

Antoine Culioli

La formalisation en linguistique

Dès l'abord, il importe de fixer la visée de cet article, afin d'éviter les malentendus et d'assurer la démarche du lecteur à travers un ensemble composite de réflexions épistémologiques et méthodologiques, de survols ou de schématisations qui supposent une bonne connaissance de la linguistique, enfin d'incursions rapides dans le domaine même de la pratique linguistique. C'est dire que les lignes qui suivent ne cherchent pas à présenter un problème technique (« Comment formaliser tel secteur, telle question, tel texte ») ou à dresser un inventaire raisonné des procédures formalisantes. Pour l'essentiel, on cherchera ici à poser des questions sur ce que le linguiste entend et ce qu'il fait quand il parle de formalisation des langues dites naturelles, car il n'est pas sûr qu'il s'entende quand il parle, tant il est le plus souvent englué dans un empirisme naïf « ou faut-il dire « spontané ») qui occulte les problèmes théoriques.

En ce sens, le présent article est une mise en garde : il ne s'agit pas ici de mettre en cause la formalisation-en-linguistique, mais de marquer les dangers d'un engouement fasciné, aux racines multiples, qui risque très vite d'avoir des effets nocifs : espoir que la machine¹ fournira une commodité expéditive dans l'analyse du langage (alors que l'ordinateur ne peut que vérifier l'adéquation d'une théorie en vérifiant la pertinence et la consistance d'un jeu de descripteurs, mais ne permettra jamais de faire l'économie du travail théorique); illusion qu'une symbolisation sténographique permettra « d'y voir plus clair » et, par là, d'établir sans trop de frais une typologie parallèle des classes de conduites et des classes de discours (qu'il s'agisse, par exemple, de pathologie ou de production littéraire); incohérence dans l'emploi des modèles, facilitée par le désir d'être inter-disciplinaire, par l'emprise de concepts mathématiques mal assimilés et par une réflexion insuffisante sur ce qui est, en droit, le thème de la science linguistique : le langage appréhendé à travers les langues naturelles.

C'est ainsi qu'au moment où la linguistique redécouvre le langage, au lieu de construire son objet, elle le clive dans des recherches aux intentions différentes, qui impliquent des modèles parfois incompatibles : la consé-

1. Par *machine*, on entend ici une machine réelle, et non une machine abstraite comme la machine de Turing.

quence, inévitable, est une *réduction* du langage, pour des raisons techniques dont on n'a le plus souvent pas conscience. En particulier, il apparaît clairement que la formalisation irresponsable — ou le refus aussi irresponsable de poser le problème théorique de la formalisation en linguistique — empêche de bien marquer la relation dialectique entre le langage et les langues. Le discours du linguiste se clôt facilement dans des jeux de réécriture qui, à la différence des mathématiques, ne sont ni rigoureux ni féconds¹, ou se bloque dans la description de langues particulières, dites irréductiblement spécifiques. Dans ce dernier cas, une langue individuelle est considérée tantôt comme un objet empirique intuitivement donné (on regarde fonctionner l'anglais, le chinois, etc.), tantôt comme une essence (la langue au sens saussurien du terme, ou le langage) dont la « pureté » aurait été brouillée.

Enfin, trop de linguistes ignorent que l'on peut, à volonté, construire des systèmes formels, et prennent pour une propriété de l'objet ce qui est une propriété du modèle ou même un expédient éphémère. Vieux problème, mais qui a une importance singulière dans la linguistique contemporaine (cette partie, du moins, qui consent à sortir du cercle magique de l'idéologie positiviste). Pour certains, la linguistique formelle, c'est d'abord une arborescence (dichotomique) et puis un système générateur syntaxique, radicalement séparé de la sémantique qui, comme dans tout système formel, est une interprétation des expressions bien formées (syntaxiquement). Le problème n'est pas, à ce stade, d'apprécier le degré de validité d'un tel modèle, mais de reconnaître tout d'abord qu'il est une caricature de la doctrine de Chomsky; ensuite, à la différence de Chomsky, qui connaît (et énonce) ses postulats, trop de linguistes épris de formalisme ne savent pas ce qu'ils font, pourquoi ils le font, et cèdent tout simplement à la fascination du bidule, de l'objet technique dont on sait qu'il a un mode d'existence privilégié. Faut-il rappeler que le problème méthodologique de la linguistique (parmi les autres sciences humaines) est de *trouver*, c'est-à-dire se fabriquer, les outils logico-mathématiques qui permettront de donner une description adéquate de l'activité langagière saisie à *travers* les langues? Rien ne nous permet de penser que les mathématiques actuelles sont forcément appropriées, que les êtres grammaticaux, même mathématisés, avec lesquels le linguiste opère, ont une valeur autre que traditionnelle.

