfondit l'étude du langage..., plus on est saisi de vertige.

Le problème des catégories est un problème d'ordre général.

D'autres, plus particuliers, mais non moins inquiétants, doivent être cependant examinés : par exemple, l'article de Paul M. Postal, (dans *Linguistic Anarchy Notes, Series F : That Much Beloved Semantics-Free Syntax, Plus 1 or, How About Arithmetic ? 1967*) pose la question : comment rendre compte de :

OK 40 cowboys en capturent un 41^e (cowboy, jaguar...)

mais * 40 cowboys capturent un 53^e cowboy
et * Un 41^e cowboy est capturé par 40
et * 6 de mes 3 amis sont arrivés

14. On peut constater, en effet, des analogies constantes de constructions entre les formes :

Nom		+ Préposition + complément :
Verbe		
Adjectif		

Exemple :

(1) il vit | en France, en 1969
le temps

(2) il écrit | de Rodez
lettres

(3) il parle | pour Marx ;
livre

(4) Ce livre est critiqué par Martial → la critique de ce livre par Martial ;

(5) St. Bruno est facile à combattre → la facilité de combattre de St. Bruno.

et * je vous présente mes deux amis : John, Paul, George et Ringo.

En général, tout ce qui a été écrit jusqu'à présent, et qui touche aux rapports de la sémantique et de la syntaxe, est insatisfaisant ; mais ce n'est pas le lieu ici d'en exposer les diverses tentatives d'explication. On peut seulement en conclure que le langage, objet immédiat très naturel, n'est certainement pas réductible à un mécanisme simpliste. On ne pourra progresser jamais dans cette étude qu'avec un maximum de modestie et de circonspection, leçon donnée par le même Bachelard dans sa définition de l'esprit scientifique : « L'esprit a une structure variable dès l'instant où la connaissance a une histoire... Sa structure est la connaissance de ses fautes historiques. »

PARIS 1969-70

Ces pages ne sont, de toute évidence, qu'une approche asymptotique de l'objet qu'elles devaient traiter.

Notre dessein n'était pas, en définitive, et quoi qu'il puisse paraître, de faire l'apologie de telle ou telle Théorie, mais plutôt de rappeler, pour un domaine particulier, la loi générale de l'histoire universelle de la pensée scientifique : chaque recherche ne progresse qu'en s'opposant aux précédentes, et qu'en « concurrence » avec les contemporaines.

La situation en France, dans nos limites nationales et locales, est, pour le moins, assez ambiguë. D'une part, la linguistique structurale n'est enseignée à la Sorbonne que depuis une dizaine d'années, une anthologie de textes des Formalistes Russes n'est parue, en français, qu'en 1966, et un enseignement véritable des grammaires transformationnelles et génératives n'existe que depuis un an, à Vincennes. La prolifération, inévitable (et à ne pas éviter), des « vitrines Chomsky » dans les drugstores, tout le bruit fait autour de ses activités politiques, n'ont pas contribué

spécialement à une meilleure compréhension de son travail scientifique.

D'autre part, la quasi-inaccessibilité de la production des linguistes de l'Europe de l'Est, non traduite et non diffusée, pour qui n'a pas accès aux rayons reculés des bibliothèques spécialisées, la condamne à un certain ésotérisme et la livre aux utilisations d'ordre idéologiques qui conduisent au scientisme, (ou à la science-fiction).

Abonnez-vous !

(bulletin d'abonnement en fin de numéro)
