

## La sémiotique documentaire

*William A. Turner<sup>+</sup>, Boris Borzic<sup>+</sup>,  
Mathilde de Saint Leger<sup>+</sup>, Yves Muller<sup>o</sup>*

### Introduction

Une nouvelle forme du travail scientifique est en émergence sur Internet. La notion de "collaboratoire" pose comme allant de soi une recherche médiatisée par ordinateur. Elle a été définie en 1989 et reprise comme l'un des concepts clés du programme national des Etats-Unis visant la mise en place d'infrastructures informationnelles (NII, 1993). Les collaboratoires sont les "centres de production scientifique et technique" sans murs "dans lesquels les chercheurs réalisent leurs recherches sans être limités par leur localisation géographique. Grâce aux réseaux [du type RENATER en France], ils peuvent tout à la fois interagir avec leurs collègues dans d'autres universités ; accéder à des instruments à distance ; partager leurs données et leurs ressources informationnelles et, enfin, accéder aux informations réunies dans les bibliothèques électroniques" [Lederberg, Uncapher, eds, 1989].

Ce texte présentera nos recherches visant le développement d'une sémiotique documentaire adaptée aux nouvelles formes du travail scientifique médiatisé par ordinateur. Le document est défini dans cette démarche comme un tableau de la réalité, représenté d'une part par un certain nombre d'éléments — les mots du texte notés a, b, c,... — et, d'autre part, par les relations entre ces éléments — notées aRb. Nos recherches visent à mieux comprendre la formation des règles de composition du tableau, celles qui confèrent un sens aux mots et aux relations représentés, dans différents types de situation du travail collectif.

L'importance du contexte dans lequel se met en œuvre le travail collectif est souvent sous-estimée en sémiotique où deux perspectives s'opposent : la première traite la relation entre les mots et les choses comme allant de soi conformément à la tradition positiviste ; l'autre considère cette relation, au contraire, comme étant socialement construite. D'un côté, le langage est envisagé comme une copie de la réalité et de l'autre côté, il est vu comme l'expression d'un comportement social

[Fodor, Katz, 1965]. Fait singulier, L. Wittgenstein peut être considéré comme l'un des pères fondateurs des deux approches à la fois. Lors de la publication de ses études sur la logique du langage dans *Tractatus Logico-Philosophicus*, il a employé la notion de "tableau de la réalité" que nous avons reprise ci-dessus [Wittgenstein, 1961a, 2. 1.]. Trente ans plus tard, son intérêt pour le langage a évolué et, dans *Investigations philosophiques*, il l'a abordé comme un comportement social organisé autour des "jeux du langage" [Wittgenstein, 1961b, p. 66]. La première notion permet d'explorer les conditions de vérité des signes (a, b, c, ...) dans les propositions qu'ils forment (aRb), alors que la deuxième conduit à l'idée résumée par J. F. Lyotard : les jeux de langage sont "le minimum de relation exigé pour qu'il y ait société (...) chaque partenaire du langage subit, lors des «coups» qui le concernent, un «déplacement», une altération (...)" [Lyotard, 1979, p. 32-33].

La divergence entre ces deux voies est nette : la première s'intéresse au pouvoir de représentation des mots alors que la deuxième débouche sur la conclusion qu'une telle propriété dépend du pouvoir tout court. Si l'on veut comprendre les *déplacements* que subissent les destinataires de messages, les attributs économiques, politiques et sociaux des partenaires du langage ne sauraient être ignorés. Comprendre les contenus des documents diffusés, c'est non seulement étudier les mots et leur relation avec le monde ; c'est également et peut-être surtout analyser les rapports de force qui organisent les jeux du langage. La circulation documentaire, le fait d'émettre des messages, constitue un effort de modification d'un rapport de forces, soit dans le sens d'une consolidation des pouvoirs en place, soit dans le sens de leur contestation. C'est pourquoi il convient de s'intéresser à l'impact des communications médiatisées par ordinateur sur le développement des jeux du langage de la science : les nouvelles technologies sont un facteur d'accélération de la circulation documentaire et cette circulation peut modifier l'organisation sociale de la science.

J. Bouveresse a montré que les deux voies de recherche ouvertes par Wittgenstein, et qui renvoient à des démarches apparemment opposées comme nous l'avons vu, s'inscrivent en fait dans la continuité d'une même interrogation que nous reprenons à notre compte : comment éviter le recours aux métalangages [Bouveresse, 1973, p. 53-61] ? Wittgenstein n'admettait pas l'idée selon laquelle une proposition, pour être comprise, doit être expliquée : s'il faut donner le sens d'une proposition par une deuxième proposition visant à la clarifier, des régressions à l'infini s'enclenchent qui sont destructrices de toutes connaissances ; autrement dit, le langage scientifique ne saurait parler du langage, il parle des choses qui peuvent être notées a, b, c, ... dans les tableaux de réalité décrivant les relations aRb, comme nous l'avons vu. Des recherches en sémiotique documentaire doivent aider à mieux comprendre la façon d'éviter des régressions à l'infini lors de la fabrication collective des tableaux de la réalité.

## «Montrer sans expliquer»

“Montrer sans expliquer” est, selon Wittgenstein, la propriété essentielle du langage scientifique : “ce qui peut être montré *ne peut pas* être dit” [Wittgenstein, 1961a, 4. 1212, italiques de l’auteur]. Le langage scientifique ne dit rien, il montre la réalité et nous comprenons le sens du tableau projeté, sans besoin d’explication. Comment caractériser le pouvoir de projection du langage scientifique ? L’expression suivante nous permettra de préciser les éléments de construction d’une sémiotique documentaire : “a est un X selon la règle R dans le contexte C”.

Dans cette notation, “a est un X” est l’expression d’une prédiction empirique. Elle établit une relation entre un terme singulier “a” et un concept “X”. Autrement dit, les propriétés de l’objet “a” doivent être semblables à celles de “b”, “c”, “d”, etc., qui sont déjà classés comme appartenant à la catégorie d’objets désignée par le concept “X”. Si tel est le cas, l’énoncé prédictif est vrai, sinon il est faux, mais quel est le fondement de ce classement dans une situation donnée ? Dans l’expression canonique ci-dessus, cette nouvelle question trouve sa réponse dans la règle “R” qui est bâtie sur une relation d’identité telle que “X est Y”. Les théories scientifiques sont tautologiques [Wittgenstein, 1961a, 6. 124], mais sans elles, il serait impossible d’interpréter les résultats d’un travail scientifique. Elles fixent les limites des pratiques classificatoires : la règle “X est Y” permet de considérer “a” sous l’angle de ces relations avec “b”, ou, autrement dit, aRb ; le monde est infiniment complexe et seul le fait de se donner un tel point de vue sur la nature des choses permet de dresser un tableau de la réalité.

K. Popper, par exemple, reconnaît aux chercheurs une liberté totale de conjecture et propose une méthode de mise à l’épreuve technique des prédictions qui en découlent. Le problème qu’il pose est le suivant : dans l’expression “a est un X selon la règle...”, le mot *selon* est source d’ambiguïté. Comment établir une relation entre les pratiques normative et prédictive de la science dans un contexte donné ? Popper répond en mettant en avant l’infrastructure socio-technique qui détermine l’état de l’art à un moment donné. En effet, les compétences des chercheurs évoluent constamment grâce au travail collectif d’amélioration des concepts, des méthodes et des techniques de la recherche. Une prédiction n’est vraie que dans les limites d’une marge d’erreur permises par un état de l’art donné : l’abandon définitif d’une théorie n’est jamais obligatoire. Cela dit, si une prédiction est fautive, les limites empiriques d’une conjecture sont établies, montrant la nécessité de modifier la direction des recherches en cours. Selon Popper, les chercheurs comprennent cette nécessité : leur travail est un effort constamment renouvelé de falsification d’hypothèses débouchant sur des nouvelles conjectures [Popper, 1968].

