

SATO : un philologue électronique
SATO: An Electronic Philologist
SATO: Un filólogo electrónico

Jean-Guy Meunier

Volume 38, numéro 2, avril-juin 1992

Analyse et gestion de l'information textuelle

URI : <https://id.erudit.org/iderudit/1028610ar>

DOI : <https://doi.org/10.7202/1028610ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

[Découvrir la revue](#)

Citer cet article

Meunier, J.-G. (1992). SATO : un philologue électronique. *Documentation et bibliothèques*, 38(2), 65–69. <https://doi.org/10.7202/1028610ar>

Résumé de l'article

Le système d'analyse de textes par ordinateur SATO est un logiciel fondé sur des théories spécifiques de l'information où celle-ci est conçue d'un point de vue sémiotique, c'est-à-dire comme un système qui traite des données symboliques interprétables. En conséquence SATO ne peut être vu comme un système expert effectuant un traitement automatique. Il opère plus comme un adjuvant au travail de lecture et d'interprétation par un être humain. En ce sens, il est vu comme une aide dans la lecture philologique des textes.

SATO : un philologue électronique

Jean-Guy Meunier

Professeur au Département de philosophie
Directeur du Centre d'ATO
Université du Québec à Montréal

Le système d'analyse de textes par ordinateur SATO est un logiciel fondé sur des théories spécifiques de l'information où celle-ci est conçue d'un point de vue sémiotique, c'est-à-dire comme un système qui traite des données symboliques interprétables. En conséquence SATO ne peut être vu comme un système expert effectuant un traitement automatique. Il opère plus comme un adjuvant au travail de lecture et d'interprétation par un être humain. En ce sens, il est vu comme une aide dans la lecture philologique des textes.

SATO: An Electronic Philologist

The Système d'analyse de textes par ordinateur (SATO) is a software based on the specific theories of information and conceived from a semiotic point of view; that is, a system that processes signs. Consequently, SATO cannot be seen as an expert system with automatic processing. It assists, instead, the human being in reading and interpretation. As such, it is an aid to the philologic reading of texts.

SATO: Un filólogo electrónico

El sistema de análisis de textos por computadora, SATO, es un programa basado en teorías específicas de información y está diseñado desde un punto de vista semiótico, es decir como un sistema que trata datos simbólicos interpretables. Por lo tanto, SATO no puede considerarse como un sistema especializado en un tratamiento automático de datos, ya que opera más bien como un colaborador en el trabajo de lectura e interpretación humanas. Es en este sentido que se lo considera una ayuda a la lectura filológica de textos.

Un des résultats inattendus de l'introduction de l'ordinateur dans l'environnement de travail tant des individus que des institutions est la prolifération de l'information de type textuel. En effet, l'accès de plus en plus facile aux micro-ordinateurs a passablement modifié la saisie, l'archivage, la conservation, le rappel de la documentation textuelle. On voit maintenant le chercheur en littérature travailler sur un corpus transcrit directement sur support électronique ou encore le contentieux d'une entreprise saisir ses documents sur lecteur optique.

Très rapidement cependant, la masse de cette documentation augmente tellement qu'elle en vient à prendre l'allure d'une immense base de données textuelles. Elle prend alors vite la forme d'une surcharge informationnelle. Tant sur le plan du rappel que sur celui de la précision, les manipulations de ces bases de données textuelles génèrent des résultats qui, tel l'apprenti sorcier, en viennent presque à noyer le requéreur.

Aussi réclame-t-on des systèmes de traitement de l'information textuelle où celle-ci est désormais vue, non comme des fiches indexées qu'il faut repérer mais comme un objet textuel qu'il faut lire, fouiller, analyser et dont on veut extraire des connaissances, des synthèses, etc. Bref, on réclame des systèmes de traitement d'information textuelle qui répondent à la nature particulière qu'est un texte et qui surtout épousent au plus près le cheminement de lecture et d'extraction d'information que peut offrir un texte.

