

**Communication
de Madame Marion Créhange**

∞ ♦ ∞

Séance du 17 juin 2005

∞ ♦ ∞

**Apports réciproques
entre informatique et sciences humaines**

Je suis bien consciente de la grande diversité que vous représentez face à l'informatique. Je vais m'efforcer d'être compréhensible pour chacun, sans être ennuyeuse pour quiconque ; mais je serais heureuse que vous n'hésitez pas à m'interrompre pour des questions qui vous semblent bloquer votre compréhension immédiate !

L'informatique est une science en soi, c'est-à-dire un ensemble cohérent de connaissances^[1] : elle a ses théories, ses méthodes, ses outils. De plus, en elle-même ou dans sa mise en œuvre, elle est pluridisciplinaire : pour nombre de ses réalisations, le domaine d'application est extérieur à l'informatique et, même, la démarche d'informatisation fait souvent appel à d'autres disciplines.

Parmi ce que j'ai appelé «autres disciplines», les sciences humaines revêtent une importance capitale. Je vais donc évoquer ici les apports réciproques de ces sciences et de l'informatique.

Mais le domaine est large... et même infini ! Il est bien évident que j'en laisserai sous silence des pans entiers, peut-être essentiels pour vous ; j'essaierai de combler certaines lacunes en répondant à vos questions.

Je vais développer ma présentation selon le plan suivant :

1. Brèves définitions : *informatique, sciences humaines, ...*;

2. Evocation de quelques exemples d'apports de l'informatique aux sciences humaines ;
3. Méthode de travail en informatique ;
4. Typologie des apports. Ici, je dégagerai une liste de **types d'apports** qui nous servira de référence et pourra, je l'espère, vous rester longtemps en mémoire pour vous aider à comprendre ou imaginer des applications de l'informatique en sciences humaines ;
5. Description d'applications de l'informatique en sciences humaines, en référence aux types d'apports ;
6. Réciproquement, les apports des sciences humaines à l'informatique ;
7. Conclusion.

1 - Brèves définitions : informatique, sciences humaines, ...

Informatique : science du traitement automatique de l'information par des ordinateurs. Le terme *informatique* vient de la contraction des mots *information* et *automatique*. Mon fils Alain, sur la trace de notre regretté confrère Gérard Gorcy, dirait qu'il s'agit d'un « mot-valise »^[2]. Il a été proposé en 1962 par Philippe Dreyfus et accepté par l'Académie française en 1966^[3].

J'ai eu la très grande chance, grâce à mes professeurs Jean Legras, dès 1957, puis Claude Pair, de vivre la naissance de l'informatique. Et, bien sûr, son évolution a plutôt été une explosion ! Je ne parlerai pas de ces aspects historiques, ni des développements actuels qui voient tous les jours émerger de nouvelles idées à la frontière de l'informatique, de la téléphonie et du traitement de l'image et du son.

Il faut bien sûr préciser aussi que l'informatique, ce n'est pas uniquement la bureautique. D'ailleurs, je n'aborderai pas les applications de celle-ci, ni les applications courantes d'Internet. Disons seulement que ce sont les recherches en informatique qui ont permis l'explosion de la bureautique, conformément à leur objectif fréquent de permettre aux utilisateurs de... se passer des informaticiens !

Mais qu'est-ce que l'**information** ? L'information peut être définie^[4] comme un support de connaissances, c'est-à-dire un substitut manipulable de connaissances. Prenons un exemple : un électeur, avant un référendum, indépendamment de tout sondage. Ce que l'on peut nommer « connaissance », c'est tout ce que l'on sait sur lui. L'« information » que l'on va prendre en compte, est le résultat d'un **choix**, c'est chacun des renseignements qui nous semblent

utiles pour les traitements prévus (par exemple son état civil et le quartier où il habite). Ce qui est traité en informatique est plutôt la **donnée** : représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement.

Aux débuts de l'informatique, les seules données prises en compte étaient des nombres ; maintenant, une donnée cela peut être un nombre, une image, un film, la partition d'une sonate pour violoncelle et piano, une interprétation de cette sonate, une radiographie pulmonaire, un programme d'ordinateur...

Le passage de la connaissance à l'information repose, sciemment ou non, sur une phase de «modélisation».

Modéliser : c'est concevoir un **modèle**, représentation simplifiée (et d'une certaine manière standardisée) de la réalité, destinée à la comprendre, l'employer et éventuellement agir sur elle. Nous reparlerons bientôt de ce point essentiel. Mais dès maintenant remarquons que c'est une activité que nous avons à chaque instant : lorsque vous racontez à un voisin votre voyage à Québec, vous choisissez les détails que vous lui exposez en fonction de l'intérêt probable que ce voisin y trouvera et de l'utilisation qu'il devra pouvoir en faire ; et le récit du même voyage à un autre voisin sera différent. Vous avez appauvri la réalité pour n'exposer que ce qui vous semblait utile et opératoire. Autrement dit, un modèle est une image opératoire qu'on se fait de la réalité, dans une situation donnée.

Sciences humaines : le Petit Larousse donne la définition suivante : «Disciplines ayant pour objet l'homme et ses comportements individuels et collectifs, passés et présents». Il n'est pas dans mon propos de discuter des limites de ce domaine ; précisons simplement que mon champ d'étude comportera en gros les disciplines traitées à l'université de Nancy 2.

Nancy a été à l'avant-garde de l'informatique ; elle l'a été aussi dans l'application de l'informatique en sciences humaines, avec le CRAL (1966) pour la traduction automatique, l'étude des textes médiévaux, avec le Trésor de la Langue Française (1965), avec la MIAGE, etc, grâce à des hommes créatifs comme le doyen Jean Schneider et le recteur Paul Imbs, qui furent membres de l'Académie de Stanislas.

2 - Trois exemples d'apports de l'informatique aux sciences humaines

Je décris ici de façon synthétique et intuitive un échantillon d'applications assez représentatif des apports de l'informatique aux sciences humaines.

Apprentissage de la parole chez les sourds

Dans les années 1980, ma collègue Marie-Christine Haton, au laboratoire CRIN, a réalisé le système SIRENE permettant à des sourds d'apprendre ou réapprendre à parler. Le principe était simple : le sujet sourd lisait un fragment de texte (lettre, syllabe, mot, ...) et cherchait à le prononcer. Pour remplacer son audition déficiente, le système lui faisait **voir** sur l'écran de l'ordinateur une représentation de ce qu'il avait prononcé, sous forme d'un spectre ; il voyait conjointement le spectre «idéal» qu'il aurait dû obtenir. Puis il re-prononçait le fragment en essayant d'approcher le plus possible le spectre idéal ; ceci plusieurs fois. En fin de séquence, le système informatique «donnait une récompense» en fonction de la qualité de la séquence et du résultat.

Dans une telle application, le système informatique doit au moins : capter le signal vocal émis par le sujet (c'est une *entrée*) ; je vais comme ici faire quelques «références en avant» vers les types d'apport) ; le transformer en spectre^[5] et le visualiser (*sortie*) ; établir (c'est un *calcul*) le spectre idéal ; et le visualiser (*sortie*). Pour déterminer le spectre idéal, le système peut utiliser un tableau de correspondance entre les fragments écrits et leur prononciation, compte tenu de différents facteurs (sexe, âge, ...) : ce tableau peut être considéré comme de la *connaissance* explicite. Le système doit aussi évaluer l'écart entre les deux spectres (*comparaison de formes, calcul*) ; à partir de là, il doit estimer si au cours de la séquence l'écart diminue, chercher à caractériser cet écart^[6] (*manipulation explicite de connaissances*), pour aider le sujet sourd ou son thérapeute. S'il y a un retour d'information de la part de celui-ci, le système pourra éventuellement améliorer sa propre compétence (*apprentissage du système*).