Gardons de Popper cette idée selon laquelle l'intérêt d'un tableau de la réalité est reconnu dans les limites d'un état de l'art. L'analyse sémiotique la reprend généralement à son compte : les infrastructures socio-techniques sont souvent tenues pour données lors de l'étude du pouvoir de projection du langage scientifique. Notre optique est inverse : peut-on étudier l'écrit scientifique comme un facteur d'évolution des infrastructures de la recherche ?

Cette question est motivée par les recherches de I. Lakatos sur le développement des programmes scientifiques. Selon cet auteur, une théorie n'est pas uniquement à l'origine d'une logique de classification ; elle est également et peut-être surtout une structure d'organisation du travail scientifique. En effet, ce qui retient l'attention de Lakatos est le conservatisme des chercheurs, c'est-à-dire leur détermination à défendre leurs programmes contre les critiques suscitées par des contre-exemples empiriques. Il considère que le poids du passé est aussi important que la réfutation des conjectures pour comprendre le développement scientifique. Au lieu d'abandonner leurs théories, les chercheurs s'attachent à reconfigurer leurs infrastructures socio-techniques afin de mettre le cœur de leurs programmes à l'abri des critiques. Ce cœur est défini par des règles constitutives d'une pratique scientifique notées "X est Y" dans l'exemple ci-dessus, alors que la périphérie d'un programme concerne la réalité des faits (notés "a est un X"). Lakatos montre par des études de cas historiques qu'en règle générale, les chercheurs s'efforcent de faire évoluer leurs dispositifs d'observation empirique afin de laisser le cœur théorique de leurs programmes de recherche intact [Lakatos, 1970, p. 91-196].

Lakatos fait entrer le social dans ses analyses sous la forme d'une hypothèse relative aux motivations des chercheurs : les programmes de recherche montrent le conservatisme des chercheurs et leur détermination à défendre les hypothèses qu'ils tiennent pour acquises contre vents et marées. Si ceci est sans doute vrai dans bien des cas, la situation inverse se présente également : des chercheurs n'hésitent pas à changer de problématique et à se lancer dans la construction de nouveaux programmes. Afin de sortir d'une dichotomie qui range les conservateurs d'un côté et les innovateurs de l'autre, il nous semble préférable d'éviter des hypothèses qui ne sont pas directement accessibles à l'analyse. Comme nous l'avons dit, les documents montrent la réalité, ils ne montrent pas les motivations des auteurs qui les ont écrits. Il faut éviter de faire entrer le social dans les analyses du développement scientifique et technique par voie de motivations, de valeurs ou d'intérêts des partenaires du langage. Une sémiologie de l'écrit scientifique ne dit rien à ce propos, par contre, elle peut servir à dévoiler l'insuffisance des infrastructures socio-techniques, comme nous allons le montrer maintenant.

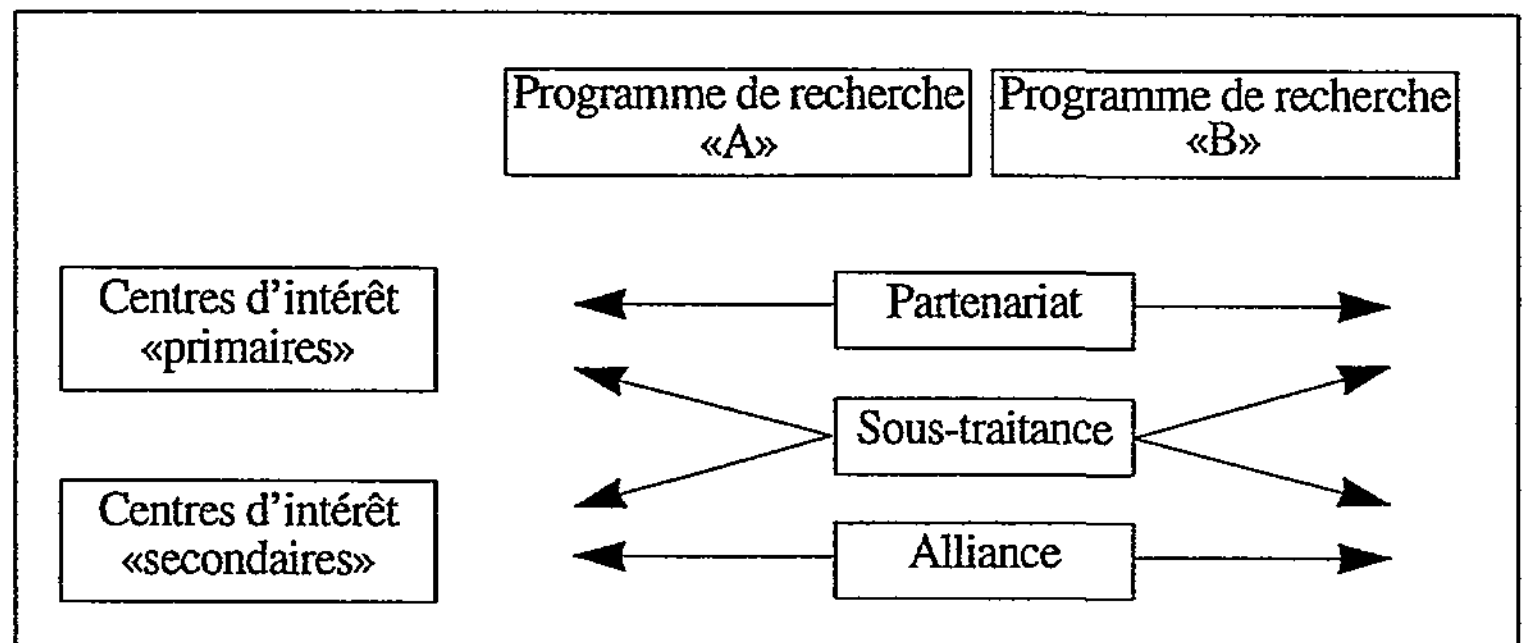
### Le transfert des expertises dans les laboratoires de la science

M. Callon a employé des outils conceptuels de l'analyse économique pour examiner les conditions d'une réfutation empirique des conjectures [Callon, 1980, 1989]. L'analyse économique porte sur les interactions sociales qui s'organisent autour de l'échange d'objets : un consommateur et un producteur entrent en relation par le produit que l'un offre et que l'autre demande. La preuve de la qualité scientifique réside dans le fait que les biens intellectuels produits par un programme de recherche (notés "aRb") circulent, c'est-à-dire qu'ils sont demandés et que les conditions de leur transfert vers des contextes d'usage sont réunies. En effet, seule l'adéquation entre une offre et une demande permet aux défenseurs d'un programme de recherche de réunir les moyens humains, matériels et financiers nécessaires à la construction, au maintien et au développement de leurs infrastructures socio-techniques. Les ressources de la science sont rares, et elles iront vers les programmes qui sont utiles, puisque producteurs de résultats utilisables par d'autres. Comme on peut le voir, Callon s'intéresse aux ajustements qui permettent de construire une demande pour les biens intellectuels diffusés par un programme de recherche, ajustements qui déterminent à leur tour l'évolution de l'appareil productif du programme. Cette évolution n'est pas le résultat d'une série plus ou moins réussie de prédictions empiriques ; elle dépend des moyens matériels et humains disponibles pour mener à bien le travail scientifique envisagé. Sans une infrastructure socio-technique adéquate, il est impossible de produire les biens intellectuels demandés.

Pour s'imposer, un programme de recherche doit prendre place comme un nœud de production et de circulation de savoirs dans un système d'échanges qui peut se limiter à la sphère scientifique, mais qui peut également s'ouvrir vers l'industrie et la société en général [Latour, 1989]. Écrire un texte revient à montrer l'infrastructure socio-technique permettant d'étayer une offre d'expertise ; lire un texte revient à localiser, à l'endroit d'un programme de recherche, l'existence des moyens permettant de traiter un problème. La sémantique de cette localisation est ancrée dans la ressemblance : "a" désigne un centre d'intérêt qui est le mien, mais qui est traité dans un autre texte sous l'angle de ses relations avec "b", et je comprends que pour ce dernier auteur, l'inscription de "a" dans un tableau de la réalité passe par la mise en perspective de la relation "aRb". Si je le comprends, c'est parce que son texte montre l'existence d'une infrastructure socio-technique lui permettant d'œuvrer dans ce sens : l'expertise de l'auteur est ancrée dans un ensemble d'équipements, d'instruments de mesure et d'outils conceptuels et méthodologiques bien déterminé.



