

# **Atelier Information, Communication et Connaissance**

*Ce rapport a été réalisé sous la direction de Jean-Gabriel Ganascia, Professeur à Pierre et Marie Curie (Paris VI) avec la participation active de Line Garnero et de Marie-Claude Gaudel. Il a été élaboré sous la responsabilité du Conseil Scientifique de Département des Sciences et Technologies de l'Information du CNRS présidé par Line Garnero, avec le concours de Nicolas Balacheff, Michel Beaudoin-Lafon, Pierre Beauvillain, Danièle Bourcier, Diddier Bourrigault, Philippe Breton, Jean Caelen, Francesco Cara, Pierre Chavel, Claude Chappert, Michel de Rougemont, Gilbert de Terssac, Didier Decoster, Rose Dieng, Philippe Dolfus, Serge Fdida, Pierre Flandrin, Patrick Gallinari, Christine Gaspin, Marie-Claude Gaudel, Jean-Michel Hoc, Christian Jacob, Philippe Jorrand, Gérard Loiseau, Antonio Munoz-Yague, Jean-Marie Pierrel, Brigitte Plateau, Roland Quéré, François Rastier, Paul-Allain Rolland, Joseph Saillard, Jean-Michel Salaün, Jean Pierre Sanchez, Michel Scholl et Denis Trystram.*

*Version du 12 septembre 2004*

## **1. Préliminaires**

### **1.1. Recherche sur les TIC**

Lesdites Technologies de l'Information et de la Communication, TIC, changent le monde avec une rapidité déconcertante. Depuis plus de vingt ans, elles bouleversent bien des institutions, bien des modes d'organisation, bien des rapports sociaux.

Mémoire collective, connaissance, travail... tout cela se transforme. Certes, depuis longtemps on stocke l'information et on la transmet au-delà de la portée immédiate de nos bras, de nos oreilles et de nos yeux. Avec l'écriture cunéiforme, les sumériens engrangeaient déjà le savoir sur des supports matériels. Et, grâce aux signaux de fumée, les indiens d'Amérique du Nord maîtrisaient la communication à distance avant l'arrivée des espagnols. Mais, les évolutions quantitatives sont telles qu'elles changent la donne. Ce qui prenait des mois, voire des années, comme porter un message à l'autre bout de la planète, s'effectue aujourd'hui en quelques secondes. La densité de stockage s'est accrue dans des proportions inouïes. La mémoire des sociétés se rassemble désormais sur quelques millimètres carrés et tous y ont accès, n'importe où, quasi instantanément. Songeons par exemple que le contenu

des treize millions de volume inscrits au catalogue des livres et imprimés de la bibliothèque nationale tiendra bientôt sur une surface de 12cm<sup>2</sup>.

Or, les laboratoires de recherche ont joué un rôle majeur dans cette révolution technologique. Ils ont pris une part déterminante dans la conception de ces nouveaux objets. Que l'on pense aux ordinateurs contemporains, aux réseaux de télécommunication modernes, aux langages de programmation actuels, aux interfaces homme-machine, aux procédures de validation de programmes, tout, jusqu'aux logiciels de reconnaissance d'écriture des assistants personnels, trouve son origine dans des laboratoires de recherche fondamentale. Ils sont le vivier où se recrutent les inventeurs, les chefs d'entreprises et les ingénieurs qui réalisent ces technologies. C'est en leur sein que mûrissent les idées maîtresses de ces évolutions. C'est là que l'on élabore les normes qui régiront les échanges dans la société future.

Sachant que ces transformations technologiques sont loin d'être achevées, la recherche fondamentale doit continuer à alimenter ces industries nouvelles. Qui plus est, nous sommes parfois si désemparés devant les conséquences de ces bouleversements, que seule une approche scientifique et rigoureuse nous aide à surmonter les difficultés rencontrées. Les recettes et les tours de main ne suffisent plus. Une démarche rationnelle s'impose. La science n'est plus seulement utile ; elle devient indispensable.

A titre d'illustration, nous prendrons quatre exemples.

1. Les protocoles de communication sur lesquels sont construits les réseaux de télécommunication doivent être validés, sous peine de conduire à des blocages catastrophiques. Des procédures de preuves permettent de le faire.
2. Un cryptage est nécessaire pour assurer la confidentialité des données médicales et bancaires qui se transmettent sur le réseau. Des travaux de recherche en cryptographie aident à concevoir des codes quasiment inviolables.

3. Comment s'assurer de la fiabilité des systèmes informatiques complexes ? Seuls des travaux de recherche fondamentale sur la vérification probabiliste de programmes et sur les tests permettent d'évaluer cette fiabilité.
4. La fouille de données extrait des connaissances à partir d'une grande quantité d'information stockée dans les bases de données. La reconnaissance des formes fait appel à des procédures d'induction, le traitement du signal aussi, et il en va de même pour les sciences du langage et pour la lutte contre la criminalité informatique. Quelle fiabilité accorder à ces résultats ? Les théories formelles de l'apprentissage et les théories de l'apprentissage statistique nous procurent des critères rigoureux et généraux.

En retour, avec les TIC, certains problèmes scientifiques anciens se posent en des termes si nouveaux que la science elle-même se transforme. Ainsi, la notion de démonstration mathématique, dont les fondements logiques furent posés au début du XX<sup>e</sup> siècle, est en passe d'évoluer. Plus exactement, la certitude d'une preuve mathématique est partiellement remise en cause : dans une preuve classique, la probabilité que des erreurs se soient glissées dans la démonstration et la difficulté à repérer ces erreurs s'accroissent considérablement avec la longueur de la preuve. La validation de la preuve en devient d'autant plus ardue. De nouvelles procédures de démonstration produisent des preuves dites transparentes ou holographiques qui amplifient les erreurs en sorte qu'on les détecte facilement, par simple échantillonnage. Dans le même ordre d'idées, la résolution de problèmes difficiles conduit à remettre partiellement en cause les idées anciennes : des procédures de résolution réputées être d'une complexité algorithmique rédhibitoire, c'est-à-dire, en termes techniques, d'une complexité exponentielle, deviennent admissibles, autrement dit polynomiales lorsqu'on se contente d'une approximation de la solution. Il en va ainsi pour la vérification de programmes où l'on préfère souvent disposer de preuves incertaines, plutôt que n'avoir aucune validation d'aucune sorte.

Dans un registre plus pragmatique, la modélisation et la simulation informatiques jouent un rôle si capital pour bien des disciplines scientifiques comme la biologie moléculaire, la géologie, l'économie, les neurosciences ou la physique, que les fondements de ces disciplines en sont modifiés. À titre d'illustration, certains biologistes parlent d'expériences *in*

*silico*, qu'ils opposent aux expérimentations classiques *in vivo* et *in vitro*, pour évoquer les tests effectués sur ordinateur, par sondage sur de grandes banques de données.

Enfin, de nouvelles questions se posent aux sociologues qui étudient les modes d'appropriation des nouvelles technologies de l'information et de la communication, aux politistes qui se demandent comment le vote à distance change le comportement des électeurs, et si ce changement est compatible avec la solennité requise par cet acte, aux économistes qui entrevoient de nouveaux types d'échanges et une nouvelle organisation du marché, aux juristes qui cherchent comment attester avec certitude de la volonté individuelle au moyen d'une signature électronique envoyée à distance sur les réseaux, aux érudits qui se transmettent depuis longtemps des textes de commentaires et des sources, et qui s'interrogent sur les véritables ruptures introduites par les nouvelles technologies etc. Bref, les disciplines scientifiques traditionnelles voient s'ouvrir aujourd'hui, avec le développement des technologies de l'information et de la communication, de nouvelles questions et de nouveaux champs d'investigation.

## **1.2. Périmètre du rapport**

Ce texte intitulé ICC, *Information, Communication et Connaissance*, aborde l'information, la communication et la connaissance, ainsi que les relations qu'elles entretiennent, sans pour autant circonscrire l'intégralité des recherches actuelles touchant à chacun de ces trois thèmes. En effet, il serait vain de prétendre embrasser tout ce qui touche à la connaissance. Si on y aspirait, les philosophes, qui ont beaucoup à dire sur ce sujet, et depuis longtemps, devraient tous être conviés. Et, plus généralement, le CNRS qui, en son entier, vise à l'extension des limites du savoir, devrait être invité, ce qui risquerait de compliquer singulièrement notre tâche.