C'est dans cette perspective que nous situons le système SATO. Ce système qu'une équipe de recherche de l'Université du Québec à Montréal (UQAM) a développé depuis déjà une vingtaine d'années est maintenant utilisé par de nombreux chercheurs dans le domaine des sciences humaines mais aussi par de nombreux professionnels des organismes publics et privés qui ont à gérer et analyser de grandes bases de données textuelles. Ce logiciel, contrairement à bien d'autres qui s'intéressent au texte,

ne fait pas partie de ceux qui se limitent à le gérer ou à le traiter sous forme de documents indexés. Il constitue plutôt un système informatique d'assistance à la lecture et à l'analyse du texte. Et il répond en cela à des paramètres théoriques influencés par la théorie sémantique de l'information, par la théorie sémiotique du texte ainsi que par la pratique philologique.

L'information textuelle

Un design adéquat d'un système d'analyse de textes doit commencer par considérer un texte du point de vue des sciences de l'information telles qu'elles se renouvellent aujourd'hui. À leur origine, surtout avec C.E. Shannon¹, la théorie dite de l'information s'était définie au sein du paradigme d'une théorie physicaliste et signalétique de type probabiliste. Elle touchait

1. C.E. Shannon, « A Mathematical Theory of Communication », *Bell System Technical Journal*, vol.27 (1948), 379-423, 623-656.

essentiellement les phénomènes de production, transmission, décodage et compression d'un signal dit informatif.

The subject study of information science is empirical phenomena associated with various information processes such as information generation, transmission, transformation, compression storage and retrieval. The ultimate purpose is to gain a better understanding of the nature of information².

Cette perspective physique a permis le développement remarquable des technologies matérielles de l'information. Mais comme l'a montré la littérature technique sur cette question³, ce concept physicaliste de l'information n'a pas réussi à expliquer la véritable nature de l'information.

Information Theory has made little impact on information science which ought to be its main and natural domain of application⁴.

De fait, cette approche, comme l'a souligné régulièrement la critique, avait éludé de l'information sa caractéristique essentielle, à savoir la signification et la connaissance qu'elle véhicule toujours.

Pourtant, quand on regarde de plus près l'usage tant technique qu'ordinaire de ce concept d'information, on constate qu'il repose sur une théorie cognitive implicite. En effet, toutes les théories contemporaines de l'information en appellent ultimement à une théorie du processus cognitif. L'information, dira-t-on, est toujours une réalité plus ou moins *attendue, reconnue, mémorisée*, quelque chose qui sert à la *décision*, à l'*action*, etc. Il n'est pas d'information sans un agent cognitif qui la traite. C'est pour ces raisons que l'on aurait distingué :

- l'information du point de vue de la probabilité (incertitude) du signal (théorie de Shannon) ;
- l'information du point de vue de la décision à rendre (théorie de la décision) ;
- l'information du point de vue du contenu (théorie sémantique et sémiotique) ;
- l'information retenue (archivage, documentation) ;

- l'information transmise et communiquée (théories sociales, situationnelles et écologiques de l'information).

À l'inverse, on ne sera pas surpris de voir que de nombreuses théories cognitives intègrent naturellement ce concept. Information et connaissances sont deux concepts qui se répondent.

« Cognitive science » is the analysis of the human mind in terms of information process⁵.

Artificial Intelligence has also embraced the larger scientific goal of constructing an information processing theory of intelligence⁶.

Or, l'une des thèses dominantes que soutiendront les théories cognitives contemporaines est que le processus cognitif en est un d'ordre sémiotique. Tout traitement cognitif, soutiendra-t-on régulièrement, est de type sémiotique. Il manipule, c'est-à-dire traite des symboles.

At the root of intelligence are symbols, with their denotative power and their susceptibility to manipulation. And symbols can be manufactured of almost anything that can be arranged and patterned and combined. Intelligence is mind implemented by a patternable kind of matter⁷.