Cette conception d'une sémiotique documentaire fait abstraction des règles du type "X est Y" : les interactions sociales ne sont pas régies par des règles ; elles s'organisent autour des objets que les uns produisent et que les autres demandent. Les conditions de l'échange sont déterminées par les coûts de production d'un tableau de la réalité. Différentes stratégies existent permettant de gérer ces coûts de l'apprentissage scientifique. Celles-ci peuvent être décrites en partant de la distinction introduite ci-dessus entre les centres d'intérêt "primaires" d'un programme de recherche et des centres d'intérêt "secondaires". Le tableau suivant résume un classement de situations susceptibles de déboucher sur des collaborations spécifiques entre les membres de deux programmes de recherche "A" et "B" :



**Tableau 1**

**Différentes stratégies de coopération scientifique :  
partenariat, sous-traitance et alliance**

Ce tableau de collaborations distingue trois situations qui apparaissent clairement à la lecture des documents.

Dans un premier cas de figure, le programme de recherche "A" concerne les sujets qui sont au cœur du programme de recherche "B", et inversement. Autrement dit, les membres des deux programmes travaillent à la fabrication d'un tableau de la réalité qui comporte les mêmes éléments (notés "aRb"). De ce fait, leurs infrastructures techniques seront quasiment identiques : les deux programmes seront dotés à peu près des mêmes instruments d'analyse et des mêmes équipements de base, cependant, l'utilisation faite de ces infrastructures sera différente, compte tenu des hypothèses spécifiques sur lesquelles s'appuie chaque programme. Le partenariat serait envisageable, mais la situation peut également devenir concurrentielle. En effet, les deux programmes s'adressent à l'extérieur en offrant un même type d'expertise scientifique. L'environnement externe est un facteur puissant d'organisation du travail coopératif : par exemple, des actions publiques d'incitation à la recherche et à l'innovation peuvent aggraver la concurrence des équipes en les mettant en compétition pour l'accès aux ressources ; ou encore, elles peuvent favoriser des

collaborations en exigeant que les équipes présentent des projets coordonnés en vue d'obtenir des subventions.

La sous-traitance correspond à une situation où les membres du programme de recherche "A" travaillent sur des sujets qui sont d'une importance secondaire du point de vue du programme "B". Dans ce cas, les équipes n'ont pas d'infrastructures communes et réunissent, chacune, une compétence scientifique distincte. Une situation opposée à la sous-traitance envisage le traitement en interne des sujets jusqu'alors laissés aux autres : cette dernière option suppose des investissements en moyens techniques et en compétences humaines que la première permet d'éviter.

L'alliance concerne une situation où les équipes travaillant dans deux programmes différents décident de mettre en commun des moyens d'action techniques et humains qui n'avaient qu'une utilisation marginale en interne. Au lieu d'abandonner purement et simplement les résultats d'un certain nombre d'expériences acquises, une décision peut être prise pour essayer de les faire fructifier dans le cadre de coopérations qui impliquent la mise en commun de moyens d'action marginaux.

On peut synthétiser le tableau de coopérations présenté ci-dessus de manière à faire apparaître le lien que nous nous efforçons d'établir entre le développement d'une sémiotique documentaire et la gestion du travail coopératif.

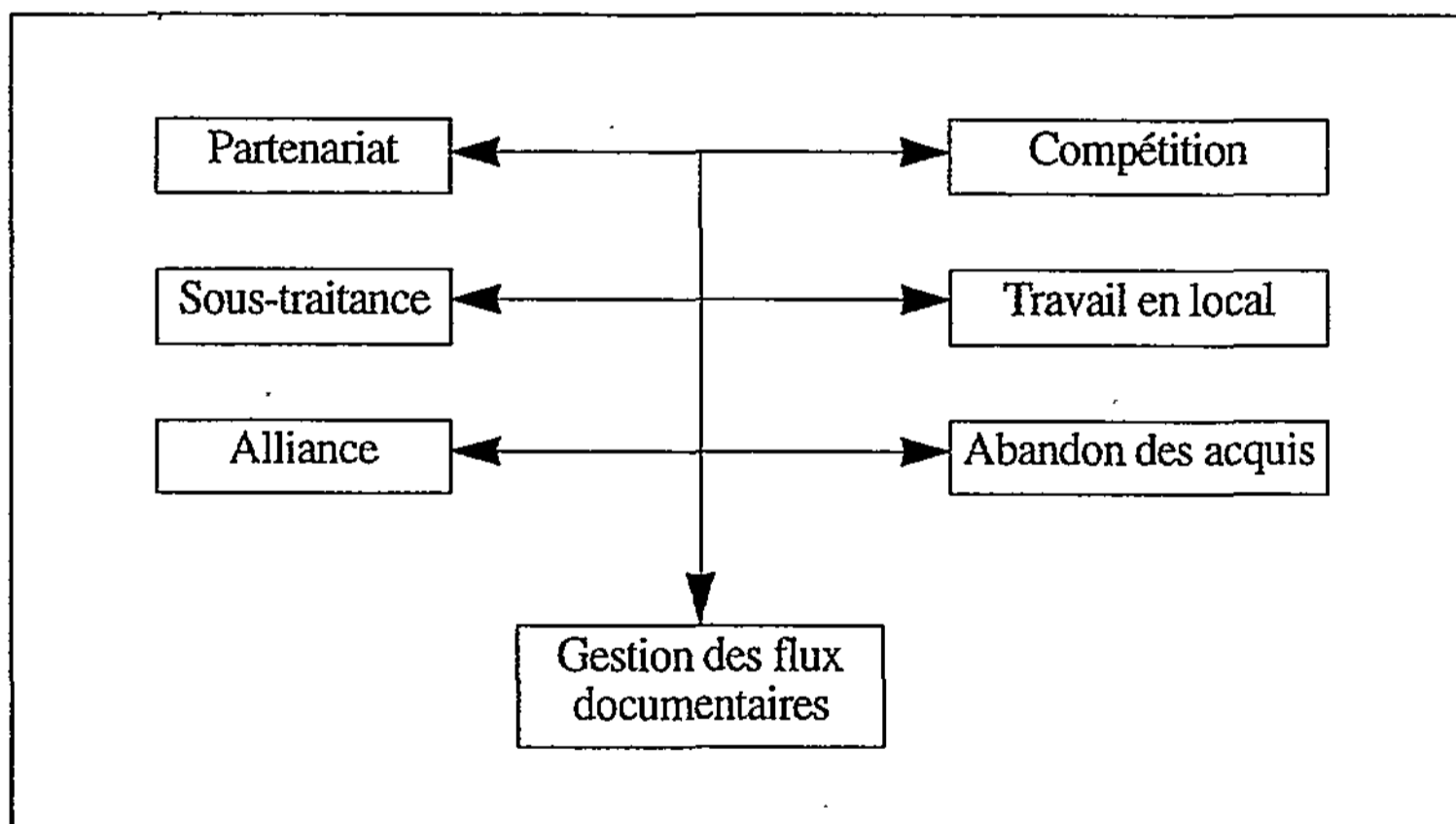


Tableau 2

D'une gestion des flux documentaires dans les laboratoires de la science à la coordination du travail collectif

Un exemple concret des travaux en cours permettra de concrétiser les informations réunies dans le Tableau 1 et le Tableau 2 ci-dessus. Il concerne la mise en place d'un système de gestion documentaire en vue de soutenir un programme de l'UNESCO d'aide à la recherche, appelé MOST (*Management Of Social Transformations*). L'exemple montrera, d'abord, comment les objets notés "aRb" sont produits par des

programmes de recherche distincts pour être automatiquement distribués, ensuite, à l'intérieur des catégories d'agrégation (partenariat, sous-traitance, alliance) qui figurent au **Tableau 1**. L'objectif est d'organiser la discussion des options de coordination du travail collectif visualisées dans le **Tableau 2**. Le lien qui doit exister, à terme, entre l'obtention de ces résultats et leur diffusion sur Internet est encore à l'étude. En effet, beaucoup reste à faire en vue d'aider au développement des partenariats, des sous-traitances et des alliances afin d'éviter que les laboratoires de la science ne soient fragilisés par une concurrence sauvage, des stratégies de repli sur soi, ou des abandons progressifs d'acquis. De nouvelles logiques de communication médiatisée par ordinateur sont à trouver, mais nous n'en sommes pas encore là, même s'il s'agit de la finalité visée par nos recherches.