Au point de vue méthodologique, remarquons que toutes les données traitées sont avant tout le fruit d'une *modélisation* de la réalité : chaque sujet est «réduit» à des informations comme son identité, des caractéristiques jugées utiles (âge, sexe, handicap, ...) ; chaque fragment à prononcer est «réduit» à son texte, plus éventuellement des consignes données au locuteur ; chaque fragment prononcé est «réduit» à un ensemble de signaux (qui seront ensuite transformés en un spectre, graphique), plus éventuellement des informations jugées utiles, comme une estimation de la difficulté éprouvée par le sujet ; les sessions sont, elles aussi, modélisées. Les méthodes de passage des signaux au spectre, du fragment à prononcer à un spectre, de comparaison des spectres,

etc. font, elles aussi, partie du modèle. En aval de cette modélisation, les informations et traitements ainsi déterminés seront représentés par des données et programmes qui seront ensuite mis en action par l'ordinateur.

Simulation du vieillissement d'un visage

Les ethnologues mais aussi la police criminelle de même que des médecins et physiologistes sont intéressés par l'évolution de l'aspect d'un visage au cours de la vie d'une personne.

Un visage peut être *modélisé* par des informations connues et jugées utiles sur le personnage, par des photos ou dessins, etc. Ces informations seront enregistrées sous forme de données dans la mémoire de l'ordinateur. Le vieillissement fait l'objet lui aussi de modélisations, avec des invariants et des facteurs variants. Exemple élémentaire et imaginaire : la distance entre les orbites est un invariant ; les plis issus des ailes du nez s'atténuent de façon régulière entre 15 et 30 ans puis se creusent ensuite, avec des coefficients estimés pour chaque tranche de 10 ans d'âge. Cette *connaissance* peut être représentée dans l'ordinateur, et peut être utilisée, manipulée (on peut par exemple à partir de deux règles en déduire une troisième, par ce qu'on peut appeler une *inférence*). Les modèles dégagés sont bien sûr le reflet de connaissances sur la morphologie, la physiologie, etc. Au fur et à mesure du fonctionnement du système, les utilisateurs pourront aussi être amenés à critiquer les résultats obtenus et à apporter des améliorations à leurs modèles, et donc aussi à leurs connaissances ; certains systèmes peuvent eux-mêmes apporter des améliorations à leurs propres connaissances (*apprentissage* du système).

Archéologie, premier domaine de développement des systèmes documentaires informatisés

Jean-Claude Gardin, archéologue connu, directeur de recherche au CNRS, qui m'a jadis accueillie en stage dans son laboratoire à Marseille, a très vite, vers 1960, ressenti l'intérêt pour ses travaux d'avoir recours à l'informatique. Puis il s'est petit à petit élevé au-dessus de sa propre discipline pour établir une méthodologie générale fort clairvoyante des systèmes de documentation automatique et plus généralement des *systèmes de gestion et recherche d'information*, et la mettre en œuvre dans un logiciel, SYNTOL. Il y a en particulier introduit les idées de thesaurus (c'est de la *connaissance explicite*, qui sera manipulée par l'ordinateur), de critères de ressemblance, de recherche progressive, etc. Il a aussi soulevé des problèmes épistémologiques sur la représentation des informations, la communication, la langue naturelle, etc... Citons quelques lignes de «La formalisation du discours savant» (vers 1990).

«... Chacun se souvient de la belle utopie conçue jadis par Leibniz, lorsqu'il imaginait une description du monde dotée de toutes les qualités nécessaires (précision, analyticit , invariance, etc.) pour servir de base aux th ories des hommes de science, pr sents et futurs, par les jeux combinatoires appropri s... L'int r t de [l'informatisation] est de soumettre   la sagacit  des experts toutes sortes de diagnostics «artificiels» qui reposent sur les m mes bases que les th ories jaillies de leur intelligence «naturelle». La critique des premiers est alors ins parable d'une remise en cause des seconds... un moyen de prendre   la lettre l'aphorisme de Cl. L vi-Strauss... : «la preuve de l'analyse est dans la synth se».

J.-C. Gardin et d'autres arch ologues, dans leurs *bases de donn es*, d crivaient les  crits sur l'arch ologie, mais aussi les objets arch ologiques pour les classer de diverses mani res, les retrouver, les exploiter. Je me souviens de travaux effectu s   Nancy sur la classification des vases antiques par Madame Friedel,  pouse du professeur Georges Friedel, qui fut pr sident de l'Acad mie de Stanislas. On peut remarquer que cette fa on de travailler, en plus de son efficacit , a fait progresser la mani re de d crire les objets, en introduisant des *mod les* et en les faisant  voluer, et a, je pense, donn  ainsi des moyens de faire avancer l'arch ologie.

Dans tous ces exemples, nous avons mis en  vidence les apports de l'informatique et le r le essentiel que joue la mod lisation.

3 - M thode de travail en informatique

Mod liser

La m thode de travail en informatique comporte toujours comme premi re phase une *mod lisation* de la situation et de la probl matique, c'est- -dire de ce qu'on appelle parfois le *r el per u*. La r alit  en elle-m me est en effet toujours trop complexe et riche pour  tre compl tement prise en compte dans le traitement des probl mes par les math matiques ou l'informatique. Il faut la r duire, en faire une abstraction dans une certaine mesure. Par exemple, lorsqu'on veut traduire en Anglais le texte d'une lettre manuscrite, on peut mod liser la lettre sous la forme d'une cha ne de caract res ; c'est un appauvrissement par rapport   la lettre elle-m me : on n'y retrouve pas l' criture, la couleur de l'encre, la nature du papier, l'odeur,...

Un autre exemple d' cole : les pr visions de r sultats d'un referendum suite   un sondage. Mod liser la probl matique de ce domaine restreint comporte deux parties : d'abord exprimer par des relations les plus simples possible la fa on dont les r sultats du vote (en eux-m mes et par rapport aux sondages)

sont influencés par la nature de la question, la situation politique, etc. ; puis expliciter un moyen d'en déduire une prévision pour le présent referendum ; sans compter le retour d'expérience par la suite.

Du modèle au programme

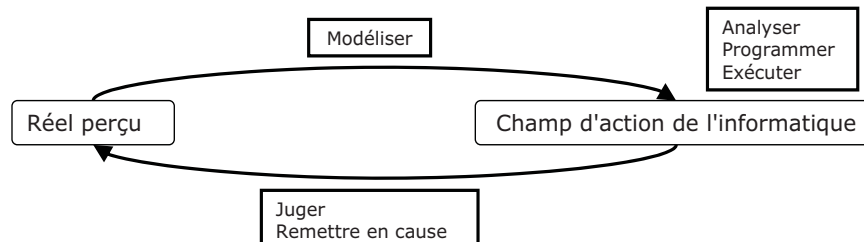
Les phases suivantes de la méthode reposent sur le modèle que l'on a conçu. La démarche consiste à définir une représentation et une organisation des données, à analyser progressivement, souvent par une démarche en «pelure d'oignon», les traitements à effectuer : décomposer les problèmes en problèmes plus simples (*modulariser*), en cherchant à *réutiliser*, chaque fois que c'est possible, des modules correspondant à des problèmes déjà traités. L'analyse des traitements a comme résultat un *algorithme*, qui, comme une recette de cuisine, est la description complète d'un procédé ; c'est un plan de travail pour l'ordinateur. Ensuite, on choisit des outils informatiques et principalement un langage de programmation (Fortran, Basic, Pascal, Java, etc.). Enfin on traduit les algorithmes en programmes puis on exploite ces programmes.