La conséquence d'une telle perspective sur les sciences de l'information est évidente : l'information n'est pas qu'un signal sans signification. Et N.J. Belkin⁸ l'avait souvent souligné, un traitement d'information devient un traitement de symboles. Dans cette perspective, un système de traitement de l'information textuelle doit être vu comme un système qui traite avant tout des objets de nature symbolique.

Dans cette perspective, SATO a été conçu comme un système de traitement d'*unités d'information* (UNIF) en tant qu'objets sémiotiques. Mais comme tout objet sémiotique, l'UNIF se présente d'abord sur un support physique spécifique, en l'occurrence comme une suite, ou mieux, des chaînes de caractères. À ce premier niveau informationnel, un système informatique de traitement textuel doit évidemment manipuler le texte dans sa dimension de porteur physique. Ce qui, malgré les apparences, n'est pas

la moindre des choses surtout quand ce support prend des milliers de pages. Aussi SATO, par un de ses modules SATOGEN, offre-t-il une gestion de la matérialité du texte. En tant que système de traitement d'information, il est en mesure de traiter d'immenses bases de données textuelles où les textes s'avèrent des porteurs structurés et prégnants. De fait, SATO opère sur la matérialité du texte, en ce sens qu'il génère une représentation informatique du texte qui permet de le manipuler, le segmenter, le partitionner, l'archiver, le rappeler et même, l'éditer ou l'imprimer. Évidemment, toutes ces opérations sont relativement classiques et on les retrouve dans tout système de traitement de textes. Mais ce qui distingue SATO des autres systèmes de traitement est l'orientation ultime de sa manipulation de textes. SATO ne se contente pas de manipuler un texte donné à des fins documentaires ou d'édition, il l'organise comme intrant à des multitudes d'analyseurs (manuels ou automatiques) qui visent à identifier certains traits, tant de sa structure organisationnelle que de son contenu signifiant. SATO, à ce niveau, est un ouvrier-assistant dans le travail de préparation matérielle préalable à toute analyse et lecture d'un texte électronique. À ce niveau de traitement, SATOGEN est neutre quant à la

2. P. Zunde, « Information Theory and Information Science », *Information Processing and Management*, vol.17, no. 6 (1981), 341.
3. N.J. Belkin, *A Concept of Information for Information Science*, PhD Thesis, University of London, 1977.
N.A. Belkin and S.E. Robertson, « Information Science and the Phenomenon of Information », *Journal of The American Society for Information Science*, vol.27, no. 10 (1971), 10-15.
B.C. Brookes, « Information Science », in H.A. Whatley, *British Librarians and Information Science, 1966-70*, London, The Library Association, 1972, p. 137-149.
F.I. Drestke, *Knowledge and the Flow of Information*, Cambridge, MIT Press, 1982.
R. Khulen, « Information and Pragmatic Value Adding », *Computers and the Humanities*, vol. 25 (1991), 93-101.
4. P. Zunde, « Information Theory ... », 344.
5. H. Simon, « Cognitive Science: The Newest Science of the Artificial », *Cognitive Science*, vol. 4 (Jan./March 1980), 35.
6. N. J. Nilsson, *Principles of Artificial Intelligence*, Palo Alto, Ca., Tioga Press, 1980.
7. H. Simon, « Cognitive Science ... », 35.
8. N.J. Belkin, « Information Concepts for Information Science », *Journal of Documentation*, vol. 34, no. 1 (March 1978), 55-85.

structure et au contenu du texte. Il ne fait que préparer le texte pour des analyses. Ce sont toujours des analystes qui définiront la nature des unités matérielles de l'information. Et en fonction de leur paradigme théorique, les analyseurs manuels ou automatiques transformeront ou catégoriseront ces UNIF tantôt en des « morphèmes », tantôt en des « lexèmes », ou encore en des « sèmes », ou en des « phrases », tout simplement: ce que SATO recevra et continuera à manipuler comme un nouvel objet textuel.