### Le Programme *Management Of Social Transformations (MOST)* de l'UNESCO

Parmi les objectifs de MOST, celui qui retiendra notre attention ici concerne les recherches sur les dynamiques d'urbanisation, et les transformations sociales rapides qu'elles engendrent dans les villes. Les équipes qui participent à ces recherches sont situées dans différentes régions du monde : en Asie et en Europe, en Russie et aux Etats-Unis, dans les pays du Nord et ceux du Sud<sup>1</sup>.

La mise en place d'un laboratoire à l'échelle mondiale pose des problèmes importants de standardisation et d'interopérabilité des systèmes techniques et, pour les régions qui n'ont pas d'infrastructures de télécommunication avancées, de marginalisation par rapport aux flux d'informations scientifiques et techniques en circulation sur Internet. A cet ensemble de problèmes techniques s'ajoutent des questions qui se rattachent aux contenus. La diversité des langues, le poids des traditions locales, la multiplicité des objets d'étude et des démarches font que, malgré des solutions de mises en communication techniques, les nouvelles technologies de l'information ne constituent pas une panacée pour la diffusion des résultats de la recherche. Par ailleurs, les équipes de recherche retenues dans le cadre de MOST s'appuient sur des hypothèses, des démarches, des méthodes différentes, malgré leur association dans un effort collectif visant à mieux comprendre les transformations sociales dans les villes.

Cette diversité peut nuire à l'objectif de MOST qui est de promouvoir des recherches susceptibles d'aider à la conception de politiques globales permettant de gérer au mieux les transformations rapides dans les villes. Sans espoir de résultats généralisables, la tentation est grande de renoncer

<sup>1</sup>Les informations concernant le programme MOST de l'UNESCO peuvent être consultées à partir du serveur de l'UNESCO : <http://www.unesco.org/most/>



aux efforts de coordination de la recherche et de se contenter d'une stratégie d'allocation des ressources fondée davantage sur des considérations d'ordre politique que scientifique. En effet, une question sensible au sein de l'UNESCO est l'égalité du traitement des équipes situées dans les pays du Nord, du Sud, de l'Est et de l'Ouest. Au lieu d'utiliser les ressources disponibles pour développer les conditions d'un apprentissage collectif, le raisonnement peut vite basculer vers une logique purement comptable visant à trouver une formule de partage des crédits disponibles qui soit équitable, et qui respecte la diversité des intérêts en présence.

Il convient de gérer les flux d'informations en circulation dans les laboratoires de MOST en vue de renforcer les dynamiques d'un apprentissage collectif. Le modèle de gestion proposé ci-dessus présente trois situations, bornées chacune par deux alternatives : le partenariat ou la compétition ; la sous-traitance ou le repli sur soi ; l'alliance ou l'abandon des acquis. L'objectif de nos recherches est de mieux comprendre le rôle que joue la diffusion de l'information scientifique et technique pour la coordination du travail collectif dans les trois contextes identifiés.

Un "Médiateur Intelligent" a été conçu comme une plate-forme d'expérimentation pour ces recherches [Borzic, 1997]. Il capitalise des recherches commencées au Centre de Sociologie de l'Innovation de l'École des Mines il y a bientôt vingt ans, et développées au CNRS dans le cadre des activités visant à cartographier les évolutions de la science et de la technique à l'aide des outils d'analyse documentaire. Ce Médiateur a été réalisé en collaboration avec la Société Euritis, qui commercialise un gestionnaire de documents, le produit AIRS (*Advanced Information Retrieval System*). Nous nous contenterons de résumer les fonctionnalités de ce système, en nous arrêtant plus longuement sur les aspects du Médiateur plus directement en rapport avec les questions de sémiotique documentaire soulevées précédemment.

Les données expérimentales rapportées ici ont été extraites d'une publication préparée par le Programme interdisciplinaire de recherche sur la Ville (PIR-Villes) [*Courrier du CNRS*, 1994]. Les articles parus dans cette publication sont courts (moins de deux pages), et regroupés par grands thèmes tels que "Espaces et lieux", "Temporalités urbaines", qui concerne les mouvements de populations et de générations, et "Ville et santé" qui traite de l'impact de l'urbanisation sur la santé physique et mentale des citoyens. Sur les 80 articles réunis dans ce volume, nous en avons sélectionné 17 qui nous semblaient en rapport étroit avec les transformations sociales dans la ville, afin de simuler l'existence d'un premier fonds documentaire diffusé par l'intermédiaire du programme MOST. L'inventaire des éléments contenus dans ce fonds est présenté dans le **Tableau 3** ci-joint (les 17 articles en question sont regroupés dans les trois premiers thèmes) ; ces éléments représentent environ 25 pages de texte qui ont été installées sur le serveur Internet du CNRS à Meudon en

se servant à cette fin de la plate-forme technique du Médiateur. L'environnement du Médiateur comporte une station d'indexation assistée par ordinateur et un gestionnaire des listes d'autorité permettant de fabriquer des répertoires normalisés d'auteurs et de laboratoires. C'est ainsi que nous avons pu produire la maquette d'un Guide des Contenus et des Compétences de MOST en matière de recherches sur la ville. Le Médiateur a permis de traiter les textes et d'en extraire un lexique des centres d'intérêt des articles publiés (notés "a", "b", "c", ...).

Si le Guide est un outil de diffusion de l'information scientifique et technique, il est également et symétriquement un outil de détection des collaborations éventuelles, à condition d'être conçu à cette fin. La sémiotique documentaire doit aider à maintenir à jour un inventaire de l'expertise réunie dans un laboratoire scientifique. Le prototype du système documentaire de MOST verra les chercheurs du Caire, de Rio de Janeiro, de Paris et de Vancouver s'équiper de sites *Web* pour devenir les premiers nœuds d'un réseau de recherche internationale sur les villes. Pour les besoins de l'expérimentation décrite ici, le serveur de l'UNESCO a été mis en communication avec le serveur du CNRS à Meudon.

Cinq articles publiés par le Centre de Recherches sur les Stratégies territoriales et les dynamiques de l'espace (STRATES) ont été installés sur le serveur de l'UNESCO et exploités depuis le serveur de Meudon à l'aide de l'environnement du Médiateur. Leurs titres et auteurs sont également précisés dans le **Tableau 3** (ci-contre). L'expérience devait permettre de :

- traiter les publications du laboratoire STRATES en vue de déterminer la spécificité de l'offre d'expertise de ce laboratoire ;
- situer cette offre par rapport aux compétences répertoriées dans le Guide de MOST, lui-même constitué à partir des publications du PIR-Villes ;
- mettre au point des stratégies de collaboration en vue de nouer des relations de partenariat, de sous-traitance ou d'alliance entre les deux communautés de chercheurs.

