Jules Duchastel
sociologue, professeur de sociologie, UQAM

(1996)

“Les procédés informatisés et le discours des sciences sociales.”

Un document produit en version numérique par Jean-Marie Tremblay, bénévole,
professeur de sociologie retraité du Cégep de Chicoutimi
Courriel: jean-marie_tremblay@uqac.ca
Site web pédagogique : <http://www.uqac.ca/jmt-sociologue/>

Dans le cadre de: "Les classiques des sciences sociales"
Une bibliothèque numérique fondée et dirigée par Jean-Marie Tremblay,
professeur de sociologie au Cégep de Chicoutimi
Site web: <http://classiques.uqac.ca/>

Une collection développée en collaboration avec la Bibliothèque
Paul-Émile-Boulet de l'Université du Québec à Chicoutimi
Site web: <http://bibliotheque.uqac.ca/>

Politique d'utilisation de la bibliothèque des Classiques

Toute reproduction et rediffusion de nos fichiers est interdite, même avec la mention de leur provenance, sans l'autorisation formelle, écrite, du fondateur des Classiques des sciences sociales, Jean-Marie Tremblay, sociologue.

Les fichiers des Classiques des sciences sociales ne peuvent sans autorisation formelle:

- être hébergés (en fichier ou page web, en totalité ou en partie) sur un serveur autre que celui des Classiques.
- servir de base de travail à un autre fichier modifié ensuite par tout autre moyen (couleur, police, mise en page, extraits, support, etc...),

Les fichiers (.html, .doc, .pdf, .rtf, .jpg, .gif) disponibles sur le site Les Classiques des sciences sociales sont la propriété des **Classiques des sciences sociales**, un organisme à but non lucratif composé exclusivement de bénévoles.

Ils sont disponibles pour une utilisation intellectuelle et personnelle et, en aucun cas, commerciale. Toute utilisation à des fins commerciales des fichiers sur ce site est strictement interdite et toute rediffusion est également strictement interdite.

L'accès à notre travail est libre et gratuit à tous les utilisateurs. C'est notre mission.

Jean-Marie Tremblay, sociologue
Fondateur et Président-directeur général,
[LES CLASSIQUES DES SCIENCES SOCIALES](#).

Cette édition électronique a été réalisée par Jean-Marie Tremblay, bénévole, professeur de sociologie au Cégep de Chicoutimi à partir de :

Jules Duchastel
sociologue, professeur de sociologie, UQAM

"Les procédés informatisés et le discours des sciences sociales."

Un texte publié dans l'ouvrage sous la direction de Josiane Boulad-Ayoub, *Former un nouveau peuple ? Pouvoir, éducation, révolution*, chapitre XVI, pp. 313-324. Québec : Presses de l'Université Laval ; Paris : L'Harmattan, 1996, 343 pp.

M Jules Duchastel, sociologue, professeur au département de sociologie de l'UQAM, nous a accordé le 5 janvier 2005 son autorisation de diffuser électroniquement toutes ses oeuvres.

Courriel : duchastel.jules@uqam.ca

Polices de caractères utilisée :

Pour le texte: Times New Roman, 14 points.

Pour les notes de bas de page : Times New Roman, 12 points.

Édition électronique réalisée avec le traitement de textes Microsoft Word 2008 pour Macintosh.

Mise en page sur papier format : LETTRE US, 8.5'' x 11''.

Édition numérique réalisée le 10 septembre 2012 à Chicoutimi, Ville de Saguenay, Québec.

Jules Duchastel
sociologue, professeur de sociologie, UQAM

“Les procédés informatisés
et le discours des sciences sociales.”

Un texte publié dans l'ouvrage sous la direction de Josiane Boulad-Ayoub, *Former un nouveau peuple ? Pouvoir, éducation, révolution*, chapitre XVI, pp. 313-324. Québec : Presses de l'Université Laval ; Paris : L'Harmattan, 1996, 343 pp.

[313]

Jules Duchastel
sociologue, professeur de sociologie, UQAM

***“Les procédés informatisés
et le discours des sciences sociales.”¹***

Un texte publié dans l'ouvrage sous la direction de Josiane Boulad-Ayoub, *Former un nouveau peuple ? Pouvoir, éducation, révolution*, chapitre XVI, pp. 313-324. Québec : Presses de l'Université Laval ; Paris : L'Harmattan, 1996, 343 pp.

