

La gestion du savoir et de l'information électronique The Management of Knowledge and Electronic Information La gestión del saber y de la información electrónica

Gracia Pagola et Réjean Roy

Volume 43, numéro 4, octobre–décembre 1997

URI : <https://id.erudit.org/iderudit/1032981ar>

DOI : <https://doi.org/10.7202/1032981ar>

[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

Pagola, G. & Roy, R. (1997). La gestion du savoir et de l'information électronique. *Documentation et bibliothèques*, 43(4), 189–196.
<https://doi.org/10.7202/1032981ar>

Résumé de l'article

L'information constitue une ressource stratégique qu'il convient de gérer de plus en plus efficacement, mais plusieurs problèmes se posent à cet égard. Les auteurs énumèrent ces problèmes avant de faire état des pistes de solution qui se dessinent aujourd'hui — solutions parmi lesquelles on compte la numérisation de l'information et la mise au point, pour traiter le document électronique, d'outils de traitement de la langue toujours plus performants.

La gestion du savoir et de l'information électronique

Gracia Pagola

Consultante auprès du Centre d'expertise et de veille inforoutes et langues (CEVEIL) et chargée de cours à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal et au Collège de Maisonneuve.

Réjean Roy

Consultant auprès du CEVEIL et d'autres organismes.

L'information constitue une ressource stratégique qu'il convient de gérer de plus en plus efficacement, mais plusieurs problèmes se posent à cet égard. Les auteurs énumèrent ces problèmes avant de faire état des pistes de solution qui se dessinent aujourd'hui - solutions parmi lesquelles on compte la numérisation de l'information et la mise au point, pour traiter le document électronique, d'outils de traitement de la langue toujours plus performants.

The Management of Knowledge and Electronic Information

Information is considered to be a strategic resource requiring efficient management. However, several problems have arisen. The authors identify these problems and then offer solutions, among which is the scanning of information and tools to codify language.

Un mercredi matin tout ce qu'il y a de banal. Vous êtes installé à votre bureau en train de lire la section sportive de votre journal favori, de manger un muffin aux bananes et de siroter votre premier café de la journée, lorsque la voix mielleuse de votre patron se fait entendre : «*Tranchemontagne ! Dans mon bureau ! Tout de suite ou vous êtes viré !*». Vous accourez au pas de course dans le bureau de votre supérieur lequel vous explique que la lecture, chez le dentiste, d'un article - «*Un article vieux de deux mois, Tranchemontagne ! Est-ce que tout le monde dort ici ?*» - vient de lui apprendre la création prochaine, sous l'autorité du gouvernement coréen et avec la participation éventuelle d'entreprises étrangères spécialisées dans la gestion hôtelière, la gestion de pistes de ski et la gestion de terrains de golf - les domaines d'expertise de la société - d'un mégacentre de villégiature dans le nord-est de la Corée du Sud.

Votre patron est survolté : il vous demande de recueillir plus de données sur

le projet sur le marché touristique coréen, sur les pratiques sportives des Coréens et sur vos concurrents éventuels, coréens et étrangers puis de rédiger, le plus vite possible, un rapport détaillé faisant état de l'opportunité, pour votre employeur, de prendre part au projet. Vous vous lancez immédiatement à la recherche de l'information demandée mais faites rapidement face à différents obstacles : les renseignements requis sont extrêmement difficiles à obtenir; les rares textes obtenus sur le sujet ne sont disponibles qu'en coréen; votre compagnie ne compte plus, depuis le départ de deux ou trois vétérans, d'employés dotés d'une forte connaissance du marché asiatique..

Au bout de trois semaines, dossier complété à la main, vous retournez voir votre patron. Ce dernier, malheureusement, tient déjà dans la sienne une manchette du *Wall Street Journal* rapportant que les dés sont déjà jetés : le gouvernement coréen a arrêté l'avant-veille la liste des entreprises étrangères appelées à

La gestión del saber y de la información electrónica

La información constituye un recurso estratégico que debemos administrar de manera más eficaz, pero encontramos muchos problemas en este campo. Los autores de este artículo enumeran estos problemas antes de señalar diferentes soluciones que podemos encontrar hoy, como la digitalización de la información y la puesta en marcha de instrumentos de tratamiento de la lengua.

collaborer au projet. Et votre société, évidemment, n'en fait pas partie...

Cinq grands problèmes

Comme l'écrivait René Mayer (1990) dans un rapport publié en France au début de la décennie - un rapport intitulé *Information et Compétitivité* - il existe deux grandes façons de concevoir l'information. La première «*comme un bien ou un service comme les autres. Sa production, sa distribution et son utilisation sont alors considérées comme des activités certes intéressantes, mais uniquement par le fait que, comme toute autre, elles sont créatrices de valeur ajoutée et d'emplois*». La seconde considère que l'information n'est pas un bien ordinaire, qu'elle est importante «*non [...] seulement en elle-même, mais principalement en raison de ses effets sur le reste de l'économie*». L'information est alors perçue comme exerçant des effets positifs sur l'innovation, la qualité, la productivité et d'autres facteurs.