La contrainte permanente est bien sûr d'obtenir un bon logiciel, c'est-à-dire que le système présente un certain nombre de qualités qui doivent faire partie intégrante, dès les premières étapes, des objectifs de l'informatisation :

- la convivialité, qualité de l'interaction entre l'homme et la machine (assistance, en particulier pour l'utilisation du système ; tolérance aux fautes ; *réponse coopérative*, qui dépasse la réponse stricte au problème posé pour guider l'utilisateur vers une solution^[7]) ;
- le respect des contraintes, en particulier de confidentialité ;
- la robustesse, qui est la résistance aux accidents ;
- l'extensibilité, qui fait qu'une montée en puissance modérée n'oblige pas à effectuer d'importants changements.

Revenons sur la modélisation

La modélisation, comme d'ailleurs les autres étapes du travail des informaticiens, est une démarche progressive : chaque étape peut entraîner une remise en cause ou un recalibrage du modèle ; la démarche revêt ainsi la forme d'un cycle. Si nous reprenons l'exemple de la lettre manuscrite, il se peut que, pour lever des ambiguïtés dans la traduction, on ait besoin d'un contexte, d'une atmosphère : on peut alors choisir d'ajouter par exemple au modèle un qualificatif de l'écriture. A des stades divers, se joue une sorte de ping-pong entre le *réel perçu* et le champ d'action de l'informaticien.



La modélisation est, aussi, pragmatique : elle est avant tout tournée vers son efficacité, pour chercher à économiser le travail humain en aval, ressource la plus déterminante et coûteuse.

J'ai pu faire mienne la remarque que lorsqu'on enseigne l'informatique ou les mathématiques, particulièrement à des non scientifiques, le plus difficile n'est pas la programmation ou l'application de théorèmes mais, de loin, la modélisation, qui consiste en une *abstraction pragmatique* à partir d'une situation réelle. Il faut en particulier bien garder à l'esprit que lorsqu'on simplifie la réalité, ce qui est inhérent à la modélisation, on fait des choix sur lesquels on doit rester sans cesse vigilant et modeste.

4 - Typologie des apports de l'informatique aux sciences humaines

Je vais, de manière assez pragmatique et en tout cas discutable, définir sept types d'apports de l'informatique aux sciences humaines, en évoquant leur présence dans les exemples cités plus haut. Ils ne sont ni exclusifs ni indépendants. Cette liste de types d'apports nous servira de référence dans la suite de cet exposé et j'espère que vous pourrez continuer à vous y référer au-delà.

(1). *Traitements numériques (statistiques, résolution d'équations,...)*

En sociologie, économie, géographie,... l'application de méthodes statistiques est courante et nécessite des calculs souvent lourds. Il y a aussi des traitements numériques importants dans des applications telles que la simulation du vieillissement, l'étude de spectres de parole, que nous avons évoqués plus haut, mais aussi dans les calculs de longueurs de courbes, de propagation de la chaleur, de circulation des nuages, etc.

(2). *Conception et mise à l'épreuve de modèles : modélisation, simulation*

Vers 1960, un ingénieur d'IBM en visite à la Faculté des Lettres dans le but de promouvoir l'introduction de l'informatique, disait à des chercheurs : «Vous travaillez à l'informatisation de vos recherches ; même si finalement vous ne

pouvez pas acquérir d'ordinateur, la démarche et le résultat de la modélisation auront fait avancer la connaissance». Cette affirmation me paraît fortement justifiée. La modélisation en elle-même est souvent un réel apport au domaine d'application, en particulier en sciences humaines : elle oblige à une certaine exhaustivité, à de la cohérence, à une compatibilité de raisonnement avec d'autres domaines, et même elle donne quelquefois un moyen d'expression de ces raisonnements. Tout cela fait d'ailleurs partie de la démarche scientifique.

De plus, du point de vue pédagogique, la modélisation de problèmes dans un domaine est souvent un moyen puissant d'enseignement des connaissances elles-mêmes et surtout de la méthodologie du domaine.

(3). Aide à l'échange d'informations (visuelles, auditives,...) entre homme et ordinateur : entrées, sorties

Le plus souvent, pour exécuter un programme, on doit entrer des données et on sort des résultats. Des efforts et progrès importants ont été accomplis dans ces échanges ; ils visent à décharger l'utilisateur de contraintes trop lourdes, mais aussi à apporter des aides, par exemple à la compréhension ou à l'assimilation des résultats, et parfois à éviter des intermédiaires. Ces échanges sont de plus en plus souvent délocalisés, avec des canaux de transmission divers pour relier les lieux de travail aux ordinateurs opérants.

Les **entrées** textuelles ou numériques peuvent être saisies par l'intermédiaire d'un clavier. Elles peuvent aussi être scannées à partir d'imprimés, ce qui transforme leur tracé en données enregistrées dans la mémoire de l'ordinateur ; mais ce n'est pas pour cela qu'elles seront reconnues comme des textes ou nombres par la machine : il faut passer par l'intermédiaire d'un logiciel de reconnaissance de caractères. En ce qui concerne les images, si elles sont numériques, comme dans un appareil photo numérique personnel ou embarqué, elles peuvent être directement entrées ; sinon, elles doivent être scannées. Il en est à peu près de même pour le son.

Mais des recherches très poussées sont menées et ont déjà des applications réelles pour accepter des données moins directement utilisables, comme des phrases en langage naturel, des images, de la parole. Ce qui entre dans l'ordinateur, ce sont alors des signaux, qui sont dès lors traités par des programmes d'analyse, de reconnaissance,... que nous citons dans d'autres types, ci-dessous.

Quant aux **sorties**, textuelles, numériques, graphiques, elles sont affichées sur un écran ou imprimées. Mais d'autres sorties sont possibles, sous forme de signaux destinés à actionner des commandes à des machines, comme l'ouver-

ture de vannes ou le déclenchement d'une sirène ; ou sous forme de parole ou d'images «synthétiques» que nous rencontrons de plus en plus fréquemment dans notre vie courante. Un exemple spectaculaire d'image synthétique animée consiste en la visualisation, en «3D», d'une promenade simulée dans un musée. Autre exemple, spectaculaire lui aussi : la production effective de volumes, par exemple par stéréo-lithographie ; un appareil basé sur la solidification locale d'un liquide polymère sous l'effet d'un courant électrique permet, à partir de la représentation d'un solide dans l'ordinateur^[8], d'obtenir une copie du solide correspondant ; on peut ainsi reconstituer la tête de Cicéron, la tour d'un château, une vertèbre manquante dans un squelette reconstitué.

De grands progrès ont été réalisés pour aider les personnes handicapées à utiliser un ordinateur : en particulier, pour les aveugles, les entrées et sorties en Braille sont à leur disposition.

(4). Recherche et gestion d'information : bases de données, informatique documentaire, fouille de données

Ce type d'apport est l'un des plus anciens dans les sciences humaines mais reste fortement d'actualité, d'abord parce qu'il correspond à un réel besoin qui est immense en sciences humaines, ensuite parce que l'avènement d'Internet et du «Web» lui a donné une puissance que l'on n'aurait jamais pu imaginer il y a quelques années.