### **De la sémiotique documentaire à l'organisation du travail coopératif**

Chaque document est indexé par le Médiateur en une liste de mots contenus dans un lexique et qui représentent les centres d'intérêt de l'auteur. Le progiciel documentaire, AIRS, permet de repérer automatiquement les paragraphes d'un texte retenus comme unités de sens pour le calcul des cooccurrences. Une relation entre deux mots "a" et "b" est comptabilisée si ces mots apparaissent ensemble dans le même

| Titre de l'article  | Auteur                     | Laboratoire  |
|---|----------------------------|--|
| <i>Espaces et lieux</i>                                   |                            |  |
| • "Ville et démographie"                                  | Hervé le Bras              | Centre de recherches historiques (CNRS/EHESS)                            |
| • "Divisions sociales de l'espace"                        | Edmond Préteceille         | Centre de sociologie urbaine (CNRS/Iresco)                               |
| • "Les quartiers en difficulté"                           | Sophie Body-Gendrot        | Institut d'études politiques (Paris)                                     |
| • "Des grands ensembles aux quartiers"                    | Michel Péraldi             | Modes de production et environnement social (CNRS-Toulouse)              |
| <i>Les temporalités urbaines</i>                          |                            |  |
| • "Villes et générations"                                 | Jean-Luc Pinol             | Cultures, arts, sociétés des villes européennes (CNRS-Strasbourg)        |
| • "Parcours urbains de deux générations de Parisiens"     | Françoise Cribier          | Géographie sociale et gérontologie (CNRS-Paris)                          |
| • "Migrations et citadinité"                              | Alain Tarrus               | Centre de recherches juives (CNRS-Toulouse)                              |
| • "La ville et les jeunes"                                | Thierry Blöss              | Socialisation, interaction et identités sociales (CNRS-Sophie Antipolis) |
| <i>Ville et santé</i>                                     |                            |  |
| • "Sociétés, environnement et santé : l'exemple africain" | Claude Raynaut             | Sociétés, Santé, Développement (CNRS-Bordeaux)                           |
| • "Accessibilité et handicap"                             | Jésus Sanchez              | Centre d'études sur les handicaps (Vanves)                               |
| • "Vieillir en ville"                                     | Serge Clément              | Centres d'études urbaines (CNRS-Toulouse)                                |
| • "Le sida, un problème urbain"                           | Bernard Paillard           | Centre d'études transdisciplinaires (CNRS-Paris)                         |
| • "Les inégalités face à la santé en Grande Bretagne"     | Sarah Curtis               | Université de Londres  |
| • "Homelessness, maladie mentale et espace urbain"        | Anne Lovell                | Université de Toulouse-Le Mirail   |
| • "Fragilités et résistances dans deux métropoles noires" | Gilles Bibeau, Ellen Corin | Université de Montréal et McGill University                              |
| • "Urbanisation et santé dans le tiers monde"             | Gérard Salem               | ORSTOM-Montpellier   |
| • "Démocratie locale et politiques publiques de santé"    | Michel Joubert             | Le Social et la Sociabilité (CNRS-Iresco)                                |
| <i>Stratégies territoriales et dynamiques de l'espace</i> |                            |  |
| • Des espaces périurbains                                 | Martine Berger             | Strates, CNRS-Paris  |
| • La notion de ségrégation et son usage en géographie     | Jacques Brun               | Strates, CNRS-Paris  |
| • La ségrégation et ses mesures                           | Catherine Rhein            | Strates, CNRS-Paris  |
| • L'espace parisien et son évolution                      | Catherine Rhein            | Strates, CNRS-Paris  |
| • Stratégies résidentielles                               | Catherine Rhein            | Strates, CNRS-Paris  |

**Tableau 3**  
Fonds documentaire analysé

paragraphe d'un document. Les paires de relations, notées "aRb", peuvent être classées par ordre décroissant en fonction du nombre de leurs occurrences dans un document donné. Cependant, ce classement doit être normalisé en contrôlant la fréquence d'apparition de chaque terme pris individuellement : en effet, il faut éliminer le biais statistique inhérent au fait qu'un mot fréquent dans un texte aura des associations plus nombreuses qu'un mot ayant une fréquence moindre [Michelet, 1988].

Afin de visualiser des informations contenues dans un flux de documents, des méthodes de cartographie scientifique sont utilisées [Turner *et al.*, 1984]. Ces méthodes exploitent (entre autres) les techniques d'agrégation dites du "saut minimal" afin de réunir deux objets (notés "aRb") dans un nouvel ensemble, dit *cluster* ou agrégat. Le problème est d'obtenir une liste d'agrégats de tailles comparables. On cherche en effet à maximiser l'information obtenue dans cette partition, c'est-à-dire essentiellement à obtenir des agrégats nombreux et de taille raisonnable : trop grand, l'analyse manque de finesse et les agrégats risquent d'être hétérogènes, trop petits, l'information obtenue est trop atomisée pour que les résultats soient exploitables.

Le programme utilisé permet d'obtenir des découpages d'agrégats ajustés en fonction de leur taille désirée. Ceci est obtenu en considérant en priorité les liens aRb les plus forts, en poursuivant l'ajout de liens et de mots à chaque agrégat tant que la taille n'est pas suffisante, et en "saturant" l'agrégat dès qu'un ajout supplémentaire entraînerait un dépassement du seuil. Les agrégats ainsi obtenus peuvent être caractérisés par l'indice de la première liaison refusée, que nous appelons indice de saturation. Ce seuil est donc variable d'un agrégat à l'autre, et est déterminé par la taille de l'agrégat en cours de calcul.

Le résultat de la partition est une série d'agrégats, chaque agrégat réunissant un groupe de mots qui apparaissent plus fréquemment ensemble qu'avec les autres mots du lexique, ces derniers étant distribués sur les autres agrégats de la série calculée. Le groupe de mots réuni dans un agrégat définit les liens aRb entre les termes internes à l'agrégat. Cependant, sachant que cette utilisation des partitions ne correspond pas réellement à une coupure brutale entre agrégats, mais plutôt au regroupement de termes les plus caractéristiques d'un usage social des mots pour parler des choses, il est indispensable de préserver l'information portant sur les liaisons entre termes faisant partie d'agrégats différents : afin de pouvoir chaîner ces agrégats, le programme fournit donc une liste de liaisons permettant de passer d'un agrégat à l'autre, soit par des liaisons tombant sous le seuil de l'agrégat (liaisons inférieures), soit par des liaisons de seuil supérieur au seuil d'arrêt de l'agrégat (liaisons supérieures). La prise en compte de ces liaisons externes (inférieures et supérieures) se fait jusqu'à un nombre fixé *a priori* (généralement, vingt).

Chaque agrégat peut maintenant être classé en fonction de deux échelles [Turner *et al.*, 1985, 1988], la première correspondant à la densité

interne des agrégats, la seconde à la force de ses liens externes. La somme des valeurs des liens internes à l'agrégat donne l'échelle de densité : les agrégats caractérisés par les plus fortes sommes de valeurs internes sont placés en haut de l'échelle, et ceux associés aux plus faibles sont placés en bas. De même, pour ce qui concerne la force des liaisons externes d'un agrégat, la somme des valeurs de ces liaisons offre une mesure de la centralité des agrégats les uns par rapport aux autres : plus la somme des valeurs des liaisons externes est élevée, plus l'agrégat sera lié à d'autres agrégats, et plus il sera placé haut sur l'échelle de la centralité. Si la densité mesure le degré d'institutionnalisation d'une pratique sociale d'usage des mots pour parler des choses, la centralité mesure le pouvoir d'attraction que cette pratique peut avoir sur le rapprochement des centres d'intérêt dans le champ. L'emploi simultané de ces deux échelles permet de situer les agrégats dans un espace bi-dimensionnel, et donc de synthétiser les résultats de l'analyse sémiotique à l'aide de quatre catégories descriptives :