Par ailleurs, le traitement de l'information recourt à des composants électroniques dont la conception résulte des efforts d'une très grande partie de la communauté des physiciens, en partant des spécialistes de la physique quantique, jusqu'à ceux de la physique des solides. Qui plus est, les astrophysiciens traitent de signaux et d'information en provenance de systèmes physiques ; la communication et le traitement de l'information leur tiennent donc à cœur. Et,

il en va de même des botanistes qui étudient les signaux chimiques échangés par les plantes ou des éthologues qui s'intéressent à la danse des abeilles ou à la communication chez les fourmis. Faut-il pour autant associer physiciens, astrophysiciens, philosophes, éthologues et botanistes à l'atelier *Information, Communication et Connaissance* ? C'est bien évidemment une question ouverte. Cependant, si l'on devait rassembler l'ensemble des communautés scientifiques touchant de près ou de loin au traitement de l'information, à la compréhension des mécanismes de la communication et à l'étude des connaissances, on risquerait de vider la juxtaposition des trois mots *information, communication* et *connaissance* de toute signification, sans compter que cela ferait certainement double emploi avec les rapports remis par les autres ateliers ou par les sections du comité national de la recherche scientifique.

Or, pour l'instant, cette juxtaposition fait sens et donne à chacun des trois termes, *information, communication* et *connaissance*, un éclairage particulier où les technologies dites de l'information et de la communication jouent un rôle central. Plus précisément, l'*information* est entendue ici au sens de traitement de l'information par des dispositifs techniques, la *communication*, au sens d'échanges entre ces dispositifs, entre les hommes et ces dispositifs ou entre les hommes par l'entremise de ces dispositifs. Enfin la *connaissance* se réfère au statut du savoir médiatisé par les techniques contemporaines, à la fois à l'incidence de ces nouvelles technologies sur la production du savoir, au stockage du savoir et à sa mise en œuvre dans les machines contemporaines. En somme, l'information, la communication et la connaissance apparaissent ici comme indissociables : l'information, au sens qu'elle a pris dans la théorie de l'information et dans le traitement de l'information, ne saurait être envisagée indépendamment de la communication, et vice-versa. De même, la connaissance prend une signification spécifique eu égard au rôle qu'elle joue en intelligence artificielle, dans le développement des bases de connaissances, c'est-à-dire dans l'informatisation des connaissances, et, consécutivement, pour l'ingénierie, la transmission, autrement dit la communication, et la capitalisation des connaissances.

Notons que la place centrale que prennent les dispositifs techniques va de paire avec une étude systématique des dimensions humaines et sociales liées à leur dissémination. L'approche scientifique des technologies ne se limite aucunement à l'étude des conditions mathématiques et physiques de leur mise en œuvre. Les questions d'apprentissage humain,

d'ergonomie ou d'usage, aussi bien que l'incidence des dites TIC sur l'organisation du travail, sur le droit ou sur la politique sont au cœur de cet atelier, au même titre que les questions d'ordre matériel.

Au reste, précisons que nous avons centré nos réflexions sur l'évolution, sur les transformations et sur les ruptures. En quoi la situation présente se distingue de celles qui la précèdent et en quoi la suivante, d'ici quatre ans, est-elle susceptible de différer ? Telles sont les questions auxquelles nous avons voulu répondre. Cela ne signifie pas que seuls les champs nouveaux ont droit de cité ; il est bien évidemment nécessaire d'asseoir dans la durée certaines recherches et d'approfondir des travaux sur de longues années. Mais ce rapport de conjoncture n'est pas destiné à faire un inventaire exhaustif ; il vise à mettre l'accent sur les mutations en cours, afin de prévoir les changements, d'alerter parfois sur leurs conséquences, et le plus souvent d'indiquer les mesures d'accompagnement nécessaires pour permettre aux équipes nationales de rester compétitives.

## **2. Synthèse des contributions**

### ***2.1. Organisation***

Suite à ce qui vient d'être dit, il apparaît légitime de cerner l'ensemble de la thématique *Information, Communication et Connaissance*, sans disjoindre ce qui relève de l'information, de la communication et de la connaissance, sans dissocier non plus les dimensions mathématiques ou physiques de ces technologies, des conditions sociales et humaines de leur diffusion. C'est ce que nous nous sommes efforcés de faire ici en prenant comme fil conducteur une progression vers l'abstraction : partant du plus concret, c'est-à-dire de la matière physique dont sont faits les systèmes de traitement de l'information, nous sommes allés progressivement aux questions posées par le traitement, en particulier à la validation, aux preuves et à la sûreté de fonctionnement des logiciels, puis à l'échange de signaux entre les machines et les hommes et à la communication médiatisée par les technologies de l'information, jusqu'au plus abstrait, à savoir, aux modalités d'acquisition, de construction et de stockage de savoir, avant d'aborder les transformations sociales consécutives à la diffusion des machines qu'elles soient d'ordre juridique, économique ou politique. Cela fait un plan en quatre parties : une première partie consacrée au traitement de l'information sous ses aspects matériels et logiciels, une deuxième partie qui aborde la communication par l'intermédiaire des systèmes de traitement de l'information, une troisième qui traite des nouvelles modalités de production du savoir et d'accès aux connaissances, enfin, une quatrième partie toute entière dévolue aux conséquences sociales du développement des technologies de l'information et de la communication..

### ***2.2. Traitement de l'information***

Le premier chapitre, consacré à la matérialité des dispositifs, envisage d'abord la physique des supports, puis les techniques de calcul et leurs évolutions. Viennent ensuite les questions de programmation, de validation des programmes et de sécurité des systèmes. Enfin, nous abordons la façon dont les problèmes scientifiques anciens sont traités au moyen

de ces techniques. C'est tout particulièrement le cas pour les sciences du vivant, pour les sciences de la terre et pour les sciences de l'univers où la modélisation joue un rôle central, et de plus en plus important, dans la production des connaissances.

### ***2.2.1. Matérialité des supports***

Pour l'essentiel, la miniaturisation s'est longtemps résumée à une réduction homothétique des dimensions physiques et ce tant pour la technologie des composants individuels que pour les architectures de circuits et l'intégration des systèmes. Jusqu'à présent les progrès technologiques ont toujours permis une telle évolution, dite de type top-down, sans que les principes physiques de fonctionnement et la conception des composants et circuits ne soient bouleversés.

Les perspectives envisagées pour les prochaines générations semblent changer radicalement cette vision simple et confortable, car on atteint désormais ce que l'on appelle le « mur du silicium » (dimensions inférieures à 10nm).

A moyen terme, des évolutions considérables se profilent. Quelques chiffres : pour 2010 (resp. 2016), la feuille de route de l'industrie électronique fixe pour la technologie CMOS des objectifs de trait de grille de 25 nm (resp. 13 nm), de longueur effective de 18 nm (resp. 9 nm) et de densité d'intégration de plusieurs milliards de transistors par puce intégrant des systèmes complets de circuits fonctionnant à plus de 20 GHz. Des efforts devront porter sur de nouvelles architectures des composants afin de repousser les limites à la miniaturisation et de nouvelles technologies pour les interconnexions entre composants. Les stratégies les plus innovantes dans ce domaine envisagent l'utilisation de nanotubes de carbone, ou bien les interconnexions optiques, qui pourraient à terme assurer au moins la transmission rapide du signal d'horloge sur la puce. En parallèle, l'architecture et la conception de circuits doivent continuer d'évoluer considérablement : actuellement, la conception consiste à manipuler des blocs fonctionnels ; on se dirige maintenant vers l'assemblage de blocs virtuels complexes (à l'échelle du processeur complet, par exemple), ce qui nécessite le développement de nouvelles méthodes et de nouveaux outils.

Pour certains paramètres technologiques, les limites physiques, et donc le besoin de nouveaux matériaux ou de nouveaux concepts, apparaissent ou apparaîtront bien avant ce terme de douze années. Pour dépasser les limites de miniaturisation des composants traditionnels, il apparaît séduisant d'exploiter des nano-objets plus ou moins auto-organisés qui présentent *intrinsèquement* une taille nanométrique : boîtes quantiques, nanotubes de carbone, molécules. Cette stratégie, dite bottom-up puisque l'on part des composants et que l'on recourt à des principes d'auto-organisation, est peut-être la seule viable ou "raisonnable" pour franchir le "mur" du 10 nm et fabriquer des circuits ultra-denses comprenant plusieurs milliards de composants élémentaires.

