



DUCHASTEL, J. (1992). " La sociologie et l'analyse de texte par ordinateur ". *Technologies, Idéologies, Pratiques: Sciences sociales et intelligence artificielle*, 10 (2-4), 1992: 253-264.

---

## LA SOCIOLOGIE ET L' ANALYSE DE TEXTE PAR ORDINATEUR

par Jules Duchastel, Département de sociologie, UQAM

**Introduction** Le présent colloque se donne pour objectif de réfléchir sur les rapports entre les sciences sociales et l'intelligence artificielle (ci-après IA). Bien que, comme nous le verrons plus loin, le terme "intelligence artificielle" connaisse une certaine disgrâce, il n'en demeure pas moins qu'il permet de désigner une réalité bien présente dans les transformations actuelles de nos sociétés. Il nous faut donc nous interroger sur la place où ou devrions-nous dire l'absence où de la sociologie dans les développements de cette approche, mais également sur le désintérêt du sociologue lui-même pour ces questions. Il sera beaucoup question, dans ce colloque, des connaissances et de leur exploitation dans l'entreprise. Ma perspective est indirectement reliée à cette question puisque je m'intéresse au traitement et à l'analyse de textes, lesquels constituent l'un des supports privilégiés de cette connaissance. Je ne me placerai pourtant pas du point de vue du sociologue des organisations, mais je m'interrogerai sur la meilleure stratégie informatique pour rencontrer les exigences de l'analyse sociologique de textes à travers leur mise en valeur informatique.

**L'intelligence artificielle en question** Il est possible de dire que le projet de simuler les processus intelligents de manière artificielle a été à la fois la condition d'énormes avancées scientifiques et technologiques et un leurre visant à convaincre bailleurs de fonds et utilisateurs potentiels d'investir massivement dans son développement. La critique de l'IA porte à deux niveaux. Le premier consiste à rappeler l'histoire de ses désillusions réitérées au terme de chacune des phases de son développement. Cette critique, très bien menée par Dreyfus (1979, 1984), montre comment l'IA s'est développée par étapes successives, présentées chacune comme devant résoudre les échecs cumulés jusqu'alors. Dreyfus reprend de Bar-Hillel l'idée du "sophisme du premier pas", cet enthousiasme débordant au point de départ d'une phase de développement de l'IA. On assiste, à ce moment, à des prédictions sur les résultats attendus qui sont délirants : "la simulation de l'intelligence est à portée de main...". Mais, devant des résultats décevants, on initie une nouvelle direction de recherche qui est proclamée encore plus prometteuse. Cette conviction, naïve ou intéressée, est encore aujourd'hui le fait de chercheurs de pointe, mais il existe de plus en plus de critiques venant aussi bien de l'intérieur que de l'extérieur de l'IA.

En effet, le second niveau de la critique s'intéresse au fondement de l'IA. Critique de la raison instrumentale (Weizenbaum, 1976, 1981), critique de la raison artificielle (Dreyfus, 1979, 1984), critique du modèle de la représentation (Flores et Winnograd, 1986, 1989), toutes contestent le projet de reproduire ou de simuler des processus intelligents complexes. Weizenbaum critique d'emblée la pensée scientifique rationaliste et mécaniste et remet en question la prétention de celle-ci à penser tout le réel. La connaissance humaine échappe, selon lui, au formalisme logique et l'IA, en tant qu'aboutissement ultime de cette pensée,



devrait s'interdire de se substituer à certaines fonctions humaines qui ne peuvent relever de la technique. C'est donc l'éthique qui fonde sa critique de la science et de l'IA. Dreyfus adopte un point de vue phénoménologique qu'il oppose à la position plus moralisante de Weizenbaum. S'appuyant sur le fait que l'être humain est toujours déjà en situation, il en déduit un recours nécessaire au contexte, ce dernier se manifestant dans les desseins, le corps, les besoins et les désirs. Tout cela étant muable en permanence, aucun fait n'est pourvu d'une signification stable ou universelle. Aux deux questions qu'il formule : est-ce que l'être humain traite l'information à l'aide de règles formelles? est-ce que toute conduite humaine peut faire l'objet d'une formalisation?, il doit répondre non. La faculté de comprendre renvoie à l'être humain dans son corps, inséré dans le monde matériel et formé à sa culture. Aucun système de règles n'est concevable dans ce contexte évolutif. Par ailleurs, la simulation formelle de cette complexité appellerait une régression à l'infini de systèmes de règles ó ou de contextes ó dont on peut penser qu'aucun algorithme ne puisse rendre compte.

