

J. P. BENZÉCRI

## Combattre pour la linguistique

*Mathématiques et sciences humaines*, tome 35 (1971), p. 5-18

[http://www.numdam.org/item?id=MSH\\_1971\\_\\_35\\_\\_5\\_0](http://www.numdam.org/item?id=MSH_1971__35__5_0)

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1971, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## COMBATTRE POUR LA LINGUISTIQUE

par

J. P. BENZÉCRI <sup>1</sup>

Sur la science comme sur les arts, la Renommée impose à notre génération quelques grands thèmes sous quelques grands noms. Ainsi, en linguistique, l'écho amplifie et déforme la voix de Saussure ou la voix de Chomsky. Ceux mêmes qui, amoureux de la seule vérité, se retirent dans le fond d'un puits, ne peuvent purifier leurs oreilles du brouhaha structuraliste. Suivrai-je en singe docile la procession des docteurs ? Broderai-je sur ces étendards « Arbitrarité du Signe » <sup>2</sup> ou « Indépendance de la syntaxe d'avec la sémantique » <sup>3</sup>, à Dieu ne plaise ! Le critique est-il moins servile ? On en peut douter. Mettre « non » partout où un autre a mis « oui », c'est être encore moins qu'un singe : l'image d'un singe dans un miroir. Corriger avec discernement les erreurs les plus criantes d'une œuvre connue de tous, est-ce mieux que de déblayer et d'orner les voies rocailleuses qu'un tyran de l'esprit nous impose de suivre ? Ignorer enfin, n'est-ce pas seulement feindre d'ignorer : fuyant jusqu'aux noms des maîtres, éviterai-je bien de les pasticher ? Croyant parler de la langue, peut-être ne parlerai-je encore que des linguistes.

Il serait plus sage de me taire, d'ourdir en secret mes calculs jusqu'à ce que, nouveau Le Verrier, je puisse, sortant de ma retraite, pointer en vainqueur vers une nouvelle planète, je veux dire : paraître avec une théorie assez avancée pour qu'une application tangible — par exemple à la traduction, ou à la composition automatique de textes — témoigne sans conteste d'un succès positif. Et si cette application ne devait jamais venir, m'honorer du moins d'avoir gardé pour moi le silence ; et de n'avoir point ajouté au vacarme du siècle.

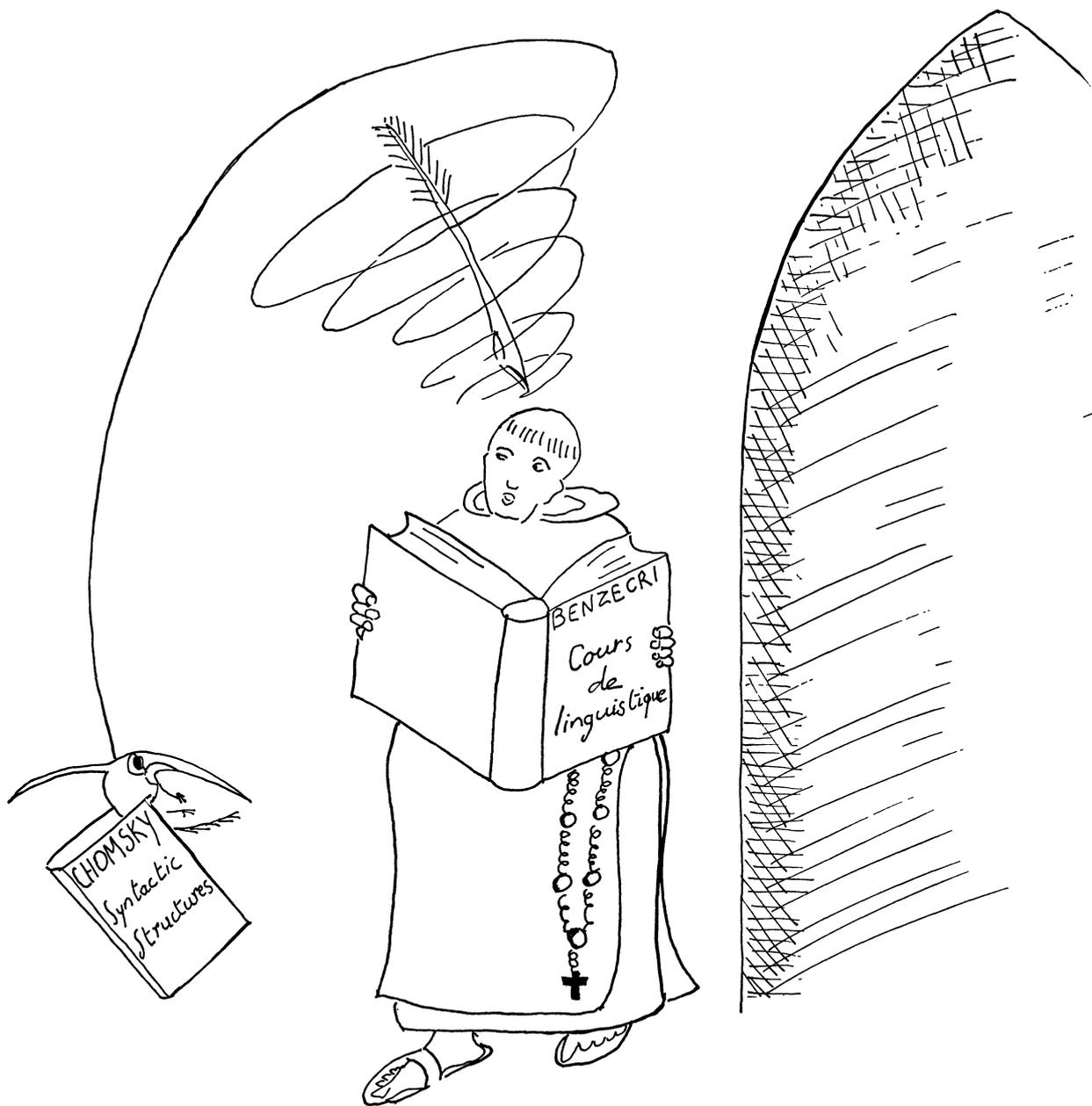
Mais il est tard. J'ai trop parlé... Voici plusieurs années déjà qu'un vénérable religieux à qui je faisais l'hommage d'un recueil de travaux où je croyais avoir concilié avec la modestie non feinte du savant, la réfutation ferme et motivée de thèses en vogue, me répondit innocemment qu'il admirait mon savoir et allait lui aussi, quoi qu'il en coûtât, se mettre au structuralisme. Hélas, fallait-il que par ma plume pénétrassent dans ce qu'il reste d'ombre sous les cloîtres les larves immondes de doctrines honnies ! Le papier dut-il en crever, force est, pour être fidèle, de piquer des points sur les i. Quelque démesure qu'il y ait pour un mathématicien qui se mêle de linguistique, à se dire antistrukturaliste, je me promis de prendre un jour ce titre scandaleux.