Les apports que je classe dans ce type comprennent la constitution et l'interrogation de bases de données bibliographiques, de textes, d'images, multimédias ; elles sont centralisées ou réparties sur différents ordinateurs ou sur Internet. Une partie importante de l'enseignement d'informatique donné aux étudiants, surtout proches des sciences humaines, porte sur la conception des bases de données, très importante en soi, mais aussi qui implique une phase, essentielle, de modélisation.

Ce type d'apports inclut aussi ce que l'on appelle «fouille de données» ou, en anglais, «data mining», exploration d'une mine, d'un gisement de données. Cette technique relativement nouvelle n'a de sens que depuis que l'on peut accéder efficacement à des masses énormes de données. Elle consiste, indépendamment ou presque d'un besoin défini, à observer une grande masse de données et à chercher à en tirer des leçons innovantes, par des outils statistiques et structurels puissants ; on retrouve ici l'idée qui guide depuis longtemps la veille technologique, en particulier à l'INIST (Institut National de l'Information Scientifique et Technique) du CNRS, installé à Nancy.

Tous ces outils comportent l'utilisation de thesaurus, de réseaux sémantiques et mettent en œuvre de nombreux concepts et techniques de l'intelligence artificielle, des sciences cognitives, de la linguistique.

(5). Représentation et manipulation d'objets complexes ou imparfaits ; reconnaissance des formes

Ces objets complexes peuvent bien sûr être des textes, des courbes, des photos, des images de manuscrits, des sons, des partitions ou extraits musicaux, des films,...

Il existe, pour les entrer dans l'ordinateur, pour les construire, pour les manipuler, pour les sortir, de nombreux outils sur le marché. Citons des outils assez courants, comme les logiciels de PAO (Publication Assistée par Ordinateur), DAO (Dessin Assistée par Ordinateur). Les applications de CAO (Conception Assistée par Ordinateur) comportent la représentation d'objets complexes, comme le plan d'une maison ou d'un avion, mais aussi tout ce qui permet de représenter et faire évoluer les étapes de conception, de gérer les contraintes.

Lorsqu'on manipule des textes longs, une technique que vous connaissez sans doute est l'hypertexte : elle permet de faire dans le texte un parcours non linéaire, opportuniste, grâce à la notion de *lien*, qui permet, à partir d'un mot du texte et avec possibilité de retour, d'atteindre d'autres informations, comme une autre partie du texte donnant des détails, ou comme un dictionnaire, un atlas, etc.

Parmi les informations complexes, une catégorie offre des débouchés et aussi des difficultés et un intérêt tout particuliers : les informations imparfaites. En médecine, principalement pour les dossiers médicaux, cette notion est de la plus haute importance ; en effet, dans la masse énorme d'informations contenues dans un dossier, un bon nombre sont incertaines, d'autres imprécises, d'autres incomplètes ; et il faut exploiter au mieux cette information. Il y a des techniques pour cela, à base de statistiques, d'algèbre du flou, etc, utiles aussi lorsqu'on doit raisonner sur des hypothèses ou des découvertes, par exemple archéologiques.

Mais sur les objets complexes, la problématique essentielle est la *reconnaissance des formes*. Exemple : dans SIRENE, comparer deux spectres de voix ; de même tout essai d'identification d'une partie de spectre, d'une séquence ADN, etc. La compréhension du langage oral ou écrit, la recherche d'un motif musical dans une partition ou même dans un enregistrement, l'identification d'un sujet dans une image sont du ressort de la reconnaissance des formes. Vaste sujet...

(6). Manipulation explicite de connaissances

Les techniques de l'Intelligence Artificielle (IA) ont pour principe de séparer l'expression de la connaissance (faits et règles) d'avec sa mise en œuvre. Celle-ci est réalisée par des programmes écrits une fois pour toutes, comme les *moteurs d'inférence* qui réalisent l'enchaînement du recours aux connaissances pour constituer des raisonnements. On réalise ainsi des *Systèmes à base de connaissances* ou des *Systèmes experts* (plus riches en connaissance d'experts). Cette séparation permet de considérer la connaissance comme une ressource que l'on peut faire évoluer, expérimenter,... Cela permet aussi de procéder par essais successifs pour l'approche d'une solution, de faire en sorte que le système, après avoir pris des décisions sur la marche à suivre, les explique, les retienne et si possible les synthétise pour pouvoir s'en resservir et même les améliorer (c'est un processus d'*apprentissage*). Exemples de connaissances : thesaurus, réseaux sémantiques ; règles, telles que «si le sujet de la phrase est un être vivant et si le récit est de type ... alors le verbe être se traduit par ...».

Il s'agit de calcul formel, dans lequel la machine traite non des données factuelles mais des formules ou des règles, sortes de méta-données. On va même plus loin dans la profondeur de «méta», en modélisant puis réalisant dans l'ordinateur des modes de raisonnement divers. Un exemple courant et intéressant est ce qu'on nomme l'*intelligence distribuée*, mise en œuvre dans les *systèmes multi-agent*. Un tel système décompose les problèmes et les fait traiter par différents agents, spécialisés ou non, programmes qui prennent chacun en charge une petite partie du traitement et échangent de l'information quand nécessaire. Voici deux exemples du mode de coopération des agents : - un agent confie un traitement à un agent spécialiste et attend la réponse ; - deux agents travaillent sur la même phase de traitement avec des méthodes différentes et un troisième reçoit, compare leurs résultats et en déduit un final ; on fait même quelquefois voter les agents et on en arrive à organiser leur ensemble selon des règles de la vie en société, appliquant des principes dictés par la sociologie ou les sciences politiques. On a d'ailleurs simulé le comportement de colonies, en particulier d'araignées^[9], par ce type de structure.

(7). Autres

Cette catégorie «fourre-tout» traduit le caractère non exhaustif de notre typologie.

5 - Description d'applications de l'informatique en sciences humaines

Pour décrire ces applications, je me référerai, par leurs numéros, aux types d'apport que j'ai énumérés. Ces exemples s'ajoutent aux trois du chapitre 2.

En géographie :

– Un SIG (Système d'Information Géographique) est un système informatisé dont le but est de centraliser, d'organiser, de gérer, d'analyser les données géographiques. Les SIG permettent d'établir des liens complexes entre plusieurs types de données (géologiques, géomorphologiques, pédologiques, ...). La restitution, sous formes de cartes, de tableaux et de statistiques, de la synthèse des données est l'un de leurs principaux atouts. Leurs différents domaines d'application sont la cartographie, l'aménagement du territoire, l'occupation des sols, l'environnement, les analyses topographiques, géologiques, agricoles, démographiques.

Les apports de l'informatique sont multiples : avant tout, (4) ; (5) manipulation d'objets complexes : cartes, photos numérisées, etc ; (2) modélisation : par exemple, le vent est souvent modélisé par une direction et une intensité, mais si cela ne suffit pas il faut introduire d'autres caractéristiques, dont l'introduction peut être l'objet de tests et faire avancer la connaissance des vents ; (3) pour l'entrée mais surtout pour la sortie (cartes avec fausses couleurs pour représenter la répartition géographique de phénomènes, représentations 3D). A Nancy, en coopération entre l'école de géologie et le laboratoire CRIN-LORIA, Jean-Laurent Mallet a conçu GEOL, un dispositif de représentation réaliste d'objets géographiques, et plus généralement d'objets naturels.