Avec le livre de Terry Winnograd et Fernando Flores (1986, 1989), la discussion s'est précisée. Dans la foulée du développement des sciences cognitives et de l'importance grandissante accordée à la représentation et au traitement de la connaissance, c'est la notion de représentation qui est désormais questionnée. Les auteurs ne rejettent pas le point de vue rationaliste "pour la génération de règles et le traitement d'opérations au moyen de représentations symboliques" (Winnograd et Flores, 1989, p. 110) dans le cas d'applications particulières. Le modèle scientifique permet de développer des applications en réponse à des besoins, mais on ne doit pas se faire d'illusion sur sa capacité d'expliquer le sens. Ils contestent l'hypothèse, à la base des théories de la cognition, selon laquelle les objets du monde auraient des propriétés indépendantes susceptibles d'être "représentées" dans un quelconque formalisme, pour être l'objet d'un traitement subséquent. La connaissance ne résulte pas d'opérations formelles, chez l'individu, dans le but de saisir des propriétés d'un monde objectif. Elle est sociale et provient de la participation des individus à des modèles orientés d'opinions, d'actions et de croyances. La cognition peut utiliser des représentations, mais elle n'est pas fondée sur des représentations. L'informatique actuelle, essentiellement fondée sur la manipulation de symboles ó représentations ó, ne saurait se substituer à l'intelligence.

Chez tous ces auteurs, il y a donc remise en question des prétentions de l'IA. Aucun ne conteste la pertinence ou l'utilité de la formalisation de certaines tâches "intelligentes". Cependant, l'IA se voit confisquer son projet global de simuler ou de reproduire des comportements intelligents complexes. Doit-on, pour autant, se tourner exclusivement vers une perspective d'ingénierie et abandonner le terrain de la modélisation? En effet, certaines nouvelles perspectives ne se dessinent-elles pas avec la venue du néo-connexionnisme, du traitement parallèle massivement distribué et de théories comme celle de l'énaction (Varela, 1988, 1989)? Évitant de succomber au sophisme du premier pas, il me semble que l'avenir de l'IA se joue dans cette tension entre redéfinition perpétuelle de nouveaux modèles et développement d'outils et de techniques de plus en plus évolués.

***Méconnaissance du social ou absence du sociologue*** La littérature en IA comporte très peu d'allusion au social, si ce n'est à travers l'évocation de contextes plus ou moins élargis ou, encore, de la métaphore du globe terrestre. Cette métaphore permet de schématiser le contexte global ó ou social ó sous la forme d'une sphère dont on présume qu'elle évoque la réalité du monde. La sociologie n'est que très rarement mentionnée dans l'énumération des disciplines impliquées en IA. Les auteurs cités plus haut, tout en rappelant inlassablement la dimension



sociale de la connaissance, s'appuient avant tout sur une phénoménologie du sens commun ou une théorie des actes de paroles. Varela (1989), quant à lui, représente une exception quand il élargit le cadre de référence à Foucault dont les travaux ont eu beaucoup d'influence en sociologie. Mais celle-ci, en tant que discipline vouée à la théorisation et à l'analyse du social, est absente du débat.

Il y a deux façons de concevoir l'interaction entre l'IA et la sociologie, du dehors ou du dedans. Le sociologue a privilégié la première, c'est-à-dire l'étude des impacts ou des effets du développement des nouvelles technologies informatiques, principalement dans les situations de travail. Cette approche est éminemment nécessaire, ne serait-ce que du fait de l'idéologie à l'oeuvre derrière la redéfinition des rapports sociaux induite par ces transformations. Une seconde approche serait pourtant tout aussi nécessaire. Il s'agit de la place que pourrait prendre la sociologie dans la réflexion sur la nature du langage et de la connaissance, objets maintenant privilégiés de l'IA. C'est en ce second sens que je dis que le sociologue est absent.