L'idée de cette communication m'est venue d'une double réflexion sur l'évolution des sciences sociales et la place grandissante que prend l'ordinateur dans notre environnement de travail. Pour ce qui est des sciences sociales, je me surprend de constater que les chercheurs font de plus en plus appel au discours comme source de savoir sur le social. Ces derniers ne donnent pas toujours une portée théorique à cette quête du sens à travers le discours. Il s'agit assez souvent d'une méthodologie conventionnelle s'intéressant au contenu de documents, d'entrevues ou de témoignages. Mais, pour plusieurs, ces objets conventionnels prennent une nouvelle dimension épistémologique. Le discours n'est plus réceptacle commode du sens à donner à des actions ou à des situations, mais devient fait social à part entière². Ce qui a attiré mon attention, ce n'est pas tant l'explosion considérable des ana-

¹ Une version transformée et approfondie de cette communication a été publiée sous forme d'article dans *Sociologie et sociétés*, vol XXV, no. 2, automne 1993. Voir Duchastel, Jules, « Discours et informatique : des objets sociologiques ? ».

² Angenot, Marc, « Présentation du Centre de recherche », *Discours social/Social Discourse : Analyse du discours et sociocritique des textes*, Vol. IV, 1 & 2, Hiver 1992.

lyses de discours que la centralité actuelle du discours dans l'interrogation sur le social.

L'autre phénomène qui a retenu mon attention est l'envahissement progressif de notre environnement par la technologie informatique. Je n'ai qu'à me rappeler l'origine de ma propre implication, au début des années quatre-vingts, pour constater une transformation totale des pratiques et des attitudes vis-à-vis l'utilisation de l'informatique. À cette époque, en dehors des cercles initiés, l'idée d'utiliser l'ordinateur à des fins de calcul symbolique soulevait les plus sérieux doutes. Pour tout dire, même l'usage du traitement de texte apparaissait comme une menace à la créativité. Douze ou treize ans plus tard, ceux qui résistent encore à l'intrusion de l'ordinateur dans leur environnement de travail font figure de dinosaures, même s'ils sont des dinosaures plutôt sympathiques.

C'est au croisement de ces deux observations que se loge mon interrogation. D'un côté, l'objet discours en gagnant un statut théorique - quelqu'il soit d'ailleurs - a poussé les chercheurs à repenser le problème méthodologique de son analyse. De l'autre, les [314] développements de la recherche informatique sur le langage et la connaissance, mais surtout la disponibilité de plus en plus grande de produits destinés, en principe, à en aider le traitement et l'analyse, ont suscité une attente quelque fois disproportionnée de la part des chercheurs. Ainsi, la nature de l'interpellation qui m'est faite depuis l'époque de mes premiers travaux en analyse du discours assistée par ordinateur a évolué considérablement ³. De l'incrédulité, au point de départ, devant la possibilité même de traiter et d'analyser le discours à l'aide de l'ordinateur, on est passé à une relative crédulité à l'égard des capacités d'analyse de l'informatique.

C'est autour de cette interrogation que je m'emploierai à discuter aujourd'hui. Je préciserai rapidement ce que j'entends par cette mise en évidence du discours dans le mouvement général de ce que Gyorgy Markus ⁴ a appelé le passage du paradigme de la production à celui du langage. Je montrerai ensuite que toute analyse du discours tend en dernière analyse à la compréhension du sens qu'il produit, mais qu'il

³ Les travaux que je poursuis avec Gilles Bourque sur l'analyse du discours politique par ordinateur remontent à 1981.

⁴ Markus, Gyorgy, *Langage et production*, Paris, Denoël/Gonthier, 1982.

ne peut y avoir compréhension sans prendre en compte à la fois la forme et le contenu. J'essayerai de définir quatre attitudes possibles face à l'ordinateur qui correspondent, pour une bonne part, aux réponses apportées à la question de la compréhension du sens. Je présenterai, par la suite, les principales opérations méthodologiques en analyse du discours en tant qu'elles peuvent être utilement effectuées par l'ordinateur. Je terminerai enfin en essayant de voir en quoi l'informatique peut nuire et contribuer à la fois à la démarche de recherche.

