

Développement d'un système expert d'aide à l'enseignement de l'indice matériel

Romain Voisard

23 octobre 2003

Table des matières

Table des matières	i
Liste des figures	iii
1 Introduction	1
2 Cadre théorique et pédagogique	3
2.1 L'interprétation comme pratique des sciences forensiques	3
2.1.1 Principes fondamentaux des sciences forensiques	3
2.1.2 L'interprétation de l'indice matériel	5
2.2 Un système expert comme aide à l'enseignement de l'interprétation	8
2.2.1 Notions générales sur les systèmes experts	8
2.2.2 Comparaison avec d'autres programmes standards	8
2.2.3 Fonctionnement d'un système expert	9
2.2.4 Un système expert comme outil d'enseignement	10
2.2.5 Perspectives pour l'enseignement de l'interprétation	11
3 Objectif et problématique de recherche	13
3.1 Hypothèse et but du travail	13
3.2 Méthodologie	14
4 Développement du prototype	17
4.1 Etat de la situation	17
4.1.1 Les réseaux bayesiens	17
4.1.2 Les systèmes experts	18
4.2 Choix des techniques	20
4.3 Modélisation de la connaissance	21
4.3.1 Limitation	21
4.3.2 Modélisation des principes	21
4.4 Développement du logiciel	25
4.5 Test pratique	28
5 Résultats et discussion	29
5.1 Le prototype	29
5.2 Test utilisateur	32

5.3 Perspectives	32
6 Conclusion	33
Bibliographie	35

Table des figures

2.1	Paradigme des sciences forensiques selon Inman et Rudin	4
4.1	Les différents type de connections dans un réseaux bayesien : (1) en série, (2) convergente et (3) divergente	18
4.2	Modèle graphique pour le transfert et la persistance des fibres sur une cible	22
4.3	Démarche du criminaliste dans le processus d'individualisation des fibres . .	22
4.4	Réseau bayesien de la connaissance dans le contexte d'une inférence à la source	23
4.5	Paramètres qui influence l'évaluation du rapport de vraisemblance	23
4.6	Réseau bayesien de la connaissance dans le contexte d'une inférence au contact de deux objets	24
4.7	schéma relationnel de la prise d'information du système et de son stockage .	25
4.8	Modèle de la prise de décision en fonction des données entrées par l'utilisateur	26
4.9	Arbre des possibilités de formulation du LR au niveau de la source en fonction des informations données par l'utilisateur	27
4.10	Arbre des possibilités de formulation du LR au niveau du contact. Plusieurs informations sont alors nécessaires pour pouvoir formuler un LR. L'arbre se termine lorsque le système a toutes les informations requises	27
5.1	Illustration du prototype, module de collecte des données au niveau du cas .	29

Chapitre 1

Introduction

Ce travail de mémoire présente une recherche sur l'utilisation d'un système expert comme aide à l'enseignement de l'interprétation de l'indice matériel. Ce sujet de recherche trouve son origine dans ma volonté de lier mes études entreprises à TECFA en technologies éducatives avec mon contexte professionnel à l'Institut de Police Scientifique de l'Université de Lausanne.

Nous allons dans un premier chapitre fixer le cadre général de la recherche en partant de son contexte d'insertion pour aboutir à la définition de ce que doit être un système expert pour être utilisé comme aide à l'enseignement de l'interprétation de l'indice matériel. Nous dressons une synthèse des principes fondamentaux sur lesquels se basent toutes les sciences forensiques et en particulier l'interprétation. Nous avançons, ensuite, les possibilités qu'offre la technologie des systèmes experts pour l'enseignement en général et enfin au cas particulier de l'interprétation.

A la lumière de ce cadre général, la problématique de recherche dans le chapitre 3 où nous fixons les hypothèses et but de ce travail. Une méthodologie est aussi définie.

Cette méthodologie est reprise dans le chapitre 4 pour structurer les étapes du développement du prototype. Enfin, dans un dernier chapitre avant la conclusion de ce mémoire, le programme réalisé est présenté et discuté. Des voies de développement et des axes de recherches sont définis dans le contexte actuel du domaine en question.

Chapitre 2

Cadre théorique et pédagogique

Ce chapitre apporte une vision générale du cadre de la recherche en partant de son contexte d'insertion, pour aboutir à la définition de ce que doit être un système expert pour être utilisé comme aide à l'enseignement de l'interprétation de l'indice matériel. Nous dressons une synthèse des principes fondamentaux sur lesquels se basent toutes les sciences forensiques et en particulier l'interprétation. Nous avançons, ensuite, les possibilités qu'offre la technologie des systèmes experts pour l'enseignement en général et enfin au cas particuliers de l'interprétation.

2.1 L'interprétation comme pratique des sciences forensiques

2.1.1 Principes fondamentaux des sciences forensiques

Les sciences forensiques sont définies comme l'ensemble des principes scientifiques et des méthodes techniques appliquées à la résolution de questions en matières criminelles, civiles ou réglementaires et qui aide la justice à les résoudre. Souvent, en matières criminelles, la question consiste à prouver l'existence d'un crime, d'en déterminer l'auteur et son mode opératoire. (MARGOT 2003)

Le terme forensique trouve son étymologie dans le mot latin *forum* et signifie "qui appartient au débat public" ou plus spécifiquement au tribunal en référence à la place publique où se tenaient les jugements à l'époque romaine.