– Domesday Book : la BBC, en 1986, a lancé dans les écoles anglaises une superbe application de base de données. Elle voulait ainsi commémorer le 900ème anniversaire du Domesday Book, document cadastral que Guillaume le Conquérant fit dresser pour connaître les biens et ressources du pays. Le but de la BBC était de mettre à la disposition de toutes les écoles d'Angleterre d'une part un vidéodisque contenant cette base de données, d'autre part des micro-ordinateurs (BBC computers) pour l'interroger.

Chaque élève (ou professeur) pouvait faire d'abord du « map walking » : au début d'une session, s'affichait une carte d'Angleterre sur laquelle il pouvait désigner la région qui l'intéressait, puis faire afficher une carte de celle-ci, sur laquelle il pouvait désigner un secteur plus petit ou voisin, et ainsi de suite, à une échelle de plus en plus grande. A chaque stade, il pouvait faire afficher des informations démographiques, géographiques, ... et des images sur le secteur

choisi. Inversement, il pouvait interroger la base pour savoir quelles étaient les entités géographiques qui répondaient à tel ou tel critère.

Pour réaliser cette base de données à moindre frais, la BBC a eu une bonne idée. Elle a lancé dans toutes les écoles et associations du pays une vaste enquête : chaque individu et chaque groupe devaient remplir une fiche et donner une photo au sujet de leur environnement proche ; la BBC a constitué ainsi une immense base d'images où chaque photo avait comme indexation le contenu de la fiche associée. Ici, les apports de l'informatique relèvent principalement des types (4), (3) (sortie des cartes géographiques, aide à l'entrée par map walking), (5) (représentation et manipulation des cartes).

– Systèmes experts de recherche de gisements de pétrole, une des premières applications effectives des systèmes d'intelligence artificielle (6). Les sols, tant leur surface que ce qu'on connaissait de leur sous-sol, étaient, en gros, représentés comme des matrices à trois dimensions (latitude, longitude, altitude), avec des propriétés de chaque point (appartenance à la surface, nature géologique, présence d'eau, etc). Par des calculs (1) sur ces informations complexes (5), l'on pouvait déduire les couches géologiques et certaines de leurs propriétés. Le système contenait aussi des connaissances générales (6) en géologie et pédologie sous forme de *règles* ; j'en imagine des exemples : «si une couche de terrain est argileuse et surmontée d'une couche perméable de nature NNN, alors elle est susceptible d'écoulement» ; «si une couche est susceptible d'écoulement et si la surface du sol est ou a été plantée de forêts, alors il est possible qu'on puisse trouver du pétrole». Le système comportait en outre un moteur d'inférence (6) permettant des raisonnements par enchaînement d'activations de règles.

Cette application a effectivement permis de trouver du pétrole, mais elle a aussi fait progresser la maîtrise des systèmes à base de connaissance et aussi, probablement, la connaissance des lois auxquelles répond la répartition du pétrole.

– On peut rapprocher de cela une autre application précoce, le contrôle automatique de procédé, comme l'arrosage automatique de plants de tomates. Ici, les apports de type 3 sont intéressants car il y a des entrées et des sorties *analogiques* : en entrée, les données directement issues de capteurs placés dans le sol et dans les plants ; en, sortie, la commande de l'ouverture et de la fermeture des vannes des tuyaux d'alimentation.

En histoire :

– Les médiévistes ont été des pionniers en utilisation de l'informatique, à Nancy en particulier. Dans la présentation des équipes de recherche de Nancy 2 sur son site Internet, on peut lire la description de l'UMR 7002 MOYEN AGE, dirigée par notre nouveau confrère Patrick Corbet. Un de ses quatre pôles, ARTEM, Atelier de Recherche sur les Textes Médiévaux, comporte une équipe «Textes diplomatiques», qui est très bien décrite comme suit : «Les textes diplomatiques sont une des sources les plus essentielles pour la connaissance du Moyen Age...

Ce sont ... des sources juridiques, mais aussi politiques (parce que le fait de donner un acte engage l'autorité d'une personne), religieuses (parce que dans le contexte médiéval, tout acte est religieux), sociales (parce qu'autour d'un acte s'agitent de nombreuses personnes, auteur, bénéficiaire, témoins, scribe...), culturelles (parce que l'élaboration des actes requiert un certain savoir, littéraire et scripturaire)...

Et on peut lire aussi : «L'acte de naissance de l'entreprise remonte à 1966, quand le doyen Jean Schneider créa à Nancy un Centre de Recherches et d'Applications Linguistiques (CRAL), groupe riche de nombreuses recherches très différentes, parmi lesquelles un chantier de textes diplomatiques... Le but était d'obtenir une base de données (bien que le mot n'existât pas encore)...». Puis on cite, Lucie Fossier,...

J'ai eu la chance de travailler sur le traitement de documents diplomatiques du Moyen Age avec Lucie Fossier à cette époque, au temps des cartes perforées et des ordinateurs encombrants et soufflants. Pour cette application de type «base de données» (4), nous avons œuvré dans la ligne des travaux de JC Gardin, qui avait une vue très «moderne» de la notion d'adéquation d'un document vis-à-vis d'une requête, basée sur une modélisation (2) efficace de la sémantique. Cette application a été décrite dans les *Annales* et dans *Le Médiéviste et l'ordinateur*.

– L'archéologie est depuis longtemps une grande utilisatrice d'informatique. Nous avons déjà parlé des travaux de JC Gardin. Un autre exemple est la reconstitution d'un temple égyptien ou d'un squelette, par une sorte de puzzle. Les pièces mises au jour sont représentées dans une base de données (4). On modélise (2) et on matérialise par des programmes la démarche cherchant à estimer la «ressemblance» entre deux pièces, à classer les pièces candidates à un voisinage, à constituer des îlots de pièces voisines, à gérer les essais. Vous pouvez imaginer que pratiquement tous les types d'apports de l'informatique sont mis en œuvre ici, en particulier (3) dans la reconstitution visuelle de sites ou de scènes.

– L'histoire de l'art est un champ immense et varié : s'ajoutent aux applications historiques générales tout ce qui concerne l'analyse des images et leur traitement, en particulier pour la restauration de tableaux, une importance particulière des bases de données d'images et de son, etc.

En linguistique, littérature et langues étrangères

La linguistique et l'informatique sont le cadre de très nombreuses interactions. Les principaux apports de l'informatique sont essentiellement de types (6), (5), (4), mais aussi (2) car tous les travaux dans ce domaine passent par la modélisation de nombreux aspects et problèmes de la langue.

– La traduction automatique, très tôt et en particulier à Nancy, a été un champ de recherche fertile, dans toutes ses dimensions, lexicographique, syntaxique, sémantique. Même si les résultats effectifs ont été décevants, il y a maintenant des systèmes qui rendent de grands services. C'est un sujet toujours d'actualité, d'autant plus que toutes les avancées de l'analyse de textes, de la sémantique, etc sont aussi mises en œuvre, et donc mises à l'épreuve, par les recherches et applications en reconnaissance de la parole ; elles le sont également dans le domaine des bases de données, où on manipule la sémantique, pour comprendre le besoin de l'utilisateur et pour estimer la distance (sémantique) entre une requête et un document candidat.

– Diverses expressions du temps. Un exemple intéressant de l'intérêt de la modélisation (2) dans ces problèmes est le domaine de l'étude des diverses expressions du temps : de la situation exacte, absolue ou relative, d'un événement, de la durée, de la chronologie relative de deux événements, etc. Pour des besoins de compréhension automatique des textes, des chercheurs, tels que Laurent Romary au CRIN-LORIA, ont dû étudier quelles étaient les diverses expressions de ces prédicats et comment les reconnaître, par exemple quand le temps du discours interfère sur le temps des événements décrits. L'informatique est un outil maintenant incontournable pour cela.