Du paradigme de la production au paradigme du langage

J'emprunte à Gyorgy Markus ⁵ l'idée que les théories des sciences humaines et sociales peuvent être réparties dans deux paradigmes, celui de la production et celui du langage. Ces paradigmes définissent le point d'ancrage et la perspective qui oriente l'analyse du social. Le paradigme de la production met l'accent sur les activités matérielles en tant que modalités de la production et de la reproduction des rapports sociaux. Le paradigme du langage cherche à comprendre le lien social à travers les manifestations du langage et ses formes d'institutionnalisation. Dans le premier cas, le discours, les idéologies, sont perçus davantage comme reflets, plus ou moins déformés, de la réalité dite matérielle. Dans le second, il devient la médiation première de l'action sociale, l'objet-même de l'intelligibilité. Si ces deux paradigmes continuent de coexister, il est intéressant de noter [315] que le second est en passe de devenir dominant dans les sciences humaines et sociales. Les déboires récents du marxisme ne sont certes pas étrangers à ce retournement, mais il ne faut pas sous-estimer l'importance considérable que ce paradigme occupe déjà dans les diverses traditions philosophiques, psychanalytiques, structuralistes et même dans le marxisme (je pense au marxisme hongrois, à l'école de Francfort...). La conséquence immédiate de ce retournement est l'importance croissante accordée au discours comme principe d'intelligibilité du social.

⁵ op. cit.

C'est ainsi que l'on peut témoigner d'une véritable explosion des perspectives dans ce qu'il est convenu maintenant de nommer analyse du discours. Disons, pour simplifier grandement, qu'à l'origine coexistent deux traditions, chacune trouvant son assise dans des domaines disciplinaires différents. Du côté des sciences sociales, on a toutes les formes d'analyse de contenu. Elles ont pour caractéristiques de sous-estimer le fonctionnement linguistique du discours et de ne considérer celui-ci que comme surface indicielle d'un autre niveau de réalité. Le discours permet ainsi de caractériser une réalité qui lui est extérieure, que ce soit la personnalité, l'action sociale, les rapports de classes ou une quelconque institution. Du côté des sciences du langage (à ne pas confondre avec le paradigme du même nom), émergent lentement du traumatisme Saussurien - je pense à la coupure langue/parole - des tentatives de penser les unités langagières plus larges que la phrase. Que ce soit l'analyse de l'énonciation, la pragmatique, l'analyse structurale, toutes ces perspectives cherchent à comprendre la structuration et le fonctionnement en langue de discours complexes produits dans des situations de communication relativement localisées. Cette première opposition ne permet plus vraiment de départager le champ des analyses de discours, car il n'est plus d'analyse de contenu qui ignore totalement la nature langagière du discours ni, non plus, d'analyse du discours qui ne pense la relation avec ses conditions de production. Elle permet, par contre, de montrer la double tension nécessaire à toute analyse qui ne peut renoncer ni à la forme ni au contenu du discours qu'elle entreprend d'analyser et de comprendre.

C'est pourquoi je me sens bien à l'aise aujourd'hui de parler uniquement d'analyse du discours. Cela ne signifie pas pour autant que le domaine se soit resserré. Au contraire, je répète qu'il a littéralement explosé. Il y a autant d'analyses du discours que de tentatives d'hégémoniser le champ. Je parlais à l'instant de la tradition des sciences du langage qui se déploient dans des écoles extrêmement variées ayant toutes pour objet l'analyse du discours (socio et psycholinguistique, sémantique, énonciation, pragmatique, linguistique textuelle, ...). De même, du côté des sciences sociales, on trouve plusieurs traditions d'analyse. Pour simplifier encore, à côté de l'école française d'analyse du discours qui a tenté dès l'origine de faire la synthèse entre histoire

et linguistique dans l'analyse du discours [316] politique ⁶, on trouve ce qu'on appelle, de manière un peu vague, la tradition germano-anglophone qui s'intéresse avant tout à des situations d'interlocutions localisées dans une perspective davantage intersubjective ⁷. Enfin, je passe sous silence toutes ces pratiques des sciences sociales inspirées de la phénoménologie et de la sociologie compréhensive qui cherchent dans le discours le sens des pratiques et des institutions sans toutefois définir de démarche précise pour l'analyse.

