



## Le modèle en linguistique

Bernard Victorri

► **To cite this version:**

| Bernard Victorri. Le modèle en linguistique. Encyclopaedia Universalis, 1997. halshs-00009518

**HAL Id: halshs-00009518**

**<https://halshs.archives-ouvertes.fr/halshs-00009518>**

Submitted on 8 Mar 2006

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

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

## Le modèle en linguistique

Bernard Victorri

L'activité de modélisation en linguistique a connu un essor considérable depuis le début des années 1970. Comme dans toutes les disciplines qui traitent de données empiriques, les modèles en linguistique cherchent à rendre compte de phénomènes observables: il s'agit de mettre en place un dispositif dont la conception est régie par la théorie linguistique que l'on veut illustrer et dont le fonctionnement produit des résultats comparables aux données observées. La modélisation est donc une manière de mettre en œuvre une théorie, qui permet de mesurer son adéquation aux faits que prétend expliquer cette même théorie. De ce point de vue, plus le modèle est simple, plus son pouvoir explicatif est grand, puisque l'on peut clairement identifier le rôle de chaque élément de la théorie dans son fonctionnement. Mais cet idéal de simplicité est contrebalancé par la nécessité de rendre compte du maximum de données et avec la plus grande précision possible, ce qui réclame en général de complexifier le modèle, et partant de diminuer d'autant sa lisibilité. Il faut donc trouver le meilleur compromis possible entre pouvoir explicatif et fidélité aux données empiriques. Comme dans les autres disciplines, les modèles en linguistique dépendent étroitement des théories dont ils sont issus, et d'abord de la définition des objets d'étude de la discipline. On sait qu'en linguistique, comme dans d'autres sciences humaines, les théories divergent assez vite, y compris sur la nature des phénomènes à étudier: aussi est-il important, quand on cherche à comparer des modèles, de ne pas imputer à la modélisation des qualités ou des faiblesses qui sont le fait des théories sous-jacentes. Comme dans les autres disciplines enfin, ce sont essentiellement les mathématiques qui sont mises à contribution dans l'activité de construction de modèles, soit directement, soit plus fréquemment par l'intermédiaire de l'informatique: en fait, une modélisation informatique dérive toujours d'un modèle mathématique sous-jacent, que cette dérivation soit explicite ou non. Ce rôle important de l'informatique ne doit pas conduire à confondre l'activité de modélisation dont il est question ici et le traitement automatique des langues. Même si des liens étroits se sont noués entre ces deux types d'activité, les objectifs et les méthodes restent différents: pour schématiser, on peut dire que dans le traitement automatique des langues la linguistique doit être en quelque sorte au service de l'informatique, pour aider à accomplir les tâches précises que l'on cherche à automatiser, tandis que dans la modélisation c'est l'informatique qui doit absolument rester sous le contrôle de la théorie linguistique que l'on veut tester.

Au-delà de ces généralités, il faut noter aussi que les modèles linguistiques dans leur ensemble présentent quelques spécificités propres à la discipline. D'abord, le fait qu'il s'agit d'étudier des systèmes de signes permet de différencier deux types de modèles. Les premiers, que nous appellerons *modèles internes*, cherchent à rendre compte de la combinatoire des signes produits (ou productibles) en elle-même, telle qu'on peut l'observer. Les seconds cherchent au contraire à représenter ce qui fait de ces productions des signes, en traitant de leur relation avec leur forme ou avec leur sens: nous les appellerons *modèles d'interface*. Comme on le verra, le statut épistémologique des ces deux types de modèles est différent, ce qui doit nous conduire à les distinguer soigneusement, même si une seule et même théorie linguistique peut produire des modèles des deux types ou réduire l'un à l'autre en considérant le langage comme un système clos dans lequel s'intègrent formes et sens des signes linguistiques. Une autre spécificité importante des modèles linguistiques découle du fait que le langage est une activité cognitive humaine, et donc que l'on peut prendre position sur la valeur du modèle en tant que modélisant

aussi quelque chose de l'activité cognitive liée à la possession et à l'utilisation du langage: comme on le verra, il s'agit là d'un enjeu important pour les théories linguistiques elles-mêmes.