J'avais rencontré le problème de l'expression du temps lorsque j'ai travaillé sur l'exploitation des dossiers médicaux. Dans un dossier, sont décrits divers événements qu'il faut situer dans le temps, individuellement et relativement ; exemple, imaginaire : ce malade a été soigné pour un zona à partir du 20 juin 2003 (mais depuis quand en était-il atteint ?) ; trois semaines après, il a été opéré de la vésicule biliaire ; il se souvient d'avoir déjà eu des troubles hépatiques pendant quelques semaines il y a environ quatre ans ; le service hospitalier lui a demandé de reprendre un rendez-vous dans trois mois et, huit jours avant, de commencer un régime sans sel...

D'autres études ont été menées d'un point de vue plus philosophique, en particulier, je crois, par Umberto Eco.

– Stylistique en littérature, musicologie. On peut caractériser le style d'un auteur à partir de statistiques (1) sur l'emploi des mots ou types de mots, des séquences, des longueurs et structures de phrases, etc. Ainsi, on bâtit petit à petit un modèle (2) du style de chaque auteur : ces modèles sont représentés dans l'ordinateur et servent à essayer d'identifier l'auteur de textes nouvellement entrés. Ceci peut être étendu à la musique, dans la mesure où l'on peut représenter des partitions dans l'ordinateur, éventuellement enrichies d'autres indications ; cette représentation a fait l'objet de longues recherches, vu la complexité de l'écriture musicale (5).

Ces applications donnent naissance à des bases de données d'auteurs (4) et comportent des représentations explicites de connaissances (6), avec des règles du genre «si dans un texte les mots «amour» et «moi» sont présents plus de deux fois toutes les dix lignes, l'auteur appartient avec une probabilité de 80 % au romantisme» ; «Lamartine est un auteur romantique».

En sciences de gestion

L'informatique est intervenue très tôt en gestion des organisations -entreprises, réseaux de vente, villes, etc-, et ceci a donné lieu à des développements novateurs car les problèmes soulevés étaient nouveaux, en particulier en ce qui concerne la phase de modélisation (2) et la constitution de grosses bases de données (4) ; on a même introduit le terme «informatique de gestion», qui n'a plus cours car ce type d'applications a fait école dans tous les domaines. Les traitements n'y sont en général pas très difficiles, mais les données sont complexes et lourdes. La phase de modélisation conditionne complètement l'efficacité des applications ; elle consiste à tirer de la réalité, en simplifiant son fonctionnement, quelquefois même en le caricaturant, ce qui peut être utile pour les traitements envisagés. La forme actuelle de l'informatique de gestion est l'«*intelligence économique*», qui est définie comme l'ensemble des actions coordonnées de recherche, de traitement, de distribution et de protection de l'information utile aux acteurs économiques en vue de la mise en œuvre de leurs stratégies individuelles et collectives.

En droit

Les bases de données (4), en particulier de jurisprudence, sont un outils précieux. On peut citer aussi... le débouché nouveau et les problématiques nouvelles qu'a fournis l'informatique au droit, tant pour les aspects purement professionnels que pour les problèmes liés à la liberté individuelle !

En science de la communication

Les apports réciproques sont ici aussi très importants. Parmi les apports de l'informatique, d'abord, tous les aspects de (3), les entrées et les sorties de l'ordinateur étant des actes de communication et l'informatique permettant de gérer divers modes d'expression et de perception. Elle permet aussi de modéliser et simuler (2) des pratiques collectives, comme la communication entre personnes ou le travail coopératif, en particulier la rédaction coopérative. L'explicitation des connaissances échangées et des modes d'échange et d'utilisation de ces connaissances est bien sûr un intéressant apport de type (6).

En sciences de l'éducation

A la fin des années 1980, a été créé à Nancy le CLEO (centre lorrain d'enseignement par ordinateur) qui produisait des logiciels d'enseignement ; cette création n'arrivait pas de façon spontanée : depuis plusieurs années le CUCES et l'INFA, organismes universitaires d'éducation permanente, s'intéressaient activement à l'«enseignement programmé», en particulier en mathématiques et en langues.

Les auteurs de cours doivent modéliser (2) et représenter dans l'ordinateur la connaissance à acquérir -faits et modes de pensée- (6) et aussi la façon de l'acquérir ; il y a bien sûr des problèmes intéressants d'entrée, de sortie et de manipulation d'informations complexes (3), (5), de bases de données (4), de calcul (1). Actuellement, sur Internet, existent des outils de formation à distance qui, en plus de leur utilité première évidente, permettent d'échanger et d'expérimenter des contenus et des méthodes d'enseignement ; tel est le cas, dans le domaine scientifique, de l'«Université en ligne».

L'informatique permet aussi d'illustrer et de rendre vivants les enseignements plus traditionnels : par les SIG, par la visualisation de phénomènes en changeant leur échelle dans l'espace ou dans le temps (3), comme par exemple la construction d'une cathédrale, etc. Mon équipe de recherche, EXPRIM, a travaillé sur l'utilisation pédagogique de la démarche de recherche d'informations dans une base de données.

Grandes bases de données sur Internet

J'ai indiqué que je ne parlerais pas des applications classiques d'Internet. Mais je ne peux passer sous silence le fantastique outil de travail que constituent les bases de données sur ce réseau, mis à la disposition des utilisateurs grâce à l'interface «Web», pour «World Wide Web». Il existe un nombre considérable de telles bases, soit bibliographiques soit de contenu direct, qu'il soit textuel, iconographique ou multimédia. Entre autres : Gallica, bibliothèque numé-

rique de la BNF ; Mérimée, base du Ministère de la Culture qui recense le patrimoine monumental français ; ConnectSciences, portail^[10] de l'INIST du CNRS donnant notamment accès aux bases bibliographiques Pascal et Francis ; Persée, portail de revues scientifiques en sciences humaines et sociales géré par le ministère de l'Éducation Nationale ; le site COLISciences (CORpus de Littérature Scientifique) qui rassemble des textes significatifs de la littérature scientifique du 19^{ème} siècle en langue française ; Edutemps, portail des lettres et des sciences humaines pour étudiants et enseignants.

Mais le potentiel des bases de données sur Internet est centuplé par l'existence de *moteurs de recherche*, outils informatiques qui, à partir de critères de recherche, donnent accès aux informations qui y répondent, quelle que soit leur localisation. Le plus utilisé d'entre eux, à l'heure actuelle, est GOOGLE : je m'en sers plusieurs fois par jour, pour chercher aussi bien des informations sur Guillaume le Conquérant que sur le temps de cuisson des pâtisseries.

En outre, Internet est un outil de diffusion et d'échange de littérature scientifique mais aussi de travail coopératif ; exemple : les bases de généalogie, dans lesquelles des personnes ne se connaissant pas peuvent mettre en commun leurs connaissances, permettant ainsi de relier des généalogies au départ disjointes.

EXPRIM (système EXPert de Recherche d'IMages)

Je vais décrire en quelques phrases des travaux menés par mon équipe au CRIN-LORIA de 1985 à 1997 parce qu'ils me semblent assez représentatifs d'utilisations de l'informatique en sciences humaines.

