

De la structure d'une langue aux structures de l'information dans le discours et dans les sous-langages scientifiques

In: Langages, 25e année, n°99, 1990. pp. 21-38.

Citer ce document / Cite this document :

Ryckman Thomas. De la structure d'une langue aux structures de l'information dans le discours et dans les sous-langages scientifiques. In: Langages, 25e année, n°99, 1990. pp. 21-38.

doi : 10.3406/lgge.1990.1590

http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726X_1990_num_25_99_1590

DE LA STRUCTURE D'UNE LANGUE AUX STRUCTURES DE L'INFORMATION DANS LE DISCOURS ET DANS LES SOUS-LANGAGES SCIENTIFIQUES

Cet article voudrait montrer que les travaux de Harris doivent être considérés comme l'élaboration d'une conception du langage et de la structure d'une langue comme instrument, comme moyen, pour la transmission de l'information. Dans le développement de ce programme, Harris a poursuivi une méthode de « régularisation » adaptée à une interprétation informative de la structure de la langue et de la description linguistique elle-même. Les paragraphes 2 et 3 ci-dessous ont pour but d'explicitier la relation entre cette méthode et sa conception générale du langage comme système d'information. Dans cette perspective, l'apport de Harris à la grammaire, en particulier l'introduction des transformations grammaticales et leur reformulation trente ans plus tard dans le système d'insertion lexicale et de réduction de : *A Grammar of English on Mathematical Principles* (qu'on abrégera GEMP)¹, est une propédeutique nécessaire mais subordonnée à l'analyse du discours, son intérêt premier. Cette méthode doit en particulier permettre d'analyser l'articulation des structures d'information à l'intérieur des discours d'une science. Les travaux de Harris permettent de prendre en considération de façon précise la manière dont une langue « contient » ou « convoie » l'information. Je commencerai par examiner cette métaphore dans un bref résumé de l'influence qu'a pu avoir la « théorie de l'information » sur la linguistique. En conclusion, je proposerai quelques observations sur le caractère explicatif de la conception informative de la structure d'une langue qu'a développée Harris.

1. La théorie de l'information et la métaphore du code / message en linguistique

Le terme « information » a un sens spécifique dans la théorie de la grammaire de Harris. Pour le dégager, il sera utile par contraste de survoler quelques-unes des tentatives qui ont été faites pour appliquer à la linguistique les méthodes et / ou la terminologie de la théorie de l'information.

Il n'existe malheureusement pas de compte-rendu complet des différentes tentatives d'application de la théorie de l'information à la linguistique, qui ont

1. New York, Wiley-Interscience, 1982.

fleuri dans les années cinquante². Il est certain que la présence commune au MIT à Cambridge, Massachusetts, lieu d'où se disséminent les modes intellectuelles, de Claude Shannon et de Norbert Wiener a contribué au fait que la cybernétique et les modèles de la théorie de l'information ont été à la mode dans les années d'immédiat après-guerre. Il n'est pas très surprenant que les linguistes aient embouché le clairon d'une théorie³ qui semblait être, au moins aux yeux de ses protagonistes, d'une ampleur suffisante pour promettre des applications générales à tous les problèmes de « communication ». Pourtant cette influence s'est aussi rapidement estompée qu'elle était apparue à la fin des années cinquante, à cause d'arguments considérés comme définitifs assénés par Chomsky contre les modèles probabilistes des grammaires inspirés par les travaux de Shannon et Wiener et proposés par Hockett et d'autres⁴. Les arguments de Chomsky (et Lees⁵) ont eu une telle influence que les principaux tenants des approches de la théorie de l'information en psychologie et en linguistique, comme G. A. Miller et M. Halle s'y sont rapidement ralliés⁶ bien que certains vestiges se soient conservés, e.g. la forme binaire des « traits distinctifs » de Jakobson. D'autres tentatives pour élucider la structure des textes en utilisant la fréquence des occurrences de mots ont subsisté dans les années soixante, en particulier celles de B. Mandelbrot (donc dans un cadre fractal) pour appliquer des notions de la thermodynamique à la linguistique des textes et également dans les études stylistiques de G. Herdan⁷.

L'emploi très large de termes métaphoriques empruntés au génie de la communication, comme « code », « message », ont eu un impact beaucoup plus durable sur le développement de la linguistique⁸. Il est bien connu que Saussure avait déjà introduit le terme de « code » en linguistique structuraliste, en désignant au départ la langue comme le code utilisé dans les combinaisons de

2. Voir, cependant, Gerold Ungeheuer, « Language in the light of information theory », *International Social Science Journal*, XIX (1967), p. 96-106.

3. Claude Shannon, « The Bandwagon », *IRE Transactions on Information Theory* (1956), IT 2 (1), p. 3.

4. Chomsky, *Syntactic Structures*, The Hague, Mouton, 1957 ; « Review of Hockett's *Manual of Phonology* », *IJAL* (1957), p. 23, 223-234.

5. Robert B. Lees, « Review of *Syntactic Structures* », *Language* 33 (1957), p. 375-407 ; « Review of *Logique, langage et théorie de l'information* », *Language* 35 (1959), p. 271-303.

6. George A. Miller, « Information Theory in Psychology », in F. Machlup and U. Mansfield (eds.), *The Study of Information : Interdisciplinary Messages*, NY, Wiley, 1983, p. 493-496 ; Morris Halle, « Confessio Grammatici », *Language* 51 (1975), p. 525-535.

7. B. Mandelbrot, « Information Theory and Psycholinguistics : A Theory of Word Frequencies », in P. Lazarsfeld and N. Kemy (eds.), *Readings in Mathematical Social Science*, NY, Science Research Associates, 1966, p. 350-368 ; G. Herdan, *The Advanced Theory of Language as Choice and Chance*, NY, Berlin and Heidelberg, Springer Verlag, 1966.

8. E.g., Roman Jakobson, « Results of a Joint Conference of Anthropologists and Linguists (1952) », repris dans ses *Selected Writings* (The Hague, Mouton), 2, 1971, p. 554-567, 559 ; « Linguistics and Communication Theory », *Proceedings of Symposia in Applied Mathematics*, vol. XII, Providence, R.I., American Mathematical Society, 1961, p. 245-252.

sons émis pour exprimer la pensée personnelle du locuteur ⁹. A cause de cela, on prétend souvent qu'il voulait faire revivre de façon explicite une « théorie » traditionnelle (et simpliste) sur la pensée et le langage ¹⁰. Cependant, il semble y avoir eu deux raisonnements, au moins conceptuellement séparés, soutenant le choix de ce terme. D'une part, ainsi que le montre le célèbre diagramme du circuit entre le locuteur et l'auditeur (*Cours de linguistique générale*, p. 27-28), il semble que Saussure ait simplement fait l'hypothèse d'une théorie psychologique traditionnelle du sens linguistique et de la communication, que j'appellerai « théorie de la traduction de la compréhension » ¹¹. D'autre part, le concept structuraliste de phonème, comme unité de son discrète et de traitement combinatoire, suggérait naturellement une analogie avec le code télégraphique (selon la formule célèbre de Saussure, une entité entièrement contrastive, relative et négative, *Cours de linguistique générale*, p. 164). Il est intéressant de poursuivre un peu plus en détail chacun des deux raisonnements parce que, ainsi que j'espère l'indiquer, ils procèdent de vues différentes, voire même conflictuelles, sur la « nature » du langage et la structure de la langue.