Cette absence se manifeste particulièrement dans le domaine de l'IA qui s'occupe du traitement informatique des langues naturelles (TLN). L'intérêt du sociologue pour le traitement des langues naturelles devrait découler de la nature sociale du langage ó situation sociale de la communication ou du discours, nature sociale du communiqué ó. Mais avant de revenir sur la contribution potentielle du sociologue, voyons d'abord comment le TLN s'est heurté au problème de la complexité du langage. L'évolution de ce domaine, à travers ses phases successives de développement, prouve bien que le traitement informatique des langues naturelles a dû reconnaître progressivement la complexité des aspects du langage. A chacune des étapes de développement, on a reconnu les limites de couverture des modèles élaborés et réévalué les difficultés liées à la compréhension automatique de la langue. En plus d'élargir la perspective au plan strictement linguistique, on a introduit de nouveaux modèles de représentation de la connaissance, tenté d'élargir la contextualisation des phénomènes de discours en incorporant des données extra-linguistiques et développé de nouvelles approches en logique (logique naturelle, logique floue,...).

La question demeure de savoir comment l'IA peut gérer la complexité croissante associée à l'évolution des modèles de traitement de la langue naturelle. Si l'on admet que toute démarche scientifique est avant tout une manière de réduire cette complexité afin de rendre possible la description, la compréhension ou l'explication de certains aspects de la réalité, il faut se demander s'il est alors possible de recomposer cette complexité? La démarche retenue en IA laisse croire que cette recombinaison peut s'effectuer par addition ou adjonction de ces divers niveaux aspectuels. Par exemple, Sabah (1988) reprend les divisions de la linguistique comme autant de niveaux ó morphologique, syntaxique, sémantique et pragmatique ó permettant la compréhension de la langue. Cette façon de voir implique, d'une part, que chacun des niveaux a une certaine autonomie et, d'autre part, que l'ensemble doit être, à un moment donné, recomposé pour reproduire la complexité du langage. Le postulat d'autonomie des niveaux entraîne une stratégie de recherche étagée visant la résolution successive des problèmes qui leur correspondent. Tout en reconnaissant l'efficacité de cette approche dans l'accomplissement de tâches spécifiques, la critique de l'IA la considère comme globalement erronée du point de vue d'une éventuelle compréhension du sens par une machine. En effet, elle ne peut permettre de résoudre le problème de la recombinaison de la complexité du sens.

Il est donc nécessaire de considérer les représentations correspondant à ces divers niveaux comme autant de constructions qui doivent être distinguées de celle que l'on suppose être une représentation "naturelle" que l'on voudrait simuler. Aucune théorie globale de cette



représentation naturelle n'est présentement disponible et aucun formalisme ne peut, à terme prévisible, en rendre compte, si toutefois elle existe comme système de traitement symbolique. Il nous faut donc tirer la conséquence qui s'impose, soit de concevoir l'informatique comme plateforme de modélisation partielle et de développement d'outils spécialisés pour la manipulation d'objets construits.

C'est dans ce contexte global que le sociologue doit s'interroger sur sa contribution potentielle à l'IA. Deux questions se posent alors à lui. Du côté de la modélisation, la sociologie peut-elle fournir un cadre formel permettant de programmer les aspects sociaux de la connaissance et du langage susceptible de s'ajouter aux autres niveaux de description ou de représentation? Dans une perspective davantage orientée vers les outils, la sociologie peut-elle proposer des règles de représentation, d'inférence et d'interprétation pour tirer profit dès maintenant de certaines technologies informatiques?

C'est dans cette seconde perspective que J.C. Gardin (1987) conçoit l'usage de la technologie informatique dans les sciences humaines. Se situant dans une perspective rationaliste, Gardin vise à disposer d'outils mettant à la fois à l'épreuve des représentations construites par les diverses disciplines des sciences humaines et permettant d'accroître le potentiel de description et d'explication de celles-ci. Il se situe dans le cadre restreint des systèmes experts dont il veut exploiter les ressources pour "parvenir à une meilleure maîtrise des raisonnements de sciences humaines, avec ou sans l'IA" (Gardin, 1988. page 250).