Des résultats théoriques sur l'information quantique et son traitement, c'est-à-dire sur l'utilisation du spin électronique comme vecteur d'information, puis leurs confirmations expérimentales au niveau du composant discret, montrent bien que des problèmes hors de portée de l'informatique classique pourraient être traités en exploitant des paradigmes de calcul non classiques. Ces travaux invitent à regarder plus loin. Ils donnent naissance à des courants de recherche interdisciplinaire où physiciens, mathématiciens, chimistes, biologistes et informaticiens apportent désormais des contributions majeures. Leur intégration à large échelle est cependant fortement pénalisée par, notamment, les problèmes de reproductibilité, de sensibilité aux paramètres technologiques (dimensions des objets et des barrières tunnel qui les séparent), de sensibilité à l'environnement électrostatique (charges fixes aléatoires) et de connexion d'objets nanométriques. Leur exploitation nécessitera vraisemblablement de nouveaux concepts d'architecture de circuits et de calcul. Il en va de même avec d'autres paradigmes de calcul issus de la biologie moléculaire ou de la chimie.

### ***2.2.2. Traitement de l'information***

À ces questions d'ordre purement physique, s'ajoutent des questions essentielles sur les nouveaux modes de calculs, sur l'algorithmique du traitement de l'information, sur le rôle de ce que l'on appelle l'intergiciel (*middleware* en anglais), sur la sûreté de fonctionnement et la sécurité des systèmes, sur les procédures de validation et de preuves. En effet, les évolutions technologiques récentes, et en particulier le rôle pris par INTERNET, montrent que ces questions se posent désormais sous un jour radicalement nouveau.

Avec la disponibilité de très grands nombres de processeurs, sites de stockage de l'information, et d'utilisateurs, toute une partie de la science informatique est en train de vivre un changement majeur.

Les approches logiques, qui tenaient une place prépondérante, montrent leurs limites quand il s'agit de passer à de très grandes échelles. Les approches probabilistes et statistiques cantonnées souvent aux évaluations de performance ou de fiabilité sont en train de se développer sous de nombreuses formes : face aux puissances de calcul actuelles et au nombre de calculateurs accessibles, il est plus efficace parfois de « tirer au sort » à bon escient que d'énumérer les choix possibles pour être sûr de trouver le bon. Pour les mêmes raisons, il devient souvent nécessaire d'observer, de modéliser, voire de concevoir, les systèmes informatiques très complexes sur des bases stochastiques. Algorithmes, protocoles de communication, méthodes de validation et de vérification, sont en train de bénéficier de ces nouvelles approches qui posent des problèmes théoriques difficiles, mais ouvrent des perspectives extrêmement prometteuses.

Un autre élément introduit par l'énorme complexité de l'environnement informatique actuel est la nécessité de gérer celle-ci et de la masquer à un utilisateur qui n'interagit plus avec son poste de travail individuel mais bel et bien avec le réseau mondial. Cela remet en cause profondément l'organisation des systèmes. Au niveau de la conception, la réutilisation de composants logiciels (dont il existe plusieurs définitions qui ne font pas l'unanimité) et les méthodes de développement basées sur des modèles se généralisent dans la pratique, mais manquent cruellement de fondements théoriques, donc de possibilités de certification. Au niveau de l'organisation, il a fallu introduire des couches logicielles intermédiaires entre les systèmes d'exploitation et les applications, afin de permettre de développer et d'utiliser celles-ci sans avoir à gérer les problèmes dus à la distribution des données et des calculs, sans se préoccuper non plus de la localisation des ressources employées. Ces couches de « middleware » (*interlogiciels* en français) posent de nombreux problèmes techniques et méthodologiques, sans parler des enjeux de standard sous-jacents. Elles sont d'autant plus stratégiques que c'est essentiellement à ce niveau que se traitent la sûreté de fonctionnement et la sécurité des systèmes. Or ces problèmes ont pris une actualité nouvelle du fait que parmi

des milliers de sites, dysfonctionnements et malveillances sont à la fois inévitables, ... mais récupérables si on sait tirer parti de la multiplicité des ressources saines.

### 2.2.3. « *Eclairages neufs* »

Grâce aux progrès enregistrés, des problèmes anciens reçoivent des éclairages neufs. La puissance de calcul autorise désormais des simulations fines, ainsi que le traitement de grandes masses de données.

Dans les sciences du vivant, il devient désormais possible, par l'emploi d'ordinateurs et de techniques de simulation, de s'engager dans l'élucidation de grandes questions : évolution des familles de gènes, évolution des espèces, organisation du génome, différences entre génomes, simulation de tout ou partie du « programme génétique »... De même, en météorologie, dans les sciences de l'univers, en physique des particules, dans les sciences de l'ingénieur, en électronique, pour le contrôle commande ou les télécommunications, dans les sciences de l'homme et de la société, en formation professionnelle, pour l'apprentissage du geste et des compétences, en économie ou en sciences politiques, pour la modélisation au moyen de systèmes complexes, c'est-à-dire de systèmes constitués d'un grand nombre de composants, et pour la simulation de ces modèles, l'informatique ouvre des perspectives neuves. Et, ces nouvelles applications de l'informatique ont grandement bénéficié de l'accroissement de la puissance de calcul et des tailles de mémoire : grâce aux progrès matériels, les simulations atteignent un tel degré de finesse qu'elles peuvent de plus en plus souvent se substituer aux expérimentations réelles.

Cependant, bien des questions d'ordre épistémologique et informatique doivent faire l'objet d'investigations poussées pour que la simulation puisse prendre la place qu'elle mérite dans le champ scientifique : Comment confronter modèles et données ? En quoi et quand un modèle constitue-t-il une preuve ? Quels critères rigoureux fondent les procédures de validation ? Quels composants doivent intervenir dans les systèmes complexes ? Quels facteurs d'échelle entre les composants et l'objet de la simulation sont-ils admissibles ? Quelles décompositions, quels modules et quels algorithmes sont souhaitables ? – on parle beaucoup aujourd'hui d'algorithmes de proximité, d'algorithmes probabilistes, d'algorithmes

d'auto-organisation – Quelles architectures matérielles et logicielles sont requises ? Autant de questions qui se posent aujourd'hui avec une acuité de plus en plus grande et qui laissent entrevoir l'émergence d'une véritable « science de la complexité » qui les aborderait toutes de façon systématique.

### ***2.3. Communication médiatisée par les TIC***

Deuxième dimension du rapport, la communication, tout au moins, la communication par l'entremise des dispositifs matériels de traitement de l'information.

#### ***2.3.1. Réseaux***

En poursuivant selon la logique annoncée précédemment, c'est-à-dire en partant du plus concret, les caractéristiques physiques des réseaux, nous irons ensuite au plus abstrait, à savoir aux protocoles de communication et à leurs évolutions, puis aux logiciels et aux services.

En deçà des architectures de réseaux, la mise au point de leurs composants, de leurs circuits et de leurs systèmes, constitue l'une des clés du succès. Ces recherches sont bien sûr guidées par les exigences de débit, d'adaptabilité, de fiabilité et de sécurité du réseau, support matériel pour l'information et la communication. De plus, l'accès « itinérant » (ou « nomade »), interactif, rapide et sûr à l'information, la communication et le stockage de données multimédia, les consommations énergétiques compatibles avec l'itinérance, constituent autant de défis lancés à la communauté scientifique.

Citons comme pistes importantes explorées dans ce domaine :

- la transmission et le traitement de l'information sur fibres optiques dont la dernière innovation majeure, la fibre microstructurée, est disponible commercialement mais encore trop jeune pour avoir un impact significatif sur les systèmes,

- le développement de systèmes mixtes fibre-radio qui représentent une alternative intéressante aux systèmes tout optique pour la transmission de l'information chez l'utilisateur ou à l'intérieur d'un bâtiment,
- le développement de nouvelles antennes actives miniaturisées et la gestion des systèmes énergétiques en utilisant des techniques de conversion d'énergie et des microtechnologies pour la fabrication de nouvelles sources et pour l'intégration de capteurs, enfin
- le développement de nouveaux systèmes de traitement et de stockage faisant appel à de nouveaux systèmes de mémoires magnétiques qui permettraient d'avoir simultanément accès à de grandes densités d'enregistrement et de grandes vitesses d'accès.

Au niveau logiciel, le domaine des réseaux informatiques a motivé l'intérêt de nombreux chercheurs de profils différents : théorie des graphes, de l'information, modélisation, statistiques, algorithmique, système, télécommunication. Ces connaissances ont montré comment construire un système complexe de grande dimension. L'objectif principal des recherches développées ces dernières années a visé à mieux comprendre l'INTERNET (établissement de lois fondamentales régissant son fonctionnement, élaboration de moyens de mesure performants etc.) et à le faire évoluer vers des services pouvant répondre à l'émergence des nouveaux besoins tenant à l'utilisation du réseau pour les activités industrielles, pour le divertissement, pour l'apprentissage ou pour le travail coopératif. Le second champ important de développement fut l'évolution vers les réseaux cellulaires de seconde et troisième génération (GPRS, UMTS). Au-delà de la conception d'ensemble de son architecture, il fallut étudier les caractéristiques du trafic de façon à établir précisément le lien entre la qualité du service et les propriétés des réseaux cellulaires. Par ailleurs, initiées à l'origine par un important programme de recherche de la NSF (National Science Foundation), de nouvelles pistes concernant le fonctionnement même des réseaux furent développées. Le fondement de ces nouveaux réseaux dits « actifs » est de rejeter le principe de bout-en-bout de l'INTERNET et d'introduire de l'intelligence dans les équipements chargés d'acheminer les paquets dans le réseau (« routeurs »).

