

Les applications de l'informatique à l'étude des textes philosophiques

André ROBINET

La première génération des procédures informatiques dans les sciences de l'homme a entraîné des applications précises dans un domaine que l'on pouvait estimer indemne des conquêtes de "l'ordinateur". Si l'on ne peut parler pour cette première génération d'application au domaine de la philosophie prospective, on doit cependant retenir que la philosophie rétrospective en a largement fait usage. Si l'histoire de la philosophie consiste à établir et à commenter les documents qui la jalonnent, la textualité de son objet ne pouvait échapper aux préoccupations lexicographiques et statistiques. Un champ considérable de l'activité philosophique rétrospectante tomberait donc sous le coup des procédures objectivantes, apportant un début de réponse à la question : est-il possible d'introduire la mesure et l'ordre en histoire de la philosophie ?

Il allait de soi que la mesure et l'ordre documentaires soient d'abord concernés. La bibliographie philosophique s'est rapidement armée des procédures de classification automatisées qui conduisaient à la constitution d'un "index philosophique", en entendant par là un relevé à entrées multiples permettant de tenir à jour la liste des productions philosophiques et d'en tirer les reliefs taxinomiques par mots-clés interposés. C'était la moindre des choses. Mais sur cette question simple les difficultés se faisaient jour. Notamment qu'est-ce qu'un mot-clé dans un champ lexical philosophique ? Comment réduire les processus mentaux aux termes du discours effectif ? Par quel système de constellations verbales connoter ce que le philosophe appelle concept ? Banalement, quelle grille édifier pour établir et

retenir les dix ou douze mots-clés qui serviront à caractériser un article de philosophie ? Qui sera le nomenclateur, ou quel sera-t-il ? L'auteur ? Un lecteur ? Un spécialiste en bibliographie automatisée ? Une procédure *ad hoc* ? Un système impersonnel de ramification ou de choix ?

Car, au vu des résultats de cette première génération, je définirai le discours philosophique comme celui dont le système de signe renvoie à un champ langagier non prononcé, ouvrant par cet écart largement ouvert un réseau très large de polysémie. Le système définitionnel des philosophes est précisément de n'en pas avoir, et de remettre sur le métier chaque définition dans chaque ouvrage, sinon à chaque partie de l'ouvrage. Nous avons largement montré comment cette instabilité définitionnelle affectait les discours philosophiques de type "geometrico demonstrata" qui émettaient la prétention de cadrer le langage à l'instar des mathématiciens. Si une suite est donnée aux premières tentatives portant sur l'établissement de lexiques armés à partir des dépouillements informatisés, alors il faudra disposer ces articles lexicographiques en éventail, de manière à délimiter les raies spectrales, plus ou moins continues, qui sillonnent le champ lexicographique de chaque terme. Le dictionnaire de Lalande, quoique fort classique et positiviste, quoique réducteur et normatif, en ressentait la nécessité en ouvrant sa nomenclature sur divers "sens". En fait, on trouve pour chaque terme philosophique d'usage banal et normal, une distribution des champs sémiologiques telle que le balayage en est continu et mène du propre à son contraire : la nature est définie par Dieu, elle l'est aussi par la matière; la nature est un système de relations, elle est aussi une position d'objet dans l'univers.

Cette polysémie est d'autant plus accentuée que le philosophe n'use que d'un tout petit cadran lexical. Certes le dépouillement des *Oeuvres complètes* de Malebranche révèle l'utilisation de 25 000 formes différentes, qui ne sont plus que 7 ou 8 000 quand on les lemmatise. Si l'on prend garde aux très hautes fréquences qui absorbent un large pan d'occurrences, on rétrécit encore le nombre des items. Si l'on dépouille les articles d'un sommaire de revue philosophique, quelconque ou thématique, on ne rencontre guère que 3 ou 400 mots pleins qui ont en charge le système signifiant sur 200 pages. Ainsi les mêmes formes doivent assumer un rôle différent sinon divergent. Plus un langage est pauvre, plus les ambiguïtés y deviennent nombreuses.

Pour pallier cette pauvreté, le philosophe ne se gêne pas d'enclencher des procédures de génération lexicale, particulièrement typiques dans certaines langues naturelles ou pour certaines oeuvres. Jointe à la précédente, cette remarque fait apparaître que le discours philosophique n'est pas à la recherche d'une désignation des "choses", mais que les mots s'y réfèrent à des "idées", en entendant par là des contenus mentaux à vocation langagière. Le philosophe n'est pas de ce monde, en ce qu'il instaure un réseau signifiant portant plus sur le système du monde que sur l'objet qui le constitue. En conséquence, ce méta-physique s'instruit dans une dénotation de l'idée et non pas de la chose. Or l'idée est la relation expressive d'un langage intérieur, pas d'un non-langage. En fait, le discours philosophique réel renvoie à ce champ langagier possible et l'interprétation n'est effectuable que par référence à la possession des clés d'un langage similaire. L'incompréhension des discours philosophiques par le non-philosophe, mais aussi l'incompréhension du monde philosophique donné envers un monde philosophique annoncé, dressent des barrières contre le nouveauté, se durcissent en objet au lieu de rester ouvertes en relations, se réifient dans la référence à l'acquis et s'excommunient comme incompréhensibles. Tant que la maturation langagière ne sera pas effective et que l'on ne se sera pas adonné au nouveau lexique, un malaise subsistera, jusqu'au jour où l'on s'aperçoit que l'on ne pouvait pas parler autrement.

