

# L'analyse logique des langues naturelles

Paul ÉGRÉ

Conçue pour mener des raisonnements sous forme de calculs et remédier aux imperfections du langage ordinaire, la logique formelle constitue aujourd'hui l'une des clefs de l'analyse des langues naturelles.

« Il n'y a selon moi aucune différence théorique importante entre les langues naturelles et les langages artificiels des logiciens. »

**V**olontiers provocatrice, cette déclaration qui ouvre l'article *Universal Grammar* (1970) de Richard Montague reflète l'évolution du rapport entre logique et linguistique qui s'est opérée à partir des années 1950.

La logique et la linguistique sont étroitement liées. L'une et l'autre naissent simultanément chez Aristote : le philosophe grec propose la première théorie de la structure logique des énoncés et s'en sert pour construire une théorie des raisonnements valides, la théorie du syllogisme (voir *La logique : une création de la Grèce antique*, par Jean-Baptiste Gourinat, dans ce dossier).

Au XVII<sup>e</sup> siècle, Antoine Arnauld, Claude Lancelot et Pierre Nicole, les Messieurs de Port-Royal, conçoivent la logique comme un « art de penser » et la grammaire comme un « art de parler ». Le lien entre langage et logique se précise, mais c'est à partir du XX<sup>e</sup> siècle, quand la logique mathématique prend son essor, que celle-ci contribue de façon substantielle à l'étude des langues naturelles. Il aura fallu pour cela que la logique et la linguistique deviennent des disciplines formelles.

La formalisation de la logique s'effectue sous l'impulsion de Gottlob Frege et Bertrand Russell. Du côté de la linguistique, c'est d'abord la syntaxe, étude de la structure de la phrase et de ses composants, qui est formalisée à partir des années 1950, quand Noam Chomsky introduit la grammaire générative. Puis, au début des années 1970, c'est au tour de la sémantique, étude de la signification, de devenir formelle grâce à l'œuvre de Richard Montague.

En dépit de certaines divergences théoriques et d'une opposition longtemps marquée entre syntaxe et sémantique, nous verrons qu'une même inspiration préside à ces deux avancées. Mais avant, interrogeons-nous sur la structure du langage naturel.

## La structure des énoncés

De quoi est composée une phrase ? Dans le *Sophiste*, Platon constate qu'il faut au moins associer un sujet à un verbe, comme dans la phrase « Théétète est assis ». Le sujet « Théétète », pris isolément, ne forme pas une phrase, pas plus que le verbe « est assis » : leur « entrelacement » est requis. Cette découverte constitue la première ébauche d'une analyse des phrases en constituants, et sert de soubassement à la théorie aristotélicienne du syllogisme.

Aristote prend pour objet d'analyse l'énoncé déclaratif, défini comme une phrase susceptible d'être vraie ou fausse, par opposition à une phrase interrogative ou impérative. Un tel énoncé peut être affirmatif ou négatif : c'est sa qualité ; et il peut être particulier ou universel : c'est sa quantité. Par exemple, « Tout homme est assis » est universel affirmatif, alors que « Quelque homme n'est pas assis » (« quelque » dans le sens de « au moins un ») est particulier négatif. « Aucun homme n'est assis » est universel négatif, et « quelque homme est assis » est particulier affirmatif.

Dans chaque énoncé de ce type, on distingue par ailleurs un terme sujet (ici « homme ») et un terme attribut du sujet, le prédicat (ici « est assis »). Les énoncés déclaratifs qu'Aristote décrit ont chacun la forme logique Q A B, où Q est une expression de quantité, ou quantificateur (*tout, aucun, quelque...*), A le sujet, et B le prédicat.

Aristote édifie la théorie du syllogisme sur son analyse logique des énoncés déclaratifs : un syllogisme est un argument constitué de deux prémisses et d'une conclusion, chacune étant un énoncé de la forme Q A B. Un syllogisme est valide si la conclusion suit nécessairement des prémisses.

Par exemple, le syllogisme « Tout homme est un animal, tout animal

est mortel, donc tout homme est mortel» est valide. Au contraire, « Tout philosophe est savant, quelque savant est heureux, donc quelque philosophe est heureux » ne l'est pas, car il se pourrait qu'aucun des savants qui sont heureux ne soit un philosophe.

Aristote requiert que dans un syllogisme ne figurent que trois termes, l'un d'eux, le moyen, servant à relier les deux autres et disparaissant de la conclusion. L'ensemble de ces contraintes formelles donne lieu à une combinatoire finie, et il est usuel de classer les syllogismes en quatre figures, selon la position du moyen terme dans les prémisses (voir la figure 2).