Que la linguistique doive être systématique, rigoureuse, explicite, ouverte à la vérification, tout le monde est prêt à accepter d'enthousiasme un tel programme, à supposer que les bonnes intentions constituent un programme. Mais les résistances affleurent très vite, dès qu'on veut se donner des

1. Afin d'éviter toute équivoque, il importe de dire en clair que ceci n'est pas une critique de N. Chomsky (ou de la majorité des linguistes qui travaillent avec Chomsky). Le rôle de Chomsky a été, et reste, capital sur le plan tant épistémologique que strictement linguistique, et l'on souhaiterait partout aussi peu de dogmatisme et autant de prudence. Mais cette attitude ouverte se ferme vite dans le discours pseudo-scientifique et caricatural de certains épigones.

moyens à la mesure de ses ambitions. Énumérons en vrac quelques problèmes que l'on doit à tout le moins poser, de façon liminaire :

1. *Observables et modèles* : le problème de la relation entre un modèle, l'objet et l'observateur, n'est certes pas propre à la linguistique! mais il a pour le linguiste une importance capitale, car, aux dangers habituels, viennent s'ajouter des difficultés supplémentaires.

a) La métalangue est la langue d'usage (dans le meilleur des cas, il faudra toujours utiliser la langue U pour parler sur la métalangue ou le système formel; en fait, le plus souvent, l'intrication entre terminologie et langue d'usage est telle que le linguiste se trouve pris au piège dont il voulait explorer le fonctionnement).

b) Le langage est une activité qui suppose, elle-même, une perpétuelle activité épilinguistique (défini comme « activité métalinguistique non consciente », ainsi qu'une relation entre un modèle (la *compétence*, c'est-à-dire l'appropriation et la maîtrise acquise d'un système de règles sur des unités) et sa réalisation (la *performance*) dont nous avons la trace phonique ou graphique, des *textes*¹.

c) L'activité langagière est *signifiante* : c'est parce qu'il y a, dans la communication, des opérations aux deux bouts que les énoncés prennent un sens (opérations complexes, car tout émetteur est en même temps, c'est-à-dire *au même moment*, récepteur, et réciproquement); mais on ne peut affirmer que les mots *ont* un sens sans être ramené à une conception outillière du langage, conçu comme un instrument dont la finalité explicite serait la communication entre des sujets universels qui, comme on le sait, se partagent le bon sens. Or, on peut montrer que le langage n'est pas extérieur au sujet (terme employé, faute de mieux, pour éviter *locuteur* ou *parleur*), mais est dans une relation complexe d'extériorité-intériorité; en outre, le code (même dans sa partie collective ou, plutôt, transindividuelle) a besoin d'un support, doit donc coder quelque chose, mais ne saurait être bijectif, car s'il y avait correspondance biunivoque, on ne saurait s'expliquer l'existence des malentendus, voire d'une certaine classe de métaphores. De même, une conception instrumentale du langage évacue le lapsus, considéré comme simple raté. Quant à la modulation du discours (rhétorique, style), elle risquerait de devenir les vêtements qui habillent la pensée, un luxe surajouté à l'automate syntaxique, alors qu'elle est inhérente au système même : une des propriétés du langage humain est de se prêter à l'axiomatique euclidienne et à l'image poétique. En fait, le langage fonctionne à des niveaux différents (dénotatif/connotatif; extrinsèque/intrinsèque; univoque/équivoque; cognitif/affectif; système de signes discrets/système symbolique c'est-à-dire analogique; etc.)

1. On évitera ici toute assimilation de *compétence* à *langue* et de *performance* à *parole*! Chomsky lui-même est fort circonspect sur ce point.

d) On ne saurait ramener les problèmes de catégorisation à de simples généralisations fondées sur la fréquence. Ceci a été clairement montré dans une étude de D. McNeill sur l'apprentissage du japonais par une enfant de deux ans : il existe en japonais deux particules postposées, *wa* et *ga*, la première correspondant en gros à « quant à x, il... » ou encore « il y a x, et il... », la seconde à « c'est x qui... » ou simplement « x est à l'origine de tel procès ». La mère utilise deux fois plus de *wa* que de *ga*, mais l'enfant emploie cent fois *ga* pour six *wa* (sur huit heures d'enregistrement), et montre qu'elle sait utiliser *ga* avec ses deux valeurs (« C'est x qui, c'est x que », d'un côté, et « x sujet », de l'autre). Ainsi, l'enfant a compris le système, a repéré celle des deux particules qui correspondait à la fonction centrale de prédication, bien qu'elle soit la moins fréquente, et en a appris l'usage correct. De même, aucune étude de fréquence ne permettra d'expliquer le statut du masculin par rapport au féminin dans tant de langues, ou de l'animé par rapport à l'inanimé.

