



Chaire de Recherche du Canada en Mondialisation, Citoyenneté et Démocratie
<http://www.chaire-mcd.ca/>

DUCHASTEL, J.; LABERGE, D. (1988). " L'analyse assistée des trajectoires judiciaires ". In *Actes du colloque 'Représentations du réel et informatisation'*. Saint-Étienne (France): I.U.T., 1988. 21 p.

L'ANALYSE ASSISTÉE DES TRAJECTOIRES JUDICIAIRES

Communication présentée au Colloque du CREIS

Représentations du réel et informatisation

Saint-Étienne

Jules DUCHASTEL

professeur au département de sociologie

et chercheur au Centre d'ATO

et

Danielle LABERGE

professeur au département de sociologie

UNIVERSITÉ DU QUÉBEC A MONTRÉAL

AVRIL 1988



Travaillant en sociologie sur un problème particulier d'analyse de données, nous avons été amenés à réfléchir non seulement sur les modalités d'utilisation de l'informatique, mais aussi sur le sens de l'informatisation d'une démarche sociologique. Notre projet de recherche soulève plusieurs difficultés qui se rapportent directement au thème du présent colloque, essentiellement sur le représenté, le représentable et les modalités de la représentation. Il nous amène d'abord à rappeler la nature de la démarche sociologique et sa place dans l'ensemble des pratiques scientifiques. Il nous conduit également à interroger le rôle de l'informatique et des techniques évoluées, dites d'intelligence artificielle, dans la production de modèles orientés vers la généralisation et l'interprétation de données sociologiques. Après avoir esquissé ces questions générales dans les deux premières sections, nous présenterons notre problématique de recherche et la stratégie informatique que nous avons retenue pour élaborer un modèle de représentation, de traitement et d'analyse des données. Nous conclurons sur l'apport de notre démarche à la résolution partielle du problème de la représentation et de la modélisation des données sociologiques.

1. QUELQUES CARACTÉRISTIQUES DE LA RECHERCHE SOCIOLOGIQUE

Nous supposons ici que la recherche sociologique ou, de manière plus générale, la recherche dans le domaine des sciences sociales soulève un certain nombre de questions qui lui sont propres ou qui y revêtent un intérêt particulièrement marqué. Ces caractéristiques affecteront nécessairement la place de l'informatique et son mode d'utilisation dans la recherche.

Roy D'Andrade (1986) distingue trois groupes différents de sciences. Le premier groupe correspond aux sciences physiques et se caractérise par ce que Hempel (1965) dénomme le modèle des lois scientifiques. Ces sciences produisent des énoncés quantifiés, tendent à réduire le plus possible le nombre d'objets et de relations et à formuler des lois. Le second groupe correspond aux sciences naturelles qui, s'intéressant à des niveaux de réalité plus complexes, produisent des énoncés quantifiés correspondant plus à des généralisations qu'à des lois. Le troisième groupe, qualifié de sciences sémiotiques, étudie des ordres construits faisant nécessairement appel au sens et produit des interprétations. Ce groupe s'intéresse à la question de savoir comment ces divers ordres se génèrent, se superposent et se transforment. Il n'est pas question de discuter ici la part des diverses sciences sociales ou des diverses approches de ces disciplines qui revient au modèle naturel ou au modèle sémiotique. Nous rappelons ces distinctions pour mieux indiquer notre conviction de l'aptitude décroissante des systèmes informatiques à gérer des domaines scientifiques qui étudient des phénomènes dont la complexité et l'indétermination s'accroissent.

Au départ, les sociologues travaillent sur des questions dont les unités de base sont toujours construites. Il n'existe pas, en sociologie, d'unités naturelles, c'est-à-dire qui seraient immédiatement saisissables sans recours à une opération d'attribution de sens. L'élargissement de l'informatisation à des domaines les plus divers en sciences sociales nous interroge à nouveau sur le statut de ces unités sujettes à traitement. L'idée familière de la décomposition vers les formes élémentaires de tout objet se concilie facilement avec le principe de base du traitement de l'information à l'aide de l'ordinateur. Pourtant, la quête d'unités simples, finies, ayant des frontières incluses dans l'objet même apparaît comme une stratégie qui, bien qu'elle semble s'adapter idéalement au langage de la machine, constitue une erreur puisqu'elle se base sur une représentation inadéquate de l'objet social.