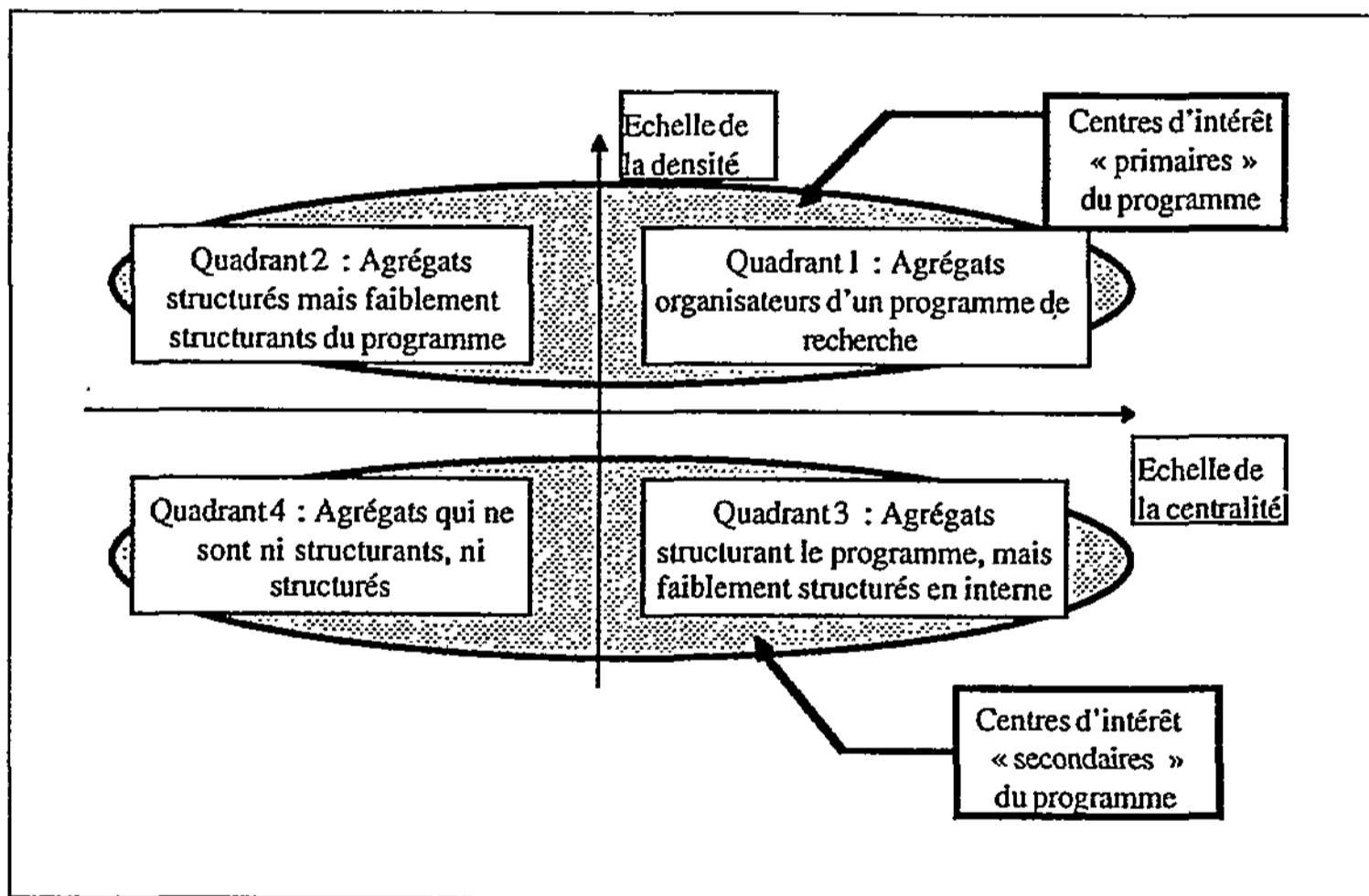


Figure 1

Les catégories d'agrégation de la sémiotique documentaire

Le regroupement des mots en agrégats est un premier résultat de l'analyse sémiotique, résultat qui est d'autant plus facile à comprendre que la densité des agrégats est élevée. Les agrégats, qui ne sont pas denses, sont rarement des axes thématiques d'un programme de recherche : en d'autres mots, la densité sert d'indicateur pour détecter des infrastructures socio-techniques plus ou moins développées. C'est ainsi que la zone en haut de la **Figure 1** sert à désigner les centres d'intérêt "primaires" d'un programme, alors que la zone inférieure réunit plutôt les centres d'intérêt "secondaires". Ce classement automatique des centres d'intérêt peut être

exploité afin de faire apparaître le lien entre la sémiotique documentaire et l'organisation du travail collectif dans le **Tableau 4** ci-dessous.

|         |                                 | Pir-Villes   |  |
|---------|---------------------------------|--|--|
|         |                                 | Centres d'intérêt «primaires»  | Centres d'intérêt «secondaires»  |
| STRAIES | Centres d'intérêt «primaires»   | Partenariat ou compétition   | Sous-traitance ou repli sur soi<br><br><i>Offre d'expertise par Strates PIR-Villes potentiellement demandeur</i> |
|         | Centres d'intérêt «secondaires» | Sous-traitance ou repli sur soi<br><br><i>Offre d'expertise par PIR-Villes Strates potentiellement demandeur</i> | Alliance ou abandon des acquis   |

**Tableau 4**

**Les options d'interprétation de la sémiotique documentaire**

La finalité de la sémiotique documentaire est d'aider au développement d'un travail collaboratif. Afin d'y arriver, une stratégie de pondération a été adoptée qui exploite les quadrants de la cartographie des flux documentaires illustrée dans la **Figure 1** ci-dessus : si un centre d'intérêt  $M_i$  appartient à un agrégat  $Q_j$ , son poids  $P_{ij}$  dans ce quadrant doit être supérieur à son poids dans les autres quadrants. Dans le cas où cette hypothèse est confirmée, le centre d'intérêt en question sera considéré comme un facteur d'organisation des collaborations. Nous disons par convention qu'il s'agit d'un centre d'intérêt "remarquable" du flux documentaire [Saint Leger, 1997]. La détection des mots remarquables tient compte de trois variables.

Le poids d'un centre d'intérêt dans un quadrant dépend de sa fréquence. En effet, plus cette fréquence est élevée, plus les documents auxquels il appartient se trouvent indexés par ce quadrant. Cependant, cette mesure dépend initialement du nombre de documents dans les flux modélisés : le poids des mots caractérisant deux programmes de recherche ne saurait être comparé sans que la taille des flux documentaires correspondant à chaque programme soit exactement identique, ce qui est rarement, ou jamais, le cas. Il convient, donc, de normaliser le calcul du poids en tenant compte de la fréquence maximale dans le quadrant afin que le poids d'un mot soit proportionnel à sa fréquence relative dans un quadrant et reflète ainsi son rang plutôt que sa fréquence.



Un centre d'intérêt est d'autant plus typique ou spécifique à un quadrant qu'il ne se trouve pas dans les trois autres. On peut mesurer cette typicalité par le rapport de la fréquence du terme dans le quadrant sur sa fréquence maximale dans la carte des flux documentaires (cf. **Figure 1**).

La présence de certains mots dans un quadrant peut ne pas être due à leur participation effective à la construction des agrégats mais plutôt à leur appartenance aux documents indexés par ces derniers. Pour mesurer la part de ce hasard dans le calcul du poids, on introduit la notion de participation effective qui est le rapport de la fréquence du terme dans le quadrant sur sa fréquence totale. Plus ce rapport est faible, plus la participation effective du terme à la construction du quadrant est improbable.