Il me semble pouvoir parler d'analyse du discours à deux conditions principalement. D'une part, l'on doit reconnaître les dimensions sociale et historique du discours ⁸ (Angenot) non seulement dans sa détermination mais en tant qu'il est lui-même production du social et constitue ainsi un plan d'intelligibilité en tant que tel. D'autre part, le discours doit être conçu comme comportant un ensemble de fonctionnements linguistiques et cognitifs. L'opposition n'est plus tellement entre analyse de contenu et analyse du discours, mais entre les diverses formes d'analyse du discours qui privilégieront des situations particulières de production et de communication et des modalités différentes de fonctionnement.

⁶ Maingueneau, Dominique, *Nouvelles tendances, en analyse du discours*, Paris, Hachette, 1987.

⁷ Vandijk, Teun A., *Handbook of Discourse Analysis*, 4 volumes, London, Academic Press, 1985.

⁸ Op. cit.

Compréhension et/ou explication du sens

Michel Pêcheux ⁹ invoquait le caractère incontournable de la langue dans toute analyse du discours. J'ai rappelé à mon tour ¹⁰ le caractère incontournable du contenu. Ces deux injonctions apparemment contradictoires me semblent se rapporter à la différence du projet des sciences du langage et de celui des sciences sociales vis-à-vis de l'intérêt respectif qu'elles portent au langage. Les premières, profondément marquées par la linguistique moderne, s'intéressent aux manifestations langagières d'abord en tant qu'elles sont des systèmes - sociaux - de signification plus ou moins complexes répondant à des règles de formation et de transformation. Les secondes, lorsqu'elles s'intéressent au langage, demeurent avant tout des sciences de l'interprétation. Elles cherchent dans le langage le sens des pratiques et des institutions. Rappeler le caractère incontournable de la langue consiste à rappeler aux sciences sociales que le discours ne peut être compris indépendamment de sa forme linguistique, au sens le plus général du terme. Réaffirmer l'inéluctabilité du contenu rappelle que tout [317] discours est avant tout une pratique signifiante. Nul ne saurait faire l'économie ni de l'une ni de l'autre dimension.

La question qui est posée à l'informatique est bien celle de la compréhension plus ou moins automatisée du sens des manifestations langagières, le plus souvent matérialisées dans des textes. Cela implique que la machine serait appelée à décrypter le sens d'un message à travers l'analyse des formes, d'abord graphiques, puis linguistiques, cognitives, pragmatiques,... L'ordinateur impose donc que l'accès au sens se fasse à travers l'étude des formes. Cela implique ensuite que ces formes puissent être analysées dans des états de complexité croissants. La question n'est pas tant de savoir si l'ordinateur peut s'acquitter des

⁹ Pêcheux, Michel, *Lire l'archive aujourd'hui*, CNRS, Projet ADELA, Paris, 1981.

¹⁰ Bourque, Gilles et Duchastel, Jules, « Analyser le discours politique duplesiste : méthode et illustration », *Cahier de recherche sociologique*, Vol 2, no 1, avril 1984.

tâches de formalisation, mais si cette formalisation appliquée à des niveaux de complexité croissants peut permettre d'atteindre la compréhension du sens produit par le langage. En d'autres mots, l'ordinateur n'introduit rien de nouveau par rapport aux questionnements de l'analyse du discours, sinon qu'il nous force à discuter d'une question cruciale qu'on éviterait peut-être autrement. Il interpelle, en quelque sorte, les chercheurs afin qu'ils réfléchissent à la portée de leur propre travail de compréhension du langage.

Des réponses apportées à la question de la compréhension du sens, découlent plusieurs attitudes vis-à-vis l'informatique. Je présenterai quatre attitudes possibles qu'il faudra comprendre comme autant de types idéaux.

La première attitude est celle du rejet. Celle-ci est invariablement motivée par l'argument de la complexité du langage qui ne saurait être formalisé dans son ensemble. Mais, cet argument de la complexité sera défini de manière diamétralement opposée selon que l'on favorise une approche herméneutique ou que l'on appartienne à la tradition générale d'analyse du discours.

On ne saurait ramener l'herméneutique à une posture épistémologique unique. J'insisterai ici davantage sur les approches compréhensives qui ont tendance à sacrifier à la critique de l'objectivisme toute procédure explicite de lecture. De leur point de vue, la complexité est le produit de la profondeur et de la multiplicité des couches significantes sédimentées dans le discours. À la limite, le sens n'est compréhensible qu'à travers une projection plus ou moins empathique. Je n'ignore évidemment pas d'autres approches herméneutiques fondées sur des procédures rigoureuses de description et d'explication. Je ne veux pas non plus sous-estimer le problème posé par le questionnement herméneutique car, comme le montre Ricoeur ¹¹, il subsistera toujours une part incontournable d'interprétation, à l'entrée comme principe de construction des critères formels de l'observation et à la sortie comme nécessité d'établir une relation au monde. Je veux [318] seulement indiquer que toute démarche interprétative qui renonce à la description ne peut avoir qu'un intérêt très limité pour l'informatique.