Le domaine des sciences forensiques repose essentiellement sur six principes qui guident l'expert dans une progression logique partant de la compréhension de l'origine d'un indice et aboutissant à la signification des résultats analytiques. (INMAN & RUDIN 2001)

La divisibilité : *La matière se divise en composants plus petits si une force suffisante est appliquée. Les composants acquièrent des caractéristiques créées par le processus de division et retiennent des propriétés physicochimiques de la source.* (INMAN & RUDIN 2001)

Le transfert : Le principe de l'échange de matière entre deux objets a été énoncé par Locard en 1920. *Nul ne peut agir avec l'intensité que suppose l'action criminelle sans*

laisser des marques multiples de son passage,[...] tantôt le malfaiteur a laissé sur les lieux les marques de son passage, tantôt, par une action inverse, il a emporté sur son corps ou sur ses vêtements, les indices de son séjour ou de son geste. (LOCARD 1920)

L'identification : Tentative de définir la nature physicochimique de la trace (SAFERSTEIN 1998)

La classification et l'individualisation : Tentative de définir la source de la trace (KIRK 1983)(DEFOREST, LEE & GAENSSLEN 1983)

L'association : Tentative de lier la source d'une trace avec une scène de crime (OSTERBURG 1968)

La reconstruction : La compréhension du déroulement dans le temps et l'espace d'événements antérieurs (DEFOREST et al. 1983)

Un paradigme de la pratique des sciences forensiques peut être formulé en s'appuyant sur ces six principes fondamentaux. Ce paradigme inclut les processus de la formation et d'analyse de l'indice qui décrivent la profession de criminaliste [figure 2.1].

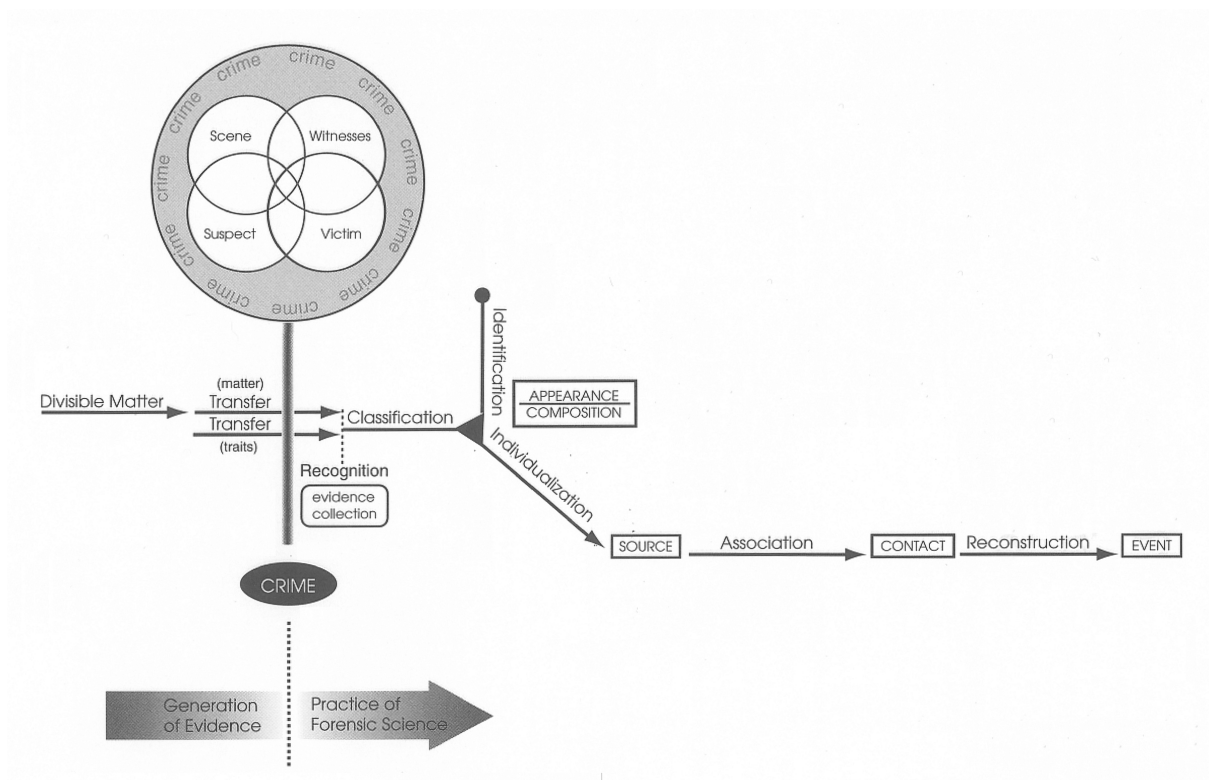


FIG. 2.1: Paradigme des sciences forensiques selon Inman et Rudin

Dans ce paradigme, on peut remarquer que les principes peuvent être classés en trois catégories principales. Ainsi, les principes de divisibilité et de transfert sont liés à l'origine de l'indice étant donné qu'ils interviennent lors de la génération de la trace. La classification, l'identification et l'individualisation sont des processus de réduction de possibilité et permettent de se prononcer sur la source grâce aux caractéristiques physicochimiques de l'indice. Enfin, l'association et la reconstruction s'intéressent aux caractéristiques d'action liées à la produc-

tion de l'indice et permettent donc de s'intéresser au contact.

Comme nous allons le constater au long de ce travail, toutes ces notions ont des répercussions sur l'évaluation et la signification de l'indice.

Pour une étude plus détaillée de ces principes généraux et de leurs implications sur la pratique des criminalistes, la lecture du livre de Inman et Rudin paru en 