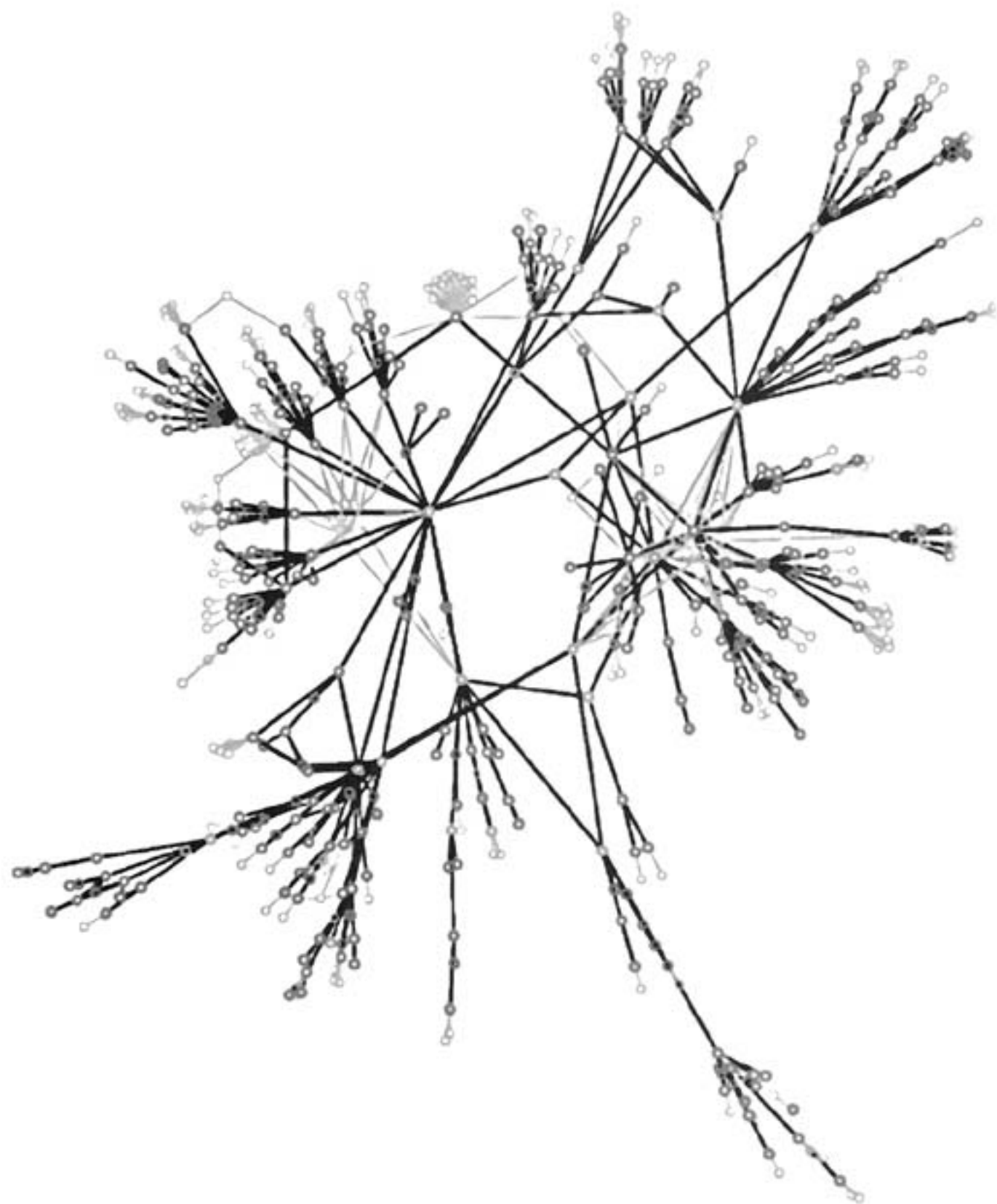


Les Cahiers de médiologie 11

N ° 11 - P R E M I E R S E M E S T R E 2 0 0 1

Com \ Trans
muniquer \ mettre

Revue publiée avec le concours du
Centre National du Livre



KARINE DOUPLITZKY

Internet : une nouvelle écologie du savoir

Une des premières caractéristiques du réseau est son ouverture à l'échelle planétaire. Dans le cadre de cette journée consacrée à la mondialisation, se pose la question des flux et stocks de savoirs, ce qu'ils deviennent dès lors qu'ils sont mis en ligne et ne restent pas confinés dans le cercle clos des bibliothèques. En résumé, qu'est ce que le changement technologique annoncé avec l'ère Internet modifie en termes de création de contenus, de classification et d'accès à ce contenu ? Cette réflexion propose ainsi de mettre en regard la fonction de médiatisation planétaire du Web et sa possible fonction de transmission culturelle.