Nous suggérerons que les unités en sociologie sont nécessairement abstraites (construites) et que leur saisie n'est possible qu'au niveau d'abstraction retenu. Cela implique qu'il ne préexiste pas de formes élémentaires de quelque réalité sociale qu'il s'agirait de repérer, dans un premier temps, pour en restituer toute la complexité d'organisation, dans un second temps.

Le sociologue doit également prendre en compte la complexité nécessaire des données en sociologie. Cette complexité se manifeste par la présence d'un très grand nombre d'éléments de niveaux différents, interreliés et sujets à des transformations constantes. Nous posons qu'il n'y a pas de limites (fixées par les objets eux-mêmes ou par les traditions disciplinaires) qui déterminent le niveau optimal des insertions successives dans des contextes plus larges, quelque soit le problème de recherche. Cela revient à dire, non seulement qu'il n'y a pas de contexte parfaitement isolable et que nécessairement ces contextes s'emboîtent à l'infini, mais que ceux-ci, pas plus que les unités, n'existent indépendamment de leur construction. Il en découle que l'emboîtement des contextes renvoie davantage à la stratégie de recherche retenue qu'à la quête d'une forme optimale de représentation. La complexité nécessite également que l'on tienne compte maximale des interrelations entre les éléments autant au plan synchronique qu'à travers leur transformation temporelle.

Enfin, puisque la sociologie travaille sur des données qui, par définition, sont de nature sémiotique, leur représentation renvoie à l'épaisseur des diverses couches de sens qui les investissent. Avant même de penser les règles de construction de l'objet, il nous faut identifier les divers processus de construction dont les objets sont déjà investis. Ainsi, les données sont souvent organisées selon d'autres logiques; par exemple, dans notre projet, les données statistiques, les dossiers administratifs, les rapports médicaux, les compte-rendus... sont déjà mis en forme dans une logique particulière qui se présente comme une évidence. Le travail de recherche consiste autant en l'identification de la logique à l'oeuvre, qu'en la réutilisation, en dépassant ou contournant leur mise en forme première, des matériaux dans une autre perspective. C'est à ce dernier niveau que se pose la question des modalités de la représentation. Nous y reviendrons à propos des procédures de formalisation associées à l'utilisation de l'ordinateur.

2. INFORMATISATION ET RECHERCHE SOCIOLOGIQUE

Lorsque l'on parle de procéder à l'informatisation d'un objet de recherche en sociologie, on peut concevoir la chose de deux manières différentes. L'usage le plus fréquent de l'informatique fait appel à sa fonction de calcul sur des masses de données avec efficacité et rapidité. Un usage plus récent, en même temps que beaucoup plus rare pour le sociologue, correspond aux capacités de modélisation de l'informatique dans le but de fournir une "représentation supposée opératoire du réel". En quoi ce deuxième usage est utile à la démarche sociologique et, plus spécifiquement, comment contribue-t-il à mettre en forme, à modéliser ou à analyser les données sociologiques? En contrepartie, une contribution de la sociologie aux développements informatiques est-elle pensable?

La présentation de notre projet illustrera la place privilégiée que nous faisons à ce second usage. Des conceptions assez différentes sont entretenues face à l'utilisation de l'ordinateur à des fins de simulation de comportements intelligents. Deux positions extrêmes définissent les pôles entre lesquels se retrouvent la plupart des chercheurs dans le domaine en formation des sciences cognitives. Ou bien on pense que la simulation de processus intelligents est indépendante de toute ressemblance ou dissemblance présumée par rapport aux comportements humains, ou bien cette simulation cognitive est l'objet d'un postulat



d'isomorphie entre les deux types de comportements. L'approche utilitariste voit avant tout l'extension des usages de l'ordinateur considéré comme un outil. La seconde approche recherche la validité de sa démarche dans la reproductibilité de l'expérience humaine. L'informatique devient de plein droit un lieu d'expérimentation de modèles hautement structurés, un simulateur de l'explication scientifique du réel. Dans les deux cas, cependant, on n'assiste plus à des débordements d'optimisme sur la venue prochaine d'une intelligence exhaustive des machines. Suite aux critiques adressées à cet optimisme des pionniers (entre autres, Dreyfus, 1984) et à l'évolution même des recherches dans des domaines d'application pour lesquels "l'homme est, aujourd'hui encore, le meilleur" (Rich, 1987), il faut conclure avec Borillo (Dreyfus, 1984) que la véritable question maintenant est plutôt celle "des limites effectives de l'autonomie décisionnelle des machines".