Cependant, on traduit en français le pamphlet fameux de Chomsky *Syntactic structures* et simultanément, du même auteur, à la même librairie, un tout autre ouvrage, *L'Amérique et ses nouveaux mandarins*, où je copie cette phrase :

« A l'aube de la nouvelle 'époque historique', où seule l'Amérique se trouve déjà engagée, l'impulsion civilisatrice de cette élite se traduit probablement par l'immense progrès accompli dans les

---

1. Laboratoire de Statistique Mathématique, ISUP.  
2. Thème Saussurien.  
3. Thème Chomskien.



*fallait-il que par ma plume...*

domaines qui embarrassaient les bourdons politiques du passé — le problème des grandes villes, de la pollution, du gaspillage et des effets destructeurs de la civilisation, de l'exploitation et de la pauvreté. Sous la direction de cette nouvelle race d'intellectuels-politiques, l'Amérique est devenue 'la société créatrice ; consciemment ou inconsciemment, les autres sociétés s'efforcent de l'imiter'. Voir, par exemple, les mathématiques, les sciences biologiques, l'anthropologie, la philosophie, le cinéma, la musique, l'histoire, etc. — les autres cultures, lamentablement distancées, se contentent d'observer et de copier ce que crée l'Amérique. Nous allons ainsi vers une nouvelle 'super-culture mondiale, fortement influencée par le mode de vie américain, transmise au moyen d'un langage-ordinateur-électronique-universel !...'

une brèche ‘psycho-culturelle’ énorme et de plus en plus large séparant l’Amérique du reste du ‘monde développé’».

Vous venez d’en avoir un exemple : il n’est pas impossible de se méprendre sur l’intention d’un texte ! Une ponctuation trop sourde nous dérobe les éclats de la voix. Pourtant, je ne puis douter que le texte de Chomsky ne soit écrit sur le mode ironique ; et parmi la liste des victoires gravée sur l’arc de triomphe grotesque du pays aux cinquante étoiles, un nom s’intercale : n’a-t-il pas sa place entre anthropologie et philosophie ? La linguistique ! Soyons sûrs que le mandarin mûrissant de Boston ne l’a pas omise sans sourire.

Et voici que le professeur Rybak entreprend de combattre pour la connaissance. Cet homme dont le nom apostolique se traduit du russe « pêcheur », jette sur moi son filet. Veut-il seulement prendre un article pour un journal, en appâtant ma vanité d’auteur par la promesse d’un livre à publier ? Peu importe : écrire un livre après un article qu’est-ce sinon surcroît de peine et de hasard ? Pourtant, je reste impassible sous le filet je ne fuis pas ; et sans trop attendre, « il faut que je me venge », je veux combattre pour la linguistique !

Combattre : mais sur quels champs ? Le nom de linguistique a conquis dans l’usage contemporain une fréquence considérable. Si La Bruyère reprenait le pinceau il se hâterait de nous donner un *caractère* de structuraliste ; Vadius et Trissotin péroreraient d’*arbres syntaxiques*, de *phonèmes* et de *morphèmes* devant des précieuses court vêtues... Mais on ne peut traiter avec quelque concision des thèses fondamentales de la linguistique générale qu’en argumentant sur des exemples, en reprenant des distinctions déjà débattues. Et il s’en faut que tous ceux qui laissent volontiers tinter à leurs oreilles les paradoxes à la mode, gardent dans leur mémoire ces données de base sans lesquelles toute démonstration n’est qu’un cliquetis de mots creux. Bien pis, il n’est que de lire ce qu’on débite aujourd’hui pour se convaincre que la paresse des lecteurs a depuis longtemps ligoté tout généreux auteur qui tenterait en proposant quelques faits précis de garder à ce qu’il affirme un sens distinct de son contraire.

Entrant dans la lice, je choisis donc de défier un chapitre de Chomsky qui, placé au centre de *Syntactic structure* est à la fois au seuil et au terme de la recherche : « on the goals of linguistic theory », les objectifs d’une théorie linguistique. Les objectifs d’une science en effet peuvent du moins être proposés à ceux qui ne cultivent pas cette science.

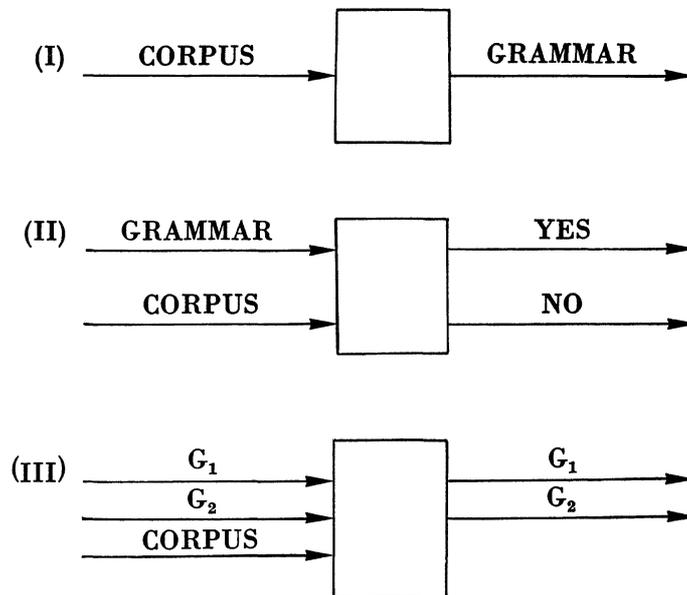


Fig. 2

La thèse de ce chapitre s'illustre d'un schéma, aussi odieux pour moi que la thèse même. Les obscènes affiches vingt fois répétées, les images grossières des écrans domestiques ont accoutumé notre siècle aux schémas. Répète qui voudra cet adage napoléonien qu'« un petit croquis en dit plus long qu'un long discours ». Je vous veux dire ici que l'abus du schéma est une offense à l'intelligence et à la nature. Offense à l'intelligence : parce que les termes et leurs rapports doivent être portés plus haut que dans les sens ; or, en s'arrêtant au plan du graphique celui qui enseigne comme celui qui apprend, sont tentés de s'accommoder de trop d'imprécision, de remplacer par un geste le mouvement peut-être impossible d'une démonstration que nul n'a scrutée. L'imagination, atelier du géomètre et jardin du poète, dessine certes dans la matière selon l'espace, des figures analogiques sans lesquelles le jeu rigoureux de l'intelligence humaine ne débiterait pas. Mais il y a loin d'une foisonnante intuition dont on ne sait pas encore décrire le faisceau, à ces sommaires assemblages de carrés et de droites étiquetés au pochoir. Offense à la nature : car en substituant aux objets des boîtes, aux relations des traits, l'homme s'accoutume à ne considérer dans les choses que les propriétés très peu nombreuses qu'il en utilise. Et placé ainsi au centre de tout, il s'exaspère en la solitude d'un univers de machines. Pourtant le plus humble objet est en lui-même, de par sa nature, dans l'œuvre de Dieu, infiniment plus que nous n'en utilisons. L'amour, parfois morbide, des ruines ne renferme-t-il pas une vérité métaphysique : neuf, un palais n'est qu'une épure à notre convenance ; ruine, il nous rappelle qu'il est pierre, rocaille, buisson... Mais considérons le schéma de Chomsky, puisque ce que nous avons dit contre tout schéma, ce que nous disons contre celui-ci, sert à la critique de sa thèse.

Première conception de l'objectif du linguiste : inventer une théorie — le carré blanc — qui, appliquée à un corpus, en fournisse la grammaire.