L'anecdote présentée en début d'article démontre bien que le deuxième point de vue présenté par René Mayer est sans conteste celui qui doit être préconisé à notre époque. Ainsi que l'ont souligné des auteurs aussi réputés que Peter Drucker ou Hervé Sérieyx, la prospérité des États a bel et bien cessé de dépendre de leur dotation en matières premières pour devenir fonction de la capacité des entreprises installées sur le sol national à produire, recueillir et traiter l'information. Rien ne prouve mieux cette assertion que le fait que le coût de l'information produite et échangée lors de la construction d'un immeuble représente 60% de la facture ou que les ressources naturelles comptent pour un maigre 2% du coût des microprocesseurs. La structure des échanges commerciaux témoigne elle aussi de la grande importance que prend aujourd'hui l'immatériel dans l'économie mondiale. Ainsi, les États-Unis exportent désormais plus de services éducatifs (6,1 milliards de dollars américains) que de maïs (5,7 milliards de dollars américains); plus de services financiers (5,4 milliards de dollars américains) que de blé (4,6 milliards de dollars américains); plus de services informationnels (2,6 milliards de dollars américains) que d'aluminium (1,2 milliard de dollars américains); plus de services juridiques (1,4 milliard de dollars américains) que d'huiles végétales (1 milliard de dollars américains); plus de services de consultation (780 millions de dollars américains) que de cacahouètes (210 millions de dollars américains) (Cornella 1997).

L'information est une ressource stratégique et on le répète partout. Rien de surprenant, alors, que les entreprises, sous les encouragements des sociétés-conseils les plus prestigieuses, attachent de plus en plus d'importance au *knowledge management*, à la gestion du savoir. Qu'est-ce que la gestion du savoir ? La réponse donnée à cette question (et le prix à payer pour l'obtenir) varie évidemment selon la boîte de consultants à laquelle elle est posée, mais pour les fins de ce court article, *gérer le savoir signifiera trouver une solution aux cinq problèmes épineux qui suivent* : 1) l'information recherchée n'existe pas ou n'a pas encore été codifiée formellement; 2) l'information recherchée existe, mais elle est difficile d'accès à un non-spécialiste; 3) l'information recherchée existe, mais il est plus difficile de la repérer

que de trouver une aiguille dans une botte de foin; 4) l'information existe, mais elle est uniquement disponible dans une langue étrangère et 5) l'information existe, mais il y a une différence énorme entre information et savoir.

L'information n'existe pas

Les responsables politiques russes abordaient récemment les dirigeants de la multinationale International Harvester pour inciter ces derniers à construire une usine de tracteurs moderne sur le territoire de la Russie. L'idée avait du sens : après tout, International Harvester avait relevé le même défi, avec grand succès, 20 ans plus tôt, en URSS. Le projet de construction a cependant échoué en bout de ligne, parce que personne chez International Harvester¹ ne savait en fait de quelle façon les choses avaient été menées la première fois (Tierney 1997).

Le problème de l'*amnésie corporative* n'est pas typiquement américain. Il est encore pire, en fait, dans des pays comme le Japon, où l'information et le savoir, fréquemment, ne quittent pas le cerveau des personnes qui en sont dépositaires : *«In Japan, you might find a kind of superman who can do a difficult job, which is convenient enough. But after he's gone, no one else can do it. In America, anyone can do the job as long as there's a manual. Documentation, the sharing of software, and building them into a business system are going to become important to manufacture the exact same products in many countries, while taking into account local conditions and culture»* (Nonaka and Takeuchi 1995). Comme on le voit, gérer le savoir consistera d'abord à assurer que l'information détenue par chaque travailleur de l'entreprise - l'information requise pour attribuer un contrat, réparer un appareil, classer un document, administrer un budget, négocier avec un fournisseur difficile, etc. - soit codifiée et stockée sous forme électronique de façon à être accessible aux autres travailleurs, notamment après un départ².

Dans un passé encore récent, il était difficile au plus gros vendeur d'une entreprise localisée à Québec de partager son expertise avec un collègue de Rimouski (surtout après sa décision de partir en croisière autour du monde). Même situa-

tion pour le mécanicien rencontrant un cas difficile dans son garage de Jonquière : il était généralement le seul à sortir enrichi de l'expérience. L'arrivée des nouvelles technologies permet cependant de combler partiellement cette lacune³.

Le cas du constructeur d'ordinateurs Compaq constitue un exemple de ce phénomène. Compaq a décidé de réinventer son service après-vente au début des années 1990 en s'appuyant sur le recours à un système-expert du nom de SMART. Grâce à SMART, chaque préposé de Compaq peut accéder à l'ensemble du savoir technique cumulé par l'entreprise au moment précis où le client requiert son aide. Pour ce faire, il fournit à l'ordinateur une description textuelle du problème informatique exposé par le client : par exemple, mon imprimante ne fonctionne pas quand j'emploie tel logiciel; cette description sommaire est ensuite traitée par l'ordinateur qui vérifie automatiquement dans une base de données textuelles si une solution au problème n'a pas déjà été trouvée par un autre technicien. Selon une étude interne, SMART a permis à Compaq de servir les clients plus rapidement, plus efficacement et avec plus de constance.

Andersen Conseil utilise aussi l'informatique pour faire une utilisation optimale des connaissances de chaque employé. Plus de 85% des professionnels de l'entreprise, dont une des 360 succursales mondiales se trouve à Montréal, sont reliés entre eux par un intranet appelé ANET. ANET permet aux spécialistes de l'entreprise de faire part sur un babillard spécialisé des problèmes rencontrés par un client et de bénéficier, pour leur résolution, de l'expertise de 82 000 employés. Le recours à ANET permet au savoir disponible

1. Même scénario chez Ford où les dirigeants ont abandonné l'idée de répéter la stratégie ayant mené au lancement de la Ford Taurus en s'apercevant que personne, en fait, ne la connaissait.
2. Les départs, à cause du vieillissement de la population, sont de plus en plus fréquents dans les pays industrialisés.
3. «Partiellement», parce que codifier dans une base de données l'information non structurée détenue par les employés d'une organisation ne pourra jamais, Dieu merci!, être réalisé à très grande échelle.

dans une succursale localisée à l'autre bout du monde de ne pas demeurer inexploité (Quinn et al. 1996).