Interroger une base d'images ? Cela peut revêtir maintes formes. Entre la requête «Image n°12937 A», le besoin pour un journaliste d'obtenir des images choc sur l'enfant et la misère, et la recherche d'idées par un publicitaire devant réaliser une affiche, il y a toute une gamme d'utilisations de banques d'images, passant par des utilisations techniques, culturelles,... EXPRIM fait appel à l'intelligence artificielle (6), avec une intensité très variée, d'autant plus grande que l'on va vers des utilisateurs plus «naïfs» (grand public) et vers des utilisations plus exploratoires, prospectives, voire créatives.

En particulier, le fait de travailler sur des images incite à rendre plus audacieuses les tentatives de relâchement de contraintes pour la comparaison entre requêtes et descriptions d'images ; ceci parce que les «bruits», pourvu qu'ils ne submergent pas le signal, ne sont pas gênants et même sont souvent fertiles, en permettant de faire du «butinage documentaire», d'explorer les environs sémantiques de la requête initiale. Dans EXPRIM, l'utilisateur est appelé, lors de la phase dite de visualisation, à donner son jugement sur les images

qui lui sont proposées suite à une requête, et le système va essayer d'améliorer sa connaissance du besoin en faisant du «bouclage de pertinence» (relevance feedback) ; celui-ci consiste à chercher des facteurs discriminants entre les documents jugés positivement et ceux jugés négativement par l'utilisateur, puis d'en tenir compte pour reformuler la requête. Notons que suite à cette phase de visualisation peuvent apparaître dans la requête des critères qui en étaient totalement absents.

On peut dire qu'il s'agit d'un *apprentissage à court terme* du système, à l'échelle d'une session de travail. Mais l'aspect le plus intéressant et le plus difficile est l'auto-apprentissage du système à l'échelle de sa vie, son *apprentissage à long terme*. Il s'agit de l'amélioration par le système lui-même, normale pour un système heuristique, sous contrôle humain cependant, de ses connaissances «passives» (indexations des images, thesaurus, connaissances sur le domaine,...) et «actives», en particulier de ses règles de stratégie et de ce qu'on peut appeler son expertise. Dans EXPRIM, ces capacités ont été, dans une certaine mesure, données au système à l'aide d'outils de l'IA, en particulier les *systèmes à base de cas*, dans lesquels le système cherche à reconnaître un cas déjà rencontré et à s'en inspirer.

On a ainsi essayé de promouvoir le rôle de l'utilisateur en un rôle de créativité et de réflexion et d'instituer une véritable coopération entre l'homme et le système ; d'ailleurs, l'utilisation d'un tel système pour de l'apprentissage humain est manifeste, et nous y avons travaillé.

6 - Réciproquement, les apports des sciences humaines à l'informatique

Ces travaux sont une bonne transition vers une évocation des apports des sciences humaines à l'informatique.

Apports de la linguistique :

Les travaux sur la syntaxe et l'analyse syntaxique, sur les liens entre syntaxe et sémantique, etc. ont eu des retombées essentielles en compilation, en théorie des langages de programmation, en calcul formel. Les théories et outils sémantiques interviennent en interrogation de bases de données, en particulier pour la confection et l'utilisation de thesaurus, en interaction Homme-Machine par le langage, etc.

Apports des sciences de la communication (et de la psychologie) :

Elles ont un rôle essentiel dans la réalisation d'interfaces Homme-Machine, l'interrogation de bases de données - pour donner par exemple au système des capacités de compréhension de ce que l'utilisateur a comme besoin réel quand il exprime une requête -, la compréhension du discours oral ou écrit, etc... Les informaticiens mettent en particulier souvent en application les idées de l'école de Palo Alto sur l'influence réciproque de l'émetteur et du récepteur dans l'émission et la compréhension d'un message.

Apports de la psychologie et des sciences cognitives :

L'IA est très à l'écoute des sciences cognitives, en premier lieu pour la conception des *Systèmes à Base de Connaissance* : modélisation et typologie des connaissances, moteurs d'inférence, notions d'apprentissage.

Les sciences cognitives interviennent fortement aussi dans l'ergonomie des interfaces Homme-Machine. Citons Catherine Garbay^[11], directeur scientifique adjoint au département Sciences et Technologies de l'Information et de la Communication du CNRS : «Le développement des interfaces de réalité virtuelle ou augmentée, l'apparition des environnements dits «immersifs» demandent de progresser dans la compréhension des bases sensori-motrices de la cognition humaine».

Les modes de raisonnement de l'homme sont bien sûr des modèles que l'on essaie d'imiter en informatique, qui à son tour donne un moyen d'explicitier, de modéliser, de simuler au moins partiellement et d'expérimenter ce qu'on sait des mécanismes de la pensée et de la connaissance (perception, communication, mémorisation, raisonnement, apprentissage).

Tout ce qui concerne le fonctionnement du cerveau et du système nerveux inspire des techniques de raisonnement : raisonnement incomplet (mis en attente et pouvant être complété plus tard ou mis en commun avec d'autres raisonnements incomplets), méthodes connexionnistes, apprentissage, notion de nouveauté, association d'idées, etc... L'intelligence distribuée, débouchant sur la confection de systèmes multi-agent, est aussi un domaine d'application très fertile des sciences cognitives et aussi de la sociologie ou même des sciences politiques.

Apports de toutes les sciences humaines, et en particulier de l'histoire et de la géographie :

Chacune de ces sciences apporte ses connaissances pour la construction d'outils sémantiques (en particulier dans l'interrogation de bases de données),

mais aussi pour fournir à l'informatique des données et des terrains d'expérimentation et d'application.

7 - Conclusion

Les apports réciproques

Entre informatique et sciences humaines, les échanges, dans les deux sens, sont, nous l'avons vu, d'une grande richesse et d'une indiscutable utilité. Nous avons en particulier mis le doigt sur l'influence de l'outil sur la façon de travailler, en particulier en ce qui concerne l'obligation de modéliser. L'informatique permet aussi d'utiliser des méthodes mathématiques complexes, de les adapter à des situations réelles elles-mêmes complexes, de les confronter à la réalité, de les vérifier...

Entre informaticiens, l'informatique est un excellent «additionneur de compétences et d'efforts», grâce à la modularité de sa démarche et de ses réalisations. Des informaticiens peuvent chacun et assez indépendamment se consacrer à des modules disjoints ; leur coopération est rendue possible par la définition de structures de données communes et de canaux d'interaction. Ceci est poussé à l'extrême en intelligence distribuée (systèmes multi-agent). Les avantages sont immenses ; mais cela comporte aussi beaucoup d'inconvénients potentiels, comme par exemple les rapprochements de fichiers de personnes, menaces potentielles à la liberté individuelle.

Entre différentes sciences humaines, maintenant ; l'informatique est un vecteur d'échanges, par communauté ou voisinage d'outils, de modèles, d'approches. Cette interdisciplinarité peut procurer un grand enrichissement réciproque, chacune des sciences pouvant s'inspirer des utilisations faites par les autres ou cherchant à avoir des approches compatibles avec celles-ci ; sans compter l'avantage d'additionner, donc de potentialiser, leurs connaissances.

L'homme et l'informatique

Les informaticiens sont souvent accusés de croire qu'ils réinventent le monde ! En fait, en obligeant et aidant à modéliser les problèmes et les situations, à les modulariser, ils sont des catalyseurs de réflexion... mais ils doivent rester modestes. A propos de modestie, justement... nous devons insister sur le fait que ce n'est pas parce qu'un résultat est fourni par un ordinateur qu'il est juste ! De nombreuses raisons en effet peuvent faire que ce ne soit pas le cas, même si le travail a été sérieux : données erronées ou biaisées, échantillons trop petits ou pas représentatifs, programmes mal conçus, hypothèses douteuses, etc.