Qu'est-ce qu'un corpus ? un texte, ou encore un ensemble de textes, ou seulement un ensemble de phrases. Jusqu'au XIX<sup>e</sup> siècle (consultez Littré) *corpus* ne s'emploie guère en langue moderne que comme un abrégé de *corpus juris* : le corps, le recueil du droit romain. Puis le terme passe à l'épigraphie. On parle du corpus des inscriptions puniques, comme du recueil de toutes les inscriptions que les fouilles faites à Carthage ou dans les colonies de son empire, nous ont livrées en langue punique ; on a encore le corpus étrusque, etc. Il eût été juste de citer d'abord le corpus combien plus riche des inscriptions grecques ou latines. Mais ces corpus n'ont pas pour le linguiste la même valeur cruciale : car une tradition ininterrompue nous relie à Athènes et à Rome dont la langue n'a jamais cessé d'être comprise au moins dans quelques cercles de savants. Tandis que d'une langue tombée dans l'oubli, il peut n'être parvenu jusqu'à nous qu'un corpus d'une centaine d'inscriptions, toutes brèves, beaucoup d'entre elles mutilées, et dont aucune ou presque n'est accompagnée d'une traduction dans une langue connue. Finalement on en est venu à appeler corpus tout recueil, clos, bien défini de fragments dans une langue donnée : de tel dialecte en voie d'extinction, on a pour corpus six bandes magnétiques enregistrées par un ethnologue ; on peut choisir de travailler sur un corpus de cent distyches chinois, etc.

Cette situation quasi désespérée — partir d'un maigre recueil pour décrire une langue — est en un sens aussi, pour le linguiste, la situation parfaite : à défaut d'être soluble, le problème est du moins défini, circonscrit. Une langue, quand on a tenté de la cerner, se dérobe dans l'infini. Elle tient à une culture, à une civilisation : convenons de ne rien dire de cela. On voudrait dire seulement : une langue est un tas de phrases ayant un sens. C'est encore bien trop pour qu'on puisse faire un inventaire ordonné, mathématique. Au cours des trois derniers siècles, on en est venu à admettre comme modèle de la science la mécanique de Newton, telle que l'expose Laplace. Dans un espace, où la position de chaque point est repérée par trois coordonnées (disons ses distances à trois plans) se meuvent des masses, que nous supposons n'occuper chacune qu'un seul point : connaissant à un instant donné, les positions et les vitesses (en grandeur et en direction) de toutes les masses on peut calculer de proche en proche tout le mouvement ultérieur qui n'est régi que par deux sortes de lois : d'une part les principes généraux de la dynamique, d'autre part les formules particulières donnant les diverses forces (électrique, etc.) en fonc-

tion des distances. En bien des manières la physique moderne diffère de ce modèle ; et il est aisé d'opposer au modèle à la fois des faits physiques particuliers et des thèses philosophiques. Mais il est difficile de trouver une conception nouvelle dans laquelle sans cesser de calculer avec rigueur, on évite de définir au départ par une sorte d'épure (en mécanique : les trois dimensions, les masses, etc.) le format, le calibre de tout ce dont on pourra jamais se permettre de parler. Voilà pourquoi les épistémologies les plus résolument novatrices, bouleversent d'abord avec fracas les points de vue idéalistes et la typographie des formules, puis reviennent aux opérations communes.

Donc il faut avouer que nous ne savons pas définir ce qu'est une langue ; sinon matériellement (en la montrant du doigt, sans dire la raison de ce qu'elle est) : le turc c'est ce que disent et écrivent les hommes qui parlent turc et se comprennent entre eux. Ainsi, on pose un objet indéfini : dans le temps et l'espace d'abord, parce qu'on ne sait pas à quelle époque, à quel dialecte s'arrêter ; dans sa nature ensuite : car même à un moment donné, dans un cercle restreint, comment délimiter les phrases correctes, que tout le monde comprend du galimatias manifestement inintelligible ; il y a tant d'inventions littéraires possibles, de pastiches de constructions étrangères, etc. Une langue offre une infinitude, une potentialité tout autre que celle de l'objet du mécanicien. Celui-ci considère à la fois tous les objets matériels possibles, mais il sait que les lois de quelques situations particulières suffiront en se composant à décrire toutes les combinaisons possibles de boules, de tiges, de ressorts. Il le sait parce que la mécanique est assez avancée pour l'en assurer ; et il croit que dans les sciences de la matière la réduction du complexe et du changeant au simple et au constant est possible. Dans les sciences humaines au contraire, particulièrement en linguistique, le réel est plastique : une société, une langue est ce qu'elle est ; expérimenterait-on sur elle, à supposer qu'on le puisse, elle deviendrait autre chose (et autre chose qui dépendrait de l'ordre, du déroulement des expériences)... Y a-t-il des lois de la matière qui régissent tout l'univers ? Je le crois, quant à moi, d'autant moins que la disposition de la matière dans un objet est plus que la mécanique (par exemple une horloge qui obéit exactement aux lois physiques est en même temps une invention). Mais n'en parlons pas. Admettons un matérialisme des plus radicaux : faits sociaux, faits linguistiques résultent de la chimie de la matière des corps humains. Si l'on veut considérer ces faits non dans le détail des molécules mais au niveau des actes humains, des mots, etc., être tout de même linguiste non chimiste, alors l'objet demeure plastique fluant. Voilà pourquoi le corpus des inscriptions puniques dans sa rigidité, son étroitesse offre une base d'étude un niveau de référence plus solide qu'une langue vivante — le français. Et c'est à partir de cette base que l'on voudrait que la théorie établisse une grammaire.

Une grammaire. Nous en avons tous compulsé dans nos études. Déclinaisons latines ; conjugaisons grecques ; table des prétérites irréguliers de l'anglais ; règles de l'accord des participes français ; quand user du subjonctif et quand de l'indicatif ; et aussi l'analyse : sujets, épithètes, attributs... S'agira-t-il en linguistique de la même chose ? Nos grammaires classiques sont des instruments conçus pour une fin pratique : se conformer à l'usage contemporain ou antique, grâce à une somme d'informations qu'une terminologie reçue des anciens permet d'ordonner. Qu'il s'exprime dans sa propre langue, ou même dans une langue étrangère, l'homme n'a besoin que d'être soutenu par des règles : il est naturellement au centre du fait même de la parole ; il ne bâtit pas sur table rase (situation imaginaire que propose par exemple la philosophie idéaliste de Descartes) ; il tente de réformer, de perfectionner une activité acquise dès la petite enfance. La linguistique contemporaine voudrait au contraire une grammaire où tout fût explicite ; où l'activité naturelle de la parole n'entrât pas comme un tout, présupposé au départ (puisque, tous, nous parlons), mais fût reproduite par une suite d'opérations élémentaires, semblables à celles qui permettent en posant les chiffres et consultant les tables, de multiplier leurs nombres comme le ferait une machine. Pour Chomsky (et pour bien d'autres) une grammaire devrait être un algorithme, un programme : pour faire des phrases (grammaires génératives), ou pour donner d'une phrase une suite d'opérations qui y conduise (grammaire d'analyse, grammaire de reconnaissance).

Est-ce le lieu de préciser ce qu'est un algorithme, un programme ? On peut le faire en quelques phrases, dans le langage de l'honnête homme. Mais y a-t-il encore de ces honnêtes hommes ? Aujourd'hui,

l'on ne se fie qu'aux formules techniques ; or celles-ci ne sont possédées (le sont-elles ?) que par le technicien qui les utilise, tandis que le non-spécialiste regarde les reproductions des formules, comme un sauvage suspend à son cou un voltmètre qu'il prend pour une amulette. Quand j'aurai mis la définition d'un algorithme, je croirai qu'on l'a lue ; mais on l'aura seulement considérée comme une perle, ou comme un grumeau, sans s'attacher à en pénétrer le sens, et finalement cette explication accroîtra le malentendu entre le lecteur et l'auteur.



