

## Analyse structurale et recherche computationnelle

*Cette communication a été présentée par M. A. PENNACINI au nom d'une équipe qui se consacre à la recherche sur la pensée politique ancienne à l'Institut de Philologie Classique "A. ROSTAGNI" de l'Université de Turin.*

Dans l'application des méthodes computationnelles à l'analyse structurale, le problème principal est de dépasser le niveau des relevés des structures de la phrase (mots et fonctions dont la phrase est constituée) et d'arriver aux relevés des structures spécifiques d'une oeuvre littéraire, c'est-à-dire, des éléments et des fonctions dont l'oeuvre est constituée en tant qu'oeuvre littéraire, douée de caractères littéraires spécifiques.

Le repérage des procédés d'art, aussi bien en poésie qu'en prose, comporte deux aspects qu'il convient de ne pas négliger : premièrement, recueillir les éléments de langue et de style, deuxièmement, évaluer les fonctions de ces éléments au sein du système constitué par la structure spécifique de l'oeuvre.

Il y a toutefois, à la base, un autre problème : celui qui concerne les rapports entre l'analyse qualitative et l'analyse quantitative exécutée par l'ordinateur.

Résoudre ce problème signifie repérer dans la structure de l'oeuvre et dans son style des éléments que la machine puisse reconnaître très facilement et par conséquent soumettre aux calculs.

Mais à ce moment-là nous avons encore un problème à résoudre : quels sont les critères objectifs pour la caractérisation de ces éléments ? Et comment apprécier leur objectivité ? Il s'agit, ne l'oublions pas, d'éléments qui ont été conçus en tant qu'ayant une fonction littéraire. Cela signifie qu'on parle de tous les outils de la langue littéraire et de ses règles, par exemple, de la rhétorique.

La réponse à cette question est jusqu'à présent négative. Il ne semble pas que l'ordinateur soit capable de distinguer, par exemple, l'emploi propre et l'emploi métonymique d'un mot. Que l'on considère le vers de Virgile (I, 177), cité par Quintilien (VIII, 6, 23) afin de donner un exemple de métonymie, *Cererem corruptam undis*. L'ordinateur n'est pas à même d'analyser *Cererem* comme métonymie; il pourra peut-être, s'il reçoit des instructions convenables, nous dire qu'un nom de divinité est suivi d'un verbe appartenant à un champ lexical d'actions qui ont lieu dans la réalité naturelle (ou, par un emploi métaphorique, au champ des actions morales), et suivi encore par un substantif appartenant au champ lexical des objets naturels.

L'ordinateur peut encore apprendre à dire, à propos de certains mots, s'il s'agit d'archaïsmes, de néologismes, de technicismes, de vulgarismes

(formes vulgaires). Il y aurait toutefois à résoudre le problème des références chronologique et culturelle; mais l'ordinateur peut considérer comme synchronique, par hypothèse, tout le matériel linguistique qu'il connaît et laisser au chercheur la tâche de discerner. D'ailleurs, déterminer si les mots, reconnus par l'ordinateur comme archaïsmes, etc. à l'intérieur du texte (de l'oeuvre ou du passage considéré) exercent une fonction littéraire et déceler quelle est cette fonction, c'est la tâche du chercheur, qui est le seul à pouvoir le dire, en examinant le texte.

En conclusion, l'ordinateur peut, au moyen des instructions et des connaissances qui sont maintenant dans sa mémoire, distinguer et relever les mots en tant que parties du discours, en donnant une description du texte qui en reflète la structure grammaticale (morphologique); et structure ici désigne l'ordre selon lequel se suivent les parties du discours. Et sans doute peut-on tirer de cette description des indications avantageuses (voir par exemple L. DELATTE, *Revue*, 1966/3, p. 12, à propos de la séquence préposition-verbe dans la *Consolatio ad Helviam*, qui renvoyait à la construction *ad* suivi par un adjectif verbal, construction absente dans les autres oeuvres de Sénèque dépouillées). Mais l'explication du sens (motivation et fonction) de la présence d'une séquence "c'est", comme M. Delatte l'a écrit, "la tâche du philologue ou du critique".