La métaphore du codage et du décodage est une réincarnation relativement récente d'une notion antique mais à la vie dure, qui tend toujours à dominer une bonne part de la pensée sur la communication linguistique ainsi, *a fortiori*, que sur le caractère et la structure du langage. Il s'agit de l'image d'un processus de traduction entre les « idées », ou plus récemment, entre des « représentations mentales », et le médium extérieur physique du langage (texte ou parole). Dans ce « modèle », la communication intervient quand un « message » ou une « croyance » existant sous une forme (propositionnelle ?) de représentation dans l'esprit du locuteur est « codée » sous une forme physique (acoustique ou orthographique) servant de médium pour la transmission à l'auditeur. Cet auditeur réalise alors, en général de façon inconsciente bien sûr, un « décodage » du message sous forme d'un ensemble d'idées ou de significations représentées mentalement. Ainsi, la communication entre un locuteur et un auditeur intervient lorsque l'idée ou la représentation mentale élaborée dans l'esprit de l'auditeur est la même ou suffisamment voisine de celle originalement produite par le locuteur ¹².

9. Par exemple : Georges Mounin, « La Notion de Code en Linguistique », in *Linguistique contemporaine. Hommage à Eric Buyssens*. Bruxelles, Editions de l'Institut de Sociologie, Université Libre de Bruxelles, 1970, p. 141-149, 141.

10. Qu'on se rappelle le commentaire acerbe de Bloomfield en 1923 dans une critique, par ailleurs positive, de la seconde édition du CLG (dans G. Hockett ed. *A Leonard Bloomfield Anthology*, Bloomington, Indiana University Press, 1970, 106-108, p. 107 : « Now de Saussure seems to have had no psychology at all beyond the crudest popular notions... »).

11. G. H. R. Parkinson, « The Translation Theory of Understanding », in G. Vesey (ed.), *Communication and Understanding*, Royal Institute of Philosophy Lectures, vol. 10, 1975-1976, London, p. 1-19 ; and Roy Harris, *Reading Saussure*, London, Duckworth, 1987, p. 204-218.

12. E.g. P. Denes and E. Pinson, *The Speech Chain : The Physics and Biology of Spoken Language*, Bell Telephone Laboratories, 1963, 6 ; J. A. Fodor, T. Bever, and M. Garrett, *The Psychology of Language*, NY, McGraw-Hill, 1974, 13-14.

On peut faire remonter l'origine du modèle du circuit de la communication à l'Antiquité, aux stoïciens, et peut-être même plus loin, mais c'est John Locke qui, d'après une opinion couramment admise, lui a donné sa version canonique dans *Essay concerning human understanding* (1690). Sa doctrine contient des idées qu'on retrouve dans la Logique de Port-Royal. D'un attrait œcuménique, cette doctrine a été adoptée ou implicitement acceptée à la fois dans les traditions empiristes et rationalistes, pour une part du fait de l'idéalisme et du romantisme allemand, mais aussi par le positiviste *Junggrammatiker* en réaction contre les vues hégéliennes et organistes du langage¹³. Une autre branche de descendance se développe avec Condillac, disciple de Locke, et parvient au vingtième siècle à travers d'éminents linguistes comme Michel Bréal et, ainsi que cela a déjà été dit, Ferdinand de Saussure. Cette tradition a été retracée par Hans Aarsleff¹⁴. Plus récemment, cette doctrine, sous ses non significatives différentes formes, la famille de vues que pour des raisons évidentes on appelle « mentalisme », jouit d'une résurgence en linguistique et en philosophie, due pour une part au génie de la communication et aux analogies avec l'informatique et pour une autre part au succès largement perçu de la linguistique ouvertement mentaliste de Chomsky. Ce développement s'est fait malgré la résistance d'anti-mentalistes aussi convaincus que Bloomfield et W. V. O. Quine.

Par ailleurs, et indépendamment du modèle codage / décodage ou traductif de la compréhension linguistique, l'hypothèse que les éléments de la langue ne peuvent être que discrets, contrastifs et que les entités ne sont individualisées que dans des relations, a amené Saussure à répéter la comparaison du langage avec des codes télégraphiques comme ceux de Morse (*Cours de linguistique générale*, p. 36), avec le jeu des échecs (p. 43, 125, 149) et avec une algèbre (p. 168).

Ensuite le structuralisme, surtout en Amérique, a mis une emphase considérable sur une structure « codage télégraphique » du langage sans aucune idée implicite de processus mental de traduction, ses protagonistes étant de fait opposés à de tels discours. Pour certains linguistes structuralistes, l'analogie du code était particulièrement bien adaptée à la caractérisation de la structure de la langue, vue comme une combinaison de différentes constructions hiérarchisées (les « molécules ») à partir d'une douzaine de phonèmes (« les atomes »)¹⁵. Mais

13. Pour la tradition humboldtienne, voir les nombreux travaux de H. Steinthal ; pour la réaction neogrammairienne, e.g. H. Paul, *Prinzipien der Sprachgeschichte*, 2. Ausg., 1886, S. 99 : « Der satz ist der sprachliche ausdrück, das symbol dafür, dass sich die verbindung mehrerer vorstellungen oder vorstellungsgruppen in der seele des sprechenden vollzogen hat, und das mittel dazu, die nämliche verbindung der nämlichen vorstellungen in der seele des hörenden zu erzeugen ». (« La phrase est l'expression linguistique, le symbole indiquant que la combinaison de plusieurs idées ou de plusieurs groupes d'idées a été effectuée dans l'esprit du locuteur et c'est ce qui permet en même temps de produire la même combinaison des mêmes idées dans l'esprit de l'auditeur. »)

14. *From Locke to Saussure*. Minneapolis, University of Minnesota Press, 1982.