... suspend à son cou un voltmètre

Je préfère montrer sur un exemple, comment faute d'être un algorithme complètement formalisé, une sorte de règle du jeu de l'oie ou du loto, la grammaire classique peut manquer même son but qui est de décrire avec ordre la pratique de la langue, en insistant sur les difficultés, les constructions compliquées ou douteuses, voire les effets de style. Prenons le problème de la coordination en français, en nous bornant à la conjonction *et*. Parfois *et* réunit deux unités de même fonction pour en faire une seule qui remplit encore cette fonction ; par exemple :

*La rivière (fraîche et limpide) ;*

*(Pierre et Paul) chantent = ils chantent ;*

*Le soleil ((réchauffe les corps) et (illumine les cœurs)).*

Parfois, au contraire, *et* réunit deux membres dont l'un qui est incomplet reçoit de l'autre ce qui lui manque ; par exemple :

*(L'erreur entre par les yeux) et (la vérité par les oreilles) ;*

*(La roi dit qu'il vienne) et (la reine qu'il parte).*

Quelles sont les limites de cette construction ? Pourrait-on omettre dans ces phrases les mots placés entre crochets :

*Pierre le dit habile et Thomas [le dit] imbécile ;*

*Pierre aime le vin sec et Jacques [aime le vin] doux.*

Je ne saurais en discuter ici. Je veux seulement dire qu'on a pu écrire sur la coordination en français une thèse qui est un chef-d'œuvre de finesse et d'érudition sans marquer la distinction entre les deux classes d'emploi de *et*, et encore moins donner des règles. (Ayant consulté souvent cette illustre thèse, je crois pouvoir assurer que ce problème, même s'il est évoqué incidemment en un lieu qui m'ait échappé n'occupe pas une place telle qu'il soit possible au lecteur de le retrouver.)

Mais alors, suis-je Chomskien, structuraliste ? Je vous dis que non... Il faudra bien qu'avant la fin de ces quelques pages je conclue là-dessus avec sincérité et clarté. Voici seulement en peu de mots ce que je trouve à redire à cette conception de la grammaire. Comme elle est relativement à la langue un idéal, justifiable certes en plusieurs de ses exigences et fécond par les problèmes qu'il impose, mais en somme inaccessible car *a priori* mal défini (on ne sait pas du tout ce qui peut en être réalisé et qui pis est, on ne sait où commence le travail utile) la conception algorithmique de la grammaire détourne de l'étude des langues vers celle des algorithmes. Au lieu d'étudier des textes, d'apprendre des idiomes, etc., les

jeunes linguistes s'appliquent de moins en moins aux faits de langues ; ils croient indispensable de s'aventurer dans les mathématiques : la plupart n'en retirent qu'un charabia rudimentaire de formules par lesquelles ils sont incapables d'exprimer des vérités linguistiques profondes et ne peuvent que mystifier leurs aînés et leurs collègues. Nous le répéterons sans cesse dans la suite : aujourd'hui en linguistique, comme dans la plupart des sciences humaines : la rigueur mathématique (de l'exposé des formules) ne s'accroît qu'aux dépens de la rigueur méthodologique (de l'acquisition des faits et de la confrontation des lois qu'on affirme aux faits qu'on observe).

Nous avons expliqué corpus, grammaire ; reste le terme moyen : théorie linguistique. En quel sens pourrait-on dire qu'une théorie permet de faire la grammaire d'une langue donnée seulement par un corpus ? Proposons des exemples.

Supposons que dans toutes les langues chaque phrase soit formée de trois mots : le sujet d'abord, puis le verbe et le complément ; et qu'il y ait seulement deux classes de mots qui sont invariables (ne se déclinent ni ne se conjuguent) : les noms (sujet ou complément) et les verbes. Il est clair que si cette théorie était vraie, l'on aurait achevé la description d'une langue (donc sa grammaire au sens pris ici) en donnant la liste de ses noms et celle de ses verbes ; d'autre part il serait facile d'établir ces deux listes d'après un corpus assez vaste. La théorie permettrait donc de déterminer une grammaire d'après un corpus.

Prenons maintenant une théorie hypothétique moins simpliste : il y a cinq classes de mots, tous invariables : les noms, les verbes intransitifs, les verbes transitifs, les adjectifs et les adverbes ; chaque phrase a un sujet (nom) éventuellement accompagné d'un ou de plusieurs adjectifs ; elle a un verbe et parfois des adverbes ; enfin, si le verbe est transitif, il doit y avoir un nom complément (avec ou non des épithètes) ; l'ordre des mots est fixe pour une langue donnée mais il varie selon les langues, toutefois le sujet précède toujours le complément. Selon cette théorie, faire la grammaire d'une langue, ce sera d'une part donner les listes des cinq classes de mots ; d'autre part indiquer quel est l'ordre des parties du discours, c'est-à-dire donner des règles telles que : l'adjectif épithète précède (ou suit) toujours le nom auquel il se rapporte ; l'adverbe est immédiatement avant (ou après) le verbe ; sujet, verbe et complément se succèdent dans l'ordre *SVC* (ou *SCV*, ou *VSC*). Sous ces hypothèses théoriques, il est encore possible de passer du corpus à la grammaire. Précisons comment le faire.

D'abord il y a dans le corpus des phrases de deux mots : ce sont soit toutes les phrases *SV*, soit toutes des phrases *VS*. Nous déterminerons plus tard quelle est la règle d'ordre des mots. D'après l'ensemble des phrases de deux mots (à supposer évidemment que celles-ci soient en assez grand nombre dans le corpus cette hypothèse va de soi, et nous n'y reviendrons pas) on peut déjà établir deux listes de mots : celle des mots qui sont toujours le premier dans une telle phrase, et celle de ceux qui y sont second (et dernier) ; l'une contient tous les noms (*N*) et l'autre tous les verbes intransitifs (ou neutres : *Vn*) ; mais on ne sait pas encore laquelle est *N* et laquelle *Vn*... Passons aux phrases de trois mots : certains contiennent, entre un nom sujet, un verbe intransitif et soit un adjectif épithète du nom, soit un adverbe se rapportant au verbe. D'autres comportent un nom sujet, un verbe transitif (*Vt*) et un nom complément : nous reconnaissons facilement ces phrases à ce qu'elles comportent deux mots (les deux noms), appartenant à l'une des deux listes déjà établies d'après les phrases de deux mots. Du coup nous savons laquelle de ces deux listes est *N* (c'est celle des mots dont deux peuvent se trouver dans une même phrase de trois mots) ; l'autre liste est *Vn* ; quant à *Vt*, c'est la liste des mots (verbes transitifs) qui avec deux noms peuvent former une phrase ; de plus nous pouvons choisir entre les trois règles *SVC*, *SCV*, *VSC*. Reste à dresser la liste des adjectifs (*Aj*) celle des adverbes (*Av*) et à trouver la liste d'emploi de ces mots auxiliaires, e.g., soit *NAj*, soit *AjN* pour les noms accompagnés d'une épithète. Supposons, par exemple, que l'on ait déjà adopté la règle *SVC* : alors, si la règle de l'adjectif est *NAj* (épithète après le nom) et dans ce cas seulement, il existera des phrases de quatre mots *NVN<sub>a</sub>* dont le quatrième ne pourra être qu'un adjectif ; tandis que si la règle est *AjN*, on aura des phrases *aNVN*. Ceci suffit à déter-