Le **Tableau 5** permet d'inscrire à l'intersection des lignes et des colonnes du **Tableau 4** les centres d'intérêts "remarquables" des deux flux documentaires analysés. La visualisation de cette intersection doit aider les responsables du programme MOST à nouer des collaborations entre les deux communautés de chercheurs.

|         |                                 | Pir-Villes                         |   |
|---------|---------------------------------|------------------------------------|---|
|         |                                 | Centres d'intérêt «primaires»      | Centres d'intérêt «secondaires»   |
| STRATES | Centres d'intérêt «primaires»   | ACTEUR q2 q1                       | GÉOGRAPHIE q4 q1<br>ORGANISATION q3 q1<br>EMPLOI q3 q1<br>ÉCONOMISTE q3 q1<br>MIGRATION q4 q1<br>CONCENTRATION q3 q1<br>EMPLOI q3 q1<br>LOCALISATION q3 q1<br>RÉPARTITION q3 q1 |
|         | Centres d'intérêt «secondaires» | STRATÉGIE q2 q4<br>URBANISME q2 q4 | ESPACE q3 q4<br>DÉMOGRAPHIE q4 q4<br>HISTOIRE q3 q4<br>MINORITÉ q3 q4<br>DYNAMIQUE q3 q4<br>CATÉGORIE<br>SOCIALE q4 q4  |

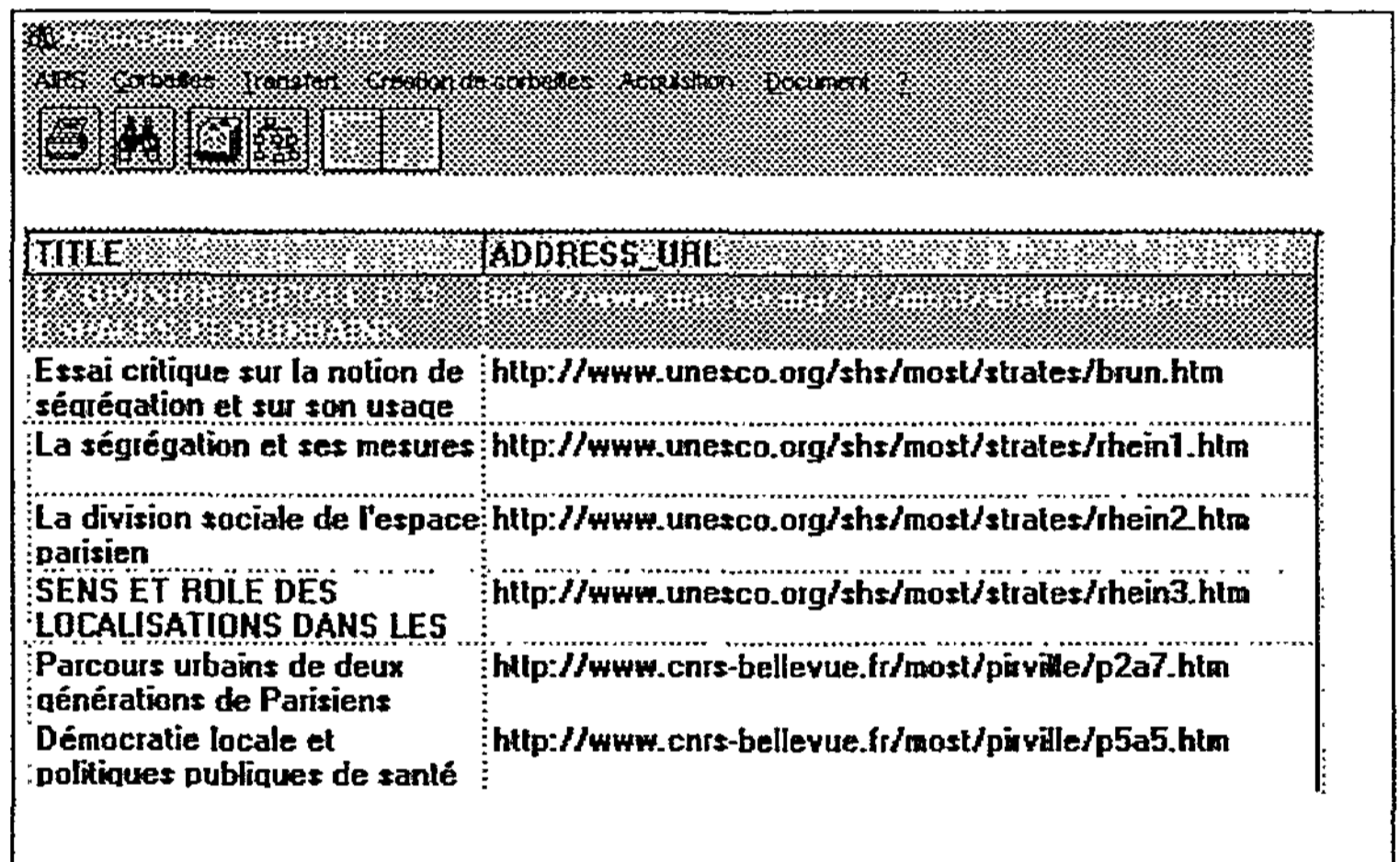
**Tableau 5**

**Les centres d'intérêt organisateurs des collaborations entre les équipes travaillant dans le cadre du PIR-Villes et de STRATES<sup>2</sup>**

<sup>2</sup>La notation q<sub>i</sub>q<sub>j</sub> renvoie aux quadrants définis dans la **Figure 1**, q<sub>i</sub> désignant l'intérêt des équipes du PIR-Villes pour les sujets visualisés ; q<sub>j</sub> désignant l'intérêt du laboratoire STRATES.

Cette analyse sémiotique doit, en principe, donner des résultats immédiatement compréhensibles pour l'organisation du travail collectif. La **Figure 2** ci-après s'efforce d'illustrer le fondement de cette compréhension sans explication en montrant l'écran du Médiateur qui

donne les résultats d'une interrogation du fonds documentaire de MOST. Les articles traitant de la *LOCALISATION* sont identifiés. Ce mot désigne dans le **Tableau 5** ci-dessus un centre d'intérêt qui est au cœur du programme de recherche de STRATES, mais qui est secondaire dans les documents du PIR-Villes. Une collaboration pourrait donc prendre la forme d'une sous-traitance, le transfert d'expertise allant de STRATES vers le PIR-Villes. Les scientifiques qui devraient être réunis pour discuter de l'opportunité d'une telle coopération sont les auteurs des articles listés, à savoir : pour STRATES, Martine Berger, Catherine Rhein et Jacques Brun ; et pour le PIR-Villes, Françoise Cribier et Michel Joubert (voir **Tableau 1** ci-dessus).



| TITLE  | ADDRESS_URL   |
|--|---|
| Essai critique sur la notion de ségrégation et sur son usage | <a href="http://www.unesco.org/shs/most/strates/brun.htm">http://www.unesco.org/shs/most/strates/brun.htm</a>       |
| La ségrégation et ses mesures                                | <a href="http://www.unesco.org/shs/most/strates/rhein1.htm">http://www.unesco.org/shs/most/strates/rhein1.htm</a>   |
| La division sociale de l'espace parisien                     | <a href="http://www.unesco.org/shs/most/strates/rhein2.htm">http://www.unesco.org/shs/most/strates/rhein2.htm</a>   |
| SENS ET ROLE DES LOCALISATIONS DANS LES                      | <a href="http://www.unesco.org/shs/most/strates/rhein3.htm">http://www.unesco.org/shs/most/strates/rhein3.htm</a>   |
| Parcours urbains de deux générations de Parisiens            | <a href="http://www.cnrs-bellevue.fr/most/pirville/p2a7.htm">http://www.cnrs-bellevue.fr/most/pirville/p2a7.htm</a> |
| Démocratie locale et politiques publiques de santé           | <a href="http://www.cnrs-bellevue.fr/most/pirville/p5a5.htm">http://www.cnrs-bellevue.fr/most/pirville/p5a5.htm</a> |

**Figure 2**

**Copie-écran du Médiateur Intelligent servant à identifier des partenaires d'une collaboration éventuelle**

Le rôle de la sémiotique documentaire dans l'organisation d'une telle discussion est encore à l'étude, car beaucoup reste à faire en vue de développer de nouvelles logiques de communication de l'information médiatisées par ordinateur, comme nous l'avons dit. Nos recherches sur cette question s'organisent autour de la cartographie des flux documentaires. Par exemple, la **Figure 3** ci-contre montre la structure des associations établies par les chercheurs de STRATES entre les différents centres d'intérêt qu'ils considèrent en relation avec la *LOCALISATION*.