Il existe au total  $4^4 = 256$  arguments syllogistiques possibles. Seuls 24 de ces schémas sont valides (si l'usage de « tout A » et « aucun A » présuppose qu'il existe des A), et Aristote a tenté d'en donner une caractérisation axiomatique.

## Les quantificateurs

Aristote est le premier à avoir décrit l'interaction entre la négation et les quantificateurs dans les langues naturelles. Les quatre quantificateurs qu'il examine obéissent à des rapports d'opposition systématique (voir la figure 3).

L'universel affirmatif (« Tout homme est assis ») et le particulier négatif (« Quelque homme n'est pas assis ») sont contradictoires : ils ne peuvent être ni vrais ni faux ensemble. Il en va de même de l'universel négatif (« Aucun homme n'est assis ») et du particulier affirmatif (« Quelque homme est assis »). En revanche, « Tout homme est assis » et « Aucun homme n'est assis » peuvent être tous deux faux, bien qu'ils ne puissent être vrais ensemble : ils sont contraires. Enfin il

suit logiquement de « Tout homme est assis » que « Quelque homme est assis » ; la relation est dite de subalternation.

La distinction entre contradiction et contrariété reflète une ambiguïté courante de l'usage de la négation dans les langues naturelles : « Tout homme n'est pas assis » peut signifier « Tout homme est non assis » (soit « Aucun homme n'est assis »), ou encore « Il n'est pas vrai que tout homme est assis » (soit « Quelque homme n'est pas assis »). Il faut donc distinguer une négation externe, dite contradictoire, d'une négation interne, dite contraire.

La théorie du syllogisme d'Aristote représente une étape importante dans le développement de la logique, mais elle fait aussi apparaître la dépendance de la logique vis-à-vis de l'analyse des langues naturelles. La logique peut en effet être définie comme la théorie des inférences valides. Or nombre d'inférences valides en langage ordinaire ne sont pas formalisables dans le cadre syllogistique d'Aristote.

Par exemple, l'inférence « Aucun étudiant ne connaît tous les professeurs ; quelque étudiant connaît tous les assistants ; donc quelque professeur n'est pas un assistant » est valide, mais sort du cadre syllogistique. Pour la formaliser, il serait nécessaire d'analyser le prédicat « connaît tous les professeurs », où apparaît un quantificateur. Cela implique de donner à l'énoncé « Aucun étudiant ne connaît tous les professeurs » une structure plus fine que la structure canonique Q A B.

En particulier, Aristote ne fait pas de distinction logique entre les verbes transitifs (comme « connaître »), qui expriment des relations binaires, et les verbes intransitifs, adjectifs ou noms communs (tels « dormir », « assis » ou « homme »),

qui expriment chacun une propriété. Cela explique qu'il ne rende pas compte des énoncés où un quantificateur apparaît en position d'objet du verbe (comme dans « connaît tous les professeurs »).

Il ne fait pas place non plus aux termes d'individus (tel « Socrate »), ni aux quantificateurs non restreints (tel « quelqu'un »), comme dans les inférences valides : « Pierre admire tous les philosophes ; Socrate est un philosophe ; donc Pierre admire Socrate (quelqu'un) ».

Enfin, Aristote ne prend pas en considération les expressions de coordination comme « et », « ou ». Il ignore aussi les conditionnels du type « si... alors... », sur lesquels reposent nombre d'inférences logiques. Par exemple, l'inférence « Si tous les hommes sont honnêtes, alors tous les banquiers sont honnêtes ; quelque banquier n'est pas honnête ; donc quelque homme n'est pas sage » est valide, mais faute d'un traitement des expressions conditionnelles, elle ne peut être traitée dans la syllogistique.

Ces différentes limitations du cadre syllogistique témoignent d'un manque de pouvoir expressif de la logique d'Aristote. Elles ont été occasionnellement soulignées, en particulier par Leibniz, mais elles ne seront réellement mises en cause qu'en 1879, date de la parution de l'*Idéographie* de Frege.

## L'idéographie de Frege

Alors jeune *Privatdocent* en mathématiques (l'équivalent de maître de conférences) à l'Université d'Iéna, Frege recherche pour les mathématiques un « langage formulaire de la pensée pure », une idéographie permettant de mener à bien les démonstrations mathématiques, de façon transparente et rigoureuse. Frege poursuit l'idéal leibnizien : comme Leibniz, il se propose d'élaborer une langue caractéristique (*lingua characteristica*), destinée à contrôler les démonstrations et à en faire apparaître les lacunes.