Kao, un conglomérat japonais spécialisé dans la fabrication et la commercialisation de produits de beauté, de nettoyants, de médicaments, de plastiques, etc., utilise une tactique similaire pour mieux servir les besoins de son équipe de vente et de sa clientèle. Parce que Kao emmagasine dans une base baptisée Écho le contenu des 50 000 appels de consommateurs reçus chaque année par ses employés, de même que les réponses qui y sont associées, il suffit à un téléphoniste de taper un ou deux mots clés pour fournir à une mère dont le bébé vient d'ingurgiter le contenu d'une bouteille de détergent l'aide dont elle a besoin de toute urgence. Dans la même veine, il suffit à une préposée de formuler en japonais courant le problème d'un homme dont la cravate neuve arbore une vilaine tache graisseuse pour trouver rapidement la solution qui sauvera sa soirée du désastre. Enfin, l'employé désireux d'en savoir plus sur l'attitude de la clientèle face à un nouveau produit est en mesure d'interroger Écho pour connaître les réactions que son lancement a suscitées. Écho renferme plus de 350 000 descriptions de problèmes, lesquelles peuvent être repérées ou classées à l'aide de plus de 8 000 mots clés. La base est considérée par les hauts dirigeants de Kao comme un instrument permettant de servir le client de mieux en mieux (Nonaka 1995). En résumé, pour paraphraser le p.-d.g. de Texas Instruments, grâce aux nouvelles technologies, il pourra devenir plus facile aux entreprises de savoir ce qu'elles savent vraiment... et ce qu'elles ne savent pas (Tierney 1997).

Un problème de convivialité

Un autre défi consistera à rendre de plus en plus direct et de plus en plus instantané le lien existant entre le travailleur et l'information requise par ce travailleur (ou encore, alors que le client se transforme de plus en plus souvent en employé⁴, le lien existant entre le client et le renseignement désiré par le client)⁵. Malheureusement, la communication homme-machine n'est pas toujours facile. Trop souvent, l'utilisateur doit en effet se plier aux exigences de la machine pour accéder au document électronique recher-

ché, plutôt que le contraire. Il s'agit d'un problème important dans le cas des travailleurs les moins qualifiés et dans celui des personnes peu rompues à l'emploi des technologies. Au Canada, plus d'un étudiant sur trois appartient aux deux groupes dont les aptitudes de lecture et de rédaction sont les plus faibles, alors qu'en Allemagne, aux Pays-Bas, en Suède et en Suisse alémanique, cette proportion varie entre 18 et 23%. Au Québec, à peine 40% des élèves du secondaire maîtriseraient la grammaire et la syntaxe de manière convenable (Roy 1997).

Le hic, c'est qu'à l'heure où l'ordinateur pénètre de façon massive dans les milieux de travail, l'incapacité de la machine à comprendre la langue est précisément ce qui le rend inflexible : à titre d'exemple, contrairement au bibliothécaire qui traite les requêtes de ses collègues avec intelligence, l'ordinateur traditionnel ne peut, le plus souvent, trouver un document dont le nom est mal orthographié. Pour que l'arrivée des ordinateurs permette l'amélioration de la performance des travailleurs les moins qualifiés, il faudra donc rendre l'ordinateur plus intelligent. On pourra ici déplorer que la réponse proposée à un problème linguistique soit technologique plutôt que pédagogique, mais, force est d'admettre qu'à leur meilleur, les outils de traitement de la langue permettront à certains d'atteindre des niveaux de productivité qui autrement auraient été hors de portée. L'utilisation que fait General Motors de la correction automatique le démontre bien.

Afin d'assister les travailleurs appelés à réparer ses véhicules, General Motors a créé une énorme base informatisée renfermant la description de tous les problèmes rencontrés par l'ensemble des mécaniciens d'Amérique et des solutions apportées pour les régler. Le mécanicien en présence d'un bris qu'il ne peut réparer contacte par téléphone un des techniciens du centre d'appel de GM. Le technicien n'a plus qu'à écouter le mécanicien faire le récit des symptômes qu'il observe pour interroger le système et obtenir de l'ordinateur la meilleure solution apportée à ce problème dans le passé. La difficulté, c'est que les techniciens sont généralement de mauvais dactylos et que la langue qu'ils utilisent pour extraire les données de la base est farcie d'abréviations, d'erreurs

typographiques et orthographiques et de fautes de syntaxe.

Lexfix est un correcteur dont l'utilisation vise à pallier ces difficultés; l'outil intercepte le texte entré dans l'ordinateur et corrige aussi bien les fautes d'orthographe commises par l'utilisateur que l'utilisation erronée de certains termes et abréviations non normalisés. Il présente ensuite une version révisée à l'utilisateur qui la revoit ou l'entérine. Lexfix permet aux techniciens du centre d'appel de se concentrer sur le diagnostic à poser plutôt que sur la saisie de données. La correction automatique et la normalisation des entrées favorisent l'utilisation plus efficace de l'information de la base par les techniciens dont le jugement reposera en effet sur l'ensemble des renseignements significatifs qui y sont contenus plutôt que sur les seules données auxquelles une saisie incorrecte aurait permis d'accéder.