La réflexion de Gardin sur l'application de la technologie des systèmes experts au raisonnement archéologique nous permet peut-être de comprendre pourquoi la sociologie est absente du développement en IA. Certes, elle peut offrir un cadre général pour comprendre les aspects sociaux de la connaissance ó sous la figure des idéologies, de la culture ou des savoirs ó, mais elle ne dispose pas de ces représentations formellement construites qui puissent être traduites en symboles et traitées par des règles de production. Contrairement aux disciplines linguistique, psychologique et même archéologique, la sociologie ne dispose pas d'un symbolisme de représentation standardisé des objets qu'elle construit. Elle est, de plus, souvent silencieuse sur les règles de son raisonnement : comment établit-elle des inférences, comment procède-t-elle à l'interprétation? Cela explique peut-être sa timidité devant le discours envahissant de l'IA. A sa défense la sociologie peut répondre qu'elle est d'emblée dans la complexité et qu'il lui est par définition difficile de segmenter le social en autant de niveaux qui se prêteraient plus aisément à des représentations formelles. Si pour cette raison, la sociologie ne peut répondre, partiellement ou entièrement, aux exigences de formalisation de l'IA, elle devrait, tout de même, réfléchir sur la valeur heuristique du recours à la technologie informatique. Cette dernière n'oblige-t-elle pas le sociologue à mieux définir les objets qu'il construit et les relations qu'il suppose entre eux?

C'est donc du point de vue du sociologue qu'il m'intéresse d'appliquer la technologie informatique à l'analyse du discours en tant qu'il prend la forme de textes. Poser le discours comme objet d'analyse est déjà interpellé l'intelligence artificielle, en rappelant qu'au-delà de la langue comme système de codes, de la cognition comme activité de production et d'acquisition de connaissances, le discours déploie ses propres règles, non seulement au niveau des actes de paroles dans des situations de communication restreintes, mais dans la profondeur du social et de l'histoire. Cette interpellation est un rappel constant à la modestie qui devrait prévaloir dans le projet de simuler des processus de compréhension du sens.



***Système d'analyse de contenu assistée par ordinateur*** Je présenterai dans les lignes qui suivent les principales caractéristiques d'un système voué à l'analyse de textes par ordinateur. Cette initiative, sans se situer de plain pied dans la recherche en IA, contribue au développement du domaine du TLN. Le projet d'Analyse de contenu assistée par ordinateur (SACAO) vise l'intégration systématique de procédures, existantes ou à développer, de lecture assistée de données textuelles. Il offre à des utilisateurs, dans un environnement logiciel relativement intégré et convivial, le soin de paramétrer ces procédures en fonction de leurs propres hypothèses de lecture. Ces procédures ne comportent qu'un minimum de préconstruction théorique et facilitent un maximum d'itérativité entre leur application et l'analyse du texte. Cet environnement répond ainsi aux besoins différents de diverses catégories d'utilisateurs confrontés aux problèmes d'analyse de données textuelles.

Ce projet subventionné a vu le jour dans le cadre de travaux de recherche et de développement d'environnements logiciels au Centre d'Analyse de Textes par Ordinateur. Il a été développé par une équipe de chercheurs dont la formation disciplinaire et les domaines de spécialisation sont hétérogènes, mais dont les intérêts convergeaient vers l'analyse de données textuelles. Le but principal était de développer des outils fonctionnels pouvant assister, dans leurs analyses de données textuelles, les chercheurs en sciences humaines possédant chacun des objectifs particuliers. La stratégie retenue fut donc critique, empirique et orientée vers la réalisation technologique d'une approche méthodologique. L'aspect critique de la démarche s'est manifesté d'une double manière : nous avons renoncé aux systèmes clé-en-main et nous avons privilégié l'examen critique des procédures de traitement déjà disponibles. Nous avons donc mis davantage l'accent sur une approche méthodologique réfléchie, mettant en oeuvre des outils technologiques dans une démarche d'analyse qui associait le lecteur. Refusant donc une stratégie unique pour l'analyse, nous avons orienté notre travail vers l'optimisation des conditions pour une application des outils dans le cadre d'une démarche d'analyse globale.

