

PREMIERS RÉSULTATS DE L'« ÉQUIPE DESCARTES »

par PIERRE COSTABEL et JEAN-ROBERT ARMOGATHE

Mes chers collègues, je veux d'abord vous donner une brève information concernant l'ÉQUIPE DESCARTES.

A l'origine, cette équipe a été formée sur l'initiative de M. Armogathe, avec quelques élèves de l'Ecole Normale Supérieure, et ce n'est qu'ensuite qu'elle s'est étendue à d'autres chercheurs. Cette équipe est informelle, c'est à dire qu'elle ne rentre pas dans des structures officielles et se définit simplement comme un groupe de personnes qui travaillent ensemble. Elle n'a vécu jusqu'ici qu'avec de très modestes revenus grâce à la fidélité et au dévouement de son secrétaire et promoteur, M. Armogathe. Professeur de Mécanique Générale à l'origine, puis historien des Sciences je ne suis pas moi-même un philosophe. J'ai été engagé progressivement sans l'avoir voulu dans la réédition de la grande édition de Descartes (Adam-Tannery); je m'étais auparavant déjà beaucoup intéressé à Descartes savant et j'avais fait un certain nombre de travaux; mon attention avait été attirée sur le problème de la bibliographie matérielle, et à cause de tout cela on me demanda, en 1969, d'assumer la responsabilité de l'achèvement de la réédition entreprise en ce qui concernait les cinq tomes de la Correspondance.

Qu'il me soit permis d'ouvrir une parenthèse pour dire que cette tâche est remplie.

Le tome V, dernier volume de la Correspondance et de la réédition, va bientôt paraître. Il a donné beaucoup de peine à cause des Tables et Index.

Mais je reviens à l'EQUIPE DESCARTES dont les membres étaient pour moi de jeunes camarades, avec 30 ans d'écart entre nos promotions à l'Ecole Normale. Ils m'ont fait l'amitié de me considérer comme leur directeur. Cela veut dire tout simplement que j'ai participé à leur travail, donné quelques avis. Mais l'essentiel, c'est M. Armogathe qui va vous l'exposer.

P. COSTABEL

Après le rappel de cette présentation de l'EQUIPE DESCARTES, je voudrais présenter un état présent de notre travail collectif, informel et pluridisciplinaire, à partir des quelques documents qui vous ont été distribués hier¹.

Les textes en langue française existent actuellement sur cartes perforées et viennent d'être classés sur ruban magnétique: *Le Discours de la Méthode*, les *Météores*, le texte français des *Méditations*.

Indexation complète sur le *Discours* et les trois *Essais*, statistiques lexicales sur le *Discours*, les *Essais*, les *Méditations*; co-occurrences étudiées sur divers pôles dans le *Discours*; et texte en langue latine avec la collaboration du Centre de M. Delatte, de Liège, les *Regulae* dont l'index va bientôt paraître au LESSICO au début de 1976. Les publications prévues sont d'une part des index (l'index du *Discours* est en cours de réalisation) et d'autre part diverses études, de MM. Costabel, Marion et Cahné.

Dans le domaine des réflexions épistémologiques que je voudrais vous soumettre aujourd'hui, je dois souligner l'état inachevé de ces réflexions, car nous avons délibérément choisi

¹ On trouvera une présentation de l'Equipe Descartes, par P. COSTABEL et J.-R. ARMOGATHE, dans le *Bulletin Cartésien* III, *Archives de Philosophie*, 1974, 3, pp. 453-458.

de présenter *le travail que nous sommes en train de faire*; alors que la mise en route du projet Descartes remonte à 1967, avec des fortunes diverses (il ne s'agit pas ici de décerner les lauriers, mais de collaborer ensemble à une recherche commune), nous avons tenu délibérément à présenter la recherche en train de se faire, sans cacher les points d'interrogation, l'autocritique que nous pouvons être amenés à faire, en vous invitant à en faire autant. Le grand débat qui divise les lexicographes entre index et concordances a reçu de notre part une option qui est la publication des '*indices locorum* lemmatisés', compte-tenu de la possibilité offerte désormais par le traitement automatique, d'obtenir des concordances à la demande. Deux critères sont ici fondamentaux: la lemmatisation, c'est à dire le regroupement de toutes les formes d'un même mot, et l'exhaustivité. Il n'y pas un index automatique possible à aucun niveau, et d'exploitation philosophique, linguistique, littéraire sans lemmatisation: ceci en termes de *publication*, l'exploitation directe non lemmatisée est possible (c'est, de fait, ce que nous faisons). Celle-ci peut être automatique pour le latin sous la responsabilité d'un philologue, malheureusement elle est encore manuelle pour le traitement d'un grand texte en français, ce qui retarde la publication des index, en particulier en ne permettant pas le recours à la photo-composition.