## 1. Les modèles internes

Les modèles de ce type cherchent donc à rendre compte d'observations sur les productions langagières en tant que telles, sans se donner l'objectif de définir ni d'expliquer les conditions (matérielles, biologiques, sociales, etc.) de leur apparition. Il ne s'agit pas d'étudier comment et pourquoi tel ou tel énoncé peut être produit, mais d'expliquer en quoi et pourquoi il constitue bien un énoncé d'une langue. Dans ce type de modèles, la notion d'acceptabilité joue donc un rôle essentiel. Mais cette notion elle-même ne relève pas du modèle: c'est la théorie linguistique qui définit la nature des critères utilisés pour décider de l'acceptabilité ou du rejet d'une production donnée. Une fois ce cadre établi, le rôle du modèle peut être plus ou moins important. Le modèle peut se contenter de représenter les données acceptables en mettant en évidence les structures que la théorie a épinglées: ainsi peut-on dire que les diagrammes syntaxiques qui émaillent les études d'un Tesnière et d'un Martinet sont déjà, en ce sens faible, des modélisations de la syntaxe. Au sens fort, on ne parlera de modèle que si celui-ci est capable de discriminer les données acceptables des données rejetées, ce qui, en principe, permet de tester de manière simple et rigoureuse sa validité: chaque production acceptable que le modèle rejette (et inversement) constitue un contre-exemple qui force à réviser le modèle. C'est ainsi que, toujours en syntaxe, le programme de recherche que s'est fixé Noam Chomsky dès ses premiers travaux consiste à construire «une grammaire générative», c'est-à-dire un modèle de la syntaxe d'une langue qui soit capable d'engendrer tous les énoncés bien formés de cette langue, et rien qu'eux. La popularité de ce programme, qui a complètement transformé le paysage de la recherche en linguistique, est en grande partie due au sentiment qu'il a pu donner, grâce à l'introduction de cette conception des modèles, de transformer la linguistique en une «véritable» science hypothético-déductive, dans laquelle la confrontation de la théorie avec les «faits de langue» pouvait se mener sur une base rigoureuse et parfaitement objective.

Plusieurs remarques s'imposent cependant à ce sujet. D'abord, les faits de langue en question ne sont pas, dans la théorie chomskyenne, des données empiriques. En effet, Chomsky distingue la *performance* des locuteurs d'une langue, dont relèvent les données empiriques, de leur *compétence*, qui définit la langue déjà comme un système intériorisé par les locuteurs, auquel ils s'efforcent de se conformer et qui est le véritable objet d'étude de la théorie chomskyenne. Pour ne prendre qu'un exemple de cette discordance théorisée entre la langue et les données empiriques, une propriété essentielle, «basique», des langues serait de pouvoir engendrer un nombre infini de phrases. Cela implique, le lexique étant fini, de pouvoir construire des phrases de longueur non bornée: ce qui, bien sûr, n'est pas le cas des énoncés que peut effectivement produire un locuteur. En fait, la propriété de récursivité qui est invoquée pour justifier cet infini ferait des productions d'une langue un ensemble «inobservable», au même titre que l'ensemble des nombres entiers: paradoxalement, la linguistique est donc dans cette optique une science *non expérimentale* par son objet même, pour les mêmes raisons que l'arithmétique.