Pour l'avenir, l'évolution des réseaux de télécommunications demande d'intégrer d'autres besoins comme la mobilité ou les calculs sur grille, mais aussi de nouvelles technologies comme les réseaux de capteurs.

Si la pression continue, elle produira certainement à long terme une évolution en rupture avec l'INTERNET actuel. Il convient donc de poursuivre l'effort engagé pour amener dans ce domaine la communauté de recherche en France à un niveau relatif comparable à celui des grands pays industrialisés, tant sur le plan des chercheurs, que des ingénieurs et des plates-formes d'expérimentation.

### **2.3.2. *Communication homme-machine***

Après avoir traité des télécommunications, c'est-à-dire des supports physiques et logiciels de transmission, venons en à la communication et à l'interaction homme-machine, autrement dit aux échanges entre hommes et machines.

Là encore, on doit distinguer les interfaces matérielles, c'est-à-dire les dispositifs de présentation de l'information, de commande ou d'échange – écrans, claviers, gants haptiques, encre électronique, etc. –, les modalités de communication – voix, image fixe ou animée, geste, etc. – et les éventuels couplages entre ces modalités.

Concernant les dispositifs physiques, on doit souligner la part de plus en plus grande prise par :

1. la notion de mobilité – les ordinateurs vestimentaires ou embarqués, et les réseaux sans-fil en sont emblématiques –
2. le caractère distribué des interfaces – dites multi surfaces et multi dispositifs, ce que l'idée d'intelligence d'ambiance traduit clairement –
3. l'intégration des objets physiques et des systèmes informatiques – on parle alors d'objets hybrides et de réalité augmentée ou mixte –

4. la prise en compte du contexte d'usage et de l'utilisateur – l'adaptation et la plasticité d'interfaces intelligentes prennent alors tout leur sens –

Depuis de nombreuses années, des spécialistes de l'homme, des ergonomes, des psychologues et des sociologues contribuent à la réalisation de ces nouveaux dispositifs ; il conviendrait maintenant d'y associer aussi des « designers » et des artistes, puisque l'on fabrique de nouveaux objets industriels, voire des philosophes, pour aborder les questions d'éthiques qui ne manquent pas de se poser.

Les différentes modalités de communication – parole, image, écriture etc. – firent l'objet de recherches approfondies pendant de nombreuses années, et continuent à le faire. On cherche depuis le début des années 1990, à coupler ces modalités sur le paradigme dit du « put that there », c'est-à-dire du « mets ça là », de Bolt pour donner naissance à ce que l'on appelle la multi-modalité. Rappelons que l'expression « mets ça là » est volontiers employée ici parce qu'elle invite à la multi-modalité : sans les deux déictiques « ça » et « là », l'expression exclusivement linguistique de l'action à accomplir s'allongerait insupportablement ; avec eux, elle ne prend sens qu'au moyen de deux gestes supplémentaires indiquant l'un, un objet, « ça », l'autre, une position, « là ». Bien des champs d'application sont susceptibles de bénéficier de la multi-modalité : aide aux handicapés, environnement pour l'apprentissage humain, recherche d'information etc. Or, l'expérience a montré que, contrairement à ce que l'intuition des latins, qui parlent avec les mains, laissait imaginer, il y a peu de synergie entre la parole et le geste. Le concept de multi-modalité n'en a pas été abandonné pour autant, mais il s'est enrichi dans deux directions principales :

1. On ne s'en tient plus uniquement au couplage entre modalités, mais on cherche aussi à comprendre leur complémentarité et à jouer de leurs éventuelles redondances.
2. La multi-modalité ne se restreint pas à la seule perception ; elle porte désormais aussi sur l'action et sur l'interaction

Pour être pris en considération tant par la communauté scientifique que par les entreprises, les travaux sur la communication et l'interaction homme-machine doivent de plus

en plus être validés par des expérimentations conduites en situation, et non par de simples prototypes. En effet, à défaut d'une théorie générale de l'homme, de ses capacités cognitives et de ses limites, qui guiderait les ingénieurs, on en est réduit à observer l'utilisation effective des nouvelles interfaces et à tirer parti de ces observations au cours de la conception. Or, de telles expérimentations demandent que l'on réunisse des équipes interdisciplinaires, avec des ingénieurs, spécialistes des machines, des ergonomes, des psychologues et des sociologues, spécialistes de l'homme, que l'on réalise des démonstrateurs, grandeur nature, et que l'on mette en place des plateformes d'étude des usages.

### ***2.3.3. Communication entre hommes par l'intermédiaire des machines***

Envisageons maintenant, dans la dernière section de ce chapitre sur la communication médiatisée par les TIC, la communication entre les hommes par l'intermédiaire des machines, ou plus exactement les modifications induites par la présence des machines sur les communications et sur l'organisation du travail dans l'entreprise ou dans les institutions d'enseignement et de formation. Soulignons d'abord que les Technologies de l'Information et de la Communication jouent un rôle central dans les organisations non seulement parce qu'elles influent sur les modes de communication ou sur la mémoire collective, mais aussi, et surtout, parce qu'elles transforment la structure même de ces organisations.

Ainsi, la présence des ERP – *Enterprise Resource Planning* –, qui rassemblent toutes les informations relatives aux hommes et aux projets sur lesquels ils travaillent, bouleverse les rapports entre individus et sociétés, et ce à au moins trois égards :

1. La structure organisationnelle devient de plus en plus souple ; elle est susceptible de modifications rapides, presque continues, au point que l'on peut parler de « réorganisation permanente ».
2. Les connaissances se formalisent et s'informatisent, ce qui permet de rationaliser de plus en plus souvent les choix et les échanges en fonction de ces connaissances.

3. La coopération entre acteurs fait elle aussi l'objet d'une formalisation et, en conséquence, d'une rationalisation de plus en plus grande.

Bien des questions se posent aujourd'hui aux spécialistes de sciences sociales qui étudient l'évolution des organisations industrielles consécutive à l'introduction des Technologies de l'Information et de la Communication. Parmi celles-ci, on en notera trois :

1. Comment approfondir les théories de l'organisation, en rapport avec les notions d'information, de connaissance, de cognition, de coopération, de décision, de concurrence, d'autonomie et de cohérence ?
2. Comment rationaliser la coopération homme-machine, de façon à mieux maîtriser les transformations engendrées par ladite « société de l'information » et à élaborer des outils d'aide à la décision plus adaptés ?
3. Enfin, comment penser la transformation des relations d'autorité et de pouvoir au sein de l'entreprise ?

#### ***2.4. Connaissances, ingénierie et capitalisation des connaissances***

La troisième partie de ce rapport est consacrée aux connaissances, à leur représentation, à leur stockage et à leur transmission par les technologies modernes.

##### ***2.4.1. Intelligence artificielle***

L'étude de la connaissance commence tout naturellement par l'intelligence artificielle. Plus exactement, la connaissance restreinte à l'acceptation technique qui nous intéresse ici se définit d'abord en référence à la signification originelle qu'elle a prise en intelligence artificielle dans les termes « base de connaissance » et « représentation des connaissances ». Le succès des systèmes experts, dits aussi systèmes à « base de connaissance », conduisit à développer des méthodologies de développement et des outils, qui se constituèrent en une « ingénierie des connaissances ». De même, le souci de conservation du savoir technologique

amena les spécialistes d'intelligence artificielle à réfléchir aux modalités de stockage du savoir à l'aide des techniques de représentation des connaissances qu'ils avaient développées, ce qui a donné naissance à ce qu'il est convenu d'appeler aujourd'hui la capitalisation des connaissances, branche de l'intelligence artificielle destinée aussi bien à la constitution de mémoires d'entreprises qu'à la veille technologique ou à l'intelligence économique.