Dans une autre veine, la complexité actuelle des ordinateurs explique que la plupart des gens ignorent ou n'aiment guère l'informatique. Ainsi selon une étude de Dell Computers, citée récemment dans le magazine *Fortune*, 55% des Américains éprouvent une forme plus ou moins avancée de phobie face aux ordinateurs et aux appareils électroniques (Roy et Georgeault

4. Que ce soit au moment de remplir un formulaire de commande ou encore de configurer le produit recherché.

5. Ce lien n'est pas toujours facile à établir, ce qui entraîne des coûts en temps et en argent. À titre d'exemple, les entreprises doivent parfois recourir aux services d'experts capables d'utiliser des langages informatiques aussi complexes que SQL pour obtenir de leurs ordinateurs la réponse à une question aussi simple que : Combien de nos clients ont acheté ce modèle d'imprimante le mois passé? En comparaison, dans l'entreprise recourant à des interfaces de bases de données en langue naturelle, chaque vendeur capable d'utiliser le français pourra formuler la bonne question et rapidement obtenir de l'ordinateur la réponse correcte. Cette méthode évitera que le renseignement recherché soit devenu périmé au moment de sa localisation par les services de recherche, une situation trop fréquente à l'heure actuelle, selon Larry Harris (1994), *Commercial natural language : Critical success factors*, présentation faite durant le colloque *Innovative Applications in Artificial Intelligence* qui s'est tenu du 31 juillet au 4 août 1994 à Seattle, enregistrée sur cassette par First Tape, n° E1047.

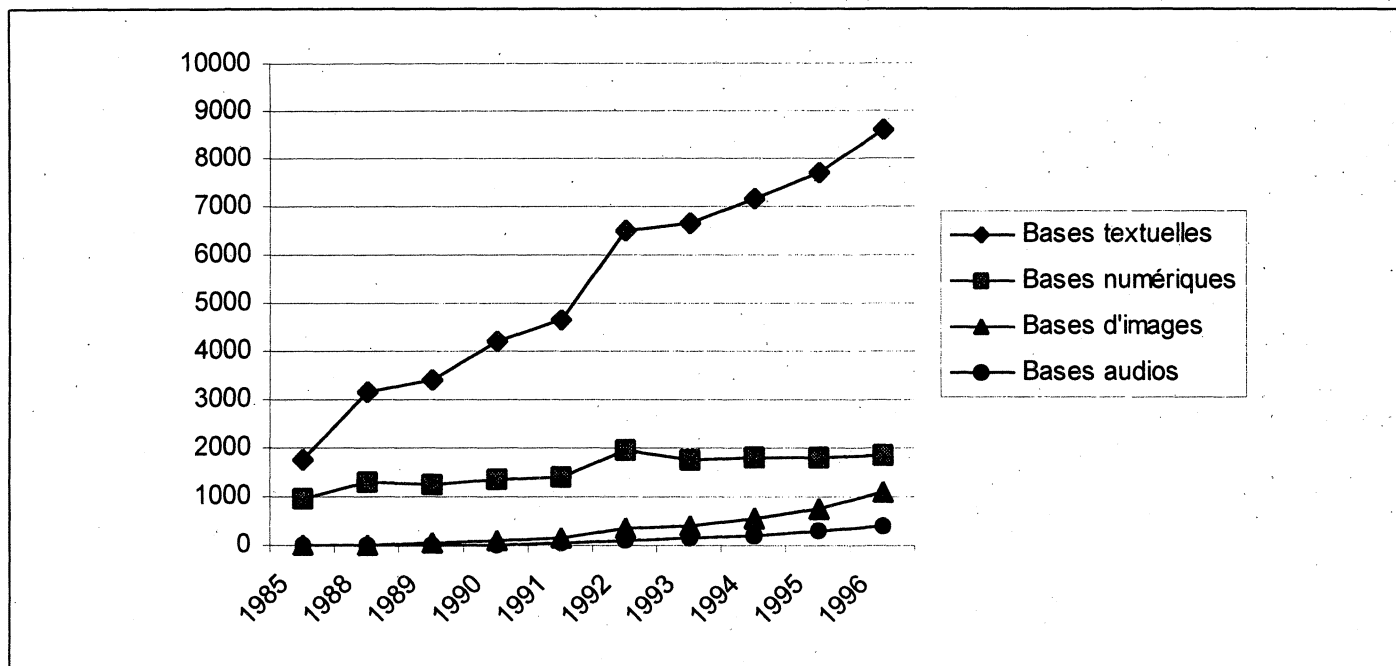


Figure 1. Les bases de données : la montée du texte

Source : Martha E. Williams. *The State of databases today : 1996*. Gale directory of databases, volume 1 : Online databases (s.l., 1996), p. xvii-xxiv.

1995). Si une large tranche de la population est technophobe et se fait tirer l'oreille, de nombreux problèmes surgissent aussi du côté des convertis. À preuve, le journal *Le Soleil* rapportait récemment qu'à cause de la complexité des ordinateurs, les usagers ne seraient en mesure d'exploiter qu'un maigre 3% de leurs potentialités. En fait, selon le directeur au marketing de la filiale canadienne de Microsoft, trois des quatre millions d'ordinateurs présents dans les foyers québécois et canadiens seraient entre les mains de personnes ne sachant pas vraiment comment les utiliser (Le Soleil 1996). Encore une fois, une grande proportion de ces personnes seront appelées à utiliser l'ordinateur dans le cadre de leur travail. Cependant, elles ne pourront le faire de façon productive qu'à condition que la machine soit plus facile à utiliser. Cette idée est de mieux en mieux comprise par les producteurs d'information électronique, comme le démontre l'importance accordée par les organisateurs du *Online Information Meeting* (www.online-information.com) à des thématiques comme la simplification de la recherche, la convivialité des interfaces ou encore la personnalisation de

l'information. L'idée est aussi de mieux en mieux comprise par les entreprises utilisatrices lesquelles accordent une attention croissante à la qualité des interfaces mises à la disposition de leurs employés. Par exemple, le fabricant d'électro-ménagers Whirlpool décidait récemment de se doter de 1 000 exemplaires du logiciel de gestion des documents de la société Documentum afin d'améliorer son service à la clientèle et de favoriser l'amélioration des performances de ses vendeurs. Selon les dirigeants de Whirlpool, «c'est la qualité de l'interface de Documentum, la plus conviviale de l'industrie, qui a milité en faveur de son acquisition.»