Le CENTRE LEXICOLOGIQUE de l'E.N.S. Saint-Cloud a levé cette difficulté depuis plusieurs années déjà, en distinguant automatiquement entre formes fonctionnalisées données avec le total d'occurrences, et formes lexicalisées fournies en *index locorum*.

Deuxième point: les statistiques. L'analyse automatique doit être plus qu'une façon commode d'aligner des listes de mots. Elle doit permettre des calculs d'occurrences et de statistiques. Nous n'insistons pas sur les divers coefficients qui ont déjà été présentés à diverses reprises par l'équipe du Centre de St. Cloud; ils ont été utilisés dans le domaine philosophique par André Robinet et par nous. Leur usage reste encore expérimental et

ne pourra être décisif, à notre avis, que sur l'horizon linguistique que fournira le TRÉSOR DE LA LANGUE FRANÇAISE pour le XVII^e et XXVIII^e siècles. L'usage et l'étude des hautes fréquences semble se généraliser et c'est un des aspects immédiatement utiles de l'entreprise. Néanmoins, la courbe de distribution des occurrences est très banale et commune à tous les usages cohérents des langues modernes; elle ne saurait être significative pour un auteur ni pour un niveau de langage ou un style de discours. Notre recherche se dirige donc désormais vers les hapax legomena. L'ordinateur permet en effet de repérer aisément les hapax d'un corpus. Or le nombre d'hapax dans une oeuvre et la répartition de ces hapax nous paraissent des éléments fondamentaux pour caractériser un style d'auteur. L'étude que Pierre Costabel est en train de mener sur les hapax de la *Géométrie* et ceux des *Météores* aboutit à des conclusions provisoires intéressantes, pour l'histoire des idées et de la langue de Descartes. Là encore, la lemmatisation est importante pour faire apparaître les hapax isolés de toute constellation sémantique: « isolement » se rencontre une fois dans le texte, il est au sens strict, un hapax, mais s'il y a « isolé » adjectif, participe passé, « isoler » verbe, « isolément », il y a une *constellation sémantique* de l'*isolement*.

Néanmoins, c'est là que travaille le regroupement; il permet de faire apparaître l'hapax. Je cède la parole à M. Costabel qui va nous présenter ses recherches sur la *Géométrie* et les *Météores*.

J.-R. ARMOGATHE

C'est en effet surtout sur la *Géométrie* que j'ai travaillé, en ce qui concerne les hapax, mais aussi sur les *Météores*. Quelques mots suffisent pour caractériser nos motivations. Si nous nous sommes intéressés aux hapax, c'est d'abord parce que les singularités sont très importantes pour les mathématiciens. C'est aussi