L'ambition première de Frege est donc mathématique, mais sa contribution est aussi décisive pour l'analyse logique du langage naturel. En regard de la logique aristotélicienne, son apport est triple : on doit à Frege une analyse entièrement nouvelle de la quantification ; l'abolition du paradigme sujet-prédicat et l'invention d'une logique des relations ; enfin, un traitement

	Figure 1	Figure 2	Figure 3	Figure 4
Prémisse majeure	Aucun animal n'est immortel Q B C	QCB	QBC	QCB
Prémisse mineure	Tout homme est un animal Q A B	QAB	QBA	QBA
Conclusion	Aucun homme n'est immortel Q A C	QAC	QAC	QAC

2. LES QUATRE FIGURES DU SYLLOGISME sont définies selon la place du moyen terme dans les prémisses (en gras). Dans chaque figure, Q désigne l'un des quatre quantificateurs tout, quelque, aucun, pas tout, et peut varier d'un énoncé à l'autre.

logique des énoncés composés, qui repose sur une représentation explicite des connecteurs logiques.

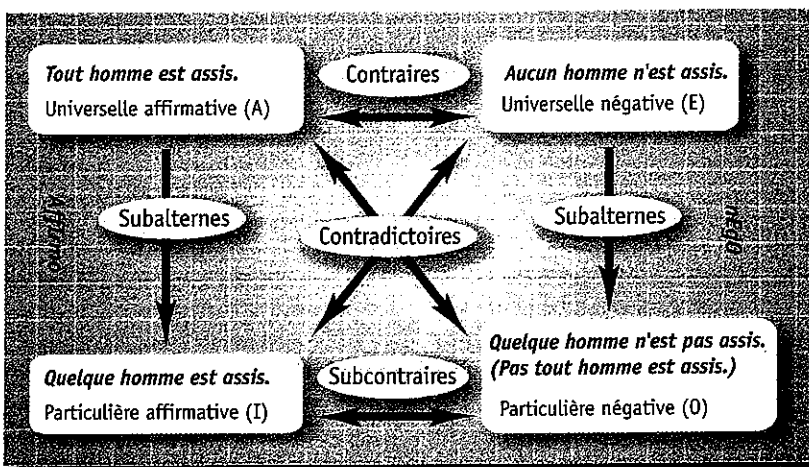
Pour analyser la quantification, Frege emprunte aux mathématiques la distinction entre fonction et argument, qu'il substitue à la distinction aristotélicienne entre sujet et prédicat. Frege admet qu'un énoncé tel que « César conquiert les Gaules » peut se décomposer en un sujet, « César », et un prédicat, le groupe verbal « conquiert les Gaules ». Il propose toutefois de considérer le prédicat « conquiert les Gaules » comme une fonction qui, lorsqu'elle prend pour argument un individu spécifique, donne une certaine valeur de vérité. Par exemple, si l'argument de « conquiert les Gaules » est « César », la fonction donne pour valeur le vrai. En revanche, si l'argument est Marc-Aurèle, la fonction donne pour valeur le faux.

Suivant l'*Idéographie*, on peut représenter le prédicat « conquiert les Gaules » par le symbole de fonction  $P(\cdot)$ , et l'argument par la variable indéterminée  $x$ :  $P(x)$  se lit «  $x$  conquiert les Gaules ». Si  $c$  symbolise « César », alors  $P(c)$  est une expression qui symbolise l'énoncé « César conquiert les Gaules », et qui dénote le vrai. Autre exemple: l'énoncé quantifié universellement « toute chose est belle » signifie pour Frege que la fonction  $\text{Beau}(x)$  prend toujours pour valeur le vrai, quelle que soit la valeur de  $x$ . Dans la notation moderne équivalente à celle de Frege, ceci se note  $\forall x \text{ Beau}(x)$ , ce qui se lit « pour tout  $x$ ,  $x$  est beau ».

La distinction entre fonction et argument fonde la logique des relations. Le verbe transitif « connaître », dans « Pierre connaît Marie », est considéré comme une fonction à deux arguments  $C(x, y)$ , et l'énoncé est symbolisé par  $C(p, m)$ . Il n'est plus pertinent de parler ici de sujet et de prédicat. On représente ainsi des relations ternaires ou  $n$ -aires, comme en géométrie: « Le point  $a$  est entre  $b$  et  $c$  » s'écrira  $E(a, b, c)$ .

Troisième innovation, Frege introduit dans son formalisme un symbole de négation et un symbole pour le conditionnel. Dans la notation moderne, « si  $A$ , alors  $B$  » se note  $(A \Rightarrow B)$  et « non  $A$  » se note  $\neg A$ .