La conclusion de cette énumération est que l'on ne peut poser le problème des observables sans se donner une théorie de l'observation, en particulier, sans se demander où l'on poste les observateurs. C'est ce que marque bien la distinction entre *surface structure* et *deep structure* chez Chomsky : la configuration de surface est la trace d'opérations sous-jacentes. Or, étudier le processus de production, cela signifie que l'on quitte le domaine de l'observation illusoirement immédiate pour opérer abstraitement. (Il ne saurait y avoir de formalisation de la surface¹; on ne peut décider *a priori* qu'il n'existera que deux niveaux, superficiel et profond, sauf pour poser une distinction rudimentaire; il est peu probable que l'on puisse se donner comme objets (termes et relations) primitifs ceux que l'on trouve en surface, simplement transportés à l'étage au-dessous). Ensuite, il faudra retrouver par le calcul les énoncés possibles, et ceci jusque dans le détail, puisque la linguistique formelle ne se donne pas pour tâche de survoler les langues dans leurs généralités, mais de rendre compte de ce que l'on trouve, dans toute sa diversité, sans aucune exception (des exceptions elles-mêmes, on doit pouvoir justifier le caractère exceptionnel).

Se construire une théorie de l'observation, cela implique en outre que l'on classe les modèles selon les questions auxquelles ils répondent, exigence fondamentale dès que la recherche s'articule sur plusieurs champs, par exemple en psycho-linguistique ou dans l'analyse littéraire. Il sera indispensable en outre d'avoir une théorie des représentations (au sens des procédés graphiques : toute formalisation est une écriture) : les modèles sont-ils équivalents, compatibles ? Les représentations sont-elles isomorphes ? Tel mode de représentation est-il opératoire, c'est-à-dire, sait-on l'utiliser pour calculer ? Ici devrait se greffer une théorie de l'approximation qui permettrait d'évaluer la force et la régionalité d'un modèle. Mais longtemps, la linguistique n'a pas éprouvé le besoin d'une telle théorie, car son empirisme épis-

1. C'est pourtant l'entreprise théorique qu'a tentée la linguistique « structurale ».

témologique lui donnait bonne conscience et lui faisait imaginer qu'on finirait bien par grignoter l'objet arbitrairement réduit; bel exemple de confiance en l'inexhaustible exhaustivité de la connaissance. Au contraire, formaliser devrait amener à reconnaître qu'aucun modèle n'est exhaustif et à en tirer les conséquences scientifiques. Ajoutons que l'on ne doit pas confondre une théorie de l'approximation (observateur) et une théorie des approximations et de l'erreur dans l'activité langagière (observable), de même que l'on ne devra pas confondre l'analyse d'énoncés ambigus et l'étude de l'ambiguïté foncière du langage.

2. *Concepts, termes et symboles* : d'emblée, on est frappé de la confusion de nombreux travaux, provoquée et entretenue par la labilité du langage et son double statut¹. On peut projeter de parler sur une langue et finalement toujours parler dans cette langue, sans même s'en apercevoir grâce au double fond du langage. Ainsi, on voit s'introduire de graves confusions entre opérateurs et méta-opérateurs, entre schéma-noyau (abstrait) et phrase empiriquement présente, entre un énoncé et une phrase, etc. Ce ne sont point ici remarques grincheuses d'académicien de la linguistique ou de puriste de la formalisation; ce qui est en cause, c'est, d'un côté, la possibilité de systématiser la linguistique naïve, afin de pouvoir ensuite formaliser une linguistique axiomatisée. Faut-il rappeler ici que l'on ne peut formaliser que quelque chose (à moins de vouloir construire des systèmes ininterprétables) et qu'on ne peut pas formaliser du naïf²? D'un autre côté, l'absence de rigueur risque de produire un système de ré-écriture qui se ferme sur lui-même, alors que l'exigence formelle contraint à ne pas se contenter de faux-fuyants : ainsi, l'assimilation de *syntactique* à *formel*, d'origine mathématique et licite en soi, est seulement permise en linguistique à titre de première approximation. Sinon, on en revient à l'automate, instrument aux mains d'un sujet libre qui crée et interprète, à la forme et au fond, au signifiant et au signifié, et la suite.