parce que la notion de fréquence, telle qu'elle est utilisée pour les mots dont les occurrences sont multiples, présente un défaut: elle n'est pas applicable à des fractions de texte, elle n'a de valeur que par rapport à un corpus constitué. Et si l'on fractionne le texte, il n'y a pas de raison a priori pour découvrir, par là et par là seulement, des résultats significatifs. Or les perspectives sont différentes lorsqu'on considère le nombre des hapax rapporté au nombre des items. Il s'agit là d'une espèce de fonction monotone décroissante lorsqu'on passe des différents chapitres d'un corpus à l'ensemble de ces chapitres. Il y a une borne pour le rapport indiqué, on le sait à l'avance. Et il y a lieu de se demander si cette borne inférieure peut avoir une signification. Pour la *Géométrie*, la borne inférieure est de 9 pour 1000, pour les *Météores* elle est de 10 pour 1000. Je me suis demandé pourquoi il y avait une différence et en regardant de plus près, je me suis rendu compte que la différence tenait essentiellement au *Discours VII* des *Météores*: ce discours est relatif aux tempêtes, à la foudre et aux autres feux qui se produisent dans le ciel. Et ce qui est clair, c'est que le sujet oblige Descartes à faire ce qu'il n'aime guère d'ordinaire, c'est à dire faire état de ses lectures et de l'information correspondante. Bien entendu les citations sont aussi brèves que possible. Descartes mentionne des noms de Constellations, des pays, des récits populaires, etc. et il se hâte. De sorte que le nombre des hapax dans le *Discours VII*, considérable en lui-même puisqu'il atteint 90 pour 13 pages, s'explique par la nature du genre littéraire auquel l'auteur est contraint de se soumettre un instant. Lorsqu'on dépouille systématiquement, pour ne pas en tenir compte dans le dénombrement définitif, les hapax qui relèvent de la chronique d'information, on retombe exactement sur le rapport 9/1000 èmes, comme pour la *Géométrie*. J'ai donc l'impression, mais ceci sera à vérifier soigneusement sur les autres textes, que la langue scientifique de Descartes est une langue qui admet le rapport 9/1000, comme caractéristique.

Je me suis intéressé par ailleurs, à la répartition des hapax, car je me suis aperçu qu'il ne suffisait pas d'en faire le relevé; évidemment, ce dernier permet de savoir combien il y a d'hapax dans une page, et après avoir fait le dépouillement dont parlait Armogathe tout à l'heure. Mais si on se confie simplement à ce résultat numérique: nombre d'hapax par page, on n'a pas du tout une vision exacte des choses. Il faut faire ce que j'ai fait, c'est à dire encadrer en rouge les hapax, et on s'aperçoit alors qu'en réalité les hapax ont une répartition très curieuse. Il y a un certain nombre d'hapax qui sont isolés et dans ce cas relèvent très souvent d'un simple souci de style: l'auteur ne veut pas répéter et varie la forme. Mais chaque fois qu'il y a une nécessité par rapport au contenu, pour exprimer le résultat d'un cheminement difficile, il y a un amas d'hapax dans un même paragraphe. Nous sommes encore en train de rechercher quelle serait la représentation la plus adéquate de cette répartition des hapax qui évoque finalement une sorte de respiration de l'auteur dans le labeur de sa rédaction. Je dois vous dire que ces amas attirent l'attention en dehors de toute préoccupation de constellation sémantique ou autre. Ils situent les passages particulièrement importants et qu'une lecture cursive ne permet pas toujours de distinguer avec cette netteté.

J'en viens à un troisième point, celui qui concerne la nature des hapax. D'après ce que je viens de dire, il y a des variétés à cet égard. Je m'intéresse surtout, évidemment, aux hapax qui sont relatifs au contenu du propos scientifique. Il est certain que ces hapax fournissent des indications intéressantes. Par exemple, l'adjectif « linéaire » apparaît une seule fois et dans un sens tout à fait différent de celui qui est le nôtre aujourd'hui. Le « linéaire » chez Descartes est ce que Leibniz appellera le « transcendant ». Il est très important de faire cette constatation, d'une part pour se garantir à l'avance contre toute mauvaise lecture de Descartes, d'autre part pour situer les raisons qu'avait l'auteur de chercher un terme mathématique nouveau. Il avait con-

science de l'existence de courbes non algébriques échappant à l'application de la méthode qu'il proposait en géométrie et ces « lignes » l'orientaient vers le qualificatif de « linéaire » pour les problèmes soustraits au domaine de l'algèbre. Autres hapax, le mot « chiffre » et le mot « lettre ». Le mot « chiffre » intervient dans le titre, ou plutôt le sous-titre marginal d'un paragraphe, et dans le paragraphe lui-même c'est le mot « lettre » qui est employé. Ces hapax sont très significatifs. Le « chiffre » qui est en définitive une « lettre » n'est pas un « chiffre » au sens de l'arithmétique, c'est un chiffre au sens des Services secrets et l'on sait combien à l'époque, au début du XVII^e s., il y a eu abondance de l'intérêt porté aux diverses manières de crypter un texte. Le « chiffre » c'est en définitive un code, et on voit très bien l'intention de Descartes: coder la géométrie avec un certain nombre de signes, fournir les règles de leurs manipulations et apprendre à décoder pour découvrir la solution. Dans le cas que je viens de citer il est très évident que le fait même de l'hapax, mis en évidence par le traitement du texte à l'ordinateur, est ce qui a permis de porter attention à des situations chargées de signification.