### ***Ontologies***

Depuis une dizaine d'années, sont apparues des terminologies enrichies de relations de synonymie, d'hyponymie et d'hyperonymie, que l'on range, un peu abusivement sans doute, sous le terme d'ontologie. Même si leur formalisation pose de nombreux problèmes, même si elles apparaissent à la fois réductrices et contraignantes, ces ontologies jouent un rôle central tant pour la représentation des connaissances que pour les bases de données, pour l'ingénierie des connaissances, pour la constitution de mémoires d'entreprises ou pour l'exploitation des données circulant sur le réseau par les tenants de ce que l'on appelle le « semantic web ». Au reste, il existe des couplages forts entre la linguistique de corpus qui exploite automatiquement les textes de façon à construire des thesaurus et des terminologies à l'aide de techniques de fouille de données, le traitement automatique des langues qui facilite cette extraction de termes, la documentation qui utilise ces thesaurus et les ontologies qui se construisent directement à partir de ces terminologies.

### ***Bases de données intelligentes***

Les bases de données ne se contentent plus de stocker et d'interroger de l'information : elles visent à retenir une information interprétable, partageable, accessible en tous lieux et non localisée, sur laquelle il serait possible de faire des inférences, ce qui conduit à évoquer les notions de mobilité et de répartition et à parler de bases de données intelligentes. De ce fait, les distinctions entre données et connaissance, ou entre bases de données et bases de connaissance s'estompent au point que beaucoup de problématiques des bases de données se rapprochent des problématiques de l'intelligence artificielle.

Au surplus, les bases de données contemporaines contiennent des documents, des textes, de la musique, des paroles, des images, de la vidéo etc. Elles sont donc de plus en plus multimédias et de plus en plus volumineuses, ce qui suscite l'essor de nouveaux modèles, de nouveaux langages d'interrogation, de nouvelles architectures, de nouvelles techniques d'optimisation et d'indexation adaptées à ces différents médias et enfin de procédures d'accès dit « par le contenu », qui permettent, par exemple, de trouver des images à partir de leurs ressemblances avec une autre image ou un dessin, ou des musiques à partir d'un chant ou d'un fredonnement...

### ***Fouille de données***

Par ailleurs, la puissance de calcul des machines actuelles autorise désormais le traitement de très grandes masses de données, en vue d'extraire des connaissances nouvelles et utiles. La discipline résultante, que l'on nomme « fouille de données » (*data mining* en anglais) ou « extraction de connaissances dans les bases de données » (*knowledge discovery from data bases*) recourt essentiellement aux techniques et aux théories de l'apprentissage machine ainsi qu'aux travaux sur les bases de données.

Les fondements théoriques de cette discipline remontent aux années quatre-vingt lorsque les techniques d'apprentissage machine se sont développées et que l'on a jeté les bases des théories formelles de l'apprentissage et des théories de l'apprentissage statistique. De plus, des liens étroits ont été tissés avec des disciplines proches : bases de données, analyse de données, statistiques etc. Depuis une dizaine d'années, l'évolution de ce domaine est directement liée aux champs d'application potentiels : exploitation de données multimédia, extraction de connaissances à partir du web, remplissage de bases de données, linguistique de corpus, philologie électronique, investigation du génome, etc. Autant d'applications prestigieuses qui assurent à la fouille de données une dynamique forte, mais qui obligent à traiter des masses très considérables de données, de l'ordre de  $10^5$  à  $10^6$  unités. Et, comme la difficulté tient à l'effet de taille, les équipes doivent fournir un investissement technologique important avant d'être en mesure de prétendre apporter une contribution effective.

### ***Agents Autonomes Adaptatifs***

L'intelligence artificielle poursuit aussi ses investigations dans la tradition ancienne des pionniers, en particulier d'Alan Newell et d'Herbert Simon, avec la notion d'*agent rationnel*. Rappelons qu'hérité de l'économie, un agent rationnel est une entité fictive construite initialement pour représenter un consommateur qui cherche à maximiser son « utilité ». Progressivement, en intelligence artificielle, cela a désigné un programme susceptible de simuler une entité abstraite que l'on qualifie d'agent parce qu'elle agit, et de rationnelle parce qu'elle possède des buts, des désirs ou des affects et que ses actions sont asservies à leur satisfaction.

Les agents servent à la simulation de processus complexes, en particulier de processus cognitifs, émotionnels ou biologiques, ce qui a donné naissance, entre autre, aux avatars affectifs et à la vie artificielle. On parle alors d'agents logiciels pour désigner les outils de simulation ou d'interaction homme-machine ; dans ce dernier cas, l'interaction réussit lorsqu'il y a coïncidence entre l'agent rationnel que l'utilisateur projette sur une machine et le comportement induit par l'agent logiciel sur cette même machine.

Des techniques d'apprentissage, apprentissage dit évolutionniste – algorithmes génétiques –, apprentissage par renforcement ou par rétro propagation de gradient sur les réseaux de neurones formels, dotent les machines de capacités d'adaptation à leur environnement. Enfin, les agents sont capables d'échanger des informations et de se coordonner pour former des systèmes dits multi-agents.

Toutes ces recherches ont des applications ludiques, en particulier à la conception de jeux vidéos – dont les enjeux économiques sont loin d'être négligeables – et des applications sur la toile, pour la fabrication de robots virtuels – les « webbots », « taskbots », « sofbots », « knowbots », etc. – conçus à des fins d'assistance commerciale ou de collecte d'information. A cet égard, notons que ce que l'on appelle le « semantic web » repose sur la possibilité de disposer de tels agents autonomes capables d'exploiter automatiquement le contenu de la toile à l'égal des Hommes.

### ***2.4.2. Documentation, linguistique et didactique***

A ces questions techniques, viennent s'adjoindre trois thématiques qui bénéficient grandement des progrès enregistrés dans la mise en œuvre et l'exploitation informatique des connaissances.

- a. La science des bibliothèques s'est totalement transformée avec les nouveaux outils issus des TIC. Grâce à eux, l'indexation et l'accès se formulent en des termes résolument neufs, ce qui remet partiellement en cause la notion de thésaurus.
- b. Les sciences du langage renouent, depuis peu, avec une longue tradition d'exploitation minutieuse et systématique de longs corpus, grâce à l'introduction d'outils de traitement de l'information.
- c. La transmission des connaissances fait appel à de nouveaux outils, au service de toutes les formes d'enseignement, que ce soit à l'école, en présence d'un maître, à distance, au cours de la vie professionnelle ou en autoformation. A cet égard, rappelons, ce qui peut sembler une évidence, que la conception d'environnements informatiques dédiés à l'apprentissage humain touche aussi à la didactique et aux sciences de l'éducation.

### ***2.5. La société de l'information, de la communication et de la connaissance***

Le dernier chapitre recouvre les dimensions sociales liées à la dissémination des technologies dites de l'information et de la communication. Notons que si les sciences de l'homme et de la société ont une part majeure dans ce dernier chapitre, cela ne signifie pas qu'elles y sont cantonnées (nous avons vu qu'elles intervenaient beaucoup dans les chapitres précédents), et cela ne veut pas dire non plus qu'elles seules concourent à la compréhension de ces phénomènes sociétaux. Ainsi, les études d'usages, les modèles économiques et les questions d'ordre juridique recourent nécessairement à des compétences techniques et mathématiques pour être traitées de façon rigoureuse.

### ***2.5.1. Histoire***

De prime abord, on pourrait songer que l'histoire porte ses yeux uniquement sur le passé et qu'en conséquence, elle se détourne des technologies actuelles. Or, il y a au moins deux dimensions qui touchent à l'histoire en tant que discipline scientifique et qui ont trait aux Technologies de l'Information et de la Communication.

L'une porte sur l'analyse rétrospective de l'évolution des supports de connaissances et des technologies du traitement de l'information : seules les références historiques permettent de replacer les transformations actuelles sur la longue durée de façon à corriger l'effet de myopie qui nous guette lorsque nous ne regardons que le présent immédiat. L'histoire peut donc nous aider à prendre la mesure des changements que nous vivons actuellement.

L'autre témoigne de la place que prennent les TIC dans la vie des disciplines d'érudition où les sources, nécessairement restreintes, sont pour beaucoup numérisées, où la lecture et la production scientifique se font de plus en plus souvent en ligne, sur un ordinateur, et où les spécialistes échangent incessamment par INTERNET.

### ***2.5.2. Société***

Sans aucun doute, les Technologies de l'Information et de la Communication induisent des changements sociaux majeurs.

#### ***Droit***

Ainsi, le droit évolue considérablement, non seulement parce que les supports s'informatisent et que les procédures juridiques passent par le truchement d'ordinateurs, mais aussi parce que des questions relatives à la notion d'auteur, ou à la signature électronique, qui témoigne de l'assentiment de la personne, se posent en des termes inédits.