La création des sciences cognitives constituent un tournant décisif dans l'identification de la nécessité de représenter un monde complexe et indéterminé. Cependant, selon le niveau d'appréhension des phénomènes où l'on se place, une certaine naïveté persiste dans la définition de la représentation. Ainsi, la linguistique ou la psychologie identifieront de manière précise les unités propres à leur domaine, mais auront tendance à rejeter le contexte dans l'univers indéterminé des "faits" et du "réel", comme si ces notions correspondaient à des données indépendantes de leur construction. Pour le sociologue le réel existe en fonction des paramètres de la théorie retenue par lui. Les faits n'ont pas d'indépendance par rapport à ce réel construit.

Même si la sociologie a surtout utilisé jusqu'à présent la puissance de calcul numérique des machines, ne pouvons-nous poser la question de son inclusion dans ce nouveau groupe de sciences dites cognitives? La sociologie travaille sur des objets qui cumulent plusieurs couches de représentations et se placent à travers plusieurs emboîtements de contextes. Pourtant, nous constatons l'absence relative des sociologues, jusqu'à maintenant, dans l'élaboration des "mondes" qui servent de toiles de fond aux systèmes "intelligents".

Par ailleurs, une certaine confusion persiste sur l'idée même de la formalisation nécessaire à l'usage de l'informatique. Autant la tradition des sciences positives que les principes mêmes du fonctionnement de l'ordinateur plaident en faveur de modèles formalistes et de procédures qui se prêtent au calcul. Ce modèle correspond davantage au groupe des sciences physiques et, dans une moindre mesure, au groupe des sciences naturelles: l'identification des éléments les plus petits (particules, primitifs) et de leurs relations; l'élaboration de règles systémiques; la production de lois ou de généralisations.

Par contre, certaines disciplines, à cheval entre les sciences naturelles et les sciences sémiotiques, produisent des modèles théoriques hautement structurés et, par là même, recourent à des formalismes de représentation et d'opération (i.e. formalismes logiques). D'autres disciplines, en raison même de la complexité de leur objet, ne peuvent atteindre un même degré de structuration théorique et auront plutôt recours à des représentations structurées et à des procédures de formalisation des divers opérations sur leur objet.

La sociologie appartient à ce dernier groupe et adoptera donc le point de vue utilitariste de l'informatique outil. Cela ne la dispense pas de s'interroger sur le degré de formalisation des représentations et des procédures qu'elle devra atteindre pour utiliser avec profit l'ordinateur dans sa fonction d'aide à la modélisation. C'est là probablement le plus grand profit du recours à l'informatique: la nécessité de procéder de manière rigoureuse au découpage de notre objet, à l'identification de nos stratégies de traitement et d'analyse. Inversement, la nécessité de



traiter la complexité à son niveau le plus grand ne peut que bénéficier aux recherches sur les nouveaux modèles de représentation et les procédures de traitement de l'information.

L'informatique peut, cependant, se présenter comme un obstacle au travail du sociologue. L'informatique impose nécessairement ses contraintes: elle propose des modalités particulières de représentation et de traitement formalisé des données. Cela a des conséquences autant sur le type de données qui sont disponibles au sociologue que sur le type de travail de construction des données qu'il peut lui-même produire. En premier lieu, l'accroissement du potentiel de traitement des machines ainsi que l'accessibilité accrue de ces mêmes machines ont favorisé l'accumulation de l'information d'une façon phénoménale, complexifiant ainsi la tâche de décodage et de sélection à laquelle doit faire face tout chercheur. L'informatique produit donc une nouvelle couche de représentation des données qu'il nous faut décrypter. En deuxième lieu, dans son usage, l'informatique nous contraint à des façons de faire en recherche qui ne s'imposeraient pas autrement. Au pire, cette difficulté s'accroît lorsque le sociologue ne contrôle pas bien cet usage, au mieux, cette difficulté est analysée et l'informatique redevient ce qu'elle doit être pour le sociologue, un outil.