miner la liste *Aj* des adjectifs, comme la liste des mots qui peuvent se trouver à l'extrémité (droite ou gauche selon la règle) d'une phrase contenant deux noms et un verbe. Le résidu, les mots qui ne sont ni dans *N*, ni dans *Vn*, *Vt*, *Aj*, forme la liste *Av* des adverbes ; il est facile de déterminer si la règle est *AvV* ou *VAv* : la grammaire est achevée. Si on a la règle *SCV*, la procédure n'est guère différente : en bref, ou bien il y a des phrases *aNNV* dont le premier mot est nécessairement un adjectif, et la règle est *AjN* ; ou bien il n'y a pas de telles phrases (avec un seul mot devant *NNV*) et alors la règle est *NAj* ; il y a des phrases *NaNV* et c'est par elles que l'on détermine la liste *Aj* puis la liste *Ad*, etc. Et semblablement dans le cas *VSC*.

Voilà logé dans le carré blanc de la théorie linguistique, un algorithme (car c'en est un) déjà un peu compliqué. Ceux d'entre vous qui aiment les rébus et les mots croisés auront peut-être suivi avec attention. Et d'aucuns, en qui survit le rêve des alchimistes de faire de l'or avec de la pierre, se voient peut-être déjà perfectionnant la théorie pour déchiffrer l'étrusque ou le cypriote linéaire A. Chomsky, quant à lui rejette cette première conception, si ambitieuse, de la linguistique. Nous ne pourrions que louer la modestie et la prudence de l'auteur de *Syntactic structures*, si la conception à laquelle celui-ci se rallie, ne risquait, sous couvert de modestie, d'engager le linguiste à légiférer loin de la langue : ce qui est certes la suprême imprudence. Nous y reviendrons après avoir exposé plus complètement les vues de N. Chomsky.

Voici une deuxième conception de la théorie linguistique : décider si telle grammaire est, oui ou non, la meilleure pour une langue représentée par tel corpus. Comme la perfection n'est pas de ce monde, on peut sans crainte de perdre, parier que la théorie n'aura jamais que l'occasion de répondre non : de ce point de vue, on aura donc une théorie parfaite en supprimant la flèche YES qui sort du carré blanc. Nous croyons que la critique que Chomsky oppose à cette deuxième conception n'est au fond qu'une forme enveloppée de notre boutade. Mais la condamnation s'assortit d'attendus qui méritent l'attention.

Nous avons dit qu'on entend par grammaire un système de règles opératoires par lesquelles on puisse composer les phrases d'une langue, et rien qu'elles. C'est prétendre par le moyen d'une description finie, maîtriser l'infini, ou plutôt (*cf. supra*) l'indéfini de la langue. Pour concevoir dans quelle mesure cela est possible, revenons à notre théorie linguistique avec cinq classes de parties du discours et quelques règles. Choisissons par exemple les règles *SVC*, *AjN* et *AvV*, et prenons pour vocabulaire des mots français. On aura par exemple les phrases :

*Lion manger chameau : SVC.*

*Grand lion toujours manger petit chameau : Aj S Av V Aj C.*

*Grand méchant lion toujours toujours vite manger gentil mignon petit chameau : Aj Aj S Av Av Av V Aj Aj Aj C.*

Puisque la théorie laisse indéterminé le nombre d'adjectifs ou d'adverbes se rapportant à un même mot, la grammaire peut conduire à des phrases arbitrairement longues, si l'on fait jouer itérativement (on dit récursivement), la règle de qualification. Dans une langue réelle, comme le français les règles de subordination des propositions peuvent de même s'appliquer récursivement : la subordonnée ayant une subordonnée qui elle-même, etc. ; et de même la construction du nom complément de nom, etc. On dira que la grammaire conduit à un infini récursif. Pour Chomsky ce caractère est essentiel : on y voit à la fois comment une grammaire finie peut prétendre à l'infini ; et comment dans cette conquête elle passe nécessairement la mesure que l'on doit garder en parlant. On doit reconnaître que ce qu'on appelle la mesure, se laisse mal définir en mathématique : on songe à imposer une borne fixe : jamais plus de trois adjectifs ; à proposer des probabilités : une fois sur quatre un adjectif, une fois sur huit deux adjectifs, etc. La réalité est tout autre.

Lisons le texte grec de l'Évangile selon Saint Luc en cherchant des exemples de récursivité : il n'y a pratiquement pas de chaîne au delà de 2. Soudain vient la généalogie du Christ : qui « était à

ce qu'on croyait fils de Joseph d'Héli de Matthat de Lévi de Melchi de Jannai de Joseph, etc.» Sur presque une page les ancêtres se suivent, chaque nom étant le complément (au génitif, introduit par l'article « tou » = de) du précédent qui est son fils. Les Chomskiens de stricte observance affectionnent la récursivité : ils distinguent entre compétence (système de règle que l'on connaît ; cf. la règle de la subordonnée relative ; la méthode pour multiplier deux nombres comportant chacun un nombre quelconque de chiffres) et performance (on ne comprend pas une phrase où s'enchevêtrent plus de 3 propositions relatives ; on ne multiplie pas sans erreur deux nombres de 10 chiffres). Nous ne partageons pas leur enthousiasme : dans l'étude des langues réelles naturelles, la récursivité n'est qu'un problème entre tant d'autres : s'y appesantir, c'est user d'un alibi mathématique, travers trop commun des formalistes...

Encore quelques phrases bâties selon notre théorie hypothétique :

*Montagne vite regarder joli chameau ;*

*Maison manger méchant lion.*

Ici le jeu aveugle des règles a produit des phrases qui n'ont pas de sens : une montagne peut-elle regarder ? Une maison peut-elle manger ? Non : mais cela relève-t-il de la syntaxe ? Pour Chomsky, le problème de la correction sémantique (avoir un sens) peut et doit être séparé totalement de celui de la correction syntaxique (respecter les règles...). Certes, si l'on distingue parmi les noms des noms d'agent, des noms de qualité, des noms d'objets tout comme on a distingué des verbes transitifs et des verbes neutres, il deviendra possible de mettre en forme de règles, certaines conditions de corrections sémantiques, e.g., un nom d'objet ne peut être sujet d'un verbe d'action, etc. Chomsky qui entrevoit une sémantique plus ambitieuse refuse de la brouiller avec la syntaxe. Que reste-t-il, dans la pratique des langues, de cette distinction idéale ? Au premier abord, à prendre des libertés avec la syntaxe, on produit des phrases incorrectes dont le sens peut être très clair (et l'on entend souvent de telles phrases) ; tandis qu'à défier la sémantique, on joue le poète surréaliste avant de sombrer dans l'obscurité impénétrable. Il s'agit bien de deux ordres distincts. Pourtant si l'on cherche à préciser la syntaxe d'une langue au delà des règles simplistes qu'après tant d'autres j'ai proposé pour exemple, on découvre que le maître problème est celui de la syntaxe spécifique des mots individuels, surtout des verbes. Parmi les verbes français il en est des dizaines (ou plutôt des centaines) qui entrent dans une grande variété de constructions : il est fort rare qu'une construction soit propre à un seul verbe (pourtant la proposition infinitive introduite par la préposition *par* est propre aux verbes *commencer* et *finir* et à leurs synonymes : *il finira par réussir...*) ; mais il est aussi exceptionnel que deux verbes qui ne sont pas de quasi synonymes acceptent exactement le même ensemble de constructions. Ainsi, quel autre verbe que *dire* pourrait-on placer à la fois dans toutes ces phrases :