De même, si l'on étudie sa langue maternelle, on peut avoir l'illusion, étant de plain-pied, que les unités, opérations, valeurs, que l'on décrit, sont primitives, bref, qu'il n'y a qu'un rez-de-chaussée (le reste étant pour le psychologue). Mais ces unités, opérations et valeurs, on les désigne au moyen de méta-termes (par ex. *substantif, verbe, auxiliaire, sujet, complément, actif, passif, négation* etc.) qui, tous, se ramènent à une conception morphologique et distributionnelle de la grammaire. Puis, il faut ajouter, comme nous l'avons déjà signalé, que toute unité de langage, même métalinguistique, est forcément engagée et ambivalente : *passif* par exemple peut être un méta-terme, mais il est en même temps surchargé³ de contenus intuitifs. A sup-

1. Sur ce point, voir plus haut (métalangue et activité épilinguistique).

2. Entendu au sens de « intuitif, non déductif, non théorisé ». A distinguer de l'emploi que Bourbaki fait de l'adjectif, quand il parle de *mathématique naïve*, ceci pour des raisons évidentes.

3. Au sens où Bachelard parle de surcharge : « On devrait donc toujours se méfier d'un concept qu'on n'a pas encore pu dialectiser. Ce qui empêche sa dialectisation c'est une surcharge de son contenu. Cette surcharge empêche le concept d'être délicatement sensible à toutes les variations des conditions où il

poser que l'on puisse le rendre univoque, il reste que, dans toute langue où existe la passivation, cette transformation introduit (ou, plutôt, peut introduire) une modulation sémantique qui interdit de considérer le passif, sauf au plan d'une syntaxe stricte, c'est-à-dire vide, comme le simple dual de l'actif. Or, ceci vient, par rétroaction, introduire à nouveau dans le mot une surcharge incontrôlée.

En fait, tout système algébrique d'opérateurs (et de méta-opérateurs) vient s'imbriquer dans un système de variables sur lesquelles il porte. Ce second système a ses structures propres, et ces variables sont, à leur tour, insérées dans un réseau référentiel qui est en correspondance avec les situations vécues, l'organisation de nos conduites perceptives et pragmatiques, voire nos élaborations fantasmatiques.

Que l'on n'oublie pas, en outre, que nous appréhendons les universaux seulement à travers les variations de surface de chaque langue : d'où un risque accru de donner un statut scientifique à des conceptualisations faussement objectives, où se mêleraient les théorisations explicites et implicites. En conséquence, on ne saurait partir, sauf pour des raisons évidentes de commodité, d'une seule langue, où l'on ferait la description de tel phénomène (ainsi, on dresserait en français l'inventaire raisonné des emplois de *être* et *avoir*), en espérant qu'ensuite on pourrait transporter dans une autre langue ce qu'on a découvert dans la première.

C'est croire que l'on trouve des structures (au sens fort) à fleur de langue et que l'on peut échapper, par une simple intention de rationalité, au leurre terminologique; c'est sous-estimer la surcharge parasite qu'entraîne toute conceptualisation sur le langage. Pour pouvoir itérer avec une stabilité raisonnable (donc avoir un noyau d'invariants qui, seul, permet la prévision), il faut munir les êtres linguistiques d'une structure, cette structure provient d'une théorie de l'analyse qui, elle-même, doit être reliée à une théorie du langage. On devra donc, à partir des langues (c'est-à-dire, en pratique, à partir de langues) se donner une métalangue, avec ses règles, puis revenir aux langues. Au lieu de cela (même s'il ne s'agit que d'un projet, lentement réalisable), on trouve un mélange de catégories mal définies, le plus souvent héritées, de relations considérées comme allant de soi, ou une utilisation schématique de classements binaires qui démarquent la phonologie¹,

$$\text{du genre } \textit{célibataire} \left\{ \begin{array}{l} \text{humain} \quad + \\ \text{m\^a}^{\text{l}}\text{e} \quad + \\ \text{adulte} \quad + \end{array} \right\} \text{ ou } \textit{branche} \left\{ \begin{array}{l} \text{animé} \quad - \\ \text{artéf}^{\text{a}}\text{ct} \quad - \\ \text{végétal} \quad + \\ \text{dénombrable} \quad + \end{array} \right\}, \text{ etc.,}$$

prend ses justes fonctions. A ce concept, on donne sûrement trop de sens puisque jamais on ne le pense formellement. Mais si on lui donne trop de sens, il est à craindre que deux esprits différents ne lui donnent pas le même sens. » (G. Bachelard, *la Philosophie du Non*, Paris, p. 134.)