3. NOTRE OBJET DE RECHERCHE

Le projet de recherche sur lequel nous travaillons présentement et qui nous servira ici d'illustration présente un bon échantillon des problèmes qui confrontent les chercheurs en sociologie. Soulignons en premier lieu que deux types de difficultés devaient être examinés: celui de l'informatisation de la banque de données initiale et celui de l'analyse spécifique d'une partie du corpus qui pose des problèmes méthodologiques importants. Nous procéderons donc ici en deux temps qui correspondent aux deux ordres de questions mentionnées en accordant une attention particulière au second ordre qui se rapporte davantage à l'usage de l'informatique comme aide à la modélisation.

3.1. Contexte et objectifs de la recherche

Le corpus sur lequel nous travaillons est constitué d'un ensemble de dossiers de personnes ayant eu des contacts, à titre de justiciables, à la Cour du Bien-être de Winnipeg, durant la période 1910-1960. Il s'agit du seul ensemble de données historiques provinciales en matière de délinquance juvénile, actuellement connu des chercheurs au Canada, à avoir échappé à la déchiqueteuse. Cette sauvegarde est due à Len Kaminsky, Professeur à l'Université du Manitoba, qui s'est employé à recenser la masse de documents qu'il contenait et à en traiter l'information afin de la rendre accessible au plus grand nombre possible de chercheurs.

Le corpus est, dans son ensemble, diversifié puisqu'il est constitué de dossiers relevant de juridictions différentes. En effet, ce tribunal devait gérer neuf (9) types de lois, la plupart relatives aux personnes mineures ou aux familles. Parmi cet ensemble, nous nous sommes intéressés aux dossiers relevant de la Loi des jeunes délinquants, pour des raisons théoriques mais aussi à cause de l'éminence de leur destruction.

Les dossiers de délinquance juvénile sont au nombre approximatif de 3.000 et les autres types de dossiers représentent entre 40.000 et 50.000 unités. Ces autres dossiers sont plus ou moins directement reliés aux jeunes personnes. Chaque dossier, quelque soit la juridiction sous laquelle il est enregistré, est constitué sur la base d'une personne (par opposition à la constitution sur la base d'incidents spécifiques) et représente ainsi une sorte de portrait assez complet de l'histoire judiciaire de l'enfant jusqu'à sa majorité.



En pratique, chaque dossier comprend un nombre variable d'informations. On y retrouve nécessairement les transactions du tribunal comprenant chacune des infractions pour lesquelles l'enfant comparaît. A cette information de base, dont l'importance varie en fonction du nombre de délits dont sera accusé, à travers le temps, un enfant donné, on retrouve des documents divers qui ont été utilisés par le tribunal dans la gestion particulière de chaque cas. Ces documents peuvent être regroupés sous neuf rubriques principales (ex. médical, légal, police, psychiatrique, scolaire,...) et se répartissent dans plus de 166 catégories différentes. Par ailleurs, on peut retrouver un même type de document à plus d'une reprise dans le dossier, soit parce que le juge demande une contre-expertise, soit parce qu'il y a suivi, soit à cause d'un nouvel incident. Un seul dossier peut comprendre entre une et quatre-vingts inscriptions différentes (données, documents...).

La première opération d'importance dans la stratégie d'informatisation a consisté à distinguer entre ce qui est propre au repérage des documents et ce qui concerne spécifiquement l'analyse sociologique du même matériel. Cette première distinction n'est pas évidente. En effet, elle suppose d'abord que les chercheurs renoncent à l'idée selon laquelle l'informatisation n'est qu'une opération de retranscription dans un autre langage de données identiques. Cette première opération s'est traduite par un traitement de l'information à l'aide de protocoles de catégorisation et d'analyses statistiques de base. Une première grille de données recense l'information disponible concernant des variables objectives concernant le jeune ou sa famille. Une seconde grille constitue un inventaire des documents présents dans chaque dossier. Dans tous les cas, les co-références entre dossiers sont notées.

Nous travaillons maintenant sur l'exploitation d'une partie du corpus dans le cadre d'un projet de recherche intitulé "Justice juvénile et psychiatrie, 1910-1945" et financé par le Conseil de recherches en sciences humaines du Canada. L'objectif de notre projet est d'étudier les rapports entre les systèmes pénal et psychiatrique de prise en charge des personnes mineures, en tant qu'ils évoluent tout au long de la période. Mais, plus spécifiquement, il nous intéresse d'étudier les diverses trajectoires individuelles, observables dans les dossiers, en tenant compte d'un ensemble de variables indépendantes (sexe, âge, milieu socio-économique, ethnie...) et de variables dépendantes (modalités de prise en charge, trajectoires types, statut des intervenants...).