1° Assertions, information :

*Pierre dit (à Paul) du bien de Jacques,*

*... que Jacques est un ami sûr,*

*... où trouver ce renseignement,*

*... où est le dictionnaire,*

*... savoir nager,*

*... cette maison fait belle.*

*On lui dit une grande fortune.*

2° Ordres : *... qu'il ne revienne plus ; de partir à l'instant.*

3° L'acte de parler : *bien faire et laisser dire ; dire un compliment, etc.*

On peut bien distinguer syntaxe de sémantique, lexique de grammaire ; pour avancer dans la connaissance des langues, dans leur description explicite et rigoureuse, il n'y a, nous semble-t-il, pas d'autre voie que de faire un inventaire ordonné des emplois des mots, particulièrement des verbes. Or le premier principe externe de classification c'est la syntaxe : tandis qu'au fond l'inventaire s'achève dans une analyse articulée des sens.

Il faudrait encore dissiper une confusion que nous avons jusqu'ici entretenue. S'agit-il de langue écrite ou de langue parlée ? Et, dans l'un ou l'autre cas, les séparations entre les mots successifs sont-elles indubitablement marquées ? Ou la segmentation est-elle le premier problème qui se pose au linguiste ? Peut-être dans le discours parlé ne reconnaissons nous même pas une suite de sons élémentaires que nous soyons capable, par exemple, de noter au moyen d'un alphabet. L'histoire de l'écriture (et que dire de l'orthographe !) témoigne suffisamment qu'il s'agit ici de problèmes difficiles. Dans les exemples écrits nous avons seulement considérés des suites de mots distincts : mais pour être complet, il faudrait commencer l'étude d'une langue par l'enregistrement d'un corpus oral ; puis, selon une théorie linguistique segmenter en éléments de son et de sens, etc. C'est bien plus que nous ne pouvons faire ici. Pour Chomsky, la seule solution est de postuler à chaque niveau une structure hypothétique (semblable à ses grammaires), la machine à découper étant tout aussi inimaginable que la machine à inventer les grammaires (*cf, supra*). Pour nous, qui ne sommes ni idéalistes ni formaliste, la solution est certes (ici comme ailleurs) de partir d'une première approximation traditionnelle de la structure (car nous ne sommes pas à l'extérieur de la langue, nous vivons au milieu d'elle) ; mais l'amélioration progressive de cette description est l'objet même de la théorie linguistique. Disons qu'à l'entrée du carré blanc, nous mettons un corpus et une grammaire, attendant, à la sortie, une meilleure grammaire (et une meilleure segmentation, etc.).

On voit maintenant pourquoi Chomsky refuse même qu'une théorie linguistique doive décider si une grammaire convient ou non à une langue. La grammaire est une construction logique que la langue refusera toujours. Mais, et c'est le sens du dernier carré du graphique, il doit être possible de dire, de deux grammaires  $G_1$  et  $G_2$ , laquelle, avec le plus de simplicité, comporte le plus de traits de la langue que représente le corpus donné. Pour Chomsky, le critère de simplicité formelle prime tout. A telle enseigne que, quand il imagine ce que pourrait être la machine à inventer les grammaires, Chomsky la présente comme un filtre : puisque toute grammaire doit être écrite, être une suite de symboles logiques, écrivons n'importe quoi au hasard ; vérifions si c'est là une grammaire pour le corpus ; et recommençons jusqu'à avoir trouvé une grammaire adéquate qui soit de surcroît la plus courte. Cette fiction logique rappelle le paradoxe des singes dactylographes ; c'est comme si quelque éditeur fatigué de chercher un romancier mûr pour le prix Goncourt, installait devant des machines à écrire des centaines de macaques, et attendait la sortie du chef-d'œuvre. Avec suffisamment de macaques, le chef-d'œuvre serait sans doute bientôt fait ; mais quel lecteur inspiré le découvrirait-il dans la montagne des manuscrits soumis à la publication ?

Nos secrétaires à quatre mains ont introduit le thème du hasard. C'est l'occasion pour le statisticien que je suis, de défendre contre Chomsky la part prépondérante qui revient à notre science dans les études linguistiques.

L'on croit encore trop souvent que la statistique ne vise pas plus loin qu'à calculer des fréquences pour en orner des sophismes. D'aucuns en usent ainsi, avouons-le. « La proportion des gens qui lisent est plus élevée parmi les spectateurs de la télévision que parmi les non-spectateurs : l'image entrevue donne le désir d'approfondir », conclut une revue acquise aux passe-temps nouveaux mais qui se targue de culture ! Il serait assurément plus juste de dire, nous le sentons bien, que puisque seuls quelques seigneurs de l'esprit, dont la plupart lisent beaucoup (faute sans doute de s'être élevés assez haut pour que la méditation de la Révélation première suffise à les nourrir), refusent la lucarne qu'ils pourraient s'offrir, ce sont les mêmes hommes qui, nantis de quelque fortune, achètent à la fois les livres et les récepteurs. Certes, le seul calcul de quatre fréquences ne permet pas de s'élever aux causes.

Mais supposons qu'au lieu de ne considérer de chaque individu que les deux propriétés « lit ou ne lit pas de livres », et « regarde ou ne regarde pas la télévision », on recueille aussi, à supposer que cela soit possible, une série d'informations telles que : loyer payé, profession, âge..., ou mieux, dépenses annuelles en produits alimentaires, en chauffage et éclairage, en transports, en objets de luxe, etc. On

pourra à l'aide de ces observations, faire une classification des personnes interrogées (sinon d'après ce que sont vraiment leurs dépenses qu'elles auront peut-être déguisées, du moins d'après ce qu'elles se plaisent à prétendre dépenser). On verra alors l'ensemble des classes s'organiser d'abord suivant un axe, aux extrémités duquel les riches s'opposent aux pauvres ; puis apparaîtra un deuxième axe qui opposera les travailleurs de l'industrie aux artisans et aux savants, etc. Sur le premier axe, télé et lecture, progressent l'une avec l'autre au fur et à mesure que s'élève le niveau de fortune, mais l'autre dimension les séparera (la télé pour les hommes des usines, les livres pour ceux de la plume et des outils ?).