Carte
topologique
d'un nœud de
communications
sur Internet
générée par
« Mercator »,
outil développé
dans le cadre du
Scan project par
R. Govindan et
A. Reddy,
*Information
Sciences
Institute*, USA.

Internet, Bibliothèque de Babel ? Un changement d'échelle

Internet, à première vue, concrétise le projet utopique d'une bibliothèque universelle qui diffuserait partout une information produite localement. Il fournit un accès aux catalogues du monde entier et à la fois recueil de nouvelles connaissances – chaque usager pouvant déposer sa propre production dans les rayonnages de cette Bibliothèque de Babel. Cependant naissent quelques inquiétudes face à ce raz-de-marée documentaire qui nous submerge et on peut penser avec

Gide : « le nombre de choses qu'il n'y a pas lieu de dire augmente chaque jour ».

Donc première constatation, Internet suppose un changement significatif d'échelle. Quelques chiffres en témoignent (début 2000) :

- 200 millions d'internautes dont 6 millions de Français, 10 millions de sites Internet (organisme *Netcraft*), 1 milliard de pages (Étude du *Gartner Group*) et 180 millions d'images (*NEC Research Institute*) – sans compter les e-mail, newsgroups, etc.

- Une information qui double tous les neuf mois (source IDC).

- 5 000 nouveaux articles français/jour (Conseil de la langue française, 1994).

- 40 000 bases de données (restent invisibles celles qui sont non indexées)

1600 moteurs de recherche.

- 42 % seulement du Web couvert par l'ensemble de ces moteurs de recherche (*NEC Research Institute*, 1999).

L'augmentation de la taille du stock informationnel n'a pas été compensée avec la même rapidité par le développement d'outils permettant d'explorer efficacement ces nouvelles bases de données. D'où l'interrogation : comment guérir cette « infobésité » ?

Sur Internet se sont au contraire complexifiés les problèmes de nomenclature, au sein d'un corpus de documents de plus en plus diversifié – multidisciplinaire, non normalisé, et extrêmement évolutif.

On pourrait utiliser pour Internet la métaphore de la mine d'or, avec ses ramifications mondiales, dont on sait que le sous-sol est très riche mais qui doit être correctement exploité pour devenir rentable : cela suppose d'une part d'en connaître la géologie et d'aboutir à une cartographie satisfaisante des veines – comme autant de couloirs informationnels, et d'autre part d'extraire le minerai et de le traiter.

Cela va donc constituer les deux termes de ma réflexion :

- le premier concerne la géographie du réseau et sa cartographie, ce qui nous conduit à introduire des notions de proximité et de visibilité : quelle est ma distance à l'information et quelle est sa réelle accessibilité ? Enfin comment y accéder, selon quel ordre ou classement et avec quelle pertinence ?

- le second terme concerne l'exploitation des données. La bonne nouvelle, c'est qu'il n'a jamais été aussi facile d'accéder à un grand nombre de données (tel le PIB d'un pays, le CA de ses premières entreprises, etc.). La mauvaise nouvelle, c'est qu'il n'a jamais été aussi difficile d'en faire quelque chose d'intelligent ! Certains outils se développent précisément autour de la condensation d'information et de la notion de pertinence.

Proximité de l'information

Un fichier disponible localement se consulte de la même façon qu'un fichier à distance. On navigue sur le Web sans référence à la géographie réelle – sauf à regarder défiler le nom des adresses URL, sur la barre de navigation, qui indique le lieu d'hébergement du site. Face à cette perte de géographie, comment redéfinir des notions de proximité à une information ?

Une étude de *NEC Research Institute* (juillet 1999) montre que dans un échantillon de 800 millions de pages, deux pages choisies au hasard ne seraient en moyenne qu'à 19 clics l'une de l'autre, ce qui constituerait une indication de mesure du diamètre du Web. Cela signifie que pour des agents intelligents qui suivraient de manière pertinente des liens hypertextuels, ils arriveraient au document recherché au maximum en 19 coups, ce qui est assez rapide.