Il se pose actuellement des questions très profondes sur l'évolution des

rapports entre homme et informatique, avec des interactions extrêmement riches et complexes, dynamiques, ... Catherine Garbay dit encore : «Il convient ... de développer une vision intégrée des relations entre humains et artefacts techniques, qui nous conduit à penser les chaînes d'interdépendance qui lient les dimensions biologiques, cognitives et sociales de l'humain, qui lient les dimensions matérielles, logicielles, de conception et d'usage des artefacts techniques, qui lient enfin les faits humains, les faits techniques et la construction des savoirs».

A la fin de cet exposé, on peut penser que ce que je viens de développer révèle une certaine hétérogénéité, peut-être un certain décalage, dans la chronologie d'évolution des sciences et des techniques. J'ai donné des exemples tirés pour beaucoup de mon expérience, personnelle ou par osmose dans mon laboratoire ; ils datent un peu, même si je me tiens raisonnablement au courant. Mais l'évolution fulgurante des outils informatiques et de leur accessibilité pour chacun n'enlève rien, je crois, à l'actualité de l'analyse que j'ai faite avec vous. La science progresse, avec des accélérations mais aussi avec des retours en arrière, et la mise en œuvre des progrès met du temps, quelle que soit la puissance des outils. Plus particulièrement en sciences humaines, où toute application de l'informatique requiert un important travail intellectuel et matériel de l'homme.

Clamons, en conclusion, que l'informatique doit libérer et non contraindre, aider à l'imagination et à la création, et non restreindre à des schémas préétablis ! Informatiser ne devrait pas entraîner de radicalisation, de sécheresse !! Et je reprends avec grand plaisir le vœu qu'a émis notre confrère Michel Hachet à la fin de son si joli hommage aux fileuses : «on ne doit pas sombrer dans le statut d'esclave de ces admirables machines».

Discussion

Colette Keller-Didier fait remarquer que l'ordinateur est incapable d'effectuer des raisonnements logiques et se demande si, dans l'avenir, cela est envisageable.

Marion Créhange répond que les limites ont été beaucoup repoussées ces dernières années. Il est cependant évident que l'ordinateur ne peut actuellement effectuer des raisonnements tels que le cerveau humain sait en faire. Des informaticiens travaillent sur ce qu'il est possible d'informatiser ou non. Jusqu'où les limites seront-elles repoussées ? Personne ne peut le dire actuellement. Les machines semblent pouvoir posséder certaines capacités d'apprendre, mais il y a probablement des limites au-delà desquelles il ne sera pas possible d'aller.

Jean-Claude Bonnefont apporte un témoignage concernant la géographie. Il y a une vingtaine ou une trentaine d'années, les recherches en géographie basées sur l'informatique avaient plutôt mauvaise presse. Ces travaux semblaient aboutir à enfoncer des portes ouvertes. La méthodologie employée était aussi discutable. Dans le cas d'un cadastre, il y a des éléments mathématiques identifiés et des éléments plus subjectifs comme les classes de fertilité des parcelles ou le niveau de valeur des habitations. Il est difficile de traiter à la fois des données précises mesurables et des données subjectives basées sur des seuils et des nuances. Il semble qu'avec la théorie des espaces flous, il soit possible de progresser.

Marion Créhange est en accord avec les remarques de Jean-Claude Bonnefont. Il n'y a cependant pas que la théorie des espaces flous qui permettent de progresser. D'autres approches permettent de traiter de l'incertitude et de travailler sur des données subjectives. D'une manière plus générale, ce n'est pas parce qu'un résultat sort d'un ordinateur qu'il est bon. De nombreux facteurs peuvent en effet biaiser les résultats, mauvais échantillonnage taille insuffisante de l'échantillon, mauvaises méthodes de mesure, etc...

Gilles Curien demande à Marion Créhange si elle a eu connaissance des activités d'un institut international fondé à Vienne dans les années soixante et qui avait pour objet l'analyse des systèmes et qui, selon certains, aurait pu résoudre tous les problèmes du socialisme.

Marion Créhange pense qu'il faut parfois être un peu utopiste, mais ajoute qu'elle n'a pas d'information sur cet institut.

Jean-Marie Schissler s'interroge sur les systèmes de reconnaissance des formes et sur les robots japonais qui utilisent ces systèmes.

Marion Créhange pense que la dimension temps n'est pas plus difficile à intégrer que les autres en informatique, ce qui n'est pas tout à fait l'avis de Jean-Marie Schissler en raison du temps de réponse qui doit être très rapide dans le cas de ces robots.

Michel Vicq remercie Marion Créhange pour la qualité de son exposé. L'évocation de «l'intelligence économique» et du mot légal par l'oratrice lui fait rappeler que l'Etat se réserve les moyens de recueillir des informations par toutes les filières possibles, ce qui n'est pas forcément légal. Il y a donc, dans l'intelligence économique, la convergence d'une multitude de renseignements recueillis par de nombreuses sources, mais qui, la plupart du temps, ne sont pas légales. Ce mot légal devrait par conséquent être mis entre guillemets ou tout simplement supprimé de façon à ce que la définition soit plus rigoureuse et conforme à la réalité.

Marion Créhange indique qu'elle a utilisé un terme couramment utilisé dans la littérature.

Louis Philippe Laprévotte remercie à son tour l'oratrice et évoque les vicissitudes de divers organismes européens dont une Université des peuples d'Europe. Il demande à Marion Créhange les raisons pour lesquelles elle n'a pas évoqué le problème de la reconnaissance des «signaux faibles».

Marion Créhange indique qu'elle a abordé ces questions dans le type «entrée» et le type reconnaissance de formes.



Notes

- [1] Petit Larousse
- [2] Créhange Alain : «Le Pornithorynque est un salopare», Ed Mille et une Nuits, Paris 2004.
- [3] Encyclopédie Microsoft® Encarta® 2003
- [4] On pourra trouver des définitions par exemple dans le site Internet de DEFIDOC : http://www.defidoc.com/info_doc_connaissance/DocDonneeInfo.htm : DOCUMENT, DONNÉE, INFORMATION, CONNAISSANCE, SAVOIR.
- [5] Transformation «analogique», c'est-à-dire effectuée non par des calculs mais directement par un dispositif physique ayant un comportement analogue à la transformation qu'on souhaite réaliser.

- [6] Exemple, imaginaire, de connaissance : si la différence de hauteur du spectre aux pics est supérieure à celle aux creux, cela implique que le sujet n'articule pas assez.
- [7] Illustration : un système de recherche d'itinéraire ferroviaire à qui on demande l'itinéraire pour aller de A à B et qui, au lieu de répondre «il n'y en a pas», répond «il n'y a plus de gare à A mais pour aller à B essayez de partir de C, desservi par autocar depuis A».
- [8] Sous forme de coordonnées des points de sa surface et éventuellement de cavités.
- [9] L'éthologiste Bertrand Krafft en coopération avec le CRIN-LORIA.
- [10] Portail : moyen d'accéder à un ensemble de bases de données sur Internet.
- [11] Catherine Garbay : Repenser l'interdisciplinarité : Les sciences du traitement de l'information comme pivot de l'interdisciplinarité : une vision systémique (Site «Interdisciplines»).