15. Par exemple, Martin Joos, « Description of Language Design », *Journal of the Acoustical Society of America*, 22 (1950), 701-708, p. 105 : « Les linguistes disent en effet que la conception de

Harris a montré que les analogies de la structure d'une langue avec un code contiennent les bases d'une explication. Observant que les éléments d'une langue sont discrets, arbitraires et pré-assemblés par convention à l'intérieur d'une communauté linguistique¹⁶, et notant également que de tels traits structuraux sont nécessaires à une transmission sans erreurs de composition, Harris en a conclu que la structure de la langue est explicable comme instrument pour la transmission, et non la communication, de l'information¹⁷.

Dans cette perspective, on peut dire au plus qu'une langue est *analogue* à un code, et non qu'une langue est un code. En effet, un code est une bijection entre des expressions déjà bien-formées (le « message ») et les éléments du chiffage choisi, ce qui donc présuppose la structure grammaticale du message¹⁸. Une telle supposition revient à considérer qu'il existe une métalangue externe à la langue pour son analyse, or Harris montre qu'une telle supposition conduit à une régression infinie de la métalangue dans la langue¹⁹. Pour spécifier comment la langue « véhicule » l'information, la grammaire n'a pas la ressource de réduire la langue à un précédent « message » ou « langage interne » ou à quelque chose de non linguistique tels que objets, événements, ou propriétés du « monde réel ». Une telle réduction, si tant est qu'elle puisse être envisagée, ne pourrait être définie qu'en employant ce qui du point de vue de la grammaire est un métalangage. Mais ce point de vue n'est qu'une position idéale qui tombe à l'extérieur du domaine d'explication de la grammaire.

Du fait d'un malentendu très répandu concernant le structuralisme américain, on croit que beaucoup de linguistes qui ont suivi Bloomfield, et notamment Harris, du fait qu'ils développaient des méthodes formelles de description linguistique, ont été amenés à épouser les vues d'un behaviorisme extrême sur la psychologie, que Bloomfield a exprimé à l'occasion. Il est certain que les attaques virulentes de Bloomfield contre le mentalisme excluent toute référence à une explication à partir de la métaphore codage / décodage de la relation entre la « représentation mentale » et la langue. Mais ne voir dans la défiance envers le mentalisme qu'une allégeance au behaviorisme est une vision par trop simpliste à une période où il y avait une diversité assez grande d'opinions sur la psychologie parmi les linguistes structuralistes. Il ne faut pas négliger, par

tout langage est essentiellement télégraphique. que le langage a la structure d'un code télégraphique, utilisant des signaux moléculaires faits d'atomes invariants et différant du morse de deux façons : les codes appelés « langues » ont de nombreux niveaux de complexité au lieu de deux seulement et à chaque niveau il y a de sévères limitations sur les combinaisons permises ».

16. Harris insiste sur le fait que c'est ce qui fait que l'interprétation d'une séquence par l'auditeur est une répétition, non une imitation du locuteur.

17. *Mathematical Structures of Language* (désormais cité *MSL*), New York, Wiley-Interscience, 1966, 6-8 ; ceci a été également été discuté de façon plus élaborée dans *Language and Information* (désormais cité *LI*), New York, Columbia University Press, 1988, p. 87-113.

18. Ceci a également été indiqué par F. François, « Le langage », *Encyclopédie de la Pléiade*, Paris, 1968, p. 11, cité par Mounin (note 8), p. 144-145.

19. *MSL*, 11 ; *LI*, 113 ; « On a Theory of Language », *Journ. of Philosophy*, LXXIII (1976), p. 253-276, 273.

exemple, l'influence également forte de Sapir (sur Harris entre autres) qui a souvent été mis au panthéon des linguistes mentalistes. De toutes façons, il est clair que la résistance de Harris au chant des sirènes du mentalisme, avec son attrait explicatif pour des représentations sous-jacentes au langage, ne provient d'aucun engagement envers un behaviourisme psychologique. Un tel attrait, qui semble n'être rien d'autre qu'une pseudo-explication *virtus dormitiva*, est en fait incompatible avec l'esprit constructiviste de l'œuvre de Harris. Mais au-delà de la contrainte méthodologique qui consiste à exclure le recours à des entités non justifiées, l'alternative à l'hypothèse d'une organisation latente ou préexistante permettant de rendre compte de la structure d'une langue est le postulat de la nature sociale et informative de cette structure.

J'ai essayé de montrer qu'il y a deux courants ou tendances dans la prévalence de la métaphore de code en linguistique récente. Cependant, sous l'influence de la terminologie de la théorie de la communication, des linguistes structuralistes importants en Europe, mais aussi Jakobson aux U.S.A., ont évacué cette distinction, parlant indifféremment du codage de la pensée dans la langue et de la structure de code d'une langue ²⁰. Dans plusieurs tentatives pour étendre les concepts de la théorie de la communication à l'analyse phonémique, on a parfois laissé passer le fait que cette théorie ne développe pas un concept unique d'information ²¹. Dans un sens, Harris a procédé d'une autre façon. Au lieu d'utiliser telle ou telle notion d'information, il a développé l'argument que les hiérarchies de restrictions de sélection sur les combinaisons d'éléments linguistiques comprennent ce qu'on peut considérer comme la structure informative d'une phrase ou d'un texte. Harris a fait remarquer de plus que, parce qu'une langue a une structure informative, une méthodologie propre à analyser cette structure devient nécessaire.

C'est le fait fondamental de la redondance et l'existence de restrictions sur les combinaisons d'éléments lexicaux qui nous révèle qu'une langue a une structure. Etant donné l'association entre redondance et structure, il est évidemment essentiel que la formulation grammaticale de chaque restriction soit d'une efficacité maximale au sens suivant : elle ne doit pas contribuer aux redondances de la combinaison d'éléments qu'elle tente de décrire. La notion de redondance, dans ce sens, fournit un moyen pour distinguer l'information dans une langue. La situation de la grammaire devient alors comparable d'une certaine façon à la théorie de la complexité en informatique ²², où l'information d'une séquence de digits peut être définie comme une certaine fonction sur la longueur du plus petit programme qui la calcule. L'information d'une langue

20. Par exemple, E. Benveniste (1963), « Coup d'œil sur le développement de la linguistique », in *Problèmes de linguistique générale*, Paris, Gallimard, 1966, 30-31 : « ... on peut espérer des théories de l'information quelque clarté sur la manière dont la pensée est codée dans le langage » et p. 23 : « la langue étant organisée systématiquement et fonctionnant selon les règles d'un code... ».

21. A. Martinet, *Éléments de linguistique générale* ; B. Malmberg, et *Structural Linguistics and Human Communication*, 2nd revised edition, Berlin/NY, Springer Verlag, 1967, chapitre 3.