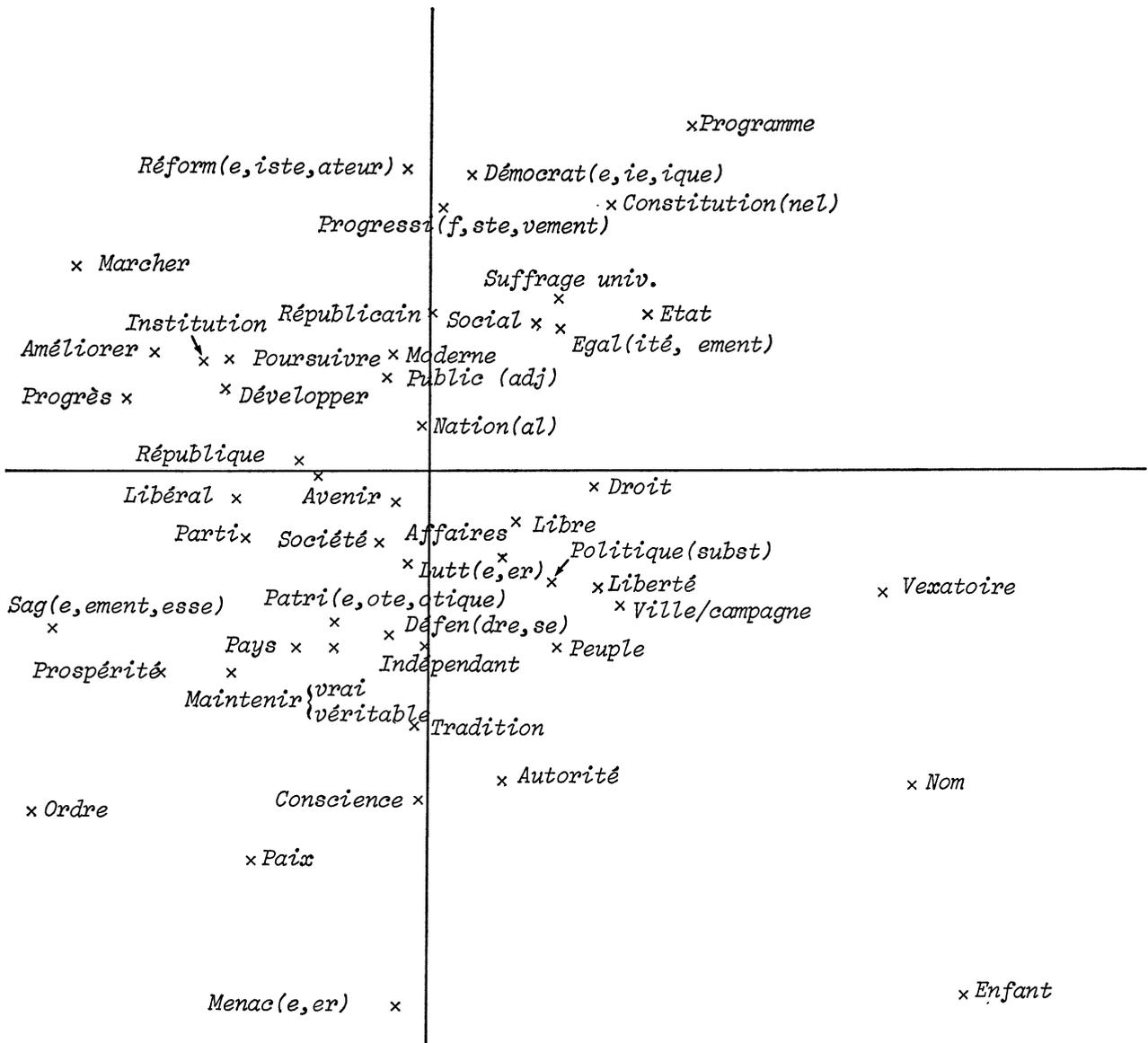
Voilà du moins ce que me suggère le sens commun. Mais seule l'expérience où plutôt une observation ample, difficile) puis des calculs statistiques (calculs si complexes que seuls les ordinateurs nous ont permis depuis peu de les aborder) permettraient de voir sans s'asservir à des idées *a priori*, ce qu'il en est en réalité ; quel équilibre, quel ordre règne sur les faits élémentaires seuls accessibles à l'observation directe. Pour montrer quelles armes le calcul statistique offre à la science, considérons un exemple qui nous rapprochera de la linguistique.

Depuis bientôt un siècle, au début de chaque législature, on réunit en un recueil les proclamations électorales des députés : ce recueil est appelé Barodet, du nom de l'homme politique qui en fit entreprendre la publication. Il y a là, on l'imagine, quelques choses, cachées sous beaucoup de mots ! L'électeur de qui l'on attend qu'un nom — pas même : un geste — n'est qu'un épiderme qui se ride ou s'assouplit à ces accords privilégiés : Liberté, Égalité, Fraternité ; Travail, Famille, Patrie. Ne valait-il pas la peine de compter ces mots ? C'est ce qu'a entrepris mon collègue l'historien A. Prost, sur le Barodet de 1881.

Mais quels mots compter ? Compter dans chaque proclamation le nombre de fois que se rencontre chaque mot de la langue française eût été une gageure presque insoutenable. D'ailleurs, qu'eût-on fait de cet inventaire... Il fallait se restreindre à quelques dizaines, voire une centaine, de mots importants. Pour choisir les mots, l'historien fit d'abord relever dans une liasse de proclamations de droite, et une liasse de gauche tous les mots pleins (noms, adjectifs ou verbes ; mais, non pas pronoms ou auxiliaires), afin d'en reconnaître les plus fréquents. Parmi ces mots, certains (tels que « député » ?) pouvaient être également chers à tous les partis : l'historien ne garda que ceux dont la fréquence d'emploi variait le plus des députés d'extrême gauche à ceux d'extrême droite : ainsi, en 1881, *radical* est un mot d'extrême gauche ; *menace* est un mot de droite, car dans la troisième république naissante, monarchistes et bonapartistes sont sur la défensive... On établit de la sorte une liste de 53 mots ; puis finalement (et non sans peine !) un tableau à 53 colonnes et autant de lignes qu'il y a de députés (près de 500) donnant pour chacun de ceux-ci le nombre de fois qu'il a employé chaque mot de la liste.

D'un pareil tableau, le statisticien de la fin du 2<sup>e</sup> millénaire de l'ère chrétienne, fait son affaire. Par un programme de calcul systématique — le même programme quelle que soit l'origine des données — histoire, linguistique ou botanique ; et un programme qui, étant exécuté par une machine, se déroule implacable, sourd aux hypothèses, aux espérances de celui qui a rempli ce tableau... On construit une carte (carte plane, ou à un plus grand nombre de dimensions : au non-mathématicien je ne puis apporter là-dessus plus de lumière que par un exemple) sur laquelle lignes et colonnes du tableau sont chacune représentées par un point. Ici cela veut dire un point par député et un point par mot sur une feuille où sont tracés deux axes. Se trouvent proches entre eux les députés qui usent à peu près des mêmes mots (en général : les colonnes qui ont des profils semblables) ; et aussi les mots qui se rencontrent souvent dans une même proclamation ; et enfin les députés et les mots qui ont des affinités entre eux.

Regardez-les donc ces mots, tels que la statistique les a rangés : n'a-t-on pas envie d'en faire un discours ! Quelques grognements imperceptibles suffiraient à les réunir en des caricatures de phrases. Et que ne pouvez-vous voir les députés à leur clavier : ceux du centre, comme on l'imagine sont au milieu ; mais, à 9 exceptions près, les 87 étiquetés « extrême gauche » ou « radicaux » sont dans la moitié supérieure de la feuille (à partir de l'axe horizontal) ; tandis qu'à 2 exceptions près, les 81 conservateurs, libéraux, bonapartistes et légitimistes sont au dessous de cet axe. Ainsi dans un tableau, patiemment rempli en comptant des mots, l'aveugle statistique devine l'ordre des familles politiques.