Une conclusion s'impose donc : pour que l'ensemble des travailleurs québécois puissent utiliser les outils leur donnant accès à l'information, il sera nécessaire de rendre ces derniers aussi simples à utiliser que la voiture, le téléphone ou le séchoir à cheveux. En d'autres mots, pour que le savoir soit bien géré, il sera nécessaire de faire en sorte que le slogan adopté en 1933 par les responsables de l'exposition universelle de Chicago, «*Le scientifique trouve, l'industriel applique, l'homme*

s'adapte», devienne plutôt : «*Le scientifique trouve, l'industriel applique et le produit s'adapte.*»

Routage de l'information

Au début du siècle, André Gide soulignait que «le nombre de choses qu'il n'y a pas lieu de dire augmente chaque jour». Gide n'aurait jamais su mieux dire. Dans son *Avis sur les industries de la langue dans la société de l'information*, le Conseil de la langue française relevait ainsi que 5 000 nouveaux articles s'ajoutent chaque jour à un stock en comptant 30 millions (Conseil de la langue française 1994). On estime que le nombre de bases de données textuelles s'est accru de 400% depuis 1985 tandis que le nombre de bases de données numériques est demeuré relativement stable (figure 1). Enfin, on prévoit qu'en l'an 2000, plus de 100 millions d'ordinateurs déverseront dans Internet une information que consommera plus d'un milliard d'internautes (Tapscott 1996). Le résultat de cette croissance : le Québécois moyen n'a aujourd'hui accès qu'à 3% de l'information qui sera mise à la disposition de la prochaine génération.

Le problème, c'est que l'augmentation de la taille du stock informationnel n'a pas été suivie de la mise au point d'outils informatiques permettant une exploration facile et efficace ou encore le classement des données recueillies. Cela, souligne Gérard Théry (1994), «*empêche de parcourir de grandes étendues de savoir au prix d'un minimum d'efforts ou de temps; les interrogations de bases documentaires restent empiriques malgré les progrès de l'informatique; le rassemblement des documents nécessaires à la préparation d'un ouvrage, d'une thèse prend un temps très long; même dans l'entreprise, la collecte des informations nécessaires à la prise de décision, à la définition d'un nouveau produit, reste longue et pénible.*»

Les coûts relatifs à l'incapacité croissante des organisations à fouiller la masse de données existantes et à y repérer l'information nécessaire à leur survie sont élevés. À titre d'exemple, il est estimé que les ingénieurs consacrent entre 45 et 75% de leur temps au repérage de données et de documents plutôt qu'à des tâches plus productives. Dans une veine similaire, au Danemark, 30% du temps de travail des

médecins et des infirmiers est consacré à la rédaction, au classement ou à la recherche de documents et seulement 16% aux soins aux malades. Enfin, 85% des responsables informatiques américains soulignaient en 1995 que les problèmes de gestion de l'information éprouvés par l'entreprise avaient des effets négatifs sur la qualité des produits et des services lancés par l'entreprise, son habileté à détecter la présence de nouveaux marchés et sa capacité à diminuer ses coûts d'exploitation et de fabrication.

Il est évident que de nouveaux outils devront être mis au point pour éviter que nous ne mourrions tous d'*infobésité* et que l'un des actifs les plus importants des sociétés et des organisations modernes devienne l'un de leurs problèmes les plus graves. Pour ce faire, il sera notamment nécessaire de mettre au point des outils de traitement du français capables de repérer aussi bien une chaîne de caractères qu'un concept ou qu'une idée. En effet, la plupart des logiciels existants traitent des unités qui, parce qu'elles ne sont pas linguistiques, sont inadéquates pour représenter le sens : les chaînes de caractères. Cette

méthode basée sur la coïncidence des formes, débouche sur l'occultation par les logiciels qui y recourent de la plupart des phénomènes morphologiques, lexicaux, syntaxiques et sémantiques à l'œuvre dans un texte. Ainsi, les logiciels présupposent que le langage est bi-univoque, c'est-à-dire qu'il n'existe qu'un seul terme pour désigner un objet et que chaque terme ne désigne qu'un seul et unique objet.

L'emploi d'outils faisant le repérage de chaînes de caractères a pu constituer un pis-aller pour la recherche dans les textes réduits, comme les titres et les résumés, mais il génère beaucoup de bruit et aussi beaucoup de silence dans le plein texte (Bertrand-Gastaldy 1991). Le fardeau repose alors entièrement sur l'utilisateur final qui doit, pour formuler une requête fructueuse, penser à toutes les formulations possibles, à toutes les variantes des chaînes à retracer et, dans le cas de corpus multilingues, à toutes les équivalences linguistiques (Lanteigne et Pagola 1994).

La faiblesse de l'approche «chaînes de caractères» explique le fait que les

Tableau 1. La valeur d'un document

Type de documents	Problèmes en cas de perte	Ampleur
Mémoire	Délais d'exécution, perte d'efficacité	Important
Message déposé sur une liste de discussion	Dédoublement d'efforts, perte d'efficacité	Important
Document de soutien à la vente (ex. : brochure)	Perte d'efficacité	Important
Rapport de recherche	Perte de revenus et de profits, perte de parts de marché	Crucial
Politiques de la société en matière de respect des normes de santé et de sécurité	Poursuites, amendes, interruption des activités, accidents	Crucial
Document requis pour l'obtention de la norme ISO 9000	Perte de ventes, interruption des activités, perte du certificat ISO	Crucial

Source : Byte (August 1997) : 72.