22. Telle qu'elle a été développée par Martin-Löf, Kolmogorov, Chaitin et d'autres.

devrait pouvoir s'exprimer comme la plus petite grammaire la caractérisant complètement en terme de ses éléments récurrents et de leur mode de combinaison. Dans la mesure où il n'y a pas ici de métalangue extérieure d'où on pourrait dériver ces éléments et leurs combinaisons, on doit en rendre compte de façon purement combinatoire.

2. Régularisation de la description linguistique

Pour Harris, le problème général de la linguistique structurale consiste à distinguer les séquences d'éléments (phonèmes, morphèmes etc.) qui peuvent occuper comme phrases de celles qui ne le peuvent pas²³. Pour une grammaire des phrases, la solution consiste à spécifier les contraintes qui gouvernent les combinaisons d'éléments lexicaux qui peuvent occuper comme phrases²⁴. Puisqu'il n'y a pas de métalangue extérieure à une langue à partir de laquelle on pourrait dériver les restrictions gouvernant les combinaisons d'éléments (ces restrictions constituant la redondance d'une langue, qui est essentielle pour véhiculer l'information), la première tâche de la grammaire est une description purement combinatoire (distributionnelle) de co-occurrences d'éléments. L'exposé des relations distributionnelles se fait sous certaines présentations de données (e.g. tables). L'exposé des résultats distributionnels a été incompris et critiqué à l'époque et plus tard, à la fois pour n'être qu'une simple taxinomie et pour être une linguistique de charabia (comme si la formulation de contraintes de plus en plus systématiques n'était pas hautement désirable dans une science). L'exposé des contraintes distributionnelles n'a jamais été considéré comme le « but » de la théorie linguistique et la présentation la plus détaillée des méthodes distributionnelles se conclut par la proposition qu'une grammaire devrait se présenter sous forme axiomatique²⁵. Précisément, ce dont une grammaire axiomatique doit rendre compte (« expliquer »), ce sont les contraintes de co-occurrences, la « distribution » lexic-syntaxique de chaque élément lexical.

La seconde tâche de la grammaire dépend de la première. Si la structure d'une langue ou la structure d'un sous-langage (voir le prochain paragraphe) est une structure de restrictions sur les combinaisons de mots, il est impératif que la caractérisation grammaticale de cette structure n'ajoute rien aux redondances des combinaisons, les porteurs de l'information, qu'elle a pour tâche de décrire²⁶. Ceci n'est pas qu'un point de méthode sur l'économie de moyens qu'on désire. Dans la mesure où il n'y a pas de métalangue disponible pour la

23. MSL, p. 13.

24. GEMP, p. 26-27.

25. (*Methods in*) *Structural Linguistics* (désormais abrégé *SL*), Chicago, University of Chicago Press, 1951, p. 372-373.

26. GEMP, p. 10-11 : « la description grammaticale doit être aussi non-redondante que possible de manière à ce que la redondance essentielle d'une langue, en tant que système porteur de l'information (...) ne soit pas masqué par d'autres redondances dans la description elle-même ». Voir aussi MSL p. 12, note 16.

description grammaticale, toute restriction sur les combinaisons contenues dans la grammaire doit correspondre à ou être corrélée avec une différence d'information, une différence reconnue par les locuteurs de la langue. Par l'intermédiaire de ce que Harris a appelé « régularisation »²⁷, dégager la structure d'une langue (ou certains emplois restreints de cette langue dans le discours ou dans les sous-langages) consiste à décomposer les contraintes multiples sur la combinatoire d'éléments en utilisant des phrases qui disent la même chose. Dès son premier livre, cette approche est présente et se manifeste à travers les différents chapitres par la recherche de classifications de plus en plus générales d'éléments linguistiques²⁸. Ainsi les transformations sont peu à peu conçues comme une sorte de « morpho-phonémique étendue ». Elles constituent une méthode de régularisation plus puissante (voir plus bas) qui permet même une dérivation du temps et des affixes. Cependant, tout au long de l'œuvre de Harris, des éléments constituent des classes d'équivalence si leurs membres occurrent dans un même environnement, qui peut être caractérisé formellement. Avec l'hypothèse d'environnements distinctement formalisables et reconnaissables de façon distincte par les locuteurs²⁹ se développent des méthodes grammaticales plus opératoires. Du fait que les éléments porteurs d'information ne sont pas définis en référence à un métalangage extérieur, leur caractérisation formelle fait appel à une justification empirique, qui met en jeu un accord des comportements linguistiques des utilisateurs de la langue.

A l'origine, les transformations ont été définies comme des relations d'équivalence entre phrases avec un domaine d'application spécifié. La caractérisation initiale des phrases qui étaient reliées transformationnellement constituaient des méthodes de « normalisation » pour régulariser la description grammaticale des textes (des discours écrits). Pour passer de l'analyse des discours à la première formulation axiomatique d'une grammaire des phrases, la relation transformationnelle a été reformulée comme une correspondance entre certaines phrases structurellement simples et le reste des phrases de la langue considérée. On appelait l'ensemble des phrases simples « les phrases noyau » dans la mesure où, les transformations réalisant une partition sur l'ensemble des phrases, chaque phrase de cette partition ne contient plus d'autre phrase comme partie propre. La classe définie par cette partition est donc appelée le noyau de l'ensemble des phrases relativement à la correspondance transformationnelle³⁰. On a fait remarquer que dans cette conception le statut des transformations est extérieur aux méthodes purement distributionnelles de la linguistique descriptive. Cependant, on remarquait en même temps que la régularisation transformationnelle

27. MSL, chapitre 6.

28. SL, en particulier chapitres 7 à 19.

29. H. Hoenigswald, « Review of John Lyons, *Structural Semantics*, *Journal of Linguistics*, 1 (1965), 191-196, 192, donne une description explicite de cette hypothèse; voir aussi Harris, « Distributional Structure », (1954) repris dans ses *Papers on Syntax* (dorénavant, DS), (Hiz, ed), Dordrecht/Boston, D. Reidel, 1981, 3-22, 13.

30. Cette notion de noyau diffère de la notion de noyau utilisé par Chomsky.

poussait plus loin l'application des méthodes distributionnelles en les étendant à des phrases qui sinon, à moins d'un coût excessif, n'auraient pu être analysées ³¹.