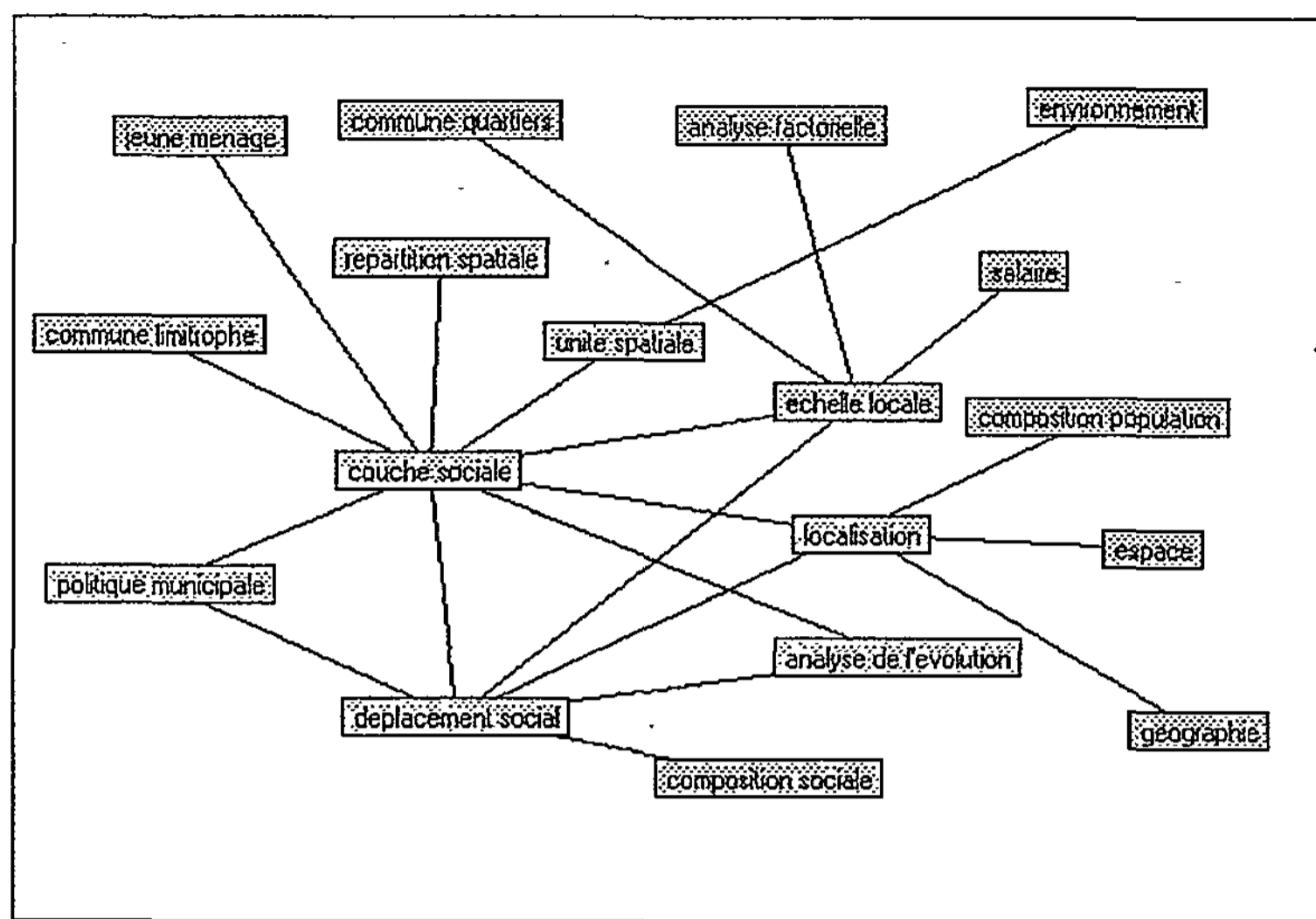


Figure 3

## Les cartes d'une sémiotique documentaire

Il s'agit maintenant de mettre au point des dispositifs expérimentaux permettant d'exploiter les informations réunies sur une telle carte. L'objectif est d'offrir aux chercheurs une lecture immédiate des infrastructures socio-techniques et des compétences humaines mises en œuvre dans les laboratoires. En effet, la question qu'il nous faudra traiter est la suivante : peut-on construire un tableau de la science telle qu'elle se fait dans les équipes de recherche, qui permettrait la mise en œuvre de collaborations scientifiques ? La réponse à cette question devra tenir compte des trois formes de collaboration identifiées ci-dessus : le partenariat, la sous-traitance et les alliances.

## Conclusion

Les laboratoires représentent une nouvelle structure d'organisation de la recherche et du développement scientifique. Certains chercheurs et gestionnaires de la recherche sont optimistes quant à la possibilité de promouvoir des collaborations scientifiques et techniques médiatisées par

ordinateurs ; d'autres le sont moins. Ce que nous nous sommes efforcés de montrer dans ce texte, c'est l'apport d'une sémiotique documentaire à l'organisation du travail intellectuel sur Internet.

*(°Ceresi, MIST-CNRS,  
1 place Aristide Briand, 92195 Meudon*

*+LOUEST/CRH,  
Ecole d'Architecture de Paris La Défense, 92023 Nanterre)*

**Bibliographie**

BORZIC (B.)

1997, *Le Médiateur intelligent : une plate-forme informatique pour la gestion documentaire sur Internet*, Thèse d'Université, Paris, CNAM-CERESI/CNRS.

BOUVERESSE (J.)

1973, *Wittgenstein : la rime et la raison*, Paris, Editions de Minuit.

CALLON (M.)

1980, "Struggles and Negotiations to Define What Is Problematic and What Is Not : The Socio-Logics of Translation", in *The Social Process of Scientific Investigation*, K. Knorr-Cetina, R. Krohn, R. Whitley, eds., Dordrecht, Reidel, Sociology of the Sciences, IV.

1989, *La Science et ses réseaux : genèse et circulation des faits scientifiques*, Paris, Editions La Découverte.

*Courrier du CNRS*

1994, «La Ville», n° 81, Paris, MIST-CNRS.

FODOR (J. A.), KATZ (J. J.)

1965, *The Structure of Language : Readings in the Philosophy of Language*, New Jersey, Prentice-Hall Inc.

LAKATOS (I.)

1970, "Falsification and the Methodology of Scientific Research Programmes" *Criticism and the Growth of Knowledge*, I. Lakatos, A. Musgrave, eds., London, Cambridge University Press

LATOUR (B.)

1989, *La Science en action*, Paris, Editions de la Découverte.

LEDERBERG (J.), UNCAPHER (K.), eds.

1989, "Towards a National Collaboratory : Report of an Invitational Workshop", Rockefeller University, March.

LYOTARD (J. F.)

1979, *La Condition post-moderne*, Paris, Les Editions de Minuit.

MICHELET (B.)

1988, *L'Analyse d'associations*, Thèse d'Université de Paris VII, Paris, CERESI-CNRS.

POPPER (K.)

1968, *The Logic of Scientific Discovery*, New York, Harper and Row, Harper Torchbook, 2nd Edition.

SAINT LEGER (M. de)

1997, *Dynatools : un outil de gestion des flux documentaires scientifiques et techniques*, Thèse d'Université, Paris : CNAM-CERESI/CNRS.

TURNER (W. A.), COURTIAL (J. P.), BAUIN (S.) *et al.*

1984, "Les Cartindex des sciences et des techniques", *Documentaliste*, vol. 21, n° 1, janvier-février, p. 65-75.

TURNER (W. A.), TURNER (G.), PECQUERIE (S.)

1985, *Les Outils Cartindex/Leximappe au Service de la Programmation de la Recherche*, Rapport de fin d'Etudes, SGDN-SERPIA/CNRS.

TURNER (W. A.), CHARTRON (G.), LAVILLE (F.) *et al.*

1988, "Packaging Information for Peer Review : New Co-word Analysis Techniques", in *Handbook of Quantitative Studies of Science and Technology*, A. F. J. van Raan, ed., Amsterdam, North Holland.

WITTGENSTEIN (L.)

1961a, *Tractatus logico-philosophicus*, Paris, Gallimard.

1961b, *Investigations philosophiques*, Paris, Gallimard.