La nature sémiotique du texte

Ce qui rend souvent la question de l'analyse de textes par ordinateur difficile à saisir est dû à l'ambiguïté inhérente de ce concept de *textes*. Ce terme est en effet polysémique, tantôt il désigne l'objet physique, matériel, imprimé que l'on range sur des rayons, tantôt il touche le contenu de ce qui est écrit. Or, ce qui intéresse soit un lecteur, soit un interprète ou un analyste, n'est que rarement le manuscrit physique, le livre, mais bien le contenu du texte, c'est-à-dire le texte abstrait.

Mais qu'est-ce qu'un « texte » abstrait? La question n'est pas simple parce qu'en fait elle touche la question philosophique de la signification ou du sens dont est porteur un texte matériel. Et lorsqu'on parle du contenu d'information d'un texte, on se réfère non pas à un *objet physique* mais à une *action* effectuée par des humains dans le but de parler de quelque chose, de raconter une histoire, de présenter une argumentation, d'établir des faits, de fixer des normes, de discuter d'un règlement, etc. Cet ensemble d'actions est ce qui est appelé communément le *discours*. Autrement dit, lire et analyser un texte, ce n'est pas chercher un quelconque objet mais retrouver l'ensemble des actions significatives dont il est le produit. Un texte matériel est le témoin d'une multitude d'actions de communication qui prennent la forme d'un discours monnayé en des propositions, des idées, des thèmes, etc. Et c'est à cet univers du discours que tente d'accéder un lecteur ou un analyste.

Dire cependant que le texte abstrait est un discours, c'est lui attribuer une propriété tout à fait spéciale. C'est le définir comme un objet sémiotique,

c'est-à-dire en regard non de ses propriétés physiques mais en relation à ses fonctions dans la production sociale de la signification.

C'est précisément parce qu'il est un tel objet signifiant et non simplement une chose physique que les questions suivantes peuvent lui être adressées: quels sont les *thèmes* retenus par l'auteur du texte?; comment se sont-ils *exprimés* dans la langue?; qui sont les *interlocuteurs* de ce texte?; quelles sont les *visées* de ce texte?; où dans le texte est-il question du *concept X*?; quelle est la *structure narrative* de ce texte?; que peut-on conclure de ce texte? etc.

Or, précisément parce que le contenu du texte est d'ordre sémiotique, le nombre et le type de telles questions sont infinis. La lecture d'un texte et son analyse relèvent d'un processus d'interprétation ou de saisie de sens qui ne peut jamais s'épuiser dans une approche unique. Chaque lecteur ou analyste, selon sa motivation, ses objectifs, peut poursuivre des enquêtes radicalement différentes. Et il est strictement impossible de penser trouver un type de questionnement qui puisse satisfaire l'ensemble de tous les lecteurs et analystes. Le contenu d'un texte appartient à cette sphère de la signification conçue comme un univers qui, on le sait, ne peut se réduire à une seule dimension.

Bref, les informations que contient un texte ne sont pas avant tout des signaux possédant uniquement un certain degré de probabilité d'occurrence et qu'il faut générer, transmettre, etc. Un texte est certes un objet matériel soutenu par un porteur physique mais il est surtout un objet structuré, signifiant et utilisé en contexte de communication humaine. Une conséquence immédiate de cette approche est qu'il n'y a pas une unique façon d'aborder un texte. Même si un texte se présente dans une langue spécifique, par exemple le français, même si l'ensemble des expressions linguistiques qui le constituent sont relativement stables et susceptibles d'être décrites selon une certaine grammaire, son contenu est plurivalent. Il est toujours relatif et à une interprétation sémantique et à une approche pragmatique. Une certaine tradition structurale a souvent laissé croire qu'un texte était un objet autonome qui épuisait son sens ou sa

signification dans sa structure interne et qu'une méthodologie rigoureuse de type distributionnel pourrait en extraire le contenu. Mais cette position s'est avérée problématique à plus d'un égard. Quelle que soit la forme d'un texte, ce dernier doit toujours, comme l'a si bien défendu M.A.K. Halliday⁹, être approché comme un objet sémiotique produit dans une situation de communication et non pas comme une pure structure autonome.