logiciels capables de reconnaître le sens des termes recherchés effectuent actuellement une percée sur le marché. Ces logiciels interviennent à plusieurs niveaux du traitement linguistique. En amont, ils analysent les textes afin de les interpréter, de produire des index intelligents et de faciliter la recherche. En aval, ils analysent la question formulée par l'utilisateur (Lubkov 1997, 24). Plusieurs logiciels de ce type sont maintenant disponibles sur le marché francophone, notamment Aleth-IR, Spirit, Sense, Search'97, Sylex et Stil. De plus, on peut noter «la multiplication des accords des éditeurs (Verity et Ever) pour se doter de fonctions de recherche en langage naturel. Synonymies, analogies, proximité sémantique, analyse syntaxique sont déjà opérationnelles. Déjà apparaissent les résumés automatiques (Verity) et la recherche floue (Excalibur de Chemdata) qui procèdent par approximation statistique sur l'orthographe des mots pour retrouver les plus proches.» (Lubkov 1997, 43)

En 1988, Bernard Pivot écrivait en préface à l'ouvrage *La Bibliothèque idéale*, que «Groupés, serrés les uns contre les autres, [les livres] ont la stabilité et la patience des menhirs. Autrefois, les souris, hardiment, les grignotaient. Mais devant la prolifération des couvertures, elles y ont à peu près toutes renoncé». Pour éviter que nos entreprises ne se retrouvent un jour face à la masse d'information véhiculée par les inforoutes telles les souris de Pivot, il faudra assurer qu'elles aient à leur disposition des outils de traitement de l'information de plus en plus puissants.

Outils de traduction

Le nombre de bases de données en français dans le monde constitue un indicateur de la vigueur de l'industrie francophone du contenu électronique : en 1995, à peine 6% des 9 000 bases de données recensées par la société Gale Research étaient disponibles en langue française tandis que 80% étaient produites en anglais. La faiblesse relative du français était aussi évidente dans l'édition électronique scientifique, technique et professionnelle. À titre d'exemple, en 1995, moins de 20% des titres de disques optiques compacts lancés sur le marché français pour répondre aux besoins des entreprises de l'Hexagone étaient produits en langue française.

La situation était particulièrement criante en sciences de la nature, où moins d'un titre sur 10 était offert en français (CEVEIL 1996).

En sondant 30 millions d'adresses à l'aide d'un logiciel d'identification automatique de langue développé conjointement par Alis Technologies et le CITI, l'Internet Society a en outre déterminé que 84% des pages Web accessibles par Internet sont rédigées en anglais. L'allemand vient en deuxième place, loin, très loin derrière, à 4 %; le japonais est troisième, à 1,6%; le français, quatrième, à 1,5%; l'espagnol vient en cinquième place, à 1,1%⁶.

On pourra regretter que la grande majorité des écrits produits dans le monde sous support électronique le soit en anglais, mais l'information étant désormais une ressource d'importance stratégique, il serait suicidaire pour les francophones de s'en tenir strictement à l'analyse de celle produite dans leur langue. Afin de demeurer concurrentielles, les entreprises québécoises devront pouvoir accéder à l'information électronique produite autant en anglais que dans les autres langues. Inversement, des passerelles devront être construites pour que les produits francophones pénètrent les marchés étrangers.

Malheureusement - ou heureusement - le traducteur automatique qui permettra de traduire parfaitement Tom Peters, *Der Spiegel* ou le *Quotidien du peuple* de Chine n'est pas encore né, pas plus, d'ailleurs, que l'interprète informatique qui permettra à un francophone de converser dans sa langue avec les 75% de gens d'affaires européens ne sachant pas l'anglais (Stamm'ler et al., 29). Malgré les ressources souvent considérables consacrées à la mise au point de ces outils par les entreprises et les gouvernements des pays les plus industrialisés, la réalité, en matière de traduction ou d'interprétation, ressemble encore trop souvent à l'exemple ci-dessous de traduction inappropriée.

· The lift is being fixed for the next day. Under that time we regret that you will unbearable. (panneau sur un ascenseur)	· Specialists in women and other diseases, (affiche sur la porte d'un médecin)
· Our wines leave nothing to hope for. (dans un resto)	· Customers will be executed in good order (dans une boutique)
Source : Stamm'ler et al. <i>La toile de Babel</i>, p. 27.	

Cependant, les bénéfices énormes susceptibles d'être retirés de l'utilisation des traducteurs automatiques font que cette solution demeure séduisante dans le cadre des activités de veille des entreprises⁷. Pour des raisons stratégiques, les forces armées américaines se servent par exemple de systèmes de traduction automatique afin de traduire les titres et les résumés d'articles diffusés en ligne par les serveurs du monde entier. C'est également dans le but d'assurer une veille technologique de qualité que de très nombreuses organisations japonaises et allemandes recourent à la traduction automatique.

Dans une autre veine, les entreprises pourront mettre à la disposition des employés disposant d'une connaissance passive de l'anglais (ou de toute autre langue étrangère) des outils de repérage capables de recevoir une requête en français, de traduire cette requête en anglais, d'explorer Internet à la recherche des textes anglophones pertinents et de renvoyer

6. Voir notamment www.alis.com.

7. Les lacunes des traducteurs automatiques n'empêchent pas, par ailleurs, certaines entreprises de les employer pour traduire leur documentation. À titre d'exemple, la société Caterpillar, célèbre pour ses véhicules lourds, mise sur Kant pour traduire en 35 langues les 100 000 pages de documentation qu'elle produit chaque année. Quant à la compagnie Siemens, elle emploie la traduction automatique pour respecter les règles de l'Union européenne voulant que tout produit vendu sur son territoire soit accompagné de documents écrits dans la langue de l'acheteur. Conscientes que les résultats de la traduction automatique demeurent très imparfaits, ces grandes entreprises travaillent cependant à l'élaboration de systèmes permettant de contrôler la langue de rédaction des textes destinés à être traduits (le rédacteur de textes techniques doit se plier à certaines règles qui faciliteront la tâche de la machine). Elles affectent également à la révision des textes traduits automatiquement des ressources hors de la portée des entreprises de petite taille.