Il faut aussi noter que l'objectif très ambitieux assigné au modèle, à savoir engendrer toutes les phrases d'une langue et seulement elles, puis, plus tard, utiliser un même modèle de base, la «grammaire universelle», paramétrée différemment pour engendrer toutes les langues, n'a jamais été considéré par Chomsky comme forcément réalisable. La position affichée dès le début du programme était de découvrir par ce biais les propriétés les plus intéressantes des langues, et de se

focaliser sur celles-ci, alors qu'une approche purement descriptive ne pouvait que se noyer dans la masse des faits, sans pouvoir hiérarchiser l'importance des problèmes à se poser. De fait, il est clair qu'après plus de trente ans de recherches l'objectif que s'était assigné le programme chomskyen est loin d'avoir été atteint, y compris pour la langue la plus étudiée dans ce contexte, l'anglais. Chomsky note lui-même, dans *The Minimalist Program* (1995), que bien des phénomènes syntaxiques, dont certains assez centraux dans les langues, n'ont pas encore trouvé de traitement satisfaisant, pas même dans les versions les plus récentes de sa théorie. Et s'il est indéniable que ce programme a permis de découvrir, ou de redécouvrir, un grand nombre de problèmes syntaxiques intéressants, on peut se demander si l'accent mis sur ces questions de manière exclusive n'a pas aussi contribué à masquer d'autres problèmes des langues dont l'étude aurait été plus féconde pour la modélisation linguistique en général.

Quoi qu'il en soit, on peut penser que depuis quelques années le programme chomskyen s'éloigne quelque peu de l'idéal d'une modélisation réussie, dans une contradiction permanente entre «un effort vers la simplicité des principes et la nécessité d'incessantes complications et coups de pouce requis par l'adéquation empirique», comme le dit Jean-Claude Milner, qui parle de «dégénérescence» du programme chomskyen en comparant cette situation à celle des dernières versions du modèle de Ptolémée, qui multiplie les épicycles pour sauver la simplicité du principe des déplacements circulaires tout en rendant compte de la complexité du mouvement des planètes.

Cela n'enlève rien au fait que le programme chomskyen a donné une formidable impulsion à la modélisation en linguistique, et particulièrement en syntaxe. Aujourd'hui, un grand nombre de modèles syntaxiques coexistent et se développent, en particulier les grammaires d'unification, sur lesquelles nous reviendrons plus loin. Généralement, ces modèles n'ont plus l'ambition de remplir un objectif aussi radical que celui du programme chomskyen. Il s'agit de couvrir le plus grand nombre possible de phénomènes syntaxiques de telle ou telle langue à l'aide d'un formalisme donné, dans une optique plus classique de modélisation telle qu'on la pratique habituellement dans les sciences expérimentales: on cherche le meilleur compromis entre la simplicité du système et l'adéquation aux faits observables, étant entendu que le modèle ne vise qu'à une approximation de ces données empiriques.

Le fait que nous n'ayons parlé jusqu'à présent que de syntaxe ne doit pas laisser croire que ce serait le seul domaine concerné par ce type de modèles. S'il est vrai que les débats théoriques et épistémologiques se sont centrés sur la syntaxe jusqu'à ces dernières années, la modélisation a touché de fait tous les niveaux traditionnels de l'étude des langues: phonologie, morphologie, syntaxe, sémantique. En particulier, en phonologie (qui a souvent été un domaine précurseur en matière de modélisation), on retrouve toute l'étendue des positions épistémologiques sur la place des modèles que nous avons décrite à propos de la syntaxe, avec une «phonologie générative», qui a joué un rôle analogue de radicalisation théorique, entraînant les mêmes bons et mauvais effets, et avec une profusion actuelle de modèles plus «locaux», s'attaquant à un phénomène précis, comme la syllabation, la position de l'accent ou encore l'intonation.

En sémantique, la situation est plus confuse, parce que l'existence même et l'autonomie d'un niveau de «sémantique linguistique» est loin d'être une évidence: pour certains, qui défendent le primat de la syntaxe, la forme sémantique de l'énoncé est directement issue du modèle syntaxique; pour d'autres, tenants de la pragmatique, la sémantique dépend presque entièrement des conditions de l'énonciation. Dans la mouvance chomskyenne, la lutte a été rude entre les partisans de l'autonomie de modèles sémantiques et les défenseurs plus orthodoxes du respect du primat de la syntaxe. Dans d'autres théories, comme certaines grammaires cognitives américaines (cf. Langacker ou Talmy) et la théorie de l'énonciation