3.2. Stratégies d'analyse

Indépendamment du traitement des données dans des séries temporelles qui tiendrait compte du processus, nous avons avant tout à résoudre des problèmes liés aux caractéristiques même de ces données. Rappelons d'abord le caractère complexe et peu standardisé de cette base de données. Nous pouvons disposer d'une variété infinie de renseignements, d'évaluations, de prises de décision à propos d'un individu dans des formats variables. On ne peut donc appréhender a priori un contenu invariant. De plus, les informations utiles peuvent ne pas être directement présentes dans le dossier sous observation, mais repérables dans la structure proposée des données. En effet, le système prévoit des interconnexions entre dossiers. Ainsi, une famille aura un numéro unique et ses individus un code alpha. Les dossiers identifient également les collatéraux: complices, adultes contribuant... Enfin, à certains égards, les informations objectives à propos d'un individu peuvent être plus complètes dans un dossier collatéral (celui du frère, par exemple) que dans le sien propre.

Les informations sont également datées et posent en tant que tel le problème de leur séquentialisation. Cependant, rien ne garantit que cet ordre temporel soit signalé de manière



standardisée ou ordonnée. Il faut mentionner, en terminant, que la construction de ces données appartient à une ou des logiques administratives indépendantes du contexte de notre propre recherche.

Il existe donc une base de données, constituée indépendamment de la volonté du chercheur, dans une logique administrative, dont le caractère est d'être complexe, variée et non standard. Compte-tenu des objectifs du projet, la constitution d'une banque de données informatisée "traditionnelle" donnant lieu à l'utilisation de progiciels de traitement statistique, nous apparaît peu efficace et ce pour plusieurs raisons. Il existe une trop grande variation dans la taille des dossiers ce qui nous forçait à déléster trop d'informations ou alors à prévoir des protocoles de saisie extrêmement lourds à gérer. Par ailleurs, la structuration de telles bases de données se prête mal à l'analyse des transformations temporelles qui constitue une dimension essentielle de notre projet. En dernier lieu, il est très difficile de procéder à des explorations systématiques sur des modèles plus complexes, compte-tenu des modes de traitement habituels. Nous nous trouvons devant la tâche de développer des stratégies de description, d'identifier des modalités d'exploration et d'analyse de ces données.

L'utilisation des bases de données traditionnelles nous semble tout à fait inappropriée compte-tenu des caractéristiques des dossiers de délinquance juvénile et de la problématique que nous développons. Pourtant, les choix informatiques que doit faire le chercheur ne peuvent se réduire à la seule question de la nature de la base de données retenue (matricielle ou relationnelle). Le véritable défi consiste alors à définir les éléments d'information de telle sorte qu'ils puissent permettre la construction subséquentes des "faits" sociologiques et leur insertion dans des "contextes" explicatifs. Puisqu'il n'existe pas, pour le sociologue, d'unités naturelles, toutes les catégories permettant la mise en forme de l'information doivent être investies théoriquement. Plus les matériaux sont complexes et multiformes, plus il existe de possibilités de définir des niveaux différents de problématisation, plus la tâche d'élaboration des catégories de données est difficile. Ce travail de définition constitue un préalable incontournable au développement d'un modèle de saisie des données dans une base de données relationnelles. A cette étape particulière de notre travail, nous avons donc opté pour une approche utilisant l'informatique comme support à l'opérationnalisation conceptuelle.

4. UN SYSTEME D'ANALYSE GRAMMATICALE ASSISTÉE (SAGA)

Nous avons retenu un système informatique nommé Système d'Analyse Grammaticale Assistée (SAGA). Ce système a été développé au Centre d'Analyse de Textes par Ordinateur (Centre d'ATO) de l'Université du Québec à Montréal. Il fait partie d'un atelier cognitif construit autour du logiciel Déredec. Déredec est un langage de programmation pour l'analyse du langage naturel. Il permet de représenter les données aux divers niveaux du fonctionnement linguistique: lexical, morphologique, syntaxique, sémantique et textuel. Il permet également de développer des grammaires pour décrire ces données à ces divers niveaux. Il donne les moyens d'explorer ces diverses descriptions. Enfin, il est aussi utilisé pour la mise au point de systèmes symboliques. Le Déredec est un cadre computationnel général pour la programmation d'analyseurs ou simulateurs linguistiques basés sur des théories algorithmiques variées. On peut le considérer, en premier lieu, comme un atelier de travail pour les linguistes computationnels. Par ailleurs, il offre aussi toutes sortes de facilités pour l'utilisation des analyseurs linguistiques dans les applications telles les analyses de contenu des textes et les systèmes experts qui questionnent les textes comme des banques de connaissances.