Sans compter que des problématiques juridiques nouvelles s'ouvrent à la sagacité des chercheurs. Elles portent à la fois sur la modélisation du droit et de la décision à l'aide de systèmes informatiques ou de concepts issus de l'informatique, sur l'écriture du droit qui peut

désormais recourir à des consultations larges, sur l'accessibilité aux lois par le plus grand nombre, grâce aux techniques modernes de communication et aux outils de traitement de l'information, enfin sur l'analyse automatique des textes juridiques et de leur cohérence, voire sur une régulation du droit par des dispositifs techniques.

### *Politique*

La politique elle aussi est concernée, car, avec les moyens modernes de communication, non seulement l'information du citoyen se trouve facilitée, mais de plus, les idées de démocratie participative trouvent un nouvel essor, ce qui conduit à reformuler en termes neufs des idéaux anciens. Des citoyens échangent directement sur INTERNET, en court-circuitant les réseaux traditionnels, ce qui donne une nouvelle vigueur au débat politique. Sans compter que l'on songe parfois à faire évoluer les procédures d'élection par le vote électronique à distance, ce qui permettrait, éventuellement, de remédier à la faible participation au risque, peut-être, de perdre la symbolique républicaine associée au scrutin populaire.

Bref, les technologies de l'information et de la communication interviennent à la fois pour informer le citoyen des actions de l'administration ou des enjeux politiques d'une consultation, pour l'inciter à participer activement aux débats, pendant la phase de formation des opinions, et pour faciliter la mise en œuvre de sa décision politique, au moment du vote.

On pourrait toutefois craindre que la constitution de communautés virtuelles déterritorialisées ne donne naissance à des groupes de pression qui viendraient interférer avec les organisations politiques traditionnelles.

Pour répondre à ces interrogations et prendre la mesure exacte de ces phénomènes, il faut évaluer l'impact réel des technologies de l'information et de la communication sur la vie politique, à partir d'études empiriques, en particulier à partir des quelques expérimentations qui ont pu être conduites tant en France qu'à l'étranger. C'est à ce travail de fond que s'attèlent les chercheurs qui abordent ces questions au CNRS.

### ***Communication***

Plus généralement, la vie des organisations tend elle aussi à se modifier du fait de l'irruption des systèmes d'information et des réseaux de communication. C'est vrai des communautés professionnelles, par exemple de l'entreprise, qui, comme nous l'avons vu, se transforme avec l'introduction des ERP. Mais c'est aussi vrai de l'ensemble de la société que l'on peut caractériser comme une « société de la communication » tout autant que comme une « société de l'information ». La recherche fondamentale devrait donc contribuer à l'élaboration d'une théorie de la communication et de l'interaction sociale qui rende compte de ces évolutions.

La communication doit être entendue ici au sens le plus large, ce qui recouvre tant l'expression de soi, à travers les œuvres esthétiques, que la persuasion politique ou commerciale, par la rhétorique, dans la publicité ou au travers des médias, et que l'échange d'information. Il s'avère toutefois que la plupart des études actuelles sur la communication demeurent si partielles qu'elles ne couvrent généralement qu'un seul de ces trois aspects, surtout le troisième. Il y a trop peu de liens entre les théories « techniques » de la communication produites au sein de la communauté des Sciences de l'Information et de la Communication et les théories « sociales » de la communication produites par les sciences de l'homme et de la société (sociologie, anthropologie, histoire). Il conviendrait donc de promouvoir des approches interdisciplinaires de la communication qui feraient la part des dimensions informationnelles, argumentatives et esthétiques de la communication, sans en privilégier une seule.

### ***Education et formation***

Annoncée depuis longtemps, la transformation de l'éducation et de la formation par l'emploi des technologies de l'information et de la communication devient aujourd'hui une réalité. En témoignent le déploiement des environnements numériques de travail dans le système scolaire, les campus numériques, le e-Learning et enfin l'émergence d'un marché concurrentiel dans les secteurs de l'éducation, de la formation et de la diffusion grand public. On constate aussi un grand nombre d'innovations matérielles avec l'apparition d'objets

communicants, de supports légers (de type assistants numériques ou téléphones multimédia), de cartables numériques, de tableaux électroniques etc.

Ces changements affectent aussi bien les contenus (avec par exemple le développement de micro mondes et de simulations) que les aspects organisationnels (avec le développement de plates-formes d'apprentissage collaboratif ou de gestion d'apprentissage). Les métiers de l'enseignement et de la formation s'en trouvent partiellement modifiés. Des questions nouvelles se posent relativement à l'organisation et au choix des enseignements, à la pédagogie et à la validation des acquis, au rôle des technologies, à leur acceptation etc.

Les défis à relever appellent la coopération étroite de disciplines relevant de plusieurs secteurs du CNRS en provenance des STIC, des SHS et des SDV. C'est dans cette voie que se sont engagés des chercheurs en informatique, didactique, psychologie ou sociologie du CNRS, par la constitution d'un réseau d'équipes et de laboratoires présent au plan national avec le RTP "Apprentissage, éducation et formation", et au plan européen avec le réseau d'excellence REX.

### *Ethique*

Ledit comité national d'éthique de la recherche s'intitule en réalité le « *comité consultatif national d'éthique pour les sciences de la vie et de la santé* ». Même s'il existe, au CNRS, un autre comité d'éthique, le COMETS – comité d'éthique des sciences –, qui aborde l'ensemble des questions posées par le développement des sciences, le prestige de cette première institution atteste de ce que, dans notre pays, les questions d'éthiques de la science sont essentiellement envisagées sous l'angle de la bioéthique, c'est-à-dire de la biologie, de la médecine et de leurs enjeux.

Or, au même titre que les sciences du vivant, les sciences et les technologies de l'information bouleversent radicalement nos sociétés. Elles modifient nos modes de communication et elles transforment les supports de nos mémoires, ce qui affecte la transmission des systèmes symboliques traditionnels sur lesquels se fondent les axiomes de la morale pratique. Pour en être plus insidieuses, les conséquences du développement des

sciences et des technologies de l'information et de la communication n'en sont donc pas moins importantes que celles des sciences de la vie. Elles portent non seulement sur la protection de la vie privée, mais aussi sur le commerce, sur l'économie, sur les échanges, sur la censure, sur l'activisme politique, sur les modes de gouvernement, sur les procédures d'élection, sur les nouvelles formes de délinquance, sur notre mémoire individuelle et collective, sur notre intégrité physique et morale etc.

Et, à l'évidence, on ne peut aborder ces questions en s'en remettant simplement au bon sens ou à l'idéologie. Les résultats des recherches fondamentales devraient donc éclairer législateurs et décideurs. On peut toutefois regretter que bien peu de philosophes et de scientifiques abordent ces questions, dans notre pays. A titre d'illustration, il existe une conférence spécialisée dans les dimensions éthiques des technologies de l'information qui se tient tous les 18 mois depuis plusieurs années. La dernière avait lieu en Grèce, courant avril 2004. Aucun français n'était présent ni parmi les intervenants, ni parmi les membres du comité de programme.

Bien évidemment, nous ne touchons ici que quelques unes des dimensions des transformations sociales induites par les technologies de l'information : il faudrait, sans aucun doute, parler d'ethnologie, de systèmes symboliques imaginaires, de frontières entre états, de renseignement, de veille technologique, d'espionnage, de « nouvelle économie », de guerre électronique, etc. Nous ne sommes pas parvenus à traiter de tous ces sujets, car, jusqu'à présent, peu de travaux les abordent à l'université ou au CNRS. Peut-être devrait-on inciter plus d'équipes à travailler sur ces questions, car les besoins sont considérables.

### **3. Conclusions**

#### ***3.1. Cohésion du champ***

La première constatation que l'on peut faire à la suite des discussions et des échanges que nous avons eus, c'est qu'il existe un ensemble de références communes à la plupart des chercheurs sollicités, du moins pour tout ce qui concerne les trois premières parties. Cela suggère deux conclusions.

D'une part, une collectivité scientifique s'est formée autour de ces nouveaux objets technologiques. On ne saurait dire, à proprement parler, qu'il y a une discipline scientifique unique, car les horizons sont parfois très éloignés et la formation scientifique originaire varie trop d'un secteur à l'autre. Mais une culture partagée autorise des échanges et des collaborations. Et, des préoccupations communes traversent l'ensemble du champ.