d'Antoine Culioli, les relations entre syntaxe et sémantique sont tout autres. Ainsi, pour ce dernier, à chaque marqueur linguistique est associée une opération (décrite par sa «forme schématique»), et ce sont les interactions dynamiques entre ces opérations qui produisent à la fois la structure syntaxique et les effets de sens d'un énoncé: il est clair qu'une modélisation de cette théorie ne peut séparer de manière séquentielle syntaxe et sémantique. Il est donc plus prudent de renoncer pour ce niveau de l'analyse linguistique à un regroupement de modèles issus de théories divergentes. De toute manière, comme on le verra, les difficultés de définition et de représentation du sens (des unités linguistiques aussi bien que des énoncés) rendent plus arbitraire à ce niveau la distinction que l'on a introduite entre modèles internes et modèles d'interface.

## **2. Les modèles d'interface**

L'ensemble de ces modèles peut à son tour être divisé en deux: d'une part, les modèles qui traitent de l'interface matérielle, c'est-à-dire des relations des signes linguistiques avec leur forme (sonore en général); d'autre part, ceux qui traitent de l'interface conceptuelle, c'est-à-dire des relations avec leur sens.

Du côté de la forme, il s'agit essentiellement de modèles dans le domaine de la phonétique, ou plus précisément à «l'interface» entre la phonétique et la phonologie. Les modélisations les plus intéressantes sont en effet celles qui cherchent à rendre compte des rapports entre la matière sonore et les systèmes phonologiques des langues. On distingue dans ce champ des modèles de production, fondés sur une modélisation de l'appareil phonatoire (le larynx, avec ses cordes vocales, le pharynx et les cavités buccale et nasale), des modèles de reconnaissance, fondés sur une analyse fréquentielle des sons (essentiellement les zones d'harmoniques constituant les premiers «formants» qui permettent de différencier les timbres vocaliques, mais aussi les variations de la fréquence fondamentale), et enfin des modèles «intégrés», qui modélisent la correspondance entre commandes motrices et trajectoires dans l'espace perceptif. Pour ne prendre qu'un exemple typique des questions traitées à ce niveau, on peut chercher à rendre compte avec ces modèles des différents systèmes de voyelles que l'on rencontre dans les langues. En effet, la diversité observée (certaines langues ne possèdent que quatre ou cinq voyelles alors que d'autres en ont beaucoup plus) recèle des régularités remarquables (la position des voyelles dans l'espace «formantique» constitué par les deux premiers formants vocaliques dépend de manière stricte du nombre de voyelles de la langue considérée) dont un modèle peut rendre compte en supposant un principe d'optimisation de la discrimination perceptive par rapport à l'effort phonatoire nécessaire pour les produire (cf. par exemple les travaux de René Carré). Ce qui caractérise ces modèles, c'est qu'il s'agit en fait d'une modélisation qui relève de la physique et dont la problématique est linguistique: il n'est pas si fréquent de rencontrer ce genre de «mixité» entre sciences expérimentales «dures» et sciences humaines... Signalons qu'appartiennent aussi à cette catégorie les modèles de la relation entre les signes des langues de sourds et les mouvements qui les réalisent, ou encore les modèles de la relation entre les signes et leur forme écrite dans les divers systèmes d'écriture (plus particulièrement pour les écritures non alphabétiques).

Du côté du sens, les modèles doivent présenter aussi une sorte de mixité, puisqu'il leur faut à la fois représenter le sens des énoncés, de manière forcément extra-linguistique, et rendre compte du passage de la structure linguistique de ces énoncés à la représentation de leur sens. On observe sur ces deux points une grande diversité, dont on se contentera ici de donner un aperçu à partir de quelques