Le langage formel de l'*Idéographie* constitue ce qu'on appelle aujourd'hui



3. LE CARRÉ DES OPPOSITIONS D'ARISTOTE constitue la première analyse de la quantification dans les langues naturelles. Les énoncés contraires (A) et (E) peuvent être faux ensemble, mais non vrais ensemble. À l'inverse, les subcontraires (I) et (O) peuvent être vrais ensemble, mais non faux ensemble. Les lettres conventionnelles (A), (E), (I) et (O) sont un héritage de la tradition scolastique (pour « Affirmo » et « nEgO »).

d'hui la logique des prédicats du premier ordre. (En réalité, le langage défini par Frege est la logique du second ordre, qui utilise des variables d'individus, mais aussi des variables de propriétés.) Ce langage est plus expressif que la syllogistique d'Aristote, puisqu'il permet de traduire formellement l'ensemble des énoncés qui posaient problème.

En effet, à l'aide du quantificateur universel  $\forall$  et de la négation  $\neg$ , on exprime tous les quantificateurs aristotéliciens, ainsi que la quantification non restreinte (quelqu'un). Ainsi, « Personne n'est venu » signifie « Chaque  $x$  est tel que  $x$  n'est pas venu », soit  $\forall x \neg \text{Venu}(x)$ , et « Quelqu'un est venu », qui en est le contradictoire, se traduira par  $\neg \forall x \neg \text{Venu}(x)$ . « Quelqu'un n'est pas venu » signifie qu'il est faux que chacun est venu, soit  $\neg \forall x \text{Venu}(x)$ .

Nous retrouvons ici, mais sous une forme explicite, le contraste aristotélicien entre négation interne et externe, selon la place de la négation par rapport au quantificateur.

Enfin, un énoncé proprement aristotélicien, comme « Tout homme est assis », qui fait intervenir deux prédicats, est considéré comme un énoncé universel restreint par un conditionnel: « Pour tout  $x$ , si  $x$  est un homme, alors  $x$  est assis », soit  $\forall x (\text{H}(x) \Rightarrow \text{A}(x))$ .

Il est du même coup possible de traiter les énoncés faisant intervenir plusieurs quantificateurs de part et d'autre d'un verbe transitif. L'énoncé « Aucun étudiant ne connaît tous les professeurs » se traduira par la formule:

$\forall x (\text{E}(x) \Rightarrow \neg \forall y (\text{P}(y) \Rightarrow \text{C}(x, y)))$ , c'est-à-dire « Pour tout étudiant, il est faux que, pour tout professeur, cet étudiant connaît ce professeur ».

En inventant le calcul des prédicats, Frege montre qu'il est possible de traiter logiquement davantage d'inférences du langage naturel que n'en traite Aristote. Les formes logiques auxquelles on parvient s'éloignent cependant de la forme grammaticale apparente des énoncés du langage naturel. Ainsi, dans l'exemple précédent, la forme logique comporte deux symboles de conditionnels, qui n'apparaissent pas explicitement dans l'énoncé « Aucun étudiant ne connaît tous les professeurs ».

### Langage formel = microscope

Pour Frege, le langage formel est au langage naturel ce que le microscope est à l'œil: le langage naturel est certes plus pratique à utiliser que le langage formulaire, mais ce dernier est, comme le microscope, plus adapté à l'observation scientifique.

En abolissant la distinction aristotélicienne entre sujet et prédicat, Frege considère ainsi que « la logique s'est jusqu'ici toujours rattachée trop étroitement à la langue et à la grammaire ». Ce même écart entre logique et grammaire se manifeste, quelques années plus tard, dans l'analyse logique que Bertrand Russell propose des descriptions définies, c'est-à-dire des groupes nominaux précédés de l'article défini. L'expression « Le roi de France est chauve » est analysée

par Russell en 1905 comme la conjonction de trois conditions : (1) Il existe au moins un roi de France ; (2) Il existe au plus un roi de France ; (3) Tout roi de France est chauve.

Ces trois conditions ont une traduction en logique du premier ordre, mais ni (1) ni (2) ni (3), ni même la conjonction de (1) et de (2) n'expriment à elles seules l'expression « le roi de France ». Russell voyait du même coup dans le groupe nominal « le roi de France » une expression dépourvue de signification propre indépendamment du groupe verbal qui la suit (voir *La logique de ce qui n'existe pas*, par Pascal Engel, dans ce dossier).

Tout en ouvrant la voie au traitement logique du langage ordinaire, Frege et Russell considèrent la logique comme un langage destiné à formaliser les mathématiques, et ne se soucient que secondairement du langage naturel. Cette attitude, qui demeure en grande partie celle des logiciens de la génération suivante, évolue considérablement à partir des années 1950, au moment où la linguistique devient à son tour une discipline formelle.

## La grammaire générative

Les années 1950, puis les années 1970 marquent deux tournants majeurs de la linguistique. Dans les années 1950, au *Massachusetts Institute of Technology*, Noam Chomsky (né en 1928) lance un programme de description de la syntaxe des langues naturelles, largement inspiré de la théorie logique des

langages formels. Au début des années 1970, à l'Université de Californie de Los Angeles, Richard Montague (1930-1971) produit une série d'articles pionniers pour la sémantique formelle des langues naturelles.