Seconde approche, topologique cette fois. Une étude des sociétés Compaq et IBM, effectuée sur les 500 millions de pages indexées du moteur de recherche AltaVista, prouve que le Web aurait la forme d'un nœud papillon, avec un nœud central, deux boucles de chaque côté, qui en seraient les entrées et les sorties, et des fils qui pendent à partir du nœud. Le nœud serait le centre ville, contenant le tiers des sites. On y navigue soit par liens hypertexte, soit à l'aide de différents moteurs de recherche. La première boucle sert à entrer dans le nœud, par des sites auxquels on a ensuite peu accès à partir du centre ville (ex. : la page d'accueil d'un portail). La seconde boucle contient des sites de fin de parcours, vers lesquels les liens du nœud central aboutissent (ex. : le site des *Cahiers de médiologie*).

Enfin il existe des fils qui pendent et représentent les sites déconnectés du contenu principal, seulement accessibles de façon périphérique, estimés à environ 20 % du Web.

Voici pour la représentation du Web, maintenant comment naviguer dans cet espace ?

Visibilité de l'information : les moteurs de recherche

85 % des internautes ont recours à un moteur de recherche ou à un annuaire pour trouver de l'information. Mais quelle est l'accessibilité effective de l'information obtenue par ces outils ? Il faut tout d'abord distinguer annuaires et moteurs de recherche. Dans les annuaires, on cherche par rubriques et sous-rubriques et on obtient ainsi l'adresse de plusieurs sites qui semblent correspondre à sa recherche. Dans un moteur de recherche, on inscrit généralement sa re-

quête dans la barre de recherche à l'aide de plusieurs mots-clés qui cernent le sujet. Les moteurs de recherche explorent ensuite leurs bases de données indexées à partir des sites Web du monde entier, et repèrent dans chaque page la position moyenne des mots clés demandés, afin d'établir des coefficients de pertinence du site et de le positionner dans une liste de réponses.

Malgré les progrès technologiques constatés ces dernières années sur les moteurs de recherche, une récente étude de l'institut de recherche NEC prouve que ces outils ne fournissent pas une représentation objective du territoire virtuel, du fait qu'ils se basent essentiellement sur l'indexation des pages, et en oublient la notion de maillage. Ils ne recensent encore qu'un faible pourcentage du Web, laissant des zones obscures, et rendent visibles les sites de façon inégale :

- d'une part, chaque moteur n'est capable d'indexer qu'une faible partie du Web (en moyenne 16 %) : AltaVista (15,5 %) ; Hotbot (11,3 %) ; Infoseek (8 %) ; Google (7,8 %) ; Yahoo (7,4 %) ; Lycos (2,5 %) ;

- d'autre part, ils n'indexent pas de manière égale les contenus, avec du retard de plusieurs mois (l'âge moyen des pages répertoriées est de six mois ; 5 % à 15 % des liens pointent vers des pages périmées ou déplacées). Les robots qui suivent les pages de lien en lien sont plus aptes à trouver celles vers lesquelles pointent de nombreux liens. Les pages « populaires » sont donc plus faciles à repérer. Il y aurait là un risque évident de pénaliser les sites les moins vus et de creuser l'écart entre sites visibles et sites invisibles.

On arrive à une première conclusion : la réussite d'un site marchand serait dépendante de sa correcte indexation par les moteurs de recherche – ou présence dans le sommaire d'un annuaire. On voit, grâce à cette tentative de cartographie des connaissances sur le Web, se dessiner un des enjeux de la mondialisation : rester visible !

Auto organisation

En quoi de nouveaux outils de navigation comme les moteurs de recherche modifient-ils notre mode d'accès aux connaissances ou nos modalités de communication, et comment la nature de l'espace virtuel influence-t-elle sur l'organisation des savoirs ? La grande révolution que l'Internet propose en terme d'organisation des données est le passage d'un système manuel et fermé de classification à un système automatique, dans un espace ouvert et en perpétuelle évolution. On a déjà été amené à distinguer en effet les annuaires (résultat d'une classification manuelle, comme Yahoo) et les moteurs de recherche qui induisent leur propre classification. Pour la première fois dans l'histoire, l'effort de classification – chère aux encyclopédistes – est délégué aux robots.