exemples. En ce qui concerne la représentation du sens, on peut relever deux grandes familles: les approches plutôt psychologisantes et les approches plutôt logicistes. Pour les premières, le sens est affaire de cognition, et sa représentation doit donc être de type conceptuel: c'est, par exemple, la position clairement affirmée de Jackendoff, dans le modèle duquel le sens d'un énoncé est représenté par une combinaison de concepts de diverses catégories (objets, substances, événements, processus, etc.). Pour les secondes, le sens est affaire de relation au monde, et sa représentation doit donc donner les conditions qui permettent de décider si cet énoncé est vrai ou faux dans un monde donné. C'est la position qu'a défendue par exemple R. Montague, en développant un modèle dans lequel le sens d'un énoncé est une proposition d'une logique intensionnelle (où, très grossièrement, un prédicat est défini comme une fonction lui associant une valeur de vérité dans chaque monde possible). Il est intéressant de remarquer que l'on rejoint par ce biais la théorie des modèles de la logique mathématique (les «mondes» en question étant essentiellement des constructions ensemblistes, dont le rattachement est problématique, que ce soit au monde vécu ou au monde des physiciens).

La manière dont ces modèles construisent la représentation du sens, quelle que soit sa nature, est aussi très diversifiée. Chez Montague, cette construction est le résultat d'un principe de compositionnalité très strict, qui permet de calculer la forme logique associée à un énoncé à partir de l'arbre issu de l'analyse syntaxique de cet énoncé: à chaque nœud de l'arbre correspond une expression logique calculée récursivement à partir des nœuds-fils et de la relation syntaxique qui les lie. Dans la D.R.T. (*Discourse Representation Theory*) de H. Kamp, les règles de construction du sens d'un énoncé font explicitement appel au contexte composé par les énoncés qui le précèdent dans le discours où il est inséré. D'autres modèles, comme la «sémantique situationnelle» de Barwise et Perry, rapprochent plus la sémantique de la pragmatique en prenant en compte d'emblée dans le calcul du sens le contexte en un sens plus large, incluant les conditions d'énonciation (qui parle? quand? où? à qui? pourquoi?). Enfin, comme on l'a vu, il existe aussi des théories pour lesquelles l'analyse syntaxique ne peut être menée indépendamment de la construction du sens, ce qui remet en cause le calcul séquentiel par niveaux (syntaxique puis sémantique) sur lequel sont fondés les modèles classiques.

### **3. Les outils mathématiques**

Les outils mathématiques les plus utilisés dans la modélisation linguistique sont, sans conteste, de type logico-algébrique. La première impulsion de modélisation des langues en termes algébriques vient sans doute de Zellig Harris (dont Chomsky fut un élève), qui a développé un modèle transformationnel dans lequel les phrases sont obtenues à partir de phrases noyaux subissant une suite de manipulations algébriques. Mais très vite c'est Chomsky lui-même qui assoit définitivement la domination des modèles algébriques en syntaxe avec le programme générativiste. La plupart des modèles syntaxiques qui suivront emprunteront les caractéristiques de base du formalisme chomskyen, en particulier la définition d'une grammaire comme un ensemble de règles de production s'appliquant à un vocabulaire constitué du lexique (vocabulaire terminal) et de symboles désignant les différents types de syntagmes. La suite des règles de réécriture utilisées pour engendrer une phrase peut être symbolisée par un arbre qui fournit donc une représentation équivalente de la structure syntaxique. Des formalismes du même type ont fleuri aussi pour les autres niveaux de l'analyse linguistique: phonologie (dès 1955, dans la thèse de Chomsky), morphologie et sémantique.

Progressivement, ces formalismes ont été remplacés ou transformés par des approches tout aussi algébriques, mais qui se révèlent plus commodes pour la description des unités en termes de traits distinctifs, et plus efficaces pour le traitement automatique: il s'agit des grammaires d'unification, qui sont assez diversifiées mais qui ont en commun d'utiliser un format unique (listes de couples attribut-valeur) pour représenter aussi bien les règles de composition que les structures qu'elles manipulent; le mécanisme qui permet cette homogénéité provient de la technique informatique de l'unification, qui opère en un sens bien précis le regroupement de deux listes de couples attribut-valeur en une seule.