ces derniers à l'utilisateur. Théoriquement, un usager québécois pourra donc transiger en français avec les ordinateurs anglophones de Gartner au moment de commander un texte. Un Mexicain n'ayant qu'une connaissance passive de la langue française pourrait quant à lui explorer les bibliothèques du Québec à l'aide de mots clés formulés en espagnol. Des outils comme ceux-là existeront très bientôt: Lerner & Hauspie, une entreprise belge dans laquelle Microsoft acquerrait récemment une participation de plusieurs dizaines de millions de dollars, a ainsi annoncé le lancement, pour la fin de l'année, d'un outil de repérage multilingue allemand-espagnol-anglais. Au Québec, une technologie similaire a pour sa part été mise au point par ZA, une filiale de la défunte SIT Technologies, pour les entreprises situées sur le territoire de l'ALENA.

Repérer ne suffit pas

Quiconque s'est servi une fois d'Altavista ou d'Infoseek pour faire un travail de recherche est en mesure de comprendre intuitivement l'importante différence existant entre des mots comme «données», «information» et «savoir». Ainsi, les 32 546 pages Web contenant les mots clés recherchés par un utilisateur sont des données, une matière brute sous laquelle il est facile de périr enseveli. De leur côté, les 28 pages présentant un intérêt pour le chercheur après une lecture rapide constituent de l'information, c'est-à-dire le produit de l'arrangement des données recueillies. Quant au savoir, il consiste finalement dans la transformation de l'information, dans son utilisation productive à l'intérieur d'un contexte donné.

Il est plus difficile d'interpréter une température que de la prendre, de jouer avec virtuosité de la trompette que de produire des sons, d'écrire un beau vers que de transcrire les mots d'un dictionnaire. De la même façon, «*Knowledge seekers are looking for insights, judgments and understanding in their searches, not the GNP of France or a list of the 20 largest U.S. banks. Knowledge seekers want answers to questions such as 'What is this particular client like ?' or 'How did we manage to win that sale ?' that require complex answers - answers imbued with all the emotional subtexts so important to*

oursensemaking.» (Davenport and Prusak 1997)

La bonne nouvelle, de nos jours, c'est qu'avec les nouvelles technologies de l'information, il n'a jamais été aussi facile de trouver le produit national brut de la France, de dresser la liste des 20 plus grandes banques américaines ou celle des 20 acteurs nains les mieux payés, de déterminer le chiffre des ventes d'une entreprise par zone postale, de déterminer la fréquence des suicides selon l'âge, le sexe, la profession, la langue maternelle ou la race ou de mettre la main sur les milliers d'articles de revues contenant les mots «secret» et «bonheur». La mauvaise nouvelle, c'est qu'il n'a jamais été aussi difficile de faire quelque chose d'intelligent de toutes ces données. En fait, contrairement à ce qu'on pourrait penser, il n'y a pas de corrélation parfaite entre la qualité d'une décision et la quantité de données sur lesquelles elle s'appuie. Au contraire, il vient un moment où obtenir plus de données nuit à la précision d'une décision, comme le montre la figure 2⁸.

À plus ou moins long terme, l'amélioration de la technologie pourra constituer une solution fort intéressante à ce problème. À titre d'exemple, de multiples entreprises et centres de recherche, dont certains sont localisés au Québec, travaillent actuellement à l'élaboration de systèmes pouvant condenser un texte automatiquement de façon à réduire le temps nécessaire à son évaluation.

Dans une veine un peu différente, des générateurs de textes pourront convertir

de simples données en information stratégique. Mis au point par la compagnie A.C. Nielsen, le système Spotlight accomplit déjà un travail de ce genre. Le logiciel interprète d'abord les chiffres recueillis à distance par les lecteurs optiques installés dans les caisses des supermarchés. Grâce à son module de traitement de la langue anglaise, il génère ensuite automatiquement les rapports faisant l'analyse des données relatives à chaque produit. Un tel outil permet aux clients d'A.C. Nielsen de se concentrer sur l'interprétation de l'information colligée plutôt que sur celle des données disponibles. Il leur permet aussi de donner un *feedback* rapide, par les réseaux, à leurs représentants (Anand and Kahn 1993). Néanmoins, il ne fait pas de doute que, pendant plusieurs années encore, l'essentiel du travail de conversion des données en information et de transformation de l'information en savoir continuera d'incomber aux travailleurs de l'entreprise.

Conclusion

Loin de prendre la place des travailleurs de l'information, les nouvelles

8. La surcharge d'information est un phénomène qui commence à affecter beaucoup de travailleurs et de cadres. Dans un article paru dans le *Times*, 9 December 1996 : 46, David Lewis, un médecin anglais allait même jusqu'à parler de l'existence d'une maladie baptisée «Information Fatigue Syndrome». M. Lewis avançait notamment l'idée que si, il n'y a pas si longtemps, on mettait un travailleur à l'écart en le coupant de l'information, on atteint aujourd'hui le même objectif en lui donnant plus qu'il ne peut en digérer.