Une telle perspective sémiotique amène alors une vue particulière de la production textuelle dans les organisations sociales et les institutions publiques. Même dans cet environnement, un texte se présente évidemment sur des porteurs physiques multiples: papiers, disquettes, fiches, filières, plans, etc. Mais il prend aussi des formes signifiantes des plus variées: lettres, rapports, factures, dessins et graphismes. Son contenu est des plus diversifiés. Il touchera l'organisation, la planification, les finances, les relations publiques, les règlements, les normes, les protocoles, les listes, etc. Comme n'importe quel autre texte, il est inséré dans un processus de communication d'information ou d'organisation, il décrit ses tâches, prend des décisions, planifie ses actions, évalue ses performances, se donne une histoire, et même une image collective, etc. Bref, dans les institutions, les textes mêmes, précisément parce qu'ils sont des objets sémiotiques, seront abordés de mille et une façons. Ils ne se laisseront jamais épuiser par une approche unique¹⁰.

Une telle conception du texte influence grandement la construction d'un système de traitement de l'information textuelle. Il est impensable qu'un tel système fournisse une lecture « automatique » des textes. Ce fut là, nous semble-t-il, une des erreurs des premiers modèles de systèmes dits *d'analyse automatique du discours*. Or, pas plus qu'on ne peut

9. M.A.K. Halliday, *Language as Social Interpretation of Language and Meaning*, Baltimore, University Park Press, 1978.

10. Suzanne Bertrand-Gastaldy, « L'indexation assistée par ordinateur: un moyen de satisfaire les besoins collectifs et individuels des utilisateurs de bases de données textuelles dans les organisations », *ICO (Intelligence artificielle et sciences cognitives au Québec)*, vol.2, no 3 (septembre 1990), 71-91.

produire par ordinateur des rituels religieux, des oeuvres d'art, des rencontres humaines, etc. - parce que ce sont des objets sémiotiques - on ne peut construire des systèmes de lecture et d'analyse automatique du discours. Ceci est une impossibilité presque a priori. Par contre, des systèmes comportant plusieurs niveaux de représentation peuvent aider à en saisir la complexité¹¹.

Cette perspective traverse SATO de part en part. En tant que système informatique de traitement de l'information textuelle, il ne peut prétendre à une lecture automatique. SATO doit être vu comme un adjuvant dans le processus de lecture et d'analyse qu'effectuent des agents cognitifs en situation de communication. Ces utilisateurs interpellent alors SATO pour qu'il les aide à gérer intelligemment les porteurs électro-physiques de ces textes et surtout à en saisir, et en analyser la structure, le sens et la portée. Autrement dit, en raison même de la nature sémiotique de l'objet textuel, un système de traitement informatique ne peut jamais être un système expert autonome et intelligent. Il est plutôt pour des agents cognitifs un *assistant* dans le processus de lecture, d'analyse d'un texte produit dans un cadre de communication.

En ce sens SATO, via certains des modules, et plus particulièrement SATOINT, est une boîte à outils dont la productivité dépend non de SATO mais bien du lecteur interprète humain. SATO doit se retirer devant le génie de la lecture humaine et non la remplacer.

L'assistant philologue

Dire qu'un système informatique de traitement de l'information textuelle doit être vu dans le cadre de l'assistance à la lecture et à l'analyse implique que l'on soit en mesure de préciser ce que sont cette lecture et cette analyse. Or, un tel système n'est pas un simulateur de la lecture humaine comme un robot d'usine peut simuler l'action de voir ou de saisir un objet. En effet, la lecture, l'analyse dont il est ici question, n'est pas celle d'un parcours visuel du graphisme des lettres ou même de leur décodage. Dans ce sens où l'on dira, par exemple, qu'un lecteur optique « lit des textes ».