Cette domination des modèles algébriques n'est pas étonnante. Elle est en fait une conséquence directe, au moins pour les modèles internes, du structuralisme tel qu'il s'est imposé, après avoir rompu tout lien avec la *Gestalttheorie*: définir les unités linguistiques des différents niveaux comme des unités discrètes s'opposant par des traits distinctifs et se combinant pour former des unités de niveau supérieur conduit naturellement à des modèles algébriques. Et cette tendance est confortée par les modèles d'interface avec le sens. Cela va de soi pour les approches logicistes, puisque l'obtention d'une forme logique est facilitée par les représentations algébriques. Mais c'est vrai aussi, en grande partie, pour les approches psychologisantes, puisque le paradigme «computationnel» dominant des sciences cognitives est aussi de type logico-algébrique, qu'il s'agisse du «langage de la pensée» de Jerry Fodor, ou des réseaux sémantiques et autres graphes conceptuels (Sowa).

Il ne faudrait pas en déduire que seules l'algèbre et la logique apportent leur contribution à la modélisation en linguistique. Car les modèles d'interface avec le son font forcément appel aux mathématiques du continu (équations aux dérivées partielles, transformée de Fourier, etc.), tout comme, quoique de manière moins évidente, les autres niveaux de modélisation. En fait, il apparaît assez vite que les langues ne répondent qu'imparfaitement aux critères algébriques de «bonne formation» qui régissent les formalismes classiques. Tant que l'on pense que ces modèles fournissent une première approximation utile des faits, on peut s'en satisfaire. Mais comme le disait dès 1967 Charles Hockett, qui fut l'un des pionniers de ces formalismes algébriques, cela n'est plus possible quand on pense que cette approximation est obtenue en écartant les propriétés des langues qui sont de fait les plus importantes.

Prenons un exemple. La polysémie, cette aptitude des mots les plus courants à avoir une multitude de sens apparentés, distingue radicalement les langues des langages formels utilisés dans ces formalismes algébriques pour représenter le sens. Si l'on pense que la polysémie est une propriété fondamentale de la sémantique des langues, qui explique leur souplesse, leur productivité et leur évolution, on devra renoncer aux modèles algébriques classiques qui conduisent à escamoter ou à marginaliser ce phénomène. Ce sont des raisons de cet ordre qui expliquent que l'on ait pu faire appel aux mathématiques du continu, et plus particulièrement à des modèles topologico-dynamiques en linguistique.

Les travaux de mathématiciens comme René Thom et Jean Petitot, de même que ceux de linguistes comme Antoine Culioli avaient depuis longtemps ouvert cette voie, qui s'est popularisée par la suite, surtout avec l'apparition d'outils informatiques adaptés à la représentation de systèmes dynamiques: les réseaux connexionnistes. Ces nouveaux outils de modélisation doivent aussi en partie leur succès à la facilité avec laquelle ils permettent l'implémentation (ou réalisation) de mécanismes d'apprentissage, qui, comme nous le verrons plus loin, jouent un rôle important dans les relations entre la linguistique et les sciences de la cognition. Signalons aussi

l'apparition, encore plus récente, de modèles que l'on pourrait qualifier d'intermédiaires, comme la «théorie de l'optimalité» (A. Prince et P. Smolensky), qui cherche à conserver l'essentiel des qualités des modèles algébriques tout en faisant une place à quelques caractéristiques intéressantes extraites des modèles connexionnistes: très grossièrement, une grammaire est constituée d'un ensemble ordonné, pas forcément consistant, de règles «violables», les énoncés engendrés devant être «optimaux» au sens où ce sont ceux qui violent le minimum de règles de haute priorité.

#### **4. Modèles linguistiques et cognition**

Pour conclure, il nous faut dire quelques mots du rôle de la modélisation en linguistique lors des débats sur la place du langage dans les activités cognitives en général. Les discussions portent principalement sur deux points: d'une part, sur les relations entre l'activité linguistique proprement dite et le reste des activités cognitives liées à la communication langagière, et, d'autre part, sur la problématique de l'apprentissage de la langue maternelle. Dans les deux cas, le type de modèle linguistique utilisé joue un rôle essentiel dans les hypothèses cognitives qui s'affrontent.