ANALYSE DU VOCABULAIRE  
DES PROCLAMATIONS ELECTORALES DE 1881

Sacré  
x  
↓

Famille  
x  
↓

Mais, une objection se présente. Assurément, il faut reconnaître au programme de calcul, le mérite de bien ordonner les lignes et les colonnes d'un tableau ; mais la construction même du tableau — plus précisément le choix des colonnes, c'est-à-dire des mots comptés — repose sur une étude historique qui contient les notions de *droite* et de *gauche* que l'on s'émerveille de retrouver ensuite (on se souvient qu'on s'est limité aux mots dont la fréquence variait beaucoup de la droite à la gauche). N'y a-t-il pas dans le vocabulaire de toutes ces proclamations une autre structure plus importante que celle que nous venons de retrouver, et que notre hypothèse politique nous aura fait perdre de vue ?

Pour répondre à cette objection, de même que, sans ostracisme, nous avons considéré tous les députés, il faudrait accepter tous les mots, ou du moins tous les mots pleins (n'éliminer que les articles, etc.) les plus fréquents. C'est ce qui a déjà été fait, mais non pour tous les députés (car on hésite devant ces dénombrements monumentaux dont l'issue est incertaine) : pour quelques dizaines d'hommes seulement, choisis au hasard. Et l'analyse d'un tableau à 36 lignes (députés) et 200 colonnes (mots) conduit aux mêmes résultats que celle du tableau à 53 colonnes (mots). Il existe encore un autre tableau à partir duquel on peut définir objectivement l'opposition droite-gauche : c'est le tableau des votes de la Chambre. L'analyse a été faite d'abord sur 13 scrutins jugés par l'historien être les plus importants ; puis cette première analyse ayant réussi, d'après 60 scrutins choisis cette fois au hasard (sans le critère d'importance politique). Les résultats sont encore plus beaux qu'avec le vocabulaire : on voit se dessiner un hémicycle avec à une extrémité (qu'on nous permettra d'appeler la droite bien que le sens dans lequel l'imprime la machine soit indifférent), séparés du reste de la Chambre, les bonapartistes, puis les monarchistes...<sup>1</sup>

On a également tenté de faire une analyse du vocabulaire des éditoriaux où certain jour des années 1880, 29 journaux débattaient devant leurs lecteurs la question tunisienne. Mais ici la carte produite par l'ordinateur n'a pas rendu la disposition politique connue des 29 journaux. Pourquoi cela ? Est-ce parce que les éditoriaux de l'époque moins superficiels que les proclamations exposent tous les faits avec les mêmes mots et n'expriment leurs opinions qu'avec mesure. Ce n'est pas le lieu d'en débattre. Je voulais seulement, pour que la discussion méthodologique fût complète, faire état aussi des obstacles <sup>2</sup>.

Avant de conclure sur un programme de recherche aux confins de la linguistique et des mathématiques, tentons d'après les exemples qui précèdent de proposer ce que nous appellerons une méthode statistique inductive.

Le programme fondamental de la science positive est de s'élever de l'observation du particulier à la vision du général ; des faits aux lois. Certes, il est peu raisonnable d'espérer que le savant disposera un jour d'une méthode transformant les données d'expériences en notions nouvelles et en lois reliant celles-ci aussi sûrement et aussi automatiquement qu'on calcule une trajectoire balistique. Chomsky est fondé à écarter d'abord de la linguistique cette conception trop ambitieuse. Pourtant si, comme nous le croyons, les notions et les lois visent à être plus que des idées forgées par le savant — ces idées seraient-elles utiles — si la science a pour objet la nature telle qu'elle est en elle-même, on ne peut pour recevoir une théorie nouvelle se contenter de vérifier qu'elle est, en un certain sens, meilleure que ses concurrentes du moment (cf. la 3<sup>e</sup> conception de la linguistique à laquelle se rallie Chomsky) : il faut, autant que possible, s'assurer qu'elle est vraie. Que toute théorie humaine soit incomplète, n'empêche pas qu'il y ait des théories vraies : la mécanique de Newton est vraie, dans son domaine propre qui est dans la nature bien défini, quoiqu'elle ne comprenne pas les faits dont traitent les théories quantiques et relativistes ; nous ne souscrivons pas au relativisme idéaliste.

---

1. Ces recherches se poursuivent sous la direction de A. Prost avec la collaboration de M. Danesh-Pajouh, M. Demonet, A. Rezvani, Ch. Rosenzweig ; cf. e.g., A. Prost et Ch. Rosenzweig, « La Chambre des Députés (1881-1885) », in *Revue française de science politique*, 21, pp. 5-50 (1971).

2. Les recherches sur le vocabulaire de la presse ont depuis la rédaction de cet article, fourni des résultats notables.

Si donc la vérité est notre objet, il faut par une méthode rigoureuse s'assurer que ce que l'on admet dans la science n'est pas un compromis entre l'hypothèse et la réalité ; mais la réalité pure que l'hypothèse nous aura seulement permis de découvrir. Tel est, nous en sommes sûrs le sens du fameux *hypotheses non fingo* (je n'invente pas d'hypothèse) de Newton. Dans les sciences du complexe (les sciences humaines, etc.) la méthode nous paraît devoir être la suivante. Ayant fait une première fois, comme par inspirations successives, le chemin des faits aux lois (du corpus à la grammaire), substituer à ce premier parcours une démarche inductive automatique qui par une suite de règles indépendantes du cas particulier traité (*e.g.*, telle langue) conduise à des lois analogues à celles qu'on obtient d'abord sans s'interdire d'« inventer des hypothèses ». L'exemple de la définition statistique de la droite et de la gauche illustre nos exigences et aussi la portée des méthodes statistiques disponibles en 1971 pour fonder l'induction : ce que l'historien pressent, on peut, dans certains cas, le cerner par approximations successives dans une définition statistique qui des données à la structure procède, à la fin, d'elle-même. Le mérite de cette définition est semblable à celui d'une mesure précise : l'appareil de mesure en introduisant la quantité là où il n'y avait qu'appréciation de qualités, permet de calculer des différences donc d'accéder à des faits nouveaux. (Par exemple, l'historien étudiera le mouvement des députés dans l'hémicycle au cours d'une législature ; il recherchera qui sont les députés radicaux qui ont un vocabulaire du centre, etc.) Mesurer les structures comme avec une balance : un programme utopique, dira-t-on ; mais s'il ne parvient à en réaliser quelque chose, le structuralisme n'évitera pas que, sous ses bannières, des impotents côtoient quelques savants inquiets ; il restera sur les voies de la science, comme un monstre ambigu, aux visions fulgurantes mais à la démarche arbitraire et incertaine.

M'écoutes-tu encore lecteur ? Il est temps que je te réveille en faisant silence. J'en profite pour te glisser une dernière phrase. Que le linguiste observe les mots, particulièrement la syntaxe si riche des verbes ; qu'il ne craigne pas d'accumuler les règles ; aujourd'hui, le programmeur peut sur la machine donner mouvement et vie à d'immenses formulaires qui naguère n'eussent pu que se résoudre en poussière, sous la poussière d'une réserve de bibliothèque ; qu'il dénombre aidé de programmes tous les faits : le statisticien pavera de formules sûres les voies qu'il explore...

Salut !