1. On trouve d'autres représentations (chemin, arbre), mais les problèmes théoriques restent.

sans que les présupposés théoriques de telles procédures soient vus clairement¹.

3. *Syntaxe et sémantique* : rassemblons ici quelques problèmes entrevus plus haut, en les énonçant sous forme de propositions :

a) rien ne permet de ramener la sémantique des langues naturelles à la sémantique interprétative des systèmes formels;

b) tout signe peut être utilisé comme symbole et les opérateurs syntaxiques n'échappent pas à cette règle : tout changement syntaxique entraîne un changement sémantique (quelle que soit l'acception de cet adjectif : régulation inter-individuelle par une communauté, pragmatique, rhétorique, etc.). Dire qu'une phrase et sa transformée sont équivalentes ne change rien au problème : tout réside dans la force de l'équivalence; certes, en termes de voisinage, on pourra établir une distance croissante, de la transformée identique à une phrase profondément remaniée, mais il est difficile, hormis dans une théorie (à expliciter) du bon sens et de la conservation du sens, de soutenir que, seule, la structure profonde reçoit une interprétation sémantique et que les transformations ne changent rien. Ou plutôt, à un tel degré d'approximation, pareille proposition peut avoir une valeur heuristique, mais n'a pas de validité théorique;

c) il est permis, au contraire, de poser, en étayant la thèse sur des arguments théoriques, qu'il existe à un niveau très profond (vraisemblablement prélexical) une grammaire des relations primitives où la distinction entre syntaxe et sémantique n'a aucun sens. On aura ensuite un filtre lexical, avec un certain nombre de règles et syntaxiques et sémantiques, y compris la modulation rhétorique (métaphores, glissements de sens), qui ne saurait être ramenée à de la syntaxe. Nous sommes ici, en effet, dans le continu, et non dans le discontinu, et aucune représentation de type syntaxique (contraintes distributionnelles sur la co-occurrence; ordre partiel) ne suffit à rendre compte du langage poétique par exemple, à moins que l'on ait recours au piètre argument de la déviance et de l'anomalie qui, de toute façon, ne règle rien.

Après un autre filtrage² on obtient une *lexis*, où les termes sont compatibles avec un ordre, mais ne sont pas encore ordonnés; en outre, la *lexis* est pré-assertive et le passage à l'assertion (au sens de « énonciation par un sujet ») implique une modalisation. *Modaliser* signifie « affecter d'une modalité » et *modalité* sera entendu ici au quadruple sens de (1) affirmatif ou négatif, injonctif, etc. (2) certain, probable, nécessaire, etc. (3) appréciatif : « il est triste que...; heureusement » (4) pragmatique, en particulier, mode allocutoire, causatif, bref, ce qui implique une relation entre sujets. Outre la modalisation,

1. Passons sur les emplois relâchés du programmeur qui parle de formaliser un texte, alors qu'il s'agit d'un codage utilisant un langage formel, du linguiste qui parle de formalisation dès qu'il emploie un symbole ou un diagramme. A ce niveau-là, il n'est même plus question de présupposés théoriques implicites ou diffus.

2. Le terme de *filtre* (ou *crible*) n'a d'autre fonction ici que métaphorique.

le passage à l'assertion s'accompagne d'un second type de modulation, que l'on pourrait appeler *stylistique*, pour le distinguer du premier type, ou modulation rhétorique. Il s'agit là d'une pondération des éléments, soit par des manipulations analogiques¹ qui se réaliseront comme traits prosodiques, soit par des permutations, etc. On a ainsi une séquence préterminale dont les éléments sont partiellement ordonnés, et pondérés. La projection de cet ordre partiel sur la chaîne aboutira à un assemblage séquentiel de termes, sur lequel est définie une relation d'ordre total, non absolu.