L'approche s'intéresse donc au développement d'outils ou d'applications qui permettent d'ores et déjà d'accroître la capacité de lecture de plusieurs manières : accès rapide et systématique au contenu de grands ensembles textuels, rigueur et régularité de la lecture, production d'informations nouvelles par rapport aux formes traditionnelles de la lecture, introduction de la mesure et de procédures de validation, etc. Cela ne veut pas dire que toute recherche fondamentale est pour autant mise à l'écart. Ce qui est rejeté, c'est l'hypothèse forte de l'IA selon laquelle les systèmes intelligents devraient avoir une autonomie décisionnelle complète et donc proposer des applications automatiques. Comme je l'ai dit plus haut, cette prétention de l'IA s'est beaucoup atténuée dans les faits, par exemple, en réduisant le contexte à des micro-mondes, mais elle en demeure un objectif consubstantiel. Pour SACAO, l'automatisation n'est recherchée que sur une base pragmatique et ne constitue pas une condition première. Le projet met de l'avant une approche hybride, alliant procédures automatiques et assistées, et substitue l'idée d'intégration maximale des outils à l'objectif de complétude et d'automatisme des systèmes.

J'ai parlé de l'itérativité entre l'application de procédures et l'analyse de texte. Cette philosophie s'appuie sur des considérations épistémologiques, quant à la connaissance des objets langagiers, et entraîne une conséquence au niveau de l'ergonomie du système. SACAO conçoit la connaissance des phénomènes langagiers comme le produit d'un processus non-univoque de construction des objets. Cela implique d'abord la coexistence de plusieurs procès de construction complémentaires (multiplication des niveaux d'analyse) et potentiellement contradictoires (coexistence d'approches non exclusivement compatibles), ensuite la nécessité



d'une démarche d'aller-retour entre la constitution des modèles et leur validation empirique. Cette démarche favorise la méthode inductive. Evitant la projection de modèles théoriques préconstruits sur le réel, nous favorisons, au contraire, l'ajout de descriptions successives du texte en alternance avec l'exploration de résultats provisoires. Enfin, la démarche impose le développement d'une ergonomie interactive du système. L'utilisateur doit être en plein contrôle du déroulement des procédures et pouvoir les modifier en fonction des hypothèses de description provenant de son modèle ou surgissant au fil de l'analyse. Cela nous éloigne évidemment de l'automatisme. Si, par contre, des procédures automatiques sont également appliquées, les instructions qu'elles produisent comme les résultats qu'elles génèrent doivent être transparents et leur mise en oeuvre être sous contrôle du chercheur.

SACAO retient une perspective générale d'analyse. Sans renoncer à la description linguistique, il ne lui accorde pas de priorité. Comme je l'ai rappelé plus haut, la définition de niveaux étagés de la réalité linguistique implique une démarche plus ou moins étagée qui donnerait successivement priorité à la résolution des problèmes liés à chacun de ces niveaux. SACAO considère les divers niveaux de description comme la résultante d'un découpage et d'une construction différentiels de cet objet, et non comme les étapes ordonnées d'un parcours obligé qui mènerait de la description lexico-syntaxique, en passant par la description sémantique et pragmatique, à la compréhension globale de la langue naturelle. Nous avons plutôt retenu une approche privilégiant la morphologie du discours. Les descriptions linguistiques du texte serviront de support à l'analyse d'un système sémiotique, par ailleurs, beaucoup plus complexe. Nous faisons l'hypothèse que le texte est un espace diversement structuré, qui se déploie selon un processus de séquentialisations multiples (structures narrative, argumentative, thématique,...) et dans lequel des objets se schématisent pour former des noyaux de sens. Nous voulons donc repérer les modes de segmentation qui caractérisent l'organisation d'un texte et les condensations de sens qui se produisent en certains lieux privilégiés.

Nous avons divisé SACAO en quatre modules qui correspondent peu ou prou aux diverses opérations qui sont produites sur les données textuelles dans le cadre de leur manipulation informatique. Le premier concerne la gestion des données. Essentiellement, il s'agit d'assurer, d'une part, la conservation et l'accessibilité à des banques de données textuelles et de fournir, d'autre part, des utilitaires pour leur exploitation. La technologie hypertexte sera d'un grand secours dans ce domaine et le projet SACAO tirera profit de ces développements.