Alors que la plupart des systèmes en Déredec ont adopté une stratégie de description automatique (Parsers), le Système d'Analyse Grammaticale Assistée (SAGA) ne fait appel à des automatismes que dans l'aide à l'écriture et à l'instanciation des SAGA. En effet, un SAGA est avant tout un système permettant la construction d'une structure de représentation des connaissances et l'instanciation de cette structure à partir d'objets réels. Ce système se présente sous forme d'un arbre de décision contenant des branches et des feuilles et servant à représenter des connaissances sur un objet donné.

Le système comporte deux moments. Le premier est celui de l'écriture durant lequel l'utilisateur construit sa structure de représentation à partir d'un système assisté qui lui propose des choix dans des menus et qui écrit pour lui les automates et les modèles constitutifs de ce SAGA. Construire un SAGA ne consiste qu'à déclarer dans l'ordre hiérarchique les noms des membres et la nature (conjonctive ou disjonctive) de leur domaine, les liens pouvant relier les membres, ainsi que le type de feuilles pouvant pousser à la fin des branches, c'est-à-dire le type d'information qui servira à l'instanciation. On retrouvera au schéma 1 les différentes caractéristiques des membres d'un SAGA. L'utilisateur d'un SAGA n'exécute que deux types d'opérations: il aiguille l'algorithme à chaque disjonction de membres et il remplit les feuilles des branches.

Le second moment consiste en l'application du SAGA sur des cas concrets. Le système assiste alors l'utilisateur dans l'instanciation à l'aide de menus. SAGA produit donc deux types d'arbres: l'arbre procédural de décision qui est construit au niveau de la conception du système et l'arbre instancié qui reflète les choix effectués lors de l'application du SAGA à des objets réels.

Sur le plan informatique les SAGA sont un ensemble d'automates Déredec dont la stratégie d'écriture est descendante; il s'agit d'une dérivation d'automates créant des structures de représentation sous forme de triplets ordonnés ((catégorie) (relation de dépendance contextuelle) (valeur)). Tous les aiguillages dans la dérivation sont automatiquement transformés en menus offerts à l'usager au moment de l'exécution du SAGA. L'orientation générale est déterministe bien qu'il soit possible de programmer des retours-arrière dans le cas de contradictions entre deux entrées pour une même question.

Trois raisons expliquent, par ordre d'importance croissante, le choix de SAGA dans le cadre de notre recherche. Premièrement, il faut souligner les fonctions d'aide à la programmation de cet environnement et la construction automatique d'une interface usager qui facilitera grandement les manipulations des données. Deuxièmement, le système offre une vraie structuration arborescente des données, une hiérarchie souple, sans contrainte sur le nombre de niveaux. Les cheminements de fouille n'ont pas à être prédéterminés et il n'existe pas de limite aux niveaux de profondeur où les explorations peuvent être étendues. Troisièmement, l'avantage fondamental est de conserver l'environnement Déredec.

Toutes les stratégies de construction (automates travaillant sur la séquence, en mémoire ou sur de nouvelles séquences) et les stratégies sélectives et prédicatives (modèles d'exploration) peuvent être appelées dans l'environnement. En effet, il est possible de bénéficier de toutes les propriétés de Déredec afin d'utiliser les relations de dépendance contextuelle pour explorer et modifier des séquences qui peuvent être associées aux divers noeuds de l'arbre. Les liens entre les divers arbres ne s'effectuent pas de fichiers à fichiers, mais d'une branche à l'autre. Les modèles d'exploration Déredec servent à établir ces liens et autorisent le transport d'automates aptes à modifier les séquences ainsi explorées. Enfin, il existe également la possibilité



d'établir un lien avec Déredec-EXPERT afin d'utiliser son moteur d'inférence au niveau de l'analyse des données.