Par ailleurs, on pourrait peut-être ajouter d'autres instructions et connaissances à celles déjà mémorisées par l'ordinateur, de façon qu'il puisse reconnaître les mots en tant qu'appartenant à différents niveaux de langue (archaïsmes - néologismes; langue haute ou illustre - langue vulgaire; technicismes hauts - technicismes humbles); ou à différents niveaux linguistiques concernant des sphères spécifiques (politique, juridique, commerciale, militaire, médicale, religieuse, magique, agricole, etc.);

ou à des champs lexicaux (l'eau, la mère, le temps, l'espace, la nature, l'amour, la reproduction, l'homme, la femme, etc.).

L'ordinateur pourrait ainsi d'un côté indiquer — par l'extérieur — des composantes linguistiques et des thèmes (autre chose sera de découvrir et d'évaluer le poids, la fonction et la position hiérarchique des thèmes à l'intérieur de la structure de l'oeuvre; à ce propos les relevés quantitatifs fournis par l'ordinateur peuvent toutefois diriger vers une hypothèse de recherche); et d'un autre côté relever les croisements et les combinaisons des composantes linguistiques et, surtout, des thèmes, c'est-à-dire proposer au chercheur des données partiellement élaborées concernant les rapports entre plusieurs éléments de la structure de l'oeuvre.

Les croisements et les combinaisons de thèmes — interactions entre éléments — jouent un rôle d'une certaine importance non seulement en ce qui concerne les thèmes pour eux-mêmes mais surtout en ce qui concerne leurs fonctions et aussi la variabilité des fonctions elles-mêmes. La situation ou position hiérarchique des thèmes à l'intérieur de la structure de l'oeuvre peut en effet changer : un même thème peut jouer un rôle tantôt prééminent, tantôt subordonné, ou disparaître ou encore être préfiguré.

Au cours d'une analyse par ordinateur d'un poème ou d'un récit (à côté des données linguistiques et morphologiques) la machine peut, si elle a reçu des instructions convenables, recueillir les noms des personnages et les mots désignant des faits et des actions (selon les champs lexicaux que l'on a insérés en mémoire) qui apparaissent dans les passages où elle a recueilli les noms des personnages.

Dans ce domaine, la question principale — et la plus difficile à résoudre —

consiste à établir la longueur du passage, c'est-à-dire le nombre de mots avant et après le nom du personnage qui doivent être rassemblés par la machine.

A ce moment-là, la machine pourra calculer les fréquences et la distribution de ces combinaisons ou, si vous voulez, de ces concordances (personnages avec situations, faits, actions) pour chaque partie de l'oeuvre (chant, livre, chapitre ou autre chose), et y ajouter un total des données.

Lorsque le chercheur aura la quantité, la distribution, la fréquence, les croisements, la concordance des thèmes, et les données qui concernent les connections entre les personnages et les états, faits et actions, il sera sans doute à même de procéder aisément à l'examen de la structure de l'oeuvre; et je pense aussi à la structure dynamique de l'oeuvre, à propos de laquelle le grand formaliste russe Juri Tynianov a écrit des pages pleines de pénétration.

Les données ainsi préparées par les machines (ordinateur et trieuse) permettront au critique — moyennant bien entendu une lecture renouvelée et attentive du texte — d'apprécier du point de vue quantitatif et distributionnel l'évolution des fonctions des thèmes, personnages, actions; et de clarifier par conséquent sur une base, je crois, objective et quelque peu solide, l'aspect qualitatif. Je pense en conclusion qu'à l'aide de fiches et de perforations additionnelles, il est possible d'étendre l'emploi des machines à l'étude de la structure spécifique d'une oeuvre littéraire.

Mais ce n'est pas seulement vers l'aspect spécifiquement littéraire des textes que l'Institut de Philologie Classique de l'Université de Turin oriente ses études et ses recherches. Comme nombre de savants ici réunis

le savent, l'Institut de Philologie Classique dispose d'une équipe dirigée par M. Italo Lana, professeur de littérature latine, dont nous sommes membres (c'est-à-dire Mlle Flocchi, Mlle Lo Manto, M. Mar none, M. Bertelli et moi-même), travaillant depuis 1964 à l'aide du Conseil National des Recherches, dans le but de composer un lexique de la pensée politique ancienne (grecque et latine, des origines — Homère — jusqu'au cinquième siècle de notre ère).