Si on récuse ce modèle, on pourra poser l'existence de deux systèmes générateurs, l'un syntaxique, l'autre sémantique, entre lesquels il existe des correspondances. On pourra encore concevoir la sémantique comme une hypersyntaxe, un passage à la limite quand on a épuisé l'analyse syntaxique. Tout vaudra mieux que la séparation *essentielle* de syntaxe et de sémantique, qui ramène, inéluctablement, à une syntaxe avec un lexique muni de règles projectives. En bref, on posera le caractère licite d'une sémantique formelle (depuis Frege et Husserl on s'en doutait); on posera qu'il existe des énoncés bien formés sémantiquement et mal formés syntaxiquement; et l'on reconnaîtra que la difficulté centrale de la formalisation en linguistique ne réside ni dans la formalisation de systèmes algébriques syntaxiques, ni dans l'étude distributionnelle des combinaisons de mots-objets en correspondance ponctuelle avec la réalité extra-linguistique, mais dans le domaine intermédiaire, spécifique des langues naturelles, où il nous faut découvrir sur quels êtres travailler, construire des types de logique inconnus à ce jour et qui ne fonctionnent sans doute pas de façon homogène, doser la force des concepts, « ces instruments d'effraction », que nous proposent les mathématiques et les adapter à nos fins.

Ainsi, nous ignorons les structures mathématiques qui se révéleront adéquates et fécondes : comme nous le disions plus haut, nous aurons vraisemblablement à les inventer avec l'aide du mathématicien, puis, encore plus vraisemblablement, à les « bricoler », du moins dans une première étape. Qu'il soit bien compris que l'on n'importe pas des techniques logico-mathématiques pour les plaquer sur un objet quelconque.

Il serait puéril d'indiquer au lecteur où puiser, puisque, justement, il n'est pas question de puiser dans un stock d'outils, mais de prendre son bien où on le trouve (combinatoire et algèbre, topologie, etc.). Disons simplement que le linguiste aura parfois des concepts-clés à portée de main (par ex. application, structure, ordre), parfois l'élaboration sera lente (ainsi en est-il de l'utilisation de la topologie en linguistique ou encore de la logique combinatoire), parfois il faudra tout faire (ainsi, dans le domaine des modalités).

Afin d'illustrer ces propos, nous voudrions, dans les lignes qui suivent,

1. En plus-ou-moins, et non en tout-ou-rien.

attirer rapidement l'attention sur quelques concepts importants, puis sur une propriété caractéristique du langage.

1. On ramènera toutes les opérations unaires de prédication (à l'exclusion ici des transformations de composition sur des lexis) à une application, ce qui n'a rien que de très banal, mais on ira jusqu'au bout de l'analyse, en y adjoignant une théorie des prédicats. On obtiendra ainsi une typologie des procès, une classification des opérations que l'on peut effectuer sur l'ensemble de départ et/ou sur l'ensemble d'arrivée, sur la flèche qui symbolise le foncteur.

On est alors en mesure d'analyser formellement les situations empiriquement rencontrées dans les langues, ainsi que des transformations telles que la passivation et autres phénomènes connexes. On est surtout à même d'accroître la complexité du modèle, en introduisant par exemple la composition de deux applications.

2. On peut donner des valeurs d'un système verbal (un système étant défini comme un réseau de valeurs) une représentation topologique qui permet de mieux poser certains problèmes concernant les systèmes de modalités, et singulièrement de relier les systèmes modaux, aspectuels et temporels.

3. On peut ramener les opérations sur les unités dans l'ensemble de départ et dans l'ensemble d'arrivée à une liste finie d'opérateurs que l'on pourra ensuite combiner (par exemple, opérateur de classe : *le chat est un félin domestique*, pour ne prendre que le cas le plus trivial; flécheur, qui distingue un élément, soit un individu, soit une portion : *le* (dans certains de ses emplois), *ce*, *mon*, etc.; extracteur, par exemple dans *il y a un chien qui aboie*; curseur, qui parcourt, ou balaie, la classe : par exemple, *tout*, *quiconque*, anglais *any*; opérateur qui fait que l'on considère la classe comme renvoyant à une « notion » : *un bruit de machine*, *une odeur de rose*).

4. On recherchera les relations de dualité qui existent entre des expressions (au sens formel du terme), mais il importe de noter que dans le langage, où tout est orienté, il se trouve des secteurs où le principe de dualité joue de façon stricte et d'autres où les phénomènes sont plus complexes. De toute façon, il importe ici de distinguer avec un soin particulier ce qui est langagier de ce qui est linguistique, pour ne s'en tenir qu'à cette dichotomie simple.