Dès leur conception, les transformations de Harris ont eu un propos et une importance qui contraste avec celles qui ont ensuite été définies en grammaire générative. Autour des années cinquante, la logique mathématique a fourni un modèle prometteur aux linguistes, en définissant récursivement les expressions bien formées et en définissant des règles de preuve permettant de dériver certaines expressions bien formées à partir d'autres. Par analogie avec les traitements de langages formels, certains linguistes ont espéré produire des méthodes effectives pour déterminer dans quelle mesure des phrases de longueur et de complexité arbitraire étaient des phrases bien formées d'une langue et montrer comment certaines phrases peuvent être reliées à d'autres via une procédure dérivationnelle. Pour les linguistes générativistes, l'existence de telles règles aurait constitué un moyen d'exhiber ce qui aurait pu être appelé « la structure » d'une langue, où une langue devenait dans cette conception un ensemble récursivement énumérable de phrases. Contrairement à la conception dite générativiste, les transformations harrisiennes ont été exclusivement conçues comme des relations entre phrases, et non comme des relations entre structures de constituants (purements formelles) sous-jacentes. Ceci va de pair avec le fait de ne pas recourir à des entités extra-linguistiques et avec l'apport sémantiquement nul ou constant des transformations. Mais ceci va également de pair avec le fait que les transformations ne sont pas seulement ni même d'abord des règles de preuve. Pour Harris, les transformations sont une méthode pour régulariser et rendre compacte la description linguistique, un moyen pour réduire la redondance de la description sur un certain domaine. Quand ce domaine est la langue dans son ensemble, les transformations sont alors analogues à des règles de preuve. Quand le domaine est celui du discours ou d'un sous-langage, les transformations établissent des relations d'équivalence qui permettent de donner la même représentation à des phrases différentes ayant le même « contenu informatif » du point de vue de ce sous-langage.

Qu'est-ce qui reste invariant sous les transformations ? Il est clair dès leur origine dans l'analyse du discours que leur intérêt descriptif réside dans leur relative invariance sémantique. Les transformations sont conçues initialement comme des relations d'équivalence en vertu de leur préservation des relations de co-occurrence. Plus précisément, on établit l'existence d'une transformation lorsque les domaines de co-occurrences sont identiques pour une séquence de mots pouvant apparaître dans deux constructions différentes. Cette condition distributionnelle sur les transformations a une valeur sémantique dans la mesure où « des éléments de sens majeurs restent constants sous une transformation » (ces éléments de sens sont appelés contenu informatif) ³². Le statut sémantique

31. « Co-occurrence and transformation » (1957) repris dans *PS*, p. 143-210, p. 200-201.

32. « Co-occurrence and transformation », p. 149-150.

particulier de la transformation posait évidemment problème, ce statut dépendant d'une évaluation convaincante du sens, mais de toutes façons le sens doit être évalué, étant donné la corrélation supposée entre les relations de co-occurrence et le sens. Les transformations préservant les relations de co-occurrence devraient également préserver le sens³³. Dans une formulation postérieure du noyau et des transformations³⁴, la condition de co-occurrence a été précisée pour permettre la conservation d'un ordre relatif de jugements d'acceptabilités. Contrairement au critère de co-occurrence pour les transformations, qui implique que certaines séquences de mots n'occurrent jamais dans une certaine forme de phrase (construction de classes de mots e.g. NVN), le fait que les transformations préservent les ordres relatifs d'acceptabilité ne présuppose pas une échelle fixe de jugements d'acceptabilité. On ne peut pas dire que les transformations préservent le sens (par exemple la négation était considérée comme une relation transformationnelle dans le cadre de MSL), mais pour Harris, si elles changeaient le sens, c'était d'une façon qu'on pouvait formuler explicitement et qui ne changeait pas le degré d'acceptabilité de la phrase avant l'application de la transformation.

En développant sa pensée, Harris a été conduit à la conception de la grammaire de « Report and Paraphrase », qui apparaît rétrospectivement comme un pont important entre les vues précédentes et la conception complètement élaborée de GEMP. Cette grammaire contient un « sous-langage de base »³⁵ où les phrases « sources » de la grammaire ne contiennent plus de transformations. Toutes les autres phrases de la langue doivent être dérivées à partir d'opérations sur ces phrases de base (ainsi que par des changements morpho-phonétiques, qui affectent la forme des mots et non la structure des phrases). Chacune de ces phrases de base a une structure prédicative qui « porte toute l'information objective de la phrase » et dont « l'interprétation la plus naturelle est celle d'un discours rapporté »³⁶. Ces phrases élémentaires à structure hiérarchisée comportent un sous-langage qui porte l'information objective, ou discours rapporté, qui est véhiculé dans une langue. « Dans la mesure où elles sont complètes du point de vue informatif, ces phrases pourraient être utilisées indépendamment du reste de la langue »³⁷. Si on considère « l'information objective » comme l'information véhiculée par les discours rapportés à structure prédicative, le programme de régularisation de la description linguistique par élimination de diverses formes qui portent la même information atteint ici sa forme la plus achevée. Cette reformulation de l'analyse transformationnelle ouvre la voie à une spécification syntaxique des mécanismes

33. *ibid.*, p. 202.

34. « Transformational theory » (1965) repris dans [PAPERS] p. 236-280.

35. Un sous-langage est défini comme un ensemble de phrases clos pour certaines opérations comme la conjonction ; voir *MSL*, p. 152 et *LI* p. 34.

36. « Report and paraphrase » (1969) repris dans *PS* p. 293-351, p. 294.

37. *Ibid.*, p. 295.

d'interprétation des combinaisons de mots, du sens (ou encore de l'information ou du discours rapporté), qui est préservé par les transformations. Le caractère syntaxique de cette spécification pourrait lui donner un caractère explicatif, parce que non arbitraire.

La grammaire de « Report and Paraphrase » a d'abord pour avantage de rendre explicite la conception d'un ensemble de phrases de base informationnellement complet. Elle préserve la double fonction des transformations préservant l'information contenue dans les phrases de base et dérivant à partir de ces phrases de base les autres phrases de la langue. Deux questions fondamentales peuvent maintenant être posées, premièrement, quelle est la structure compositionnelle des phrases de base (i.e. qui ne comportent plus de réductions) et comment l'information véhiculée par le langage se crée-t-elle ? Deuxièmement, étant donné l'effet informativement nul des transformations, y a-t-il une façon générale de définir les transformations de manière à ce qu'elles n'apparaissent plus comme des changements de formes (définies par des entités métalinguistiques) de phrases ? Pour des raisons qui ne devraient peut-être pas surprendre, il s'est avéré que ces deux questions étaient liées.