Le second module renvoie aux différentes descriptions de textes qui produisent les données d'analyse. Toute investigation suppose une intervention technique sur les unités à analyser. La notion de "donnée" implique nécessairement un processus de construction des unités d'analyse et, par là même, une intervention de re-structuration qui transforme les unités d'information en données d'analyse. Le module de description fournit l'occasion de procéder à la structuration initiale des données. Trois niveaux de description sont prévus : les niveaux lexical, morphologique et syntagmatique. Ces niveaux sont autonomes, mais peuvent être conjugués de manière différente en regard des besoins spécifiques d'une problématique de recherche ou d'analyse. Dans le premier cas, il s'agit de structurer le vocabulaire d'un texte à partir de dictionnaires ou de thésaurus. Des catégories peuvent ainsi être accolées à des mots en fonction de divers systèmes de description. Au lexique des mots, il est possible d'ajouter celui des expressions complexes (locutions prépositionnelles, adverbiales, usuelles, propres à un locuteur ou à une famille de locuteurs, techniques, institutionnelles, onomastiques, etc). Au niveau morphologique, le module recourt à LCMF ou à BDL afin d'identifier les dimensions



grammaticales (morphèmes lexicaux et grammaticaux). Au niveau syntagmatique, il est possible de faire usage d'analyseurs syntaxiques du français. ALSF construit les structures syntagmatiques projetées par les catégories majeures du français : les noms, les verbes, les adjectifs et les prépositions. Il construit les relations qu'entretiennent, entre elles, ces catégories dans des unités séquentielles. Un module d'ALSF, servant à construire le groupe nominal, permet l'analyse préalable à l'identification des synapsies (expressions complexes) par Termino. Également, dans le cadre d'une approche de grammaire fonctionnelle, des algorithmes permettent une description raffinée des structures thématique et de la détermination présentes dans un texte (Duchastel, Paquin et Beauchemin, 1991 a et b). Dans le premier cas, le système identifie le thème de chaque proposition (Halliday, 1985), détermine le caractère marqué ou non de ces thèmes et les hiérarchise en fonction de leur situation dans la structure propositionnelle des phrases. Dans le second cas, l'algorithme distingue les relations de détermination contiguës et par expansion permet d'associer, au niveau de l'exploration, des structures de détermination différentes en fonction des classes sémantiques d'appartenance des éléments déterminés.

Le module d'exploration permet un travail complémentaire à celui effectué par les unités de traitement du module de description. En effet, les données construites doivent pouvoir être manipulées à partir d'opérations de sélection, de regroupement et de classification. Ce module s'appuie largement sur les fonctionnalités du logiciel SATO. Pour les unités structurées linéairement (séquences lexicales), celui-ci permet d'obtenir : des lexiques fréquentiels d'occurrences ou de cooccurrences, des concordances basées sur la recherche de mots-clés ou sur des étiquettes symboliques ou numériques associées à ces mots. Il est possible de déterminer la forme et le nombre des chaînes de caractères comme paramètres d'extraction. L'ensemble de ces manipulations peut être obtenu sur des sous-ensembles des textes, partitionnés selon divers principes établis par les chercheurs. Cette partition des textes peut répondre aux exigences des traitements statistiques (échantillonnage) ou à la volonté de tester des hypothèses sur une portion d'un texte (principe de la maquette). Elle sert également à la comparaison systématique d'ensembles construits à partir d'un ou de plusieurs critères. Enfin, dans le cas des unités structurées à partir d'une description arborescente quelconque, des patrons de fouilles, correspondant aux besoins du chercheur, sont développés ad hoc.