¹¹ Ricoeur Paul, *Du texte à l'action. Essai, d'herméneutique II*, Paris, le Seuil, 1986.

Le rejet peut être motivé également par une posture opposée qui pose l'explication au centre de la démarche d'analyse. Je pense en cela aux analyses de discours qui, à travers leurs multiples perspectives, font la démonstration d'une complexité telle de structures enchevêtrées qu'il serait utopique d'en proposer une formalisation transposable au niveau informatique ¹². Une analyse du discours qui privilégie la multiplicité des niveaux de description s'accommode mal du recours à l'informatique.

La seconde attitude est relativement paradoxale. Il s'agit de l'enthousiasme naïf pour tout produit informatique proposant de près ou de loin des analyses de données langagières. Je ne parlerais pas de cette attitude si elle n'était si répandue. Combien de projets de recherche, en effet, expriment le besoin plus ou moins motivé de logiciels d'analyse de texte ¹³. Comme je le disais au début, à une incrédulité de départ a succédé une crédulité peu rationnelle. Je fais l'hypothèse que cet engouement est le fait des chercheurs qui ont le moins pensé au problème de la formalisation de l'objet discours et des procédures pour l'étudier. En conséquence, leur recours naïf à l'informatique est le plus susceptible de créer des effets de connaissance non maîtrisés.

La troisième attitude est celle de la croyance dans la capacité à terme de l'ordinateur de venir à bout de la compréhension automatique du sens produit par le langage. Je n'ignore pas que peu nombreux sont ceux qui affichent encore leur foi en l'intelligence artificielle. Pourtant les mêmes intérêts de recherche qui ont caractérisé ce projet dès le départ se reformulent sous de nouveaux emblèmes en référence avec les disciplines qui la fondent, soit les sciences de l'information et les sciences cognitives. Dans ce cadre renouvelé, sous des formes moins naïves, n'en subsiste pas moins le projet d'une intelligence artificielle, soit le projet pour l'ordinateur de simuler de manière relativement

¹² Angenot, Marc, *La parole pamphlétaire. Contribution à la typologie des discours modernes*. Paris, Payot, 1982. Angenot, Marc, 1889. *Un état du discours social*, Montréal, Édition du Préambule, 1989.

¹³ Il n'existe pas de statistique officielle sur la demande. Je peux tout simplement témoigner de cet accroissement à partir de deux points d'observation privilégiés : d'une part, en tant que membre actif d'un Centre de recherche (ATO-CI) dont l'une des vocations est de proposer des services spécialisés en analyse de texte par ordinateur ; d'autre part, à titre de membre de comités d'évaluation de projets de recherche en sociologie.

isomorphe des processus intelligents. Évidemment, ne peuvent se loger ici que les tenants de la formalisation absolue de l'entièreté du processus de compréhension. L'interprétation, à la limite, n'est plus que procédure interne.

La quatrième attitude est celle que j'ai adoptée. Il s'agit de considérer l'informatique comme ensemble de procédures d'aide au [319] traitement et à l'analyse du discours. Je développerai ci-après les caractéristiques d'une telle approche. J'ajouterai maintenant que cette attitude me semble la plus propice pour répondre à la situation d'explosion des perspectives d'analyse du discours, Faute d'une théorie unifiée, il me semble qu'il faut pouvoir tirer profit des capacités de l'ordinateur de formaliser certaines procédures et effectuer de manière très efficace des tâches complexes et répétitives. La philosophie d'aide vise à identifier les procédures communes à toutes les démarches de description et d'analyse afin d'en permettre l'activation par les hypothèses propres à chaque chercheur. Il n'y a pas de modèles imposés, si ce n'est au niveau des opérations méthodologique de base.