5. On représentera certaines catégories par des vecteurs de propriétés, de telle sorte que l'on pourra avoir des vecteurs de vecteurs. Ainsi, on notera Sujet (d'une phrase) = (C₀, Agent, Thème). C₀ se lit *Complément de rang zéro* dans une théorie des compléments, impliquant que le sujet (de surface), dans les langues où sa présence est obligatoire, est nécessaire comme élément de l'ensemble de départ, pour que l'énoncé soit canoniquement bien formé; *Agent* implique que très souvent (la formulation est volontairement approximative) le Sujet est Agent, soit à un niveau de surface soit, en inversant la proposition, que l'Agent était Sujet (C₀) à un niveau profond. En outre, cette notation indique que, même quand le C₀ n'est pas Agent, on tend à lui attribuer des propriétés d'Agent; *Thème* ne demande guère d'explication à première vue, puisque l'acception du terme est ici proche de l'acception

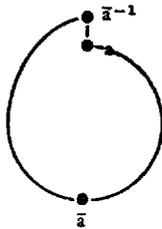
traditionnelle. Une analyse plus fine révélerait cependant des problèmes complexes.

Ce vecteur est « coulissant », c'est-à-dire que chaque terme, à l'exception de C_0 , peut prendre une valeur nulle. On peut donc avoir : $(C_0, Ag., Th.)$, $(C_0, Ag.)$, $(C_0, Th.)$, (C_0) . De son côté, Agent est représenté dans un autre vecteur (Agent, Animé, Déterminé), lui aussi coulissant¹.

Une telle notation permet d'éliminer bien des ambiguïtés et maladresses dans la métalangue, de mener certaines analyses linguistiques et psychologiques qui, sinon, tourneraient court; d'une façon générale, il s'agit ici d'une combinatoire beaucoup plus complexe que dans l'analyse dite structurale, où l'on a affaire à des structures pauvres.

6. On construira des systèmes logiques particuliers, du type 0,1 (où 0 peut être un absorbant² selon les systèmes), \star (terme neutre, ce qui signifie : « qui n'est ni 0 ni 1 ou bien est 0 ou 1 »), ω (terme qui est en dehors de $(0, 1, \star)$)³. On retrouve ici, mais sous une forme plus féconde, le concept de marque (non-marqué : - / marqué +), et cela a quelque rapport (malgré de profondes différences) avec le système de Bröndal. L'important est de bien comprendre que, seule, une décision théorique (théorie du langage) permet d'attribuer à telle unité le statut du 0 ou origine (par exemple dans masculin/féminin, actuel/non-actuel), puis le statut de successeur, etc.

7. De très nombreux systèmes sont munis d'une structure en « came »⁴, sur laquelle nous reviendrons dans un autre article, de la forme suivante :

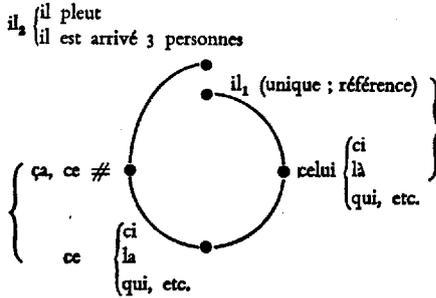


Il ne s'agit pas ici d'une involution : le schéma n'est pas à 2 dimensions, et \bar{a}^{-1} amorce une spirale, puis se projette en a , et le cycle recommence. Ce modèle, d'une grande importance dans les langues naturelles, permet de mieux concevoir certains problèmes touchant à l'ambiguïté, l'ambivalence (au sens psychanalytique du terme), et d'une façon générale fait sans doute apparaître une propriété fondamentale du langage.

Ainsi, pour fixer les idées, on pourra parmi de nombreux domaines prendre les trois exemples suivants :

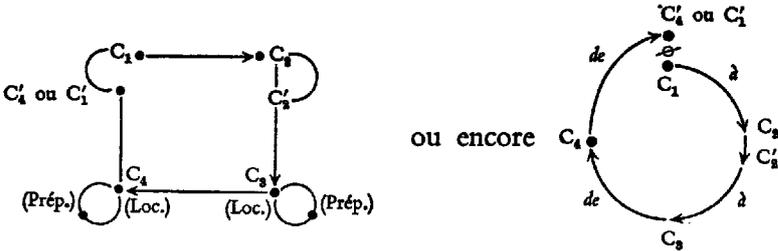
1. C'est par hasard que l'on a, à deux reprises, un triplet.
2. Ainsi, *masculin + féminin*, en français, donne *masculin* ; de même *nous + vous + eux* donne *nous*.
3. Peu importe les symboles adoptés, naturellement.
4. Le terme m'a été suggéré par F. Bresson. Il s'agit d'un terme général pour renvoyer à une catégorie ; certains des points considérés ici sont bien connus des mathématiciens.

a) On peut ramener le jeu de *il* et *ce-* à une came ¹.