L'acte de « lecture » ici est autrement plus complexe. Il touche un processus de compréhension et d'interprétation. Dans sa forme la plus technique, il participe à ce que L. C. Paquin¹² appelle une lecture « experte ». De fait, un système comme SATO a été conçu comme un assistant lecteur expert de type philologique. Et cette perspective en signe de nombreuses caractéristiques. Précisons brièvement ce que l'on entend par cette lecture philologique.

Dans son acception la plus générale, la lecture philologique est un processus d'analyse textuelle pratiquée depuis l'Antiquité par des lecteurs experts qui visaient à saisir dans les textes le génie propre d'un peuple ou d'une civilisation ou encore l'esprit ou l'intention d'un auteur.

À l'origine, la lecture philologique prenait la forme de l'examen des textes anciens, mais au cours de l'histoire, en raison du développement même de la production textuelle (multiplicité des manuscrits, imprimerie, traductions), elle en vient à se fonder avec l'étude de l'histoire des idées (Vico) et même de toute l'idéologie et de la culture (Dilthey, Marx). Aussi le lecteur philologique se confond-il souvent avec le processus technique et méthodologique d'interprétation textuelle.

Dans la pratique, la lecture philologique devient un processus d'examen et d'interprétation des documents textuels. Elle est une certaine méthode assurée de lecture des textes. Contrairement à la tradition de type structuraliste, qui voyait le texte comme une entité autonome régie par des règles, la lecture de tradition philologique en appelle à une implication intime de l'analyste qui doit construire la signification du texte au fur et à mesure de sa fréquentation avec le texte. La perspective va au-delà de la sémiotique, elle est herméneutique. Le lecteur cherche à dégager du texte son véritable sens, tant à des fins d'explication qu'à des fins de compréhension. Cette lecture touche autant la forme du texte (explanatio) que l'univers dont il parle (exégèse). La lecture philologique a ainsi développé une méthodologie spécifique qui, en général, malgré des variantes, inclut les « moments » suivants :

- *Une étude de la matérialité du texte* : Par exemple : la datation, le déchif-

frement (paléographie, codicologie, etc.), la collation des versions, l'examen des variantes, la classification, l'interprétation des fautes, le repérage des interpolations, etc. Ces opérations visent ultimement l'établissement d'un texte authentique ou d'une édition critique.

- *Une étude de la grammaire* : La grammaire touche ici autant les formes lexicales que les modalités expressives. Cette dimension prend, dans la pratique, une grande importance au point de devenir le coeur de la méthode. On y trouvera ainsi l'étude critique et comparative des manuscrits, leur filiation (méthode de Lachman), leur origine historique, l'analyse conceptuelle, rhétorique, et argumentative et même narrative.

- *Une production de commentaires* : Vu la complexité des résultats des multiples analyses, la compréhension s'opérationnalise souvent sous la forme de commentaires qui expriment l'interprétation que le lecteur expert applique au texte.

- *Un jugement qualitatif sur cette oeuvre* : Comme le processus de compréhension de type herméneutique en appelle non seulement à la rationalité du lecteur mais à ses valeurs et à son esthétisme, la méthode permet l'insertion dans le commentaire des jugements qualitatifs du lecteur. Ici s'exprimait la pertinence tant logique, sociale qu'esthétique du texte d'origine.

Selon le domaine où elle s'applique et en raison des variantes théoriques, cette méthodologie a souvent reçu des noms différents. Appliquée à la Bible, elle devient exégèse, à la littérature, elle devient analyse, au texte philosophique, elle prend la forme de l'interprétation et au discours social, elle en appelle à l'analyse de l'idéologie.

Même si les pratiques contemporaines d'analyse textuelle de la linguistique, de l'analyse littéraire, du structuralisme ont fait exploser l'approche

11. Jean-Guy Meunier et al., « A Call for Enhanced Representation of Content as a Means of Improving On-Line Full-Text Retrieval », *International Classification*, vol. 14, no. 1 (1987), 2-10.