Chomsky cherche à caractériser de façon systématique les énoncés grammaticaux d'une langue donnée. Un langage est un ensemble d'énoncés, et chaque énoncé est une suite de mots. Parmi toutes les suites possibles, seules certaines sont grammaticalement correctes. Par exemple, en français, la suite de mots « un homme embrasse une femme » constitue un énoncé grammatical, alors que la suite « homme une embrasse femme un » est immédiatement perçue comme agrammaticale. La différence réside dans le fait que le second énoncé, contrairement au premier, ne suit pas les règles syntaxiques du français.

Bien informé des travaux de ses contemporains logiciens sur les langages formels et la théorie de la calculabilité, Chomsky publie plusieurs résultats fondamentaux sur le lien entre grammaire formelle et grammaire des langues naturelles. Une grammaire formelle est un système déductif d'axiomes et de règles d'inférences pour un lexique donné ; elle engendre les énoncés du langage à la façon dont un système axiomatique engendre des théorèmes. Chomsky pose le problème suivant : quel type de grammaire formelle est le plus adéquat pour engendrer toutes les phrases d'une langue donnée, et rien que ces phrases ?

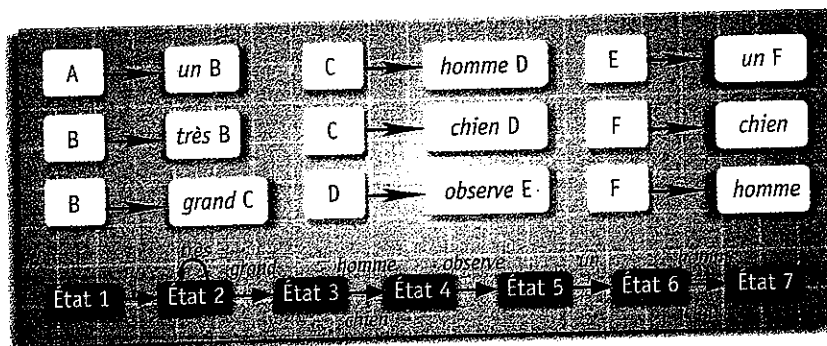
L'une des premières contributions de Chomsky fut de montrer qu'une grammaire comme celle de l'anglais ou du français n'est pas réductible à une grammaire à états finis (dite encore grammaire régulière). Une grammaire à états finis consiste en un mécanisme de transition, un automate, qui produit linéairement une suite de mots piochés dans un lexique. On la décrit de façon équivalente par un automate à états finis (une machine à nombre fini d'états et de transitions entre ces états) ou bien par un système de réécriture linéaire à droite, où toutes les règles sont de la forme  $A \rightarrow aB$  ou  $A \rightarrow a$  (« A est remplacé par a suivi de B », « A est remplacé par a »), a désignant un mot du lexique, et A et B des symboles de transition.

L'automate de la figure 4 procède à partir du lexique {un, homme, chien, observe, grand, très}. Cet automate contient une boucle qui permet d'engendrer un nombre infini d'énoncés, tel « un très grand chien observe un chien », « un très très grand homme observe un chien », et ainsi de suite.

Une petite grammaire comme celle-ci engendre des énoncés du français certes en nombre infini, mais peu diversifiés. Chomsky prouve que ce type de grammaire, même sur un lexique plus riche, ne permet pas de dériver des dépendances plus complexes entre énoncés (comme les constructions conditionnelles en *si... alors...*).

Un système de réécriture plus riche est fourni par les grammaires non contextuelles, où les symboles de transition représentent des catégories syntaxiques, et qui permet de représenter la structure d'un énoncé sous la forme d'un arbre (voir la figure 5). On admet aujourd'hui que ce type de grammaire décrit la syntaxe d'un grand nombre de langues, bien qu'il existe des constructions dans certaines langues (par exemple en suisse allemand) qu'une grammaire non contextuelle ne permet pas de dériver.

Plus généralement, on doit à Chomsky la définition d'une hiérarchie de grammaires formelles de plus en plus expressives, chacune correspondant à un type spécifique d'automate. L'exploration des grammaires formelles et de leur complexité se poursuit aujourd'hui.



**4. UNE GRAMMAIRE RÉGULIÈRE** est décrite de manière équivalente comme un système de réécriture linéaire à droite, ou comme un automate fini. Le système de réécriture (a) produit une chaîne de mots en remplaçant successivement « A » par « un B », puis chaque symbole de transition (en rouge) selon les règles de réécriture. L'automate (b) engendre une suite de mots en passant d'un état à un autre. Partant de l'état 1, il produit le mot « un » et passe à l'état 2. À chaque transition, il ajoute un nouveau mot au précédent : « un », « un très », « un très très », « un très très grand »... L'automate s'arrête dans l'état 7 en ayant produit une phrase complète.