Cela entraîne une modification du rôle de documentariste, intermédiaire chargé de traduire en langage documentaire la requête d'un utilisateur. Cette tâche est aujourd'hui de plus en plus attribuée à la machine : des systèmes informatiques jouent le rôle de courtiers en information (*knowledge brokers*) et reformulent la requête de l'utilisateur exprimée en langage naturel à l'aide de mots-clés ou autres concepts reconnaissables par la machine. Certains peuvent aller jusqu'à analyser le contenu des textes, les filtrer, ou proposer un système de veille documentaire. Ainsi cohabitent plusieurs images du réseau qui s'appuient sur différents principes de navigation et induisent différentes géographies. Quelques autres exemples permettent de montrer qu'Internet propose de nouvelles formes organisées de savoir qui ne s'excluent pas, mais contribuent à former diverses représentations du réseau.

Dans les annuaires comme Yahoo, l'architecture arborescente est abandonnée au profit du graphe : ainsi un même contenu peut être classé dans plusieurs rubriques différentes – ce qui n'avait pas de sens tant qu'on avait affaire à un objet matériel comme le livre, qui ne se trouvait que sur une étagère à la fois. Non seulement un élément virtuel peut être classé dans plusieurs rubriques, mais son contenu peut y être résumé, fragmenté et dispersé.

Le site *askjeeves.com*, plutôt que de classer les savoirs par rubriques, propose d'y donner accès sous la forme de grandes questions. Ce concept d'organisation du savoir est repris sur de nombreux sites qui proposent une rubrique FAQ (frequently asked questions), listant les questions et réponses les plus demandées. D'autre part, les *Webrings* – comme celui de Netscape ou de Lycos – proposent une nouvelle classification des savoirs basée autour du volontariat. Chaque utilisateur intervient dans son champ d'expertise et se propose comme animateur d'un anneau, qui référence un certain nombre de sites et s'enrichit des apports d'autres utilisateurs concernés par ce même anneau. On peut également citer le site *déjànews.com* qui permet d'effectuer des recherches ou d'établir des portraits d'internautes au sein des *newsgroups* – groupes qui génèrent des fils de discussion et formes de savoir anarchiques, où se reconstituent des clivages idéologiques forts au risque du bruit de fond. Enfin certains sites, comme celui de la BNF, proposent des systèmes de recherche propriétaires, autorisant un accès à des catalogues construits autour d'un corpus fini de documents. L'enjeu reste la compatibilité de ces différents systèmes universels de classification entre eux : comment établir une équivalence entre les classes de la DDC (classification décimale de Dewey), de la CDU (classification décimale universelle), de la LCC (Bibliothèque du Congrès), de la BC (*British classification*) avec les sujets de la ICC (*Information Coding Classification*) ?

La force et la faiblesse d'Internet tiennent à cette coexistence de plusieurs

systèmes de classification des savoirs, qui fournissent un ensemble de grilles d'interprétation contribuant à cartographier le Web.

Les stratégies de la visibilité

L'image de la mine des savoirs, évoquée précédemment, n'est pas innocente, car aujourd'hui la prospérité d'un état ne dépend plus tant de la richesse de son sous-sol en matières premières, que de la capacité qu'ont ses entreprises à produire et à traiter des informations. L'immatériel prend une part croissante dans l'économie mondiale (cf. Cornella, 1997) : par exemple, les États-Unis exportent plus de services éducatifs (6,1 milliards de dollars), de services financiers (5,4 milliards de dollars), de services informationnels (2,6 milliards de dollars), de services juridiques (1,4 milliard de dollars) que de maïs (5,7 milliards de dollars), de blé (4,6 milliards de dollars), ou encore d'aluminium (1,2 milliard de dollars).

On estime à 83 % la proportion de contenu commercial sur le Web. On trouve ensuite des sites relatifs à l'éducation, au monde scientifique et à la santé, puis des pages personnelles et des sites à caractère institutionnel. Un ingénieur passerait 45 % à 75 % de son temps de travail à la recherche de documents et un médecin 30 % dans la rédaction et consultation de dossiers. La technologie de la recherche documentaire devient donc un enjeu stratégique et économique prioritaire pour une entreprise. L'information triée et pertinente étant une ressource stratégique, il n'est pas étonnant que les entreprises se penchent sur les questions de gestion du savoir et que les sociétés de conseil développent des théories sophistiquées de *knowledge management*. L'enjeu est de se rendre visible et de détourner les outils de navigation à son profit.

Le risque existe d'un cybermonde taillé à l'image des entreprises, s'appuyant sur des logiciels qui enregistrent les déplacements et requêtes de la clientèle et par conséquent réfléchissent une image consommatrice du réseau. Ceci recentrerait la vision du Net informatif autour d'une vision consumériste. Par exemple les *newsgroups* – espace public de discussion – dans lesquels les encarts publicitaires sont bannis, seraient délaissés des investisseurs au profit d'un Web plus marchand et plus rentable.