Dans le paradigme encore le plus répandu aujourd'hui en sciences cognitives, que l'on appelle généralement le *paradigme cognitiviste*, défendu entre autres par Jerry Fodor, l'activité linguistique du sujet (comme la perception visuelle et autres activités «périphériques») est une activité automatique, inconsciente et modulaire qui consiste à «traduire» les énoncés de la langue dans le langage computationnel de l'esprit (le «mentaux»), et réciproquement. Cette formulation, que l'on trouve par exemple telle quelle sous la plume du psycholinguiste chomskyen Steven Pinker, est fortement inspirée par les modèles générativistes, qui présentent une architecture compatible avec ce point de vue: ils sont en effet constitués d'un système syntaxique central muni de deux interfaces, qui conduisent l'une à la forme phonétique et l'autre à la forme logique.

S'articule sur ce premier point une théorie de l'apprentissage. La conception chomskyenne d'un système génératif capable d'engendrer une infinité de phrases conduit à un problème sérieux pour l'acquisition par l'enfant de sa langue maternelle, que l'on appelle le problème de la «pauvreté de l'input»: comment l'enfant pourrait-il induire si facilement les règles de bonne formation des phrases de sa langue, alors qu'il n'est soumis qu'à un nombre fini et limité d'exemples de phrases correctes et que ses propres incorrections sont peu (ou même pas du tout) relevées par son entourage? Mathématiquement, on peut montrer qu'un apprentissage de ce type de grammaire ne peut réussir dans ces conditions que si la famille des grammaires possibles est fortement contrainte. Ce qui conduit directement à la thèse «innéiste» chomskyenne: chaque humain naît avec un organe du langage muni d'une grammaire universelle, et l'apprentissage de sa langue maternelle consisterait simplement à fixer un petit nombre de paramètres, à valeur binaire variable suivant les langues. Ce qui, à son tour, sert d'argument à une théorie modulaire de la cognition, dans laquelle cet organe du langage trouve naturellement sa place dans les mécanismes automatiques périphériques. Ainsi des modèles purement linguistiques, «internes», au sens que nous avons donné ici à ce mot, ont-ils des répercussions importantes, bien au-delà des sciences du langage, parfois même contre la volonté de leurs auteurs. Chomsky, par exemple, ne partage pas la conception classique de la modularité telle qu'on la trouve chez Fodor.

À cette conception des relations entre langage et cognition s'opposent d'autres, qui cherchent au contraire à découvrir, à travers l'étude de la diversité des langues, les mécanismes généraux à l'œuvre dans l'activité de langage, lesquels doivent à leur tour aider à mieux



appréhender la cognition en général: c'est le programme que s'est fixé la théorie de l'énonciation d'Antoine Culioli, qui utilise largement dans ses formulations des concepts issus de la topologie (ouverts, frontières, etc.) et de la théorie des systèmes dynamiques (attracteurs, bifurcations, etc.). On peut aussi citer, parmi bien d'autres, les grammaires cognitives nord-américaines (Langacker, Talmy, Fauconnier, Lakoff...), dont les analyses linguistiques se fondent sur l'abstraction de schémas perceptifs dynamiques pour rendre compte des phénomènes syntaxico-sémantiques. Dans ces courants aussi, le lien entre langage et cognition passe donc en grande partie par le choix des outils de modélisation des phénomènes linguistiques eux-mêmes. Notons par ailleurs que le problème de l'apprentissage ne se pose plus du tout dans les mêmes termes dans ces approches, puisqu'il n'est plus question d'isoler le processus de bonne formation syntaxique du sens qui est construit par les constituants de l'énoncé: tout l'appareil cognitif de l'enfant peut donc contribuer à l'acquisition de la structure de sa langue.

Ainsi la modélisation en linguistique se trouve-t-elle au centre de débats qui concernent l'ensemble des sciences cognitives, et c'est en grande partie à cette situation que l'on doit la vigueur de cette activité de recherche aujourd'hui.