Le dernier module regroupe les diverses analyses produites sur les données extraites. Il faut souligner que le concept d'analyse retenu ici n'épuise pas les possibilités de cette activité. Ainsi, certaines opérations classées dans le module description sont, sans contredit, des activités d'analyse. Ainsi, l'analyse syntaxique produite par un parser est bien de l'analyse et représente, pour le linguiste, un point d'aboutissement. Du point de vue que nous adoptons, celui de l'analyse de texte, elle correspond à une phase préliminaire de structuration des données. Ce qui nous intéresse, c'est l'analyse de certaines caractéristiques linguistiques identifiées dans ce processus de description préalable. Les analyses produites par ce module portent donc essentiellement sur des lexiques, plus ou moins qualifiés par les descriptions mentionnées plus haut, de mots ou de catégories. Ces données se présentant sous forme de matrice, il est loisible de recourir à tout système d'analyse statistique disponible (SPSS, SAS, LOTUS, SPAD\_N). SATO permet l'interface entre ces divers systèmes. Par ailleurs, il offre quelques analyses statistiques standards (moyenne, écart-type, variance, fréquence minimum et maximum, score z et distribution procentuelle des classes de fréquences et d'occurrences). Il offre également une mesure de distance inter-textuelle permettant d'identifier les éléments contribuant davantage à la distance entre deux textes. Je mentionnerai, en terminant, les nouvelles possibilités offertes par la technologie des systèmes experts, tel qu'elle est



implantée dans ACTE. Il est possible de modéliser le processus de la lecture experte, selon le genre de lecteur, et programmer dans cet environnement des règles d'inférence et d'interprétation.

En conclusion, il est possible d'affirmer que le projet SACAO permet au sociologue, comme aux autres chercheurs en sciences humaines, de poser ses propres exigences dans les stratégies d'analyse de la connaissance et des phénomènes langagiers. La primauté donnée au contrôle des diverses étapes du traitement des données textuelles, facilitée par un environnement qui donne priorité à l'interaction avec le chercheur, assure que le point de vue de la sociologie ou de disciplines soeurs ne soit pas subordonné à celui du linguiste ou du cogniticien. C'est avec une attitude de prudence que le sociologue peut alors s'associer, du dedans, à l'expérience de la simulation partielle d'opérations intelligentes.

### Bibliographie

Bourque, G., Duchastel, J., *"Restons traditionnels et progressifs". Pour une nouvelle analyse du discours politique, le cas du régime Duplessis au Québec*, Montréal, Boréal, 1988, 399 pages.

Coulon, D., Kayser, D., "Informatique et langage naturel : Présentation générale des méthodes d'interprétation des textes écrits", in *Techniques et Science informatiques*, Vol. 5., no 2, 1986, pp. 103-129

Duchastel, J., Dupuy, L., Daoust, F., "Système d'analyse de contenu assistée par ordinateur (SACAO)" in *La description des langues naturelles en vue d'applications linguistiques*, Recherche en linguistique appliquée à l'informatique, Centre international de recherche sur le bilinguisme, Université Laval, Québec, 1989, pp. 197-210.

Duchastel, J., Paquin, L.C., Beauchemin, J., "Automated Syntactic Text Description Enhancement : Thematic Structure Analysis".in *Computers and the Humanities*, no. 26.1, 1991.

Duchastel, J., Paquin, L.C., Beauchemin, "Automated Syntactic Text Description Enhancement : Determination Analysis", in *The New Medium, Research in Humanities Computing*, Oxford University Press, sous presse.

Dreyfus, H.L., *Intelligence artificielle, mythes et limites*, Paris, Flammarion, 1984, 443 pages. Version originale, deuxième édition, 1979.

Gardin, J.C. et al., *Systèmes experts et sciences de l'homme*, Paris, éditions Eyrolles, 1987, 269 pages.

Halliday, M.A.K., *An Introduction to Functional Grammar*, London, Edward Arnold, 1985, 1986, 1987, 387 pages.

Sabah, G., *L'intelligence artificielle et le langage, représentations des connaissances*, Paris, Hermès, 1988, Vol. I, 352 pages.

Varela, F.J., *Connaître, les sciences cognitives, tendances et perspectives*, Paris, Seuil, 1989, 126 pages.



Chaire de Recherche du Canada en Mondialisation, Citoyenneté et Démocratie  
<http://www.chaire-mcd.ca/>

Weizenbaum, J., *Puissance de l'ordinateur et raison de l'homme, du jugement au calcul*, Paris, Éditions de l'informatique, 1981, 195 pages. Édition originale, 1976.

Winnograd, T., Flores, F., *L'intelligence artificielle en question*, Paris, PUF, 1989, 295 pages. Édition originale, 1986.