La grammaire décrite dans GEMP résulte des tentatives précédentes et conserve le but de montrer comment l'information est structurée et véhiculée par la langue, elle illustre également la recherche d'une théorie axiomatique de la grammaire. Dans cette théorie, la grammaire de l'anglais est exposée à partir de quelques contraintes, principalement trois, qui ne sont pas indépendantes. La principale est un ordre sur le vocabulaire de l'anglais, qui décrit dans une certaine mesure quels mots doivent être présents dans une phrase pour permettre à d'autres d'y figurer. Par exemple des verbes transitifs comme *lire* et *manger* sont des opérateurs qui requièrent deux arguments appartenant à la classe des mots contenant *Max*, *poisson*, *livre* : des arguments élémentaires c'est-à-dire des mots qui n'ont pas eux-mêmes d'arguments. Un mot comme *continue* requiert un mot opérateur comme *lire* ou encore *croire* ou *imaginer* qui requièrent eux-mêmes un argument phrastique et un argument élémentaire. Par exemple, dans GEMP, une phrase comme : *Jean imagine que tu crois que Max continue à lire des livres* serait représentée par l'ordre partiel suivant, où l'ordre partiel > représente « opère-sur » et « , » représente « et » :

imagine > *Jean*, *crois* > *tu*, *continue* > *lire* > *Max*, *livres*

L'ordre applicatif sur les mots, le fait pour un mot, ou un couple de mots, d'être sélectionné comme argument par un autre mot, peut lui-même être interprété par l'expression « ce qui peut être dit à propos de » ; ainsi, dans l'exemple précédent, ce qui peut être dit à propos de *Max* et de *livres*, c'est qu'il les lit.

Les deux autres contraintes sont d'une part une hiérarchie grossière de vraisemblances d'occurrences pour la sélection lexicale des arguments syntaxiquement possibles d'un opérateur donné, d'autre part une réduction sur la forme phonémique des mots qui sont hautement vraisemblables (lexicalement « appropriés ») comme arguments d'un certain opérateur. Ainsi, *Nous attendons Max*

peut être considéré comme une forme réduite de : *Nous attendons que Max vienne* (ou *soit ici, arrive, etc.*), c'est-à-dire un argument hautement approprié de *attendre* et opérateur approprié de *Max*.

Cette théorie présente deux grands avantages. Le premier est qu'il n'est plus nécessaire de prédéfinir des classes de mots comme : verbe, nom, adverbe. Les classes de mots sont définies de façon interne à la langue par la relation opérateur / argument et c'est la satisfaction de ces relations qui doit fournir des phrases bien formées. Le second avantage est que les transformations ne sont plus des changements de la forme syntaxique des phrases mais des réductions de la forme phonémique des mots. Les réductions de forme des mots, même à des formes zéro, préservent les relations opérateur / argument de la base.

Les phrases reliées par réduction sont équivalentes d'un point de vue informatif. L'information se crée dans les phrases de base et la dérivation de phrases quelconques à partir de ces phrases de base est donc une régularisation de l'information véhiculée par une langue.

3. L'information dans les sous-langages scientifiques

Sous l'hypothèse d'une intégration de la métalangue de description dans la langue, la structure informative et grammaticale d'une phrase peut être dégagée à partir de sa comparaison avec les autres phrases du texte à analyser. Pour une langue prise dans son ensemble, c'est-à-dire considérée indépendamment des domaines particuliers dont elle permet de véhiculer l'information, la comparaison des phrases se fait sur la base des principes de combinaison des opérateurs lexicaux exposés dans GEMP, qui définissent la bonne formation et les conditions de reconstruction d'informations non explicites.

L'utilisation d'une langue dans un sous-langage, c'est-à-dire dans un domaine particulier, fait intervenir des contraintes supplémentaires qui rendent possibles d'autres régularisations. Un ensemble de textes traitant d'un sujet particulier fait apparaître des répétitions de séquences de mots identifiables.

Pour illustrer l'analyse d'un sous-langage scientifique, je présenterai brièvement une analyse d'articles en immunologie autour des années 1935-1966 *. Cette analyse ainsi qu'une grande partie des textes représentés est tiré d'un volume récemment paru ³⁸. On a montré que les contraintes sur les combinaisons de mots dans ces articles traitant d'une même question en immunologie rendait possible une autre régularisation de la description et en particulier la délimitation d'un sous-langage. A l'intérieur d'un tel sous-langage, le vocabulaire et les possibilités de combinaison des mots étant beaucoup plus restreintes que dans une langue prise dans son ensemble, l'objectif des opérations de régularisation

* NdT Les exemples tirés de textes américains seront ici traduits en français, cette traduction n'ayant pas d'incidence sur l'argumentation.

38. Z. S. Harris *et al.* *The Form of Information in Science*, Dordrecht and Boston, Kluwer Academic Press, 1988.

est d'obtenir une représentation des structures de phrases qui maximise leurs ressemblances. En établissant des séquences de classes de mots répétitives (chaque séquence formant un type de phrase), des transformations peuvent être appliquées à une phrase d'un texte de manière à ce que les mots de cette phrase aient entre eux la même relation applicative que dans une autre phrase de ce texte où les mêmes mots seraient combinés dans une structure grammaticale différente **. Ici, comme dans les premiers travaux sur l'analyse du discours, les transformations facilitent l'identification purement combinatoire des classes de mots en enlevant les restrictions non pertinentes dans l'environnement grammatical de certaines occurrences. Contrairement à ce qui se passe pour l'analyse d'une langue prise dans son ensemble, conduite en dérivant des phrases attestées quelconques à partir de phrases de base sans réductions, une des façons de régulariser un sous-langage consiste à transformer les phrases, si nécessaire, pour qu'elles se présentent sous des formes canoniques permettant de les représenter toutes de façon concise. Ces phrases canoniques sont ensuite traduites par des formules qui indiquent ce qui est équivalent ou distinct du point de vue de l'information dans les phrases reliées transformationnellement.

A l'époque où ces articles ont été publiés, le domaine était considérablement moins développé qu'il ne l'est actuellement ; de plus, il y avait alors une question centrale qui était de déterminer quel type de cellule était responsable de la production des anticorps. En suivant la première démonstration expérimentale de la formation d'anticorps dans le tissu lymphatique par McMaster et Hudack en 1935 ³⁹, la question préliminaire à la réponse était : quel type de cellule parmi celles qu'on trouve habituellement à l'intérieur du nœud lymphatique est le véritable producteur d'anticorps ? Une controverse s'ensuivit à partir des années 40, pour la majeure partie entre des savants européens et américains, les premiers considérant généralement que c'étaient les cellules du système lymphatique appelées plasmocytes (du fait de leur abondance en cytoplasme) qui étaient responsables, les derniers considérant pour la plupart que d'autres cellules du système lymphatique, les lymphocytes, généralement plus petites et probablement distinctes des plasmocytes, étaient le site cellulaire de la formation des anticorps (Dougherty, Chase et White en 1944) ⁴⁰. Au bout d'une vingtaine d'années de recherches, il a finalement été prouvé en utilisant le microscope électronique et les techniques de fluorescence que les deux types de cellules produisaient des anticorps et la controverse s'est résolue par la

** NdT On trouvera plus bas des exemples de ces transformations. La relation applicative dont il est question dans ce sous-langage est d'une nature différente de celle qui exprime la structure informative d'une phrase d'une langue n'appartenant pas à un sous-langage.