Tableau des opérations méthodologiques

Je suis de ceux qui ne croient pas que l'on puisse à terme simuler la compréhension humaine dans un système parfaitement automatisé. Cette conviction ne découle pas principalement des limites intrinsèques de l'informatique, mais des difficultés mêmes que nous éprouvons dans la formalisation de ce processus dans les sciences du langage et les sciences sociales. L'informatique ne peut résoudre des problèmes théoriques que nos propres sciences n'arrivent pas à solutionner.

Même si l'on pose le problème de l'automatisation à un niveau plus restreint, celui de descriptions partielles, je crois qu'il faut de nouveau s'interroger sur leur pertinence. Toute procédure automatisée implique une théorie. Pour le sociologue, il se peut qu'un analyseur morpho-syntaxique convienne sans pour autant qu'il soit nécessaire d'en vérifier le fondement théorique. Mais il n'en est pas nécessairement de

même de tous les analyseurs disponibles sur le marché. L'automatisme n'a de sens que si ses fondements théoriques sont connus et maîtrisés.

De manière plus positive, il me semble que tout système informatique devrait être conçu de sorte qu'il ne se substitue pas à la démarche analytique du chercheur. À partir de là, il est intéressant de tirer partie des caractéristiques propres à l'informatique, soit la nécessité qu'elle pose de formaliser et la capacité qu'elle offre d'effectuer des tâches complexes et répétitives. C'est dans ce sens que il examinerait rapidement les opérations générales qu'elle peut faciliter. Je parle d'opérations générales parce qu'il existe des opérations logiques fondamentales qui sont mises en oeuvre, indépendamment de la perspective d'analyse.

Ces opérations peuvent être regroupées dans quatre catégories. Il y a d'abord des opérations de gestion des données textuelles ¹⁴. Ces [320] opérations de gestion sont les plus connues et ont donné lieu à des applications informatiques nombreuses. Elles consistent, entre autres, en la saisie de textes (lecteurs optiques et logiciels d'interprétation), en son traitement comme suites de chaînes de caractères (traitement de texte), en système de gestion de base de données textuelles avec ses fonctions d'indexation et de navigation (SGBD, Hypertextes). L'ensemble de ces systèmes permettent de gérer des unités d'information à divers niveaux. C'est ici qu'on rencontre la première opération logique de toute analyse. En effet, quelle que soit l'approche, le chercheur devra toujours identifier le jeu des unités formelles qui feront l'objet de son analyse. Les systèmes informatisés sont aptes à reconnaître et à manipuler les divers niveaux d'unités qui peuvent intéresser le chercheur : le mot et l'ensemble des segments plus larges (phrases, paragraphes, textes, collections de textes). La capacité d'identifier divers niveaux d'unités et de les retrouver est donc la condition première des autres opérations sur le texte.

La seconde série d'opérations concerne l'ensemble des dispositifs de description des unités du texte. Il n'est pas nécessaire d'insister sur

¹⁴ Dans la mesure où les systèmes informatiques s'intéressent avant tout aux textes, j'aurai tendance à substituer, à partir de maintenant, la notion de texte à celle de discours, sans pour autant sous-estimer la restriction que cela introduit.

le caractère fondamental de la catégorisation comme processus de connaissance. Toute science procède à la catégorisation d'unités d'observation. Celle-ci est préalable à la formulation de règles et à l'établissement de relations. La seconde opération de description concerne donc l'identification de structures qui définissent des relations entre les objets décrits sur la base d'un système de règles. L'ordinateur va donc faciliter l'apposition de catégories aux diverses unités du texte. Par exemple, les mots pourront recevoir leur caractérisation morpho-syntaxique ou les segments thématiques ou argumentatifs une codification appropriée. Il permettra ensuite de marquer les diverses relations entre ces objets. Par exemple, ces enchaînements entre objets pourront être notés dans des systèmes hiérarchiques (arborescences) ou inférentiels (systèmes-experts). Dans tous ces cas, le chercheur tentera de représenter les divers aspects du texte dans une structure de catégories et de relations. Il existe un grand nombre de logiciels qui sont destinés à ces opérations d'annotation et de description structurelle. Le chercheur y trouve donc un support à sa démarche de recherche, en autant toutefois qu'il pousse l'analyse jusqu'à ce niveau de formalisation.