A notre avis l'étude de la culture et de son histoire (ou si vous voulez, du système évolutif de la culture) est, d'une part, indispensable à la compréhension et à l'interprétation correcte des faits littéraires et, d'autre part, indispensable à la connaissance de la civilisation ancienne dans son ensemble. Nous sommes tous d'accord que la civilisation d'une certaine société est constituée en forme de système complexe encadrant, en les organisant et en les hiérarchisant, d'autres systèmes ou des éléments simples.

Il y a des connections (qu'il faudra étudier moyennant des centres interdisciplinaires) entre la série de la littérature et les séries sociale, économique, politique, etc. Ces connections forment les liaisons structurales du système des sociétés des hommes.

L'étude de la pensée politique ancienne, le rassemblement et l'élaboration des témoignages qui la concernent, ont donc pour but de rendre disponibles des matériaux non pas épisodiques ou anecdotiques mais au contraire — bien que le lexique ait un ordre alphabétique — pourvus de tous les moyens permettant, dans la perspective des études interdisciplinaires, une utilisation que nous voudrions satisfaisante.

Nous avons jusqu'à présent constitué un fichier concernant la pensée politique dans les poèmes homériques. Cependant, quand nous avons commencé ce travail, nous n'étions pas encore renseignés sur le système de perforation IBM (un mot par carte). Nous avons par conséquent choisi une carte qui s'appelle Carte Décimale Universelle (30 x 20 cm environ) où l'on dispose de la place pour écrire ou bien, comme nous l'avons fait, pour coller au-dessous du mot (qu'on écrit en haut) tous les passages qui le concernent. Tous les passages de l'Iliade et de l'Odyssée, que nous avons choisis en considération de leur signification politique, se trouvent maintenant recueillis sur 1.600 fiches (= 600 lemmes). Chaque fiche peut contenir environ dix passages, mais il y a aussi des mots qui ne sont représentés que par un seul passage. Les quatre bords de cette fiche sont tous préparés pour être perforés. Il y a 32 sections (classifications) : dans chacune on peut perforer des trous qui symbolisent les chiffres de zéro jusqu'à neuf. Il est par conséquent possible de combiner les trous de façon que chaque mot soit symbolisé par un chiffre et en même temps réuni à tous les mots qui lui sont apparentés par le sens, en utilisant un autre chiffre.

Nous sommes à présent en train de constituer d'après le fichier le lexique de la pensée politique dans les deux poèmes. Chaque mot sera suivi de sa ou de ses significations, de quelques renseignements sur son étymologie, d'observations à propos de son emploi et de sa position métrique, des témoignages les plus significatifs avec traduction en italien, des références de tous les autres témoignages et finalement de la bibliographie, s'il y en a.

En ce qui concerne l'emploi des techniques mécanographiques, en 1967, nous avons adopté les fiches IBM élaborées par le L.A.S.L.A. pour l'analyse de la langue grecque afin de dresser un fichier qui constitue l'étape précédant la constitution du lexique.

Compte tenu de la recherche, qui ne concerne pas l'aspect morphologique mais la signification ou l'acception politique des mots, nous avons quelque peu modifié la fiche, c'est-à-dire que nous avons supprimé les colonnes 27-55 afin que le chercheur puisse y inscrire le passage concernant le mot. Le mot lui-même, lemmatisé, sera perforé dans les colonnes 1-26; les colonnes 56-80 seront réservées à la perforation des références.

Nous avons maintenant environ 6.000 fiches IBM (un mot par fiche) tirées des oeuvres des auteurs suivants : Alcée, Alcman, Anacréon, Ananios, Archiloque, Callinos, Hipponax, Ibycus, Mimnerme, Phocylide, Sappho, Simonide, Simonide d'Amorgos, Solon, Stésichore, Théognis, Tyrtée. En même temps, comme prolongement de cette recherche, l'équipe est en train de projeter un relevé concernant la distribution géographique et chronologique des mots politiques dans la Grèce archaïque.

Université de Turin

A. PENNACINI