39. « The Formation of Agglutinins within Lymph Nodes », *J. of Experimental Medicine*, 61, 783-805.

40. « The Demonstration of Antibodies in Lymphocytes », *Proc. Soc. Exp. Biol. & Med.*, 57, 295-298.

compréhension que les deux noms de cellule ne faisaient que dénommer deux stades de développement de la même lignée cellulaire ⁴¹.

Le propos de l'analyse qui sera résumée ici était de voir s'il était possible de donner une représentation formelle, commode à utiliser, de l'information contenue dans les articles de ce domaine et permettant également de localiser et de caractériser les désaccords entre les chercheurs du domaine et plus généralement les changements d'information intervenant au cours du temps. Cette controverse a été choisie pour tester nos méthodes linguistiques parce qu'elle avait un commencement et une résolution clairement identifiables et parce que les deux parties de la controverse pouvaient être clairement identifiées en utilisant un nombre raisonnable d'articles (environ 25) pour l'anglais et le français. Cette sélection d'articles a été faite par des chercheurs ayant participé à cette recherche.

L'étude a commencé par une analyse détaillée de 14 articles en anglais et 4 en français après qu'un premier aperçu ait permis de déterminer que les autres articles étaient peu ou pas significatifs pour le résultat global de l'analyse. Cela a permis de constituer un corpus de plusieurs milliers de phrases. Les méthodes de régularisation linguistique ont été utilisées pour normaliser les phrases sous des formes qui facilitaient la recherche de schémas de répétitions de mots. Par exemple, des phrases passives ont été transformées en phrases actives, ou vice versa, des nominalisations ont été « dénominalisées ». Ainsi *La formation d'anticorps dans les ganglions lymphatiques a été mise en évidence par McMaster et Hudack* peut être transformée en : *McMaster et Hudack ont mis en évidence que la formation d'anticorps se produit dans les ganglions lymphatiques* de manière à maximiser sa ressemblance avec d'autres occurrences de phrases comme : *McMaster et Hudack ont prouvé expérimentalement que les anticorps sont formés dans le tissu lymphatique*. Notons que la transformation met la phrase sous une forme paraphrastique de l'originale du point de vue de l'information pertinente dans ce sous-langage. Ces transformations peuvent être définies formellement mais on a également testé empiriquement que nos informants immunologues vérifiaient le caractère paraphrastique des phrases transformées.

Des mots ou des groupes de mots ont été considérés comme étant équivalents dans le discours quand ils apparaissaient dans des environnements linguistiques spécifiques. Par exemple des mots apparaissant dans le contexte : ... *a été injecté dans* ... (mais pas dans le contexte : *à été injecté de ... dans* ...) constituent une classe formellement définie de mots désignant des antigènes (représentés par une catégorie notée G). Des mots apparaissant dans le contexte *anticorps ... cellule* sont des expressions représentées par la catégorie V par exemple : *est produit par, se trouve dans, est secrété par*. On distingue des sous-catégories à partir de restrictions sur les domaines de sélection des principales catégories représentées.

41. On peut noter que, dès 1961, apparurent les premières indications d'une distinction, qui s'est depuis ramifiée, entre les lymphocytes dérivés du thymus et les lymphocytes dérivés de la moëlle épinière. Avec les macrophages qui sont une autre sorte de globule blanc, et peut-être d'autres encore, ces différents types de cellules fonctionnent ensemble pour élaborer des anticorps correspondant à des antigènes spécifiques par des moyens qui ne sont pas encore totalement compris.

Par exemple, on trouve une phrase comme : *Un anticorps est secrété par une cellule mais cet anticorps ne se trouve pas dans la cellule en question*, ce qui nous amène à distinguer deux sous-classes de V, Vi et Vs où seront représentés *est secrété par* et *se trouve dans*, respectivement. Les catégories de représentation ont été définies sur la base des propriétés de restriction de sélection de mots ou expressions linguistiques en partant des catégories définies comme élémentaires par les chercheurs du domaine. En définissant ces catégories, on a également défini les unités d'information pour le sous-langage, ainsi V a pour arguments A et C dans cet ordre où A représente des mots désignant des anticorps et C des mots représentant des cellules et la formule AVC représente des phrases exprimant la production d'anticorps dans des cellules.

La grammaire du sous-langage est commune à l'anglais et au français ***. Elle comporte environ 15 catégories de classes de mots. Les expressions ou mots de ces classes se combinent dans une douzaine de formules représentant des phrases canoniques. Chacune de ces classes de mots représente une classe d'équivalence du point de vue de l'information traitée. Bien entendu, ces équivalences ne sont pas nécessairement valides pour une langue prise dans son ensemble, par exemple, dans notre sous-langage, les deux phrases suivantes sont synonymes : *les cellules se sont multipliées* et *les cellules se sont divisées*.

Certaines unités d'information sont combinées pour constituer une unité d'information plus grande à l'aide d'une catégorie notée « : », qui joue le rôle d'une conjonction. Cette conjonction représente des expressions linguistiques exprimant en gros une relation de conséquence. Elles apparaissent de façon explicite ou implicite, dans les phrases qui mettent en jeu des expressions linguistiques représentées par deux unités informatives appartenant à deux sous-classes distinctes, par exemple GJB : AVC, ce qui permet d'exprimer qu'une injection d'antigène est suivie d'une réponse immunitaire. Cette « conjonction » représente des expressions appartenant à différentes catégories linguistiques comme : *après, provoque, à la suite de, consécutivement* à mais aussi une préposition *de* comme dans l'exemple : *la cellule contenait un anticorps spécifique de l'antigène* où l'antigène peut être reconstruit comme *l'antigène qu'on a injecté à un animal* parce que *antigène* est représenté par la catégorie G qui n'occure que comme co-argument de B (mots désignant les lieux de l'injection) sous l'opérateur J (mots désignant le fait ou l'action d'injecter). Cette conjonction « : » peut être affectée de modifieurs de temps exprimant des expressions comme : *quelque temps après, quelques jours après*. D'autres catégories d'information peuvent recevoir des modifieurs (notés par des indices) représentant des

*** NdT Ceci est dû au fait que si une représentation efficace de l'information dépend de sa formulation en langue naturelle, et même de propriétés distributionnelles fines des expressions linguistiques, la représentation de l'information pertinente pour ce domaine, et probablement pour tout domaine scientifique, ne dépend pas de sa formulation dans une langue particulière. On a vérifié pour ce sous-langage que les propriétés distributionnelles des expressions linguistiques pertinentes des textes rédigés en anglais étaient conservées pour les textes rédigés en français. On peut en fait considérer que ce sont ces propriétés distributionnelles qui définissent comme tel le sous-langage.

mots ou expressions linguistiques d'origines grammaticales très diverses pour chaque indice, comme c'était déjà le cas pour les catégories principales.