Si la locution SAGA suggère l'idée que les utilisateurs recevront assistance dans la construction d'une grammaire quelconque, cette assistance est procédurale et le chercheur est seul responsable du modèle grammatical adopté. Nous utilisons le terme de grammaire dans son acceptation la plus large, comme système de règles permettant de structurer un niveau donné de la réalité étudiée. Faute d'une intuition minimale concernant la possibilité de fournir des règles, il serait superflu de recourir à une procédure aussi lourde. Le caractère déterministe de la description en arbre n'a de profit que dans la mesure où le chercheur retient une stratégie de construction de séquences significatives d'objets ou d'événements. Nous verrons par la suite que nous avons dans notre projet un seul arbre qui répond à ce critère (le contact judiciaire), alors que les autres, dans l'état actuel de notre prototype, collectent des faits relativement indépendants (fiche lieu, fiche temps, fiche tranche de vie).

Nous avons donc considéré SAGA comme un banc d'essai pour le développement d'une grammaire d'événements. C'est donc, jusqu'à présent, dans les termes de l'élaboration d'une maquette que nous avons travaillé. Les SAGA offrent un environnement propice pour tester des hypothèses de structuration des représentations à propos d'un objet et de manipulations de ces représentations.

A une phase ultérieure d'optimisation du système, il pourrait être envisagé de transposer, en tout ou en partie, la représentation des données dans un SGBD relationnelle. Afin d'obvier à la réduction des possibilités théoriques qu'offre l'environnement Déredec pour établir des rapports avec d'autres niveaux de représentation et de traitement des données, nous pouvons prévoir des interfaces tel que le permet l'architecture de SAGA.

5. MODELE D'ANALYSE ASSISTÉE DE TRAJECTOIRES JUDICIAIRES

Le modèle que nous avons développé est conçu: 1) pour représenter plusieurs sous-ensembles d'informations à propos d'un individu qui seront mis en relation à certains moments déterminés de l'analyse; 2) pour construire à l'aide d'une grammaire d'événements des trajectoires composées d'un ou de plusieurs contacts emboîtés; 3) pour explorer les données, enrichir la base à partir de résultats et poursuivre l'exploration; 4) pour fournir un cadre d'analyse de ces trajectoires.

5.1. Description des fiches

L'information sera consignée sur des fiches, actuellement au nombre de quatre et dont le statut varie: la fiche "acteur", la fiche "lieu", la fiche "tranche de vie" et la fiche "inscription".

L'intérêt de distinguer plusieurs fiches est d'éviter de véhiculer une information non pertinente à certaines phases de la recherche. Les trois premières comportent de l'information qui pourra être fouillée à l'occasion de l'application de modèles d'exploration particuliers. La fiche "acteur" recense l'information à propos des individus, quelque soit leur statut, qui seront mentionnés dans les dossiers. La fiche "lieu" comporte des données caractérisant des lieux physiques. La fiche "tranche de vie" concerne des informations sur des bribes de l'histoire personnelle des individus.

La principale fiche est celle des "inscriptions". Elle diffère des autres puisqu'elle constitue la base à partir de laquelle seront contruits les contacts, comme nous le verrons plus loin. L'inscription constitue pour nous l'unité la plus petite d'une trajectoire individuelle. Elle est



une marque dans le dossier d'un événement à propos d'un individu. Elle comprend nécessairement certains éléments de base, mais l'importance des informations varie d'un événement à l'autre. Toute inscription comprend des informations à propos de son contenu (i.e. ce qui le spécifie comme événement); des acteurs impliqués (minimalement, l'acteur principal, le délinquant, et le juge concerné), des indications sur le temps et le lieu. L'ensemble de ces informations constituent les références pour la construction ultérieure du contact. Le schéma 2 nous montre les éléments constitutifs de cette fiche, la nature des informations que l'on peut instancier (voir exemple du schéma 3), la nature unique ou répétable, obligatoire ou facultative, conjonctive ou disjonctive de chaque membre.

5.2. Construction d'un contact

L'idée maîtresse de notre modèle est la construction d'un contact. Nous pouvons définir le contact comme une histoire, un segment complet et souvent unique d'interactions entre un enfant et le tribunal. La notion demeure difficile à définir à cause de son contenu extrêmement variable. Réduit à sa plus simple expression, un contact est composé de deux événements, une accusation et un verdict. Dans les faits, un contact est toujours beaucoup plus complexe: chaque contact est ponctué par une série d'événements (reportés sur des fiches "inscriptions") qui en constituent le profil particulier. Ainsi, on retrouve plusieurs comparutions devant le tribunal, des évaluations de spécialistes, des interventions de la part d'acteurs non judiciaires (parents, professeurs, voisins,...), des mesures temporaires et une décision du juge qui dispose du cas.