## L'interprétation par les modèles

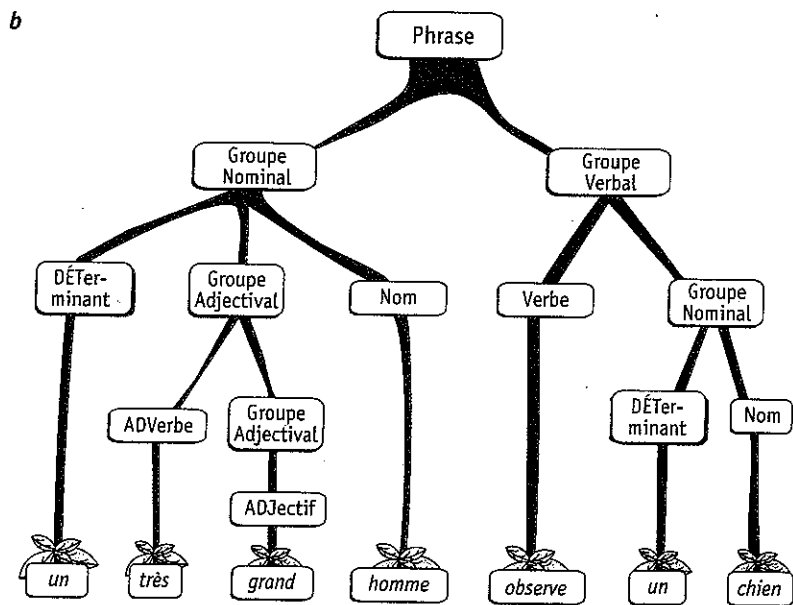
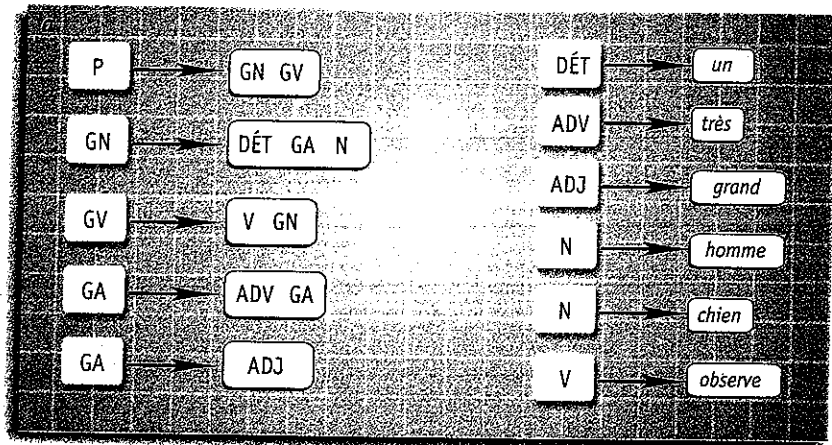
Il importe de mesurer ici l'écart entre l'approche chomskyenne et l'analyse du langage initiée par Frege, point de départ de la sémantique formelle. Suivant la définition classique due au théoricien du langage Charles Morris (1938), la syntaxe traite des expressions linguistiques indépendamment de leur sens, par opposition à la sémantique, qui traite du lien entre les expressions et leur sens. Cette distinction a d'abord été clarifiée par les logiciens à propos des langages formels. La logique formelle de Frege, par exemple, est un langage dont les expressions sont définies syntaxiquement, pas à pas. Pour l'appliquer, il faut toutefois l'interpréter, c'est-à-dire spécifier dans quelles conditions une formule est vraie ou fausse.

En 1933, Alfred Tarski publie un mémoire sur le concept de vérité dans les langages formalisés, où il définit la notion de vérité pour la logique du premier ordre, dans un cadre ensembliste. L'énoncé  $\forall x (P(x) \Rightarrow Q(x))$  est vrai si et seulement si tout individu qui a la propriété P a la propriété Q. On doit alors spécifier un domaine d'individus, de manière que, quelle que soit la valeur de la variable  $x$  sur ce domaine, si cette valeur appartient à l'ensemble dénoté par P, alors elle appartient aussi à l'ensemble dénoté par Q.