12. L.C. Paquin, « La lecture experte », *Technologies, Idéologies, Pratiques*, vol. X, no 2-4 (1992), 209-220.

philologique traditionnelle en plusieurs démarches souvent autonomes - que l'on prétend aseptiques et objectives - l'influence de cette méthode n'en est pas moins encore des plus présentes. Si dans la présentation des résultats de l'analyse textuelle, on cherche à donner un air d'objectivité au contenu et à la forme textuelle, (graphe, arbre, statistiques, listes, formalisme) le processus de découverte demeure marqué par la lecture philologique. Car ce qui caractérise le plus cette lecture experte de type philologique est la multiplicité et l'interaction des vecteurs d'analyse. D'une part, elle exige un parcours rigoureux de la matérialité du texte admis, d'autre part la saisie des multiples formes d'expressions du texte, l'extraction du contenu signifiant, l'introduction des connaissances du lecteur et l'intégration dans un horizon de culture. Et ceci dans un processus interactif constant où le texte se transforme au fur et à mesure de ce processus.

Cette perspective philologique est au coeur de la conception du système SATO. Comme nous l'avons dit plus haut, le texte est ici vu comme un objet sémiotique avec lequel le lecteur entre en interaction dynamique. Et ce processus est complexe. Un texte communique un contenu à son lecteur sous une forme structurée mais l'accès à ce contenu n'est pas simple et direct. Il résulte d'un processus d'aller-retour, de construction, d'examen, de catégorisation, de synthèse qui ne se donne pas d'emblée. En cela, il se veut plus proche du processus de découverte du sens d'un texte que de son explicitation et de son explication. Tout comme l'ancien philologue avec les manuscrits anciens, SATO doit parcourir ses « manuscrits électroniques ». Sa lecture en est une première approche presque « monastique ». Il est sensible à toutes variations, marques, formes matérielles de son écriture. Via son module SATOINT, dans son moment de fouille et d'extraction, il marque les segments, identifie les formes lexicales et phrastiques. Mais dans son moment d'analyse, via son module SATOINT, il navigue dans le texte pour en extraire des informations pertinentes relativement à un angle particulier de lecture. Qui plus est, via l'appel à des analyseurs périphériques d'appoints, manuels ou automatiques, il parvient à ajouter au texte une catégorisation et une classification qui

peuvent à leur tour être fouillées, repérées et analysées, etc. Dans certaines de ses applications, il est en interface avec d'autres progiciels qui l'aident à gérer plus ergonomiquement la multiplicité des informations extraites.

Bref, l'extraction des « connaissances » ou, dans le sens plus riche, la « compréhension » et l'interprétation du texte qui se réalise, est le résultat du processus interactif de construction du sens et non de stratégies algorithmiques définies d'avance.

Conclusion

Comparé à d'autres systèmes de traitement ou d'analyse de textes, SATO

est loin d'être un système artificiellement intelligent et surtout n'est aucunement un système expert. Au contraire, c'est un postulat du système de laisser l'intelligence et l'expertise au lecteur et à l'analyste. Mais, croyons-nous, c'est précisément cette limite qui est sa force. Dans l'horizon théorique qui fonde SATO, l'information est conçue comme étant d'ordre symbolique et non simplement de type signalétique: un texte est avant tout un objet sémiotique et sa lecture en est philologique. En conséquence, SATO ne peut être pensé comme un robot expert ou intelligent, mais comme un adjuvant au lecteur et analyste humain, seul expert et intelligent.

La plus grande source d'informations sur la société québécoise !

- Plus de 370 000 références bibliographiques sur un même CD-ROM
- Des dizaines de sources répertoriées (quotidiens, périodiques, monographies, rapports, etc.)
- Inclus tous les index produits par **Inform II - Microfor** (L'Index de l'actualité de 1982 à 1991, l'Index des affaires de 1988 à 1991, l'Index de la santé et des services sociaux de 1989 à 1991, ainsi que la base de données Médiadoq, produite par le RQIC (1958 à 1991))
- À l'achat du CD-ROM, profitez d'un rabais lors d'abonnement à nos index mensuels

Démonstration gratuite - Information: (514) 524-7722