Les consommateurs-producteurs de savoirs

Le réseau propose donc un espace ouvert des savoirs qui s'auto-organisent et évoluent avec leur environnement socio-politique, échappant au monde clos des bi-

bibliothèques. En même temps qu'une mise en commun des connaissances à un niveau mondial, on constate une interdisciplinarité croissante des sujets qui nécessite une révision des grands systèmes de classification s'appuyant sur des structures universelles de la pensée – héritage aristotélicien qui distingue le sujet de l'objet, classé par catégories abstraites. Une telle évolution exige des principes d'organisation du savoir capables de s'adapter à des connaissances de type informel.

La classification supposait jusqu'à présent la définition d'un ensemble de concepts et de structures qui articulaient ces concepts. La construction de classifications doit être aujourd'hui envisagée dans un cadre plus large, intégrant les fruits d'une évolution perpétuelle ; le savoir s'adapte à son interlocuteur et ne reste plus immuable. Un des concepts clé du réseau est cette obligation de retour sur production. L'interactivité inhérente au réseau encourage chacun à devenir acteur et producteur de formes de savoirs au sein de cette bibliothèque de Babel. Les technologies interactives de la communication intègrent dans l'univers du discours le sujet, qui devient vecteur d'expression.

Transmission des savoirs : du vertical à l'horizontal

La transmission de la culture était jusque-là prise en charge par trois institutions, l'École, la famille et l'entreprise, la famille assurant la transmission de valeurs éthiques, l'École une compétence et l'entreprise un capital économique. Cette forme de transmission liée à un héritage du passé s'apparente à une idée de verticalité dans l'espace (du maître au disciple) et dans le temps (de génération en génération). L'arrivée de réseaux informationnels comme Internet propose un mode de transmission culturel plus horizontal et modifie cette image séculaire d'une transmission hiérarchique des savoirs. Elle suppose au contraire l'autonomie d'un individu par rapport à son corps social, s'exprimant au sein d'un espace public ouvert, dans le cadre d'un savoir partagé qui n'est plus sous l'autorité d'un maître.

L'éducation n'est sans doute plus le premier vecteur de construction d'un savoir dont le mot même change de sens. Un savoir renferme aujourd'hui des significations plus larges qu'auparavant et englobe tous les composants de la vie (pluralité des sources d'information ; multidisciplinarité ; variation des points de vue). Il est envisagé comme le résultat d'une construction collective, interactive, rapide et évolutive et dont l'accès se démocratise avec Internet.

En retour de cette apparente démocratisation, il ne faut pas occulter le risque hégémonique de la technique. Car le cybermonde, expression collective de vécus individuels et subjectifs, propose une représentation subjective et culturelle du

monde réel, comme un spectacle dans lequel le sujet est inclus mais *décorporelisé*, relayé par des prothèses techniques.

Nouvelle écologie du savoir

L'exploitation des données ne s'arrête pas à l'utilisation d'outils d'indexation automatique, mais nécessite l'implication d'utilisateurs et de professionnels qui participent à une écologie du savoir.

Comment, par un enseignement ou usage approprié du réseau, conservera-t-on au Net sa fonction émancipatrice et comment contrer les publicistes tentés de s'appropriier l'espace virtuel pour le transformer en espace publicitaire, au détriment de la fonction de transmission culturelle et de libre circulation des savoirs ? Cette réflexion sur la nouvelle cartographie et écologie des savoirs nous conduit à deux types de préconisations :

- rétablir un esprit critique et ne pas entièrement dépendre de l'outil. Il ne suffit pas d'apprendre à se servir techniquement d'un outil, il faut considérer ses dimensions politiques. Le rôle du pédagogue, par exemple, peut être d'affirmer sa responsabilité dans la transmission du savoir en rétablissant la verticalité de l'information effacée par la dimension communicationnelle du net, c'est-à-dire de hiérarchiser les résultats obtenus automatiquement, de vérifier les sources et la pertinence des documents ;
- ne plus penser le savoir seulement en terme de réception, mais aussi d'action ; le processus de production fait partie intégrante de la fonction d'apprentissage et participe à une forme nouvelle de transmission culturelle à partir d'un centre qui est chacun de nous.

François Schuiten,
planche extraite de
l'album *La Route*
d'Armilia,
Casterman, 1988.