On passe ainsi de *il*₁ qui renvoie à un représentant unique, masculin, à ce qui n'est ni déterminé ni indéterminé, ni masculin ni féminin (*il*₂)². On pourra ensuite jouer sur ce système et produire aussi bien *Les chats, ça griffe* que, chez Giraudoux : « Aujourd'hui, cela a tué. Je parle de l'inoffensif. Cela va en prison pour meurtre. Cela a saccagé sa vie. Cela vous a vue. Cela a été heureux » (*Pour Lucrèce*), où *cela* renvoie à un homme.

b) De même, on montrera que la syntaxe des pronoms incorporés du français (type *je lui en donne*) obéit à des règles strictes et que l'on a, à nouveau, un système qui se boucle. Si l'on appelle C₁ *le, la, les*, C₂ *lui, leur*, C₂' *γ* dans *j'γ donne* (= *je le lui donne*), C₃ *γ* de localisation (*j'γ vais, j'γ reste*) et « prépositionnel » (*j'γ pense*), C₄ *en* de localisation (*j'en viens*) et « prépositionnel » (*il l'en frappe*), C₄' *ou*, indifféremment, C₁' *le en d'extraction* (*j'en prends*), par rapport à *je le (s) prends*), on obtient le diagramme suivant :



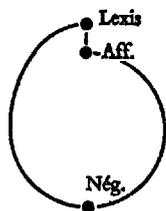
L'intérêt d'une telle représentation est qu'elle force à prendre des décisions, donc à poser un problème : ainsi, quel sera le point de départ ? Pourquoi analyser de la sorte le système des pronoms ? En outre, il importe de

1. Le choix de ces deux pronoms est didactique ; il aurait été trop long d'exposer la question dans sa totalité.

2. Il faudrait expliciter ce qu'on entend ici par *déterminé* et justifier la place de *ça* (on voit que la détermination décroît de *il*₁ à *ça*, que *il*₂ est non-déterminé, c'est-à-dire en dehors de l'opposition *déterminé/indéterminé*) ; il faudrait rattacher ceci au point (s) ci-dessus ; enfin, il serait nécessaire d'expliquer pourquoi *il*₂ qui n'est référentiellement ni un masculin ni un féminin est, morphologiquement, au masculin. Ces démonstrations ne sont pas faites ici pour de simples raisons de place.

comprendre que le diagramme n'est pas un jouet, une illustration pour soutenir l'intuition, mais un outil avec ses règles formelles d'emploi. Enfin, on peut ainsi rendre comparable les faits dans des langues diverses, comparer l'emploi de *de*, *à*, \emptyset devant un infinitif et devant un substantif, etc.

c) Si l'on fait une étude complète de l'assertion, on peut montrer les relations qui existent entre négation et interrogation, interrogation et hypothétique, situer l'injonctif par rapport à l'assertorique, etc. Contentons-nous ici de considérer la lexis par rapport à l'assertion positive (affirmation) et négative (négation). Il est remarquable que la lexis et l'affirmation aient la même forme (à l'ordre près, et quelques autres différences non pertinentes dans cette discussion), alors que la négation se signale par l'adjonction d'un marqueur. D'un autre côté, des considérations strictement linguistiques confirment nettement la thèse selon laquelle il n'existe pas de lexis négative (en d'autres termes, la négation porte sur la lexis, qui, elle, n'est ni affirmative, ni négative). On retrouve une structure en came :



Ici encore, de tels modèles permettent de résoudre des problèmes qui se posent à propos des langues ou du langage. Parmi ces problèmes, signalons l'ambiguïté de la lexis [mon père, mourir] (a) *la mort de mon père* (simple événement : « je considère que mon père meure », « si mon père meurt », « l'idée, le fait que mon père meure ») (b) désir (« je souhaite que mon père meure » ; « que mon père meure ! ») (c) rejet (« je ne veux pas que mon père meure », « je ne veux pas envisager l'idée que mon père meure ») (d) retour à la lexis, etc. On a, naturellement, reconnu ici la discussion par Freud du cas de *l'Homme aux Rats* : ce qui est important, c'est que, comme l'indique le diagramme ci-dessus, on a un chemin qui est en gros le suivant : [père, mourir ou pas mourir] → « l'idée que... » → « le souhait que » → « le souhait que ne... pas » ou « le rejet du souhait que » → finalement « le souhait que », par l'intermédiaire de [père, mourir ou pas mourir].

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