Perclus de polysémie, rebouté à coups de néologismes, le lexique philosophique s'ouvre cependant sur une *praxis*, si l'on en juge par ses répercussions historiques. Sa relation aux contenus mentaux d'une société ou de l'humanité entière reste féconde, preuve que son objet n'est pas évanescent, mais correspond à la sécrétion des concepts sans lesquels les percepts sont inassimilables. Or, si l'on aborde l'analyse des discours constitués et pas seulement leur saisie documentaire, on se trouve plongé dans le trop plein de cette fusion conceptuelle qui désigne un monde par une langue naturelle qui n'est pas faite pour lui. Cet écart marque la situation tragique du philosophe face au littéraire. Ce monde d'idée n'a pas de correspondant dans la sensation ni dans le sentiment, ni dans les conceptions que nous agrafons à nos perceptions.

Si, dans l'explication philosophique, on veut sortir de l'herméneutique, il faut créer cette distanciation que procure l'examen des reliefs sémiotiques du discours étudié. L'herméneutique n'est qu'une perception prolongée : elle enchaîne son discours

sur le discours, se met en phase avec lui, au terme d'un accommodement plus ou moins sympathique et compréhensif. Elle entraîne toute la charge des ambiguïtés qui pullulent dans le discours donné. Alors que si l'on considère ce "donné" comme un "fait", on sera conduit à une double exigence. D'abord utiliser des éditions "armées", solidement ancrées dans leurs appareils critiques où se reflètent les jeux de la créativité, pour peu que cet appareil scientifique révèle les distorsions lexicales qui jouent entre les brouillons et les copies d'un auteur (Leibniz), ou entre les éditions successives de ce même auteur (Ramus, Malebranche). Ensuite, faire donner l'arsenal de nos procédures tirées de l'analyse statistique lexicale, pour meubler cet écart de la distanciation.

C'est ce qu'il convenait de rechercher par quelles procédures de relais passer des relevés d'occurrences, des tableaux de concordances, des rangs de fréquence, à l'examen des associations verbales, les syntagmes langagiers, des reliefs qui affectent ce langage en miette du discours desossé, pour leur faire ressaisir des relations, des connivences, des vicariances, des prééminences, bref un discours induit sous le discours, portant sur le matériau utilisé et sur la façon de le faire fonctionner.

A ce compte, il ne s'agit pas de "revenir au texte" comme si l'on n'avait rien fait et en s'excusant de cette rupture d'herméneutique. Il s'agit de "dépasser le texte" en le faisant supporter par cette armature graphique et logique, d'y infiltrer cet autre sens de par-dessous le sens. Ou plutôt de constituer le sens en tenant compte des conditions dans lesquelles la sémantique apparente se réfère à cette sémiotique profonde. Alors des garde-fou nous sont procurés par les termes de l'analyse objectivante : un circuit dont on ne peut sortir, un point de rebroussement indépassable, en fonction desquels l'"erklären" devient la condition constituante du "verstehen". La compréhension passe par cette explication. Le discours ramené aux "faits" qui le constituent ne peut plus prendre n'importe quel essor sémantique : il se trouve cadré entre un pas assez et un par trop, face à un bougé lexical qui traduit l'effort philosophique. Car si ce monde de la créativité philosophique n'existe pas encore, on comprend que le processus de son dévoilement passe beaucoup mieux par l'étude du fonctionnement sur soi de ce discours que par sa référence abstraite et quelconque à l'herméneute, eût-il tous les dons de l'interprétation.

Cette première génération aura vu de nombreuses applications en des champs variés. L'apport des concordances diacritiques, les grilles des dispositions pyramidales, les rapports statistiques internes aux fréquences et à leur classement, l'affinement comparatif des discours dépouillés, permettent de comptabiliser de précieux résultats. L'un des plus spectaculaires concerne les souches lexicales ventilées par années qui se poursuivent avec l'édition des *Sämtliche Schriften* de Leibniz. L'amas des manuscrits leibniziens n'est classable qu'à la condition d'user des procédures classiques : paléographie, science papier, étude des filigranes, analyse interne de contenu, repérage des possibilités de datation par référence interne, apparition de termes ou de structures typiques d'une époque. Mais de telles mailles sont beaucoup trop lâches pour que l'on puisse établir une taxinomie diachronique valable des textes. La plus large part reste inappréhensible. Ce vaste ensemble est maintenant soumis à des procédures informatiques qui mettent en relief et expriment statistiquement les résultats portant sur les souches de références. Chaque texte est traité à part et comparé à cet ensemble raffiné. Si l'on ne peut parvenir à des dispositions chronologiques précises, du moins peut-on dès maintenant ouvrir une grille des diachronies envisageables, en rétrécissant l'écart entre *l'a parte ante* et *l'a parte post*.

La seconde génération de nos procédures s'amorce au moment même où les capacités de sortie de nos documents s'affinent et s'enrichissent, nous débarrassant des codifications lourdes de ces derniers temps, dégageant des sorties esthétiquement agréables et paléographiquement exactes. L'enjeu de cette seconde génération sera de passer de la réduction lexicale à la maîtrise interlexicale, phrastique si possible, à défaut de transphrastique. Les procédures adéquates nous viendront certainement des applications en cours qui portent sur la maîtrise du discours dans son ensemble suivi, telles que les pratiquent axiomaticiens ou traducteurs axés sur les ressources informatiques. Pour l'analyse des textes, ce sera un apport considérable que de transférer l'acquis obtenu dans ces domaines à l'analyse des textes, sinon à leur constitution.