On a également distingué l'information scientifique proprement dite et la méta-science (matériel linguistique qu'on a représenté par M indistinctement), qui spécifie la relation du chercheur avec la science. Ce matériel occure dans différents environnements grammaticaux et doit donc être distingué sur la base des propriétés combinatoires des mots et non en utilisant des notions sémantiques définies *a priori*. Il s'agit de séquences comme : *X et Y ont montré que ... ; que ... peut difficilement être mis en doute ; du fait que ..., on peut difficilement mettre en route que*. Ce matériel n'a pas été soumis à une analyse plus fine, qui aurait probablement révélé une structure interne substantielle, du fait de notre objectif qui était d'analyser les propriétés de la formulation de l'information dans ce domaine de l'immunologie.

Nous n'avons fait que donner un rapide et informel aperçu de ce travail qui va cependant nous permettre de voir maintenant quel cadre a été défini pour représenter (une bonne part de) l'information de ce domaine. La représentation d'une phrase type (une fois le matériel méta-scientifique mis à part) est une unité d'information, qui contient une combinaison applicative de catégories informatives, présentées séquentiellement. Par exemple, l'unité GJB est une séquentialisation de l'application de J sur G et B. Une unité d'information peut être simple ou constituée de deux unités, appliquées sous un opérateur « : ».

Ainsi qu'on peut s'y attendre pour une recherche qui s'est développée sur une trentaine d'années, les phrases types les plus courantes ont changé au cours du développement des textes. Les premiers articles contenaient presque exclusivement des séquences traitant des changements histologiques faisant suite à l'injection d'antigène. Les articles suivants mettent l'accent sur les différents types de cellules contenant des anticorps après l'injection répétée d'antigène. Après que le microscope électronique ait considérablement amélioré le seuil de résolution du visible, l'accent a été mis sur les changements cellulaires faisant suite à la stimulation antigénique et on a déterminé que l'anticorps était présent dans des cellules particulières.

L'analyse de ce sous-langage révèle plusieurs choses intéressantes du point de vue du domaine scientifique lui-même. Le changement continu de la connaissance apporté par de nouveaux tests immunologiques est indiqué par les changements intervenant dans les structures informatives des textes : apparition de nouvelles structures, disparition progressive ou rareté des anciennes. Les controverses se manifestent par différentes attitudes en partie localisées dans le matériel M du sous-langage et en partie exprimées à travers les structures informatives communes. Considérées dans leur ensemble, ces structures constituent un système de points de références qui permettent de notifier des résultats, de déterminer dans quelle mesure les structures d'information établies à un certain moment suffisent ou requièrent des modifications. Cette représentation de l'information révèle si un article donné dit quelque chose de nouveau. La nouveauté d'un article scientifique consiste précisément à modifier les structures

d'information connues (i.e. introduire de nouvelles catégories opérant sur le ou les arguments de celles déjà mentionnées ou introduire de nouveaux modificateurs).

En conclusion, je voudrais rapidement souligner l'apport de Harris du point de vue de l'explication en linguistique. Il y a eu récemment de nombreuses discussions sur la nécessité qu'aurait la linguistique, en tant que science, à dispenser des « théories explicatives » s'opposant aux « simples » descriptions des données ou des comportements linguistiques. Influencés par les courants réalistes qui prévalent actuellement dans la philosophie de la science, ces tenants de la linguistique « explicative » pensent que les régularités et structures linguistiques discernées ne peuvent être analysées en dernier ressort qu'en référence à des structures biologiques ou psychologiques sous-jacentes. Dans cette perspective, la linguistique, avec ses données sociales et historiques contingentes et avec ses caractérisations « abstraites » de ces réalités sous-jacentes devrait se résoudre finalement dans une science future de la « biologie cognitive ». Néanmoins, pour les temps à venir, il faudrait pousser la linguistique vers des théories de plus en plus abstraites et « profondes » reposant sur des niveaux de déduction de plus en plus éloignés des données observables des langues. Il s'agit en effet d'un programme de recherche audacieux, qui semble en grande mesure calqué sur le modèle de l'histoire récente de la physique.

Quelles que soient les difficultés internes d'un tel programme, il devrait être clair que le travail rapidement décrit ici a une orientation totalement différente tant pour ses méthodes de formalisation que pour ce qu'il veut expliquer. Comme Harris l'a fait remarquer, « la généralité n'est pas la même chose que l'abstraction »⁴². J'ai essayé de montrer qu'en défendant l'autonomie de la théorie linguistique (avec pour conséquence l'inclusion de la métalangue de description dans la langue décrite), Harris ne dépend pas des contraintes extérieures sur la façon de faire la science qu'avaient les positivistes, qu'ils soient contre ce qui ne peut être observé ou qu'ils soient pour le caractère instrumental des théories. Il a toujours cherché à rendre compte de la structure d'une langue comme système de transmission de l'information évoluant à travers un processus continu d'institutionnalisation de l'usage⁴³. Comme résultat de différents processus d'institutionnalisation, l'unicité du développement des langues est réduite ainsi que le supposé mystère du langage, qui a donné lieu à tant de nobles spéculations. Dans cette conception, le langage procède moins d'un génome humain que d'une pratique sociale partagée, point qui est particulièrement important quand on considère les langages particuliers des sciences particulières. Le langage est le meilleur moyen que nous ayons, non pas pour exprimer le sens des choses, nous avons pour cela beaucoup de moyens quasi-linguistiques ou non linguistiques, mais pour articuler, délimiter et transmettre le sens, comme information structurée, d'un individu ou d'un

42. Introduction de *PS*.

43. *LI*, p. 107-113.

groupe à une communauté plus vaste. De plus, les moyens eux-mêmes par lesquels le langage véhicule l'information ne sont pas quelque chose d'externe à la langue mais simplement la structure de ses contraintes qui agissent les unes sur les autres. Ainsi, l'aspect fondamental de ces pratiques sociales qu'on appelle le langage, qu'il s'agisse du langage naturel dans son ensemble ou des emplois particuliers du langage dans des domaines de connaissance spécifiques, est une caractéristique informative.