La troisième série d'opérations renvoie à l'exploration de l'information à l'état brut ou telle qu'elle a été structurée. Elle permet de retrouver les unités, simples ou complexes, leur(s) catégorie(s) ou encore la représentation structurelle qui en a été faite. Par exemple, il est facile d'obtenir des listes de mots et leur contexte, des propositions ou phrases contenant certaines catégories, des structures thématiques, argumentatives ou actanciennes en autant qu'elles ont fait l'objet d'une [321] description préalable. Cette exploration permet l'observation de régularités ou, inversement de faits isolés mais significatifs. Elle peut être pratiquée sur plusieurs plans comparatifs, tirant partie de la capacité de l'ordinateur de consulter de larges bases de connaissances en même temps que des structures d'une grande complexité. L'ensemble de ces opérations sont des automatismes dont l'efficacité repose sur la pertinence de la représentation informatique des données et des règles procédurales capables de les explorer. Cependant, ces automatismes ne proposent pas d'analyse. Ils sont des dispositifs de traitement de l'information, idéalement paramétrables, qui exécutent des tâches trop complexes ou trop répétitives pour être accomplies manuellement.

L'analyse commence dans le quatrième groupe de procédures. Celles-ci consistent toujours à développer des automatismes de raisonnement. Elles sont, dès lors, possibles dans la seule mesure où celui-ci peut être formalisé. Je peux donner trois exemples élémentaires de ces procédures d'analyse. Le premier relève de la mathématique et permet le calcul d'indices chiffrés pouvant rendre compte du comportement de certaines unités du texte. Cette analyse est appliquée aux résultats générés par les différentes procédures d'exploration. Ce type d'analyse ne s'applique pas toujours et doit être laissé au jugement du chercheur. Il est clair cependant qu'il n'intervient qu'à la sortie des données, une fois toutes les autres procédures exécutées. Le deuxième exemple est celui des analyseurs linguistiques. Ceux-ci interviennent souvent dès le départ de l'application des procédures. On peut penser, par exemple, à des dispositifs automatiques de catégorisation morpho-syntaxique et de lemmatisation ou encore à des analyseurs syntaxiques plus complexes. Outre le fait que ces analyseurs ne peuvent, en aucun cas, se suffire à eux-mêmes, il faut être conscient de leur forte dépendance des modèles linguistiques théoriques. Dans la mesure où on tient compte de leurs limites et de leur prédétermination théorique, ils pourront être mis à contribution dans le processus plus global d'analyse du discours. Le troisième exemple concerne l'application de règles d'inférences à des faits observés. Les systèmes-experts visent à reproduire des tâches d'interprétation interne des données sur la base de l'identification de faits et l'application de règles. Dans tous ces exemples, il y a toujours une forte détermination par le modèle théorique mis en branle. La procédure ne peut donc qu'être locale. L'idéal sera donc de permettre au chercheur de les activer dans un environnement global en raison des besoins de la recherche.

Impact de l'informatique

Il ne faut pas hésiter à poser la question de l'impact de l'usage de l'informatique en analyse du discours. J'ai toujours cru à la contribution non seulement instrumentale, mais fondamentale de l'informatique à la démarche de recherche. J'essayerai d'en donner plus loin les raisons. Mais, Je ne peux passer sous silence les effets pervers d'un tel [322] recours. J'examinerai maintenant le pour et le contre en m'en

reportant aux diverses attitudes face à l'informatique décrites plus haut.

Le rejet de l'informatique peut, comme nous l'avons vu, être parfaitement motivé par différentes conceptions de la complexité des faits de discours. A un extrême, toute tentative d'objectivation est récusée au nom de la compréhension. Il est inutile alors de penser à l'informatique si ce n'est sous l'aspect de la gestion des données. À l'autre, le choix méthodique d'une perspective d'analyses multi-niveaux empêche encore de s'appuyer sur une technologie insuffisamment développée. Dans les deux cas, la question de l'impact est marginale.

L'enthousiasme naïf est certainement l'attitude la plus susceptible d'entraîner des effets pervers. Le chercheur utilise une machine dont il ne sait pas précisément ce qu'elle fait. Le traitement informatique produira des résultats d'analyse qui, faute d'une maîtrise des principes ayant présidé à leur production, prédétermineront les conclusions du chercheur.