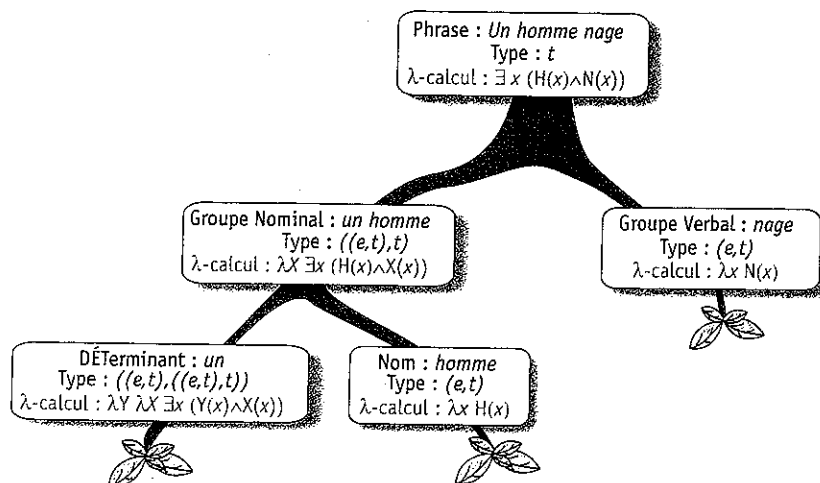
Un domaine d'individus assorti de l'interprétation ensembliste de chaque symbole de prédicat est appelé un « modèle ». Le passage de l'énoncé formel (une suite de symboles) au modèle (un univers ensembliste) sert à prouver la cohérence d'un système formel. Par exemple, l'énoncé  $\forall x (P(x) \Rightarrow Q(x))$  est vrai dans le modèle constitué des nombres entiers, où P est interprété par l'ensemble des multiples de 4, et Q par l'ensemble des nombres pairs. Dans ce modèle, l'énoncé signifie « Tout multiple de 4 est un nombre pair ».

L'idée d'interpréter les énoncés d'un langage formel dans un modèle est à l'origine de la sémantique dite modèle-théorique, que développe Montague vers la fin des années 1960.

C'est sous la direction de Tarski que Montague soutient sa thèse de doctorat en 1957 à l'Université de Berkeley, en Californie. Au cours des années



5. UNE GRAMMAIRE NON CONTEXTUELLE est un système de réécriture où les symboles de transition (en rouge) correspondent à des catégories syntaxiques (P pour phrase, GN pour groupe nominal, etc.), qu'on remplace successivement par leurs constituants (a). Les phrases engendrées par la grammaire ont une structure arborescente (b).



6. UN ÉNONCÉ TRAITÉ EN GRAMMAIRE DE MONTAGUE. À chaque catégorie syntaxique est associé un type sémantique (en bleu), par exemple t pour la phrase complète. Chaque composant possède sa traduction logique (en vert).

1960, il se met en quête d'une analyse formelle de la notion de signification en suivant le programme qu'a initié Rudolf Carnap en 1947. Comme Carnap, Montague s'intéresse d'abord à l'analyse logique des énoncés modaux ou intensionnels : les énoncés de nécessité (il est nécessaire que  $p$ ), de croyance (Pierre croit que  $p$ ), ou de temporalité (il sera / est / fut le cas que  $p$ ). Il analyse ces constructions en généralisant le formalisme de Frege, ainsi que la notion de modèle élaborée par Tarski.

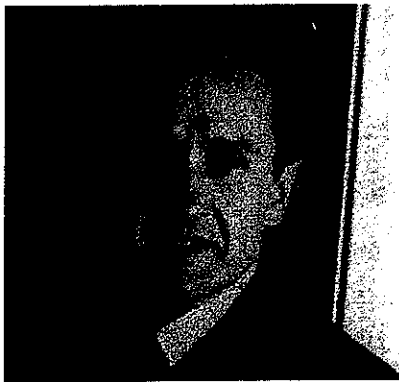
En 1970, un an avant de mourir brutalement (étranglé dans sa salle de bain, dans des circonstances mal élucidées), Montague publie trois articles fondateurs, où il propose de formuler la syntaxe et la sémantique pour des fragments de l'anglais. Comme Aristote ou Frege, Montague définit un cadre logique pour analyser les inférences qu'on peut faire dans un fragment de la langue naturelle, mais les constructions qu'il étudie sont plus complexes, et la logique qu'il utilise est bien plus expressive.

### Un outil d'analyse puissant

La logique intensionnelle que développe Montague combine trois outils logiques : les grammaires catégorielles du Polonais Kazimierz Ajdukiewicz, parentes des grammaires non contextuelles, le lambda-calcul de l'Américain Alonzo Church, et enfin la théorie des types (ou logique d'ordre supérieur), également développée par Church à la suite de Russell.