D'autre part, si cette culture est partagée par les spécialistes des composants, de l'algorithmique, de la programmation, des réseaux, de l'intelligence artificielle, des interfaces, de la communication homme-machine, de la sociologie des usages, des sciences du langage, etc. il semble qu'il n'en aille pas de même avec quelques secteurs des sciences sociales dont les contributions apparaissent relativement hétérogènes par rapport au reste. Qui plus est, certains domaines, touchant par exemple à l'éthique ou à la guerre électronique, sont peu abordés aujourd'hui par les chercheurs. Peut-être faudrait-il avoir des actions incitatives dans ces directions ?

#### ***3.2. Complexité et quantité***

Deuxième constatation, les chercheurs font face à deux défis majeurs.

Premier défi, celui de la complexité des procédures de traitement de l'information. Aujourd'hui, les systèmes informatiques se composent d'une multitude de sous-systèmes, eux-mêmes composites, de sorte que notre intelligence ne parvient pas à appréhender l'ensemble dans une vision synoptique claire. De plus, les différentes parties de ces systèmes

interagissent entre elles dans une grande multiplicité d'interactions que l'on formalise avec difficultés. D'où une remise en cause profonde de l'algorithmique et des méthodes de modélisation et de validation afin de maîtriser la complexité de tels assemblages.

Deuxième défi, celui de la masse d'information, de données et de connaissances à traiter, à stocker et à retrouver. Là encore, la grande quantité pose des problèmes nouveaux : comment se repérer dans cette profusion ? Comment appréhender de telles masses ? Comment les traiter ?

Et, ces deux défis sont d'autant plus difficiles à relever que les délais de rétroaction entre la recherche fondamentale, la R&D – recherche et développement – le transfert et l'innovation se raccourcissent.

L'énoncé de ces deux défis a un corollaire : pour apporter une contribution scientifique à la résolution des problèmes contemporains, on doit désormais valider cette contribution en « vraie grandeur », sur des problèmes tests dont les dimensions sont analogues à celles des problèmes que l'on veut résoudre, car la difficulté réside précisément dans la capacité à affronter la complexité et les très grands nombres. Par exemple, pour aborder des questions de langage naturel ou de fouille de données, il faut le faire sur des grandes quantités de données, ce qui demande d'accéder à des ressources précieuses et volumineuses, de maîtriser un ensemble de techniques et de disposer de temps assez long pour opérer les validations, toutes choses difficilement compatibles avec l'isolement de petites équipes, dans de petits laboratoires universitaires. De plus, le coût de mise en œuvre de ces ressources excède considérablement celui de la recherche elle-même.

Deux conclusions découlent naturellement de ce que nous venons de voir.

La première porte sur les fondements : une réflexion épistémologique, conduite par les scientifiques eux-mêmes, sur la nature des procédures de validation qu'il convient de mettre en œuvre, en particulier des validations expérimentales, aiderait à mieux cerner le statut de ces expériences d'un genre nouveau, à en préciser la légitimité scientifique et à mieux définir les exigences de rigueur requises.

La seconde est relative à la centralisation des moyens : une coordination nationale, voire européenne, faciliterait bien des choses en mettant à la disposition des chercheurs toutes les ressources nécessaires à la validation. Des actions ont été lancées dans ce sens pour certaines problématiques, en particulier pour quelques aspects du traitement du langage naturel ou de l'apprentissage humain. Il serait souhaitable d'amorcer d'autres actions institutionnelles dans cette direction.

### **3.3. Centralité de la communication**

Troisième constatation, parmi les trois termes, *information*, *communication* et *connaissance*, dont nous avons dit qu'ils étaient indissociables, ou plus exactement que chacun reçoit un éclairage spécifique de par sa relation avec les deux autres, on remarque aujourd'hui que le terme communication prend une position centrale. En effet, la plupart des thématiques de recherche ont été profondément renouvelées par la présence d'INTERNET ou par les réseaux mobiles.

Actuellement, le traitement de l'information se conçoit essentiellement en regard de l'échange d'information et de la distribution des calculs sur plusieurs machines distantes. Le concept de « grille de calcul » en est l'illustration la plus frappante. De même, bien des problèmes de programmation portent sur la répartition de calculs complexes entre différentes machines sur le réseau, ou sur la distribution de composants logiciels qui doivent pouvoir fonctionner sur n'importe quelle machine.

La connaissance ne peut, elle non plus, être envisagée indépendamment de son accessibilité et de sa répartition sur le réseau. Par exemple, la veille technologique se fait en grande partie à partir de données accumulées sur la toile qu'il faut rassembler et indexer en fonction de besoins spécifiques. Les bases de données ne se conçoivent désormais que distribuées et accessibles à distance, sur des objets mobiles. Les interfaces homme-machine et, plus généralement, les interactions entre hommes et machines ne sauraient s'envisager indépendamment d'un réseau d'ambiance et en relation avec d'autres objets communiquant. La notion de *mobilité* apparaît donc centrale aujourd'hui.

Enfin, tous les problèmes de preuves ou de validation sont envisagés eu égard à l'interconnexion des machines sur les réseaux et à l'incroyable complexité qui résulte de cette interdépendance.

### ***3.4. Thèmes récurrents***

Au regard des contributions reçues et des discussions que nous avons eues, certains thèmes transverses se dégagent comme saillants et novateurs. Ce sont là bien évidemment des perspectives qui demandent à être replacées dans le contexte actuel et qui ne signifient nullement qu'il faut abandonner l'ancien. Ainsi, les questions relatives aux notions de calcul, d'inférence, de modèle ou de représentation demeurent. Mais il semble que, pour les aborder aujourd'hui, il convienne de reformuler un peu les choses. En d'autres termes, il arrive que l'accent soit déplacé et que des questions anciennes cèdent, pour un temps, l'avant-scène à d'autres questions apparemment nouvelles, et éventuellement moins bien formulées.

Ces déplacements apparaissent centraux, car ils correspondent aux évolutions présentes ; ils témoignent des difficultés rencontrées et des tentatives conduites pour les surmonter. Au surplus les colloques, les rencontres, les débats, une grande part de la vie sociale des scientifiques s'anime autour de ces thèmes. C'est la raison pour laquelle il apparaît nécessaire de les mentionner ici.

Ces thèmes sont transverses en ce sens qu'ils sont mentionnés plusieurs fois, dans des contributions différentes, portant sur des questions éloignées. Ils sont saillants, car ils font l'objet de beaucoup de manifestations scientifiques. Ils sont neufs, puisque leur statut a considérablement évolué dans les dix dernières années.

#### **Réseau(x)**

Si l'on part du cœur, à savoir de la communication, on a le premier thème : le (ou les) réseau(x), ce qui comprend à la fois la toile et les réseaux privés, avec les différents supports, les réseaux sans fils, les objets communicants etc.

Indubitablement, cette thématique est au cœur des préoccupations de la plupart. Qu'il s'agisse de programmation, de preuve ou de validation de systèmes, de recherche d'information, de bases de données, de communication, tout est centré autour du réseau. A partir de cette problématique centrale, quatre thématiques périphériques semblent se dégager : les ontologies, le web sémantique, la fouille de données et les usages.

### **Ontologie**

Comme nous l'avons dit, les ontologies sont des terminologies enrichies de relations de synonymies, d'hyponymie et d'hyponymie entre les termes, voire parfois de processus d'inférence plus complexes. On mentionne les ontologies en représentation des connaissances où elles apparaissent aujourd'hui comme un élément clef, au dépend de la notion de formalisme de représentation des connaissances. Elles sont vues comme déterminante en ingénierie des connaissances, pour les bases de données intelligentes, dans le traitement du langage naturel et pour le web sémantique. Notons toutefois que l'accent fort mis aujourd'hui sur la notion d'ontologie n'équivaut nullement à un abandon des problématiques anciennes sur les formalismes de représentation des connaissances ou sur les mécanismes d'inférence. Il semble simplement que les ontologies renouvellent partiellement la façon dont les problèmes étaient posés et qu'elles permettent de les affronter sous un jour différent.

Soulignons que les enjeux sont d'autant plus importants que les systèmes d'information régissent désormais la vie de toutes les organisations (hôpitaux, administrations, etc.) et qu'ils devraient, à l'avenir, se fonder sur des ontologies. Déjà, les systèmes d'information médicaux de la plupart des pays développés se construisent sur une ontologie unique imposée comme norme.

### **Web sémantique**

Donner sens à l'information stockée sur la toile et l'enrichir de façon à l'utiliser au mieux de ses besoins, en s'aidant éventuellement de robots virtuels, tel est l'objectif affiché par les tenants du web sémantique. Derrière cette finalité, beaucoup de questions portent sur la représentation, sur la structuration et la recherche des informations, sur les formalismes de

description et sur les procédures d'accès d'une part, sur la composition, l'orchestration et la découverte de services disponibles sur la toile d'autre part.