L'inscription d'un délit signale donc le début d'un contact et la disposition par le tribunal, relative à ce même délit, signale la fermeture du contact. Chaque événement inscrit au dossier et compris entre ces deux marqueurs est considéré comme une inscription et contribue au profil particulier d'un contact. L'étude des fiches "inscriptions", ordonnées dans le temps, nous permettra de caractériser la trajectoire globale de l'individu. Celle-ci sera constituée d'un ou d'une série de contacts qui peuvent se présenter en séquence ou en emboîtement, peu importe. Le schéma 4 illustre comment, à partir d'une séquence d'événements, SAGA construit un contact en repérant l'ensemble des événements s'y rapportant. Le schéma illustre également comment, à partir de grammaires-photos, un contact peut être emboîté dans le premier. Dans l'exemple décrit, un second délit ouvre un deuxième contact avant que le premier ne soit refermé. Ici la fermeture pour les deux éléments sera commune, ce qui n'est pas une nécessité.

5.3. Analyse des contacts

Le schéma 5 présente toutes les étapes de l'analyse des contacts avec le système judiciaire. Pour l'exemple du dossier judiciaire #123, nous disposerons d'une banque de connaissances, composée d'un grand nombre de fiches de nature différente. Nous avons ajouté au schéma une fiche inférence qui pourrait contenir de l'information produite à partir d'un moteur d'inférence opérant sur les résultats de modèles d'exploration. Les contacts seront construits sur la base de l'ensemble ordonné de toutes les inscriptions à propos d'un individu. L'ensemble même de ces contacts constituera la trajectoire d'un individu. Des modèles de fouille en Déredec pourront être appliqués à ces contacts, ainsi qu'à toutes les autres fiches de manière relationnelle. Les résultats obtenus pourront être l'objet d'analyse ou considérés comme des requêtes à un système expert dont la tâche serait de générer de l'information supplémentaire pour des explorations subséquentes. Par exemple, il serait possible d'inférer des informations sur la sévérité des décisions prises pour un individu en rapport avec des contacts antérieurs, afin de tenir compte de cette information dans des modèles d'exploration subséquentes.



6. CONCLUSION

Nous reviendrons en conclusion sur les deux questions posées dans ce colloque, celui de la représentation du réel et celui de l'informatisation de la démarche de recherche. Le problème de recherche que nous avons brièvement esquissé, illustre particulièrement bien la difficulté de représenter nos objets sociologiques. Ainsi, il nous a fallu définir les unités d'information pertinentes, résoudre le problème complexe de leur distribution dans les dossiers individuels ou dans des dossiers interconnectés, tenir compte de leur présence aléatoire et prévoir leur traitement temporel. La solution informatique retenue n'est pas nécessairement définitive. Comme nous l'avons dit, plusieurs structures de représentations dans des logiciels de base de données relationnelle pourraient être envisagées. Cependant, il nous fallait d'abord opérationnaliser conceptuellement un modèle apte à représenter, à construire des descriptions, à les explorer et à produire des analyses. Le choix opéré nous a permis ce découpage rigoureux de l'objet et l'identification de ces stratégies. A cette étape, nous avons utilisé l'ordinateur comme outil servant non pas au calcul sur des masses importantes de données, mais avant tout comme aide à la modélisation.

La pleine capacité du système s'observe dans la démarche de construction des contacts et la stratégie d'analyse des trajectoires. On y voit à l'oeuvre les procédures de construction d'un objet conceptuel que nous avons nommé "contact", avec les possibilités d'interrelations entre eux, ce qui constitue une complexification supplémentaire du modèle. A partir d'une séquence non structurée d'événements discrets, on reconstruit des séquences structurées d'événements se rapportant à un même principe défini théoriquement. L'examen de la stratégie d'analyse indique la puissance d'investigation à travers des modèles d'exploration pouvant circuler dans la structure complexe de représentation de nos données. Au niveau de la maquette, l'ensemble de ces dispositifs sont fonctionnels, il nous reste à résoudre le problème de leur généralisation sur des ensembles de données plus importants.