Mais dans cet ordre d'idées j'ai rencontré la nécessité de ne pas omettre des hapax qui sont en réalité des groupes binaires. Par exemple: « toute simple », « égal à rien », « moindre que rien ». « Egal à rien » est extraordinaire dans la bouche du philosophe Descartes, et ramène bien à la notion de code. Car Descartes introduit cette expression à propos d'un calcul qui l'amène à faire passer tous les termes du second membre d'une équation dans le premier, c'est à dire à prendre par *commodité* la différence des deux membres de l'équation. On voit bien que l'expression « égal à rien » peut être en elle-même scandaleuse, en rigueur de termes, mais que cela n'a pas d'importance si elle n'est qu'une convention de langage pour caractériser le fait qu'il n'y a pas de différence entre les deux membres de l'équation.

Il est curieux de constater que si Descartes emploie la nota-

tion « o » pour désigner l'« égal à rien », il ne prononce jamais le mot de zéro et je crois que cela tient au fait qu'il veut bien d'une convention de langage assortie d'une notation de code, mais qu'il ne veut pas d'un mot susceptible de tromper l'utilisateur. « Zéro » n'est pas pour Descartes un nombre et il lui faut exclure toute confusion à cet égard.

Bien entendu cet exemple m'a amené à m'intéresser à d'autres absences dans le vocabulaire mathématique de Descartes.

Permettez-moi de quitter les perspectives ouvertes par l'étude des hapax pour dire quelques mots sur les groupes binaires. J'ai constaté par exemple que le mot « position » est pratiquement toujours associé à l'adjectif « donné » et ce phénomène de saturation de l'ensemble des occurrences de « position » est très révélateur. « Donné par position » est une expression courante de la mathématique du XVI^e siècle et de tous les commentaires d'Euclide. Autre exemple: « côté droit » qui désigne une quantité bien spéciale dans la théorie des coniques. La langue de Descartes est donc insérée dans la tradition et il n'est pas inutile de situer cette insertion de manière précise. Si la découverte de tels exemples est ici encore le fruit de la réflexion sur les données de l'indexation automatique, il faut reconnaître que nous n'avons pas trouvé encore de méthode pour intégrer ce genre de recherche dans l'automatisme.

Il faudrait pouvoir programmer les intersections d'occurrences plutôt que les cooccurrences proprement dites entre deux items (lexicalisés nécessairement dans l'état actuel de nos programmes de co-occurrence). Par exemple « arrêter » est pour 50% coordonné avec la négation (fonctionnalisée de manière diverse): c'est intéressant. Descartes ne « s'arrête pas à expliquer », à développer; ce sont les Anciens qui se sont arrêtés — dans plusieurs cas « à dénombrer » mais pas dans le sens que Descartes assigne au dénombrement — et ils ont eu tort. Descartes est un homme qui est en marche et c'est pourquoi d'ailleurs à la fin du Livre II de la *Géométrie*, il y a un hapax: « chemin

ouvert ». La signification pour la mentalité et l'intention profonde de l'auteur est très importante. Autre exemple: l'intersection des occurrences des *notions* « calcul » et « géométrie » (avec les mots, les adjectifs, les synonymes) montre qu'il n'y a qu'un seul cas où le « calcul » échappe pour se coordonner avec l'arithmétique. Ainsi est confirmé le caractère de code opératoire que Descartes conçoit pour la *Géométrie*.

Je m'en tiens là en espérant avoir montré comment nos recherches nous confirment dans l'intérêt des méthodes automatiques en même temps qu'elles soulèvent des questions auxquelles nous ne pensions pas tout d'abord.

P. COSTABEL