Le lambda-calcul est un langage fondé sur la notion de fonction, et qui constitue la base de certains langages de programmation (voir page 35). Church s'en est servi pour donner une formulation originale de la théorie des types, proposée par Russell pour éviter certains paradoxes (voir *L'antinomie de Russell*, page 56). La théorie des types hiérarchise les expressions de la logique selon le type des objets à quoi elles renvoient (individus, propriétés d'individus, propriétés de propriétés...). Une expression  $F(x)$  n'est bien formée (et la fonction correspondante n'est bien définie) que si le type de l'argument  $x$  est de degré immédiatement inférieur au type de la fonction  $F$  dans la hiérarchie.

L'association du lambda-calcul et de la théorie des types permet de géné-



**7. RICHARD MONTAGUE (1930-1971)** pionnier d'une analyse logique unifiée de la syntaxe et de la sémantique des langues.

raliser à toute expression composée l'analyse fonction-argument de Frege, au moyen d'expressions fonctionnelles plus riches. Partant d'un énoncé du langage ordinaire, par exemple « Un homme nage », la méthode d'analyse sémantique de Montague procède en quatre étapes (voir la figure 6).

Dans un premier temps on construit l'arbre syntaxique de l'énoncé en associant à chaque catégorie syntaxique le type sémantique approprié. La phrase entière a le type  $t$  (du mot anglais *truth* pour vérité), c'est-à-dire qu'elle prend pour valeur le vrai ou le faux. Le prédicat ou groupe verbal « nage » a le type  $(e, t)$  des fonctions qui associent à un individu (de type  $e$ ) une valeur de vérité (de type  $t$ ). Et ainsi de suite. En vertu du principe de compositionnalité, le type d'une expression composée se calcule à partir des types des expressions composantes. Par exemple, le type  $((e, t), t)$ , combiné au type  $(e, t)$ , produit le type  $t$ .

Dans un deuxième temps, on traduit en logique intensionnelle les expressions simples apparaissant dans l'énoncé. Par exemple, le prédicat « nage » est traduit par la fonction qui à tout  $x$  associe la valeur de  $N(x)$ , à savoir le vrai si  $x$  nage, et le faux si  $x$  ne nage pas : on note  $\lambda x N(x)$  cette fonction caractéristique de l'ensemble des individus qui nagent.

Troisième temps : on calcule la forme logique des expressions complexes à partir de la forme logique des expressions simples, selon les règles combinatoires du lambda-calcul. Enfin on évalue la forme logique de l'énoncé dans un modèle de la logique intensionnelle.

Cette procédure met en œuvre de façon systématique le principe de

« compositionnalité », formulé pour la première fois par Frege, selon lequel la signification d'une expression composée doit être une fonction des significations de ses composantes.

Respectant un parallélisme strict entre syntaxe et sémantique, la grammaire de Montague a ainsi été utilisée récemment pour le traitement automatique des inférences en langue naturelle.

De plus, cette analyse a ouvert la voie à une théorie nouvelle de la quantification, plus fidèle que celle de Frege et Russell à la grammaire « de surface » des énoncés. Ainsi, un énoncé comme « Un homme nage » aboutit à la traduction en logique du premier ordre  $\exists x (H(x) \wedge N(x))$ , qui se lit « Il existe un  $x$  tel que  $x$  est un homme et  $x$  nage ». Mais cette fois le groupe nominal « un homme », comme le déterminant « un », reçoivent une forme logique et une valeur sémantique propres (voir la figure 6).

Ce traitement s'applique aux expressions de quantité comme « tout homme », « un homme », mais aussi « la plupart des hommes », « cinquante hommes », etc., dont plusieurs linguistes et logiciens ont élaboré la théorie, dite des quantificateurs généralisés, à partir des années 1980.

L'analyse modèle-théorique de Montague a aussi été appliquée, avec succès, à la sémantique des énoncés interrogatifs (dont Aristote réservait l'étude à la rhétorique plutôt qu'à la logique), de même qu'à une variété croissante de constructions du langage naturel, telles que les expressions du temps ou le traitement des pronoms.

Depuis les années 1980, d'autres théories logiques ont vu le jour pour le traitement du langage naturel, mais le travail de Montague demeure, pour les linguistes comme pour les logiciens, une contribution décisive, et un exemple exceptionnel de progrès scientifique par la convergence et l'unification de méthodes initialement éparées.

**Paul ÉGRÉ, chargé de recherche au CNRS, est membre de l'Institut Jean Nicod, à Paris.**

ARISTOTE, *De l'interprétation*, trad. franç. J. Tricot, Vrin, 1959.

N. CHOMSKY (1957), *Structures syntaxiques*, Le Seuil, Paris, 1971.

G. FREGE (1879), *Idéographie*, trad. C. Besson, Vrin 1999.

R. MONTAGUE, *Formal Philosophy: Selected Papers of Richard Montague*, R. Thomason (ed.), Yale University Press, 1974.