### **Fouille de données**

Extraire des connaissances utiles à partir de grandes quantités d'information disséminées soit sur la toile, soit dans des entrepôts de données, voici un défi moderne qui répond aux aspirations et aux besoins de beaucoup. Les assurances, les banques, les firmes pharmaceutiques sont les premiers demandeurs. Toutefois, au cours de cette exploitation des données, on risque de se méprendre sur la portée des corrélations observées. Il convient donc d'établir avec certitude les principes logiques et statistiques sur lesquels on peut fonder les procédures d'induction.

Profitons de cette remarque sur la légitimité des procédures d'induction formelle pour dissiper d'éventuels malentendus en précisant le sens de certains termes : l'inférence, entendu au sens usuel de manipulation d'expressions formelles, doit se distinguer de l'inférence statistique qui est une forme d'induction statistique. Ainsi, si comme nous l'avons déjà dit, l'inférence, en tant que telle, n'apparaît plus au cœur des préoccupations actuelles, en revanche l'inférence statistique et plus généralement toutes les formes de preuves approximatives prennent une part de plus en plus importante tant en intelligence artificielle, pour la modélisation du raisonnement ou pour la fouille de données, qu'en traitement du signal ou dans la théorie de la preuve.

### **Usages**

Dernier thème référencé régulièrement, en particulier par les concepteurs des réseaux de télécommunication, qui se demandent à quels besoins répondent leurs techniques, ou par les spécialistes des interfaces homme-machine, ou encore par les sociologues de l'innovation : les usages, ou plus exactement, l'étude des usages et des modes d'appropriation des nouvelles technologies par leurs utilisateurs. Deux types d'approches sont plus ou moins confondus lorsqu'on parle d'usage. Les approches du premier type répondent au besoin d'« humaniser » des technologies en recourant à des principes d'ergonomie ou à ce que les anglo-saxons

appellent les facteurs humains (*human factors* en anglais), afin d'adapter ces technologies aux capacités des hommes interagissant avec elles. Les approches du second type correspondent à un mode de conception des objets industriels dans lequel on prend en considération l'appréciation des utilisateurs, ou ce que l'on appelle en termes techniques, les retours d'usage. Ce terme fait alors implicitement référence à la notion de conception centrée sur l'utilisateur qui porte ses fruits depuis plus de quinze ans dans la réalisation d'interfaces. Bien évidemment, ce mode de conception centré sur l'utilisateur a une incidence directe sur l'organisation de l'entreprise, sur ses circuits de décision et sur la place qu'y prennent les services de recherche et développement.

A cet égard, soulignons que l'étude des usages est restée relativement marginale en France, en dépit d'une politique assez volontariste des organismes de recherche depuis quelques années. Beaucoup de connaissances existent, mais elles sont difficilement accessibles ; les communautés scientifiques travaillant dans ce secteur demeurent morcelées. Il y a, à cela, des raisons structurelles que l'on tente de surmonter en créant des structures idoines. Mais les résultats ne permettent pas encore d'évaluer le fruit de ces efforts.

### ***3.4.1. Position de la France***

Au terme de cette synthèse, voici brossée à gros traits quelques remarques sur la recherche dans le secteur « Information, Communication et Connaissance », en France, telle qu'elle nous est apparue suite aux débats que nous avons eus dans le cadre de l'atelier.

En tout premier lieu, il faut insister sur l'existence d'une constellation de communautés scientifiques qui étudient les technologies de l'information et de la communication. Ces communautés se réunissent régulièrement. Elles nouent des contacts de proximité très étroits entre elles. Leurs échanges et leurs productions témoignent d'une grande vitalité. Des projets nationaux (RNRT, RNTL, RIAM, ACI, RTP ...) et européens leur ont permis d'acquérir une visibilité internationale et des contacts industriels. Les entreprises les plus avancées dans le domaine des nouvelles technologies (XEROX, Sony etc.) ont installé des laboratoires de recherche en France. On a dénombré, fin 2001, plus de 60 « start-up » en

activités créées en France par des chercheurs français travaillant dans le domaine des TIC. Il y a donc un certain nombre de sujets de satisfaction.

Toutefois, pour arriver au plus haut, se faire reconnaître et respecter de tous nos partenaires, en particulier de nos collègues nord-américains, japonais ou australiens, et s'imposer dans les institutions internationales qui fixent les normes d'usage, il convient de valider pleinement les travaux qui sont conduits, avec des expérimentations convaincantes, et d'investir les instances scientifiques internationales.

Or, dans beaucoup de domaines, la petitesse et la dispersion des équipes spécialisées y font obstacle. C'est du moins ce qui ressort de nos discussions.

Pour remédier à cet état de choses, il faudrait aider des équipes compétentes, comprenant plusieurs personnes en poste dont des chercheurs ou des enseignants-chercheurs expérimentés, à se structurer et à poursuivre dans une même direction de recherche sur plusieurs années. Aujourd'hui la plupart des équipes, surtout les équipes universitaires associées au CNRS, fonctionnent avec un ou deux permanents et quelques doctorants. De plus, les politiques incitatives conduites ces dernières années ont parfois eu tendance à disperser les efforts des chercheurs. En effet, la multitude d'appels d'offres, la création de jeunes équipes au sein de laboratoires, le manque de concertation entre les différents programmes nationaux de recherche, la remise en cause des financements avant terme, la quasi-absence d'évaluation *a posteriori*, tout ceci a eu assez souvent pour conséquence de déstructurer les noyaux existants.

Il conviendrait donc de renforcer des petites équipes déterminées, avec plus de chercheurs confirmés, et de les inciter à afficher clairement une direction de recherche sur plusieurs années et à la préserver. Pour cela, deux mesures semblent souhaitables.

La première serait de promouvoir des projets d'équipe avec des financements pluriannuels et des évaluations systématiques. Notons au passage, que la création de gros laboratoires de recherche conduit progressivement à une disparition de l'évaluation scientifique : l'évaluation globale des laboratoires ne se fait plus que sur un plan gestionnaire,

sans que les projets scientifiques des équipes constituant les laboratoires ne soient discutés de façon précise.

Une deuxième mesure passerait par la promotion de financements post-doctoraux. Grâce à eux, des jeunes doctorants auraient la disponibilité requise pour valider pleinement les hypothèses qu'ils ont énoncées dans leurs travaux de thèse, et pour les faire connaître de par le monde.

Par ailleurs, il conviendrait de structurer la recherche au plan national et européen en constituant des réseaux de compétences. Les RTP du département STIC pourraient éventuellement être vus comme la préfiguration de ces réseaux, là où ils représentent effectivement la communauté scientifique dont ils se réclament.

Il faudrait aussi conduire une réflexion sur les procédures de validation des travaux scientifiques, sur le statut épistémologique des expérimentations et sur leur coût.

Enfin, et surtout, on doit centraliser au plan national ou européen les moyens de façon à mettre à la disposition de tous les chercheurs d'une communauté les ressources nécessaires à la validation des outils développés. En effet, comme nous l'avons vu, il arrive souvent que l'investissement technologique nécessaire pour aborder les compétitions internationales, ou simplement pour valider certaines idées, soit tel que des équipes isolées sont découragées devant l'effort à fournir.

### ***3.5. Fonctionnement de l'atelier***

La direction et le conseil scientifique du département STIC sont à l'origine de l'atelier « Information, Communication et Connaissance » ; ils en ont posé l'intitulé, en ont supervisé la constitution et ont participé à ses travaux. Ce rapport leur est donc grandement redevable, même s'ils n'en sont pas directement les auteurs.

Le texte qui suit résulte d'un travail collectif ; il résume les contributions de plus de trente chercheurs qui ont échangé par voie électronique avant de se réunir une fois, le 12

décembre, puis de poursuivre leur travail de réflexion à partir des documents qu'ils se sont transmis. Ces participants ont été choisis, en concertation avec la direction du département STIC, en fonction de leurs compétences scientifiques ; beaucoup d'entre eux ont participé à l'animation des réseaux thématiques pluridisciplinaires (RTP) lancés par le département STIC. Ils témoignent donc de l'activité des communautés scientifiques relevant de la thématique STIC, de leur vitalité et du besoin qu'elles éprouvent de se structurer. Ce rapport peut donc légitimement prétendre faire un état des lieux de la recherche dans cette problématique en France.

Les participants à cet atelier ont rédigé des textes qui sont d'ores et déjà rassemblés dans un rapport plus volumineux. Ce dernier devrait faire l'objet d'une publication sous forme d'un ouvrage collectif destiné à refléter à la fois la diversité et la cohésion des approches scientifiques de la problématique des STIC.