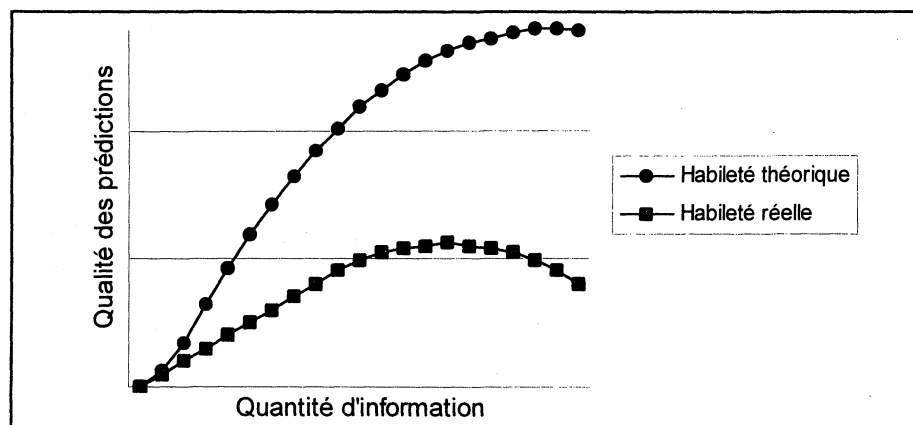


Figure 2.

Source : David, Shank. *Data smog* (San Francisco : HarperEdge) p. 38.

technologies leur permettront de s'engager dans des activités à valeur encore plus grande et partant mieux rémunérées. Les changements qui se produiront dans l'industrie du voyage permettent d'illustrer ce point de vue. À l'heure actuelle, les agents de voyage contrôlent 80% du marché de la réservation de billets d'avions. L'arrivée des inforoutes, en donnant aux compagnies aériennes la possibilité de vendre chaque billet au client le plus offrant, sans intermédiaire pourrait ainsi avoir un effet extrêmement positif sur les profits réalisés par les transporteurs mais pourrait mettre en danger un des principaux gagne-pain des agents de voyage.

Quoiqu'il soit parfois postulé que les agents de voyage deviendront une espèce en voie de disparition au fur et à mesure que les gens achèteront eux-mêmes leurs billets d'avion, on peut penser que la transformation du consommateur en concurrent les forcera à s'engager dans de nouvelles activités à valeur encore plus grande. Ainsi, pour survivre, les agents devront se transformer en véritables spécialistes du voyage et apprendre à exceller dans la cueillette d'information touristique et la création de forfaits de plus en plus alléchants.

Il y a toutes les raisons de penser que les mêmes changements frapperont éventuellement l'ensemble des secteurs économiques et que l'avenir s'annonce rose pour ceux qui, comme les experts en bibliothéconomie, se seront dotés d'une solide expertise dans la gestion poussée du savoir.

Références

- Anand, Tej and Gary Kahn. 1993. Focusing knowledge-based techniques on market analysis. *IEEE Expert* (August) : 19-24.
- Bertrand-Gastaldy, Suzanne. 1991. Comment les logiciels de bases de données bibliographiques et textuelles peuvent-ils répondre aux différents besoins de leurs utilisateurs. In *Les industries de la langue : perspectives des années 1990*. Montréal : Office de la langue française, Société des traducteurs du Québec, p. 502.
- CEVEIL. 1996. *Le français et les technologies de l'information*. Rapport préparé pour le Comité interministériel sur la situation du français au Québec. Québec : Secrétariat à l'auto-route de l'information, p. 14. Document reprographié.
- Cornella, Alfons. 1997. *La culture de la informacion como institution previa a la sociedad de la informacion*. www-acad.esade.es/~cornella/aprats.pdf.
- Conseil de la langue française. 1994. *Avis sur les industries de la langue*. Québec : Le Conseil, p. 12.
- Davenport and Prusak. 1997. Information ecology. Cité dans Peter Tierney.
- Lanteigne, Diane et Gracia Pagola. 1994. L'indexation à l'ère de la bibliothèque virtuelle. *Argus* 23 (3) : 17.
- Le Soleil* 8 janvier 1996 : C7.
- Lubkov, Michel. 1997. L'abc du langage naturel. *Archimag* n° 103.
- Nonaka, Ikujiro and Hirotaka Takeuchi. 1995. *The knowledge-creating company*. New York : Oxford University Press, p. 220.
- Quinn, James B., Philip Anderson and Sydney Finkelstein. 1996. Managing professional intellect. *Harvard Business Review* (March-April) : 71-80.
- Roy, Réjean. 1997. *Les outils d'aide à la rédaction*. Québec : Conseil de la langue française.
- Roy, Réjean et Pierre Georgeault. 1995. *Des industries de la langue et des autoroutes de l'information*. [Québec] : Observatoire québécois des industries de la langue, p. 3.
- Stamm'ler et al. La toile de Babel? In *The IPTS report*.
- Tapscott, Don. 1996. *The digital economy*. Toronto : McGraw-Hill, p. 16.
- Théry, Gérard. 1994. *Les autoroutes de l'information*. Paris : La documentation française, p. 15.
- Tierney, Peter. 1997. *Knowledge management : Using CBR to leverage what your organization knows*. www.inference.com. section des *White Papers*.



resAnet une interface installée sur le Web du catalogue de la Bibliothèque nationale du Canada (BNC). L'interface **resAnet** fournit de courtes notices décrivant des documents de la collection de la Bibliothèque nationale.

Les caractéristiques de **resAnet** comprennent entre autres :

- des liens hypertextes à des ouvrages, des vedettes-matières et des documents connexes dans la collection de la Bibliothèque nationale;
- la capacité de faire le tri des résultats par titre, par auteur et par date ;
- la fonction de recherche par mots clés ;
- la capacité de limiter les recherches par date, par langue, par lieu de publication ou par type de document.

On peut y avoir accès gratuitement par l'entremise du Service Web de la BNC à :

www.amicus.nlc-bnc.ca/resanet



Bibliothèque nationale
du Canada

National Library
of Canada

Canada