Il faut donc aborder les attitudes résiduelles avec plus d'attention. On trouvera plus de partisans de l'automatisation des procédures du côté des sciences plus formelles, comme la linguistique ou la psychologie. Dans la mesure où ces disciplines testent des hypothèses descriptives et explicatives à travers la formulation d'algorithmes, elles sont entièrement à même de contrôler la validité et d'évaluer l'efficacité de leur démarche. Le problème commence du côté d'utilisateurs naïfs qui prendraient les résultats générés par ces algorithmes comme des faits objectifs. L'informatique favorise de plus en plus l'interconnexion entre disciplines, pas nécessairement l'interdisciplinarité. La disponibilité de logiciels effectuant certaines tâches dans une direction donnée ne devraient jamais l'emporter sur la nécessité pour l'analyste d'obtenir des descriptions allant dans une direction contraire.

La perspective qui consiste à considérer l'informatique comme un ensemble de procédures d'aide au traitement et à l'analyse de données textuelles, devrait mettre le chercheur à l'abri des effets pervers. Ce n'est malheureusement pas toujours le cas. Il faut d'abord rappeler la prévention qui vient juste d'être dite. Tout recours à des automatismes devrait toujours être évalué en connaissance de cause. Mais cela est relativement facile. Ce qui l'est moins, c'est de se protéger contre la prédétermination des procédures elles-mêmes. Le système ouvert au-

quel je me suis référé jusqu'à présent demeure encore un type idéal. Tous les systèmes informatiques comportent des contraintes. Celles-ci s'imposent en quelque sorte au chercheur qui doit les adapter à sa propre démarche. Je donnerai quelques exemples. Si le système facilite la représentation des mots plus que celle des segments, on aura tendance à retraduire la stratégie dans les limites imposées par le système. Si l'on dispose d'analyseurs statistiques à la sortie, on pourra [323] être tenté de poursuivre un raisonnement probabiliste plutôt qu'analytique. Mais, si ces contraintes existent, elles ne sont probablement pas plus fortes que celles de toute méthodologie dans laquelle un chercheur pourrait progressivement s'enfermer. Il s'agit de conserver en tout temps une vigilance critique en rapport avec les dispositifs mis en oeuvre pour supporter l'analyse et non pas pour la supplanter.

En terminant, je m'en voudrais de ne pas rappeler les avantages décisifs du recours à l'informatique. L'avantage le plus évident est celui qui provient de l'efficacité de l'ordinateur. J'ai parlé depuis le début de sa capacité d'étudier en largeur des bases importantes de données et d'explorer des structures complexes. La faculté de traiter les grands ensembles peut certes avoir l'effet d'inciter le chercheur à multiplier inutilement ses corpus. Mais, là où cela s'impose, l'exploration systématique de grands ensembles de données procure connaissance et validité. Quant à l'exploration des structures complexes, l'ordinateur parvient à produire des analyses qui seraient difficilement réalisables sans son aide.

Le second avantage de l'ordinateur vient des exigences mêmes liées à son fonctionnement. Il oblige le chercheur à formaliser ses procédures. Je pourrais prendre les exemples faciles de la linguistique et des sciences cognitives qui ont connu un développement fulgurant sur la base de leur confrontation aux exigences de l'informatique. Ma propre expérience témoigne de l'enrichissement des procédures d'analyses qui sont suscitées par la simple nécessité d'explicitation liée à l'utilisation de l'ordinateur.

L'avantage le plus décisif découle d'une combinaison des caractéristiques mentionnées jusqu'ici. L'efficacité en largeur et en profondeur, le raffinement des descriptions possibles, les possibilités d'exploration tous azimuts des données dans autant d'états qu'il en existe, la présence d'analyseurs produisant de nouveaux résultats, tout cela contribue à développer les opportunités de découvertes (on parlerait,

en anglais, de *serendipity*). À l'encontre même de ce qui se passe le plus souvent dans l'application de méthodologies rigoureuses, l'utilisation de l'ordinateur, en autant que l'on s'en tienne à une stratégie d'utilisation souple, permet de découvrir des choses inattendues.

En conclusion, je dirai que s'il est vrai que les sciences sociales penchent aujourd'hui du côté du paradigme du langage, elles peuvent aisément rencontrer la science informatique qui s'est elle aussi déplacée progressivement d'une problématique des nombres vers une problématique des lettres. L'informatique devient un stimulant pour les sciences sociales leur imposant une rigueur nouvelle, mais elle ne peut [324] offrir qu'un support à leur démarche propre. En autant que les chercheurs sauront profiter de cette dynamique, il n'y a pas de mal pour les sciences sociales.

Jules Duchastel

Département de sociologie

Université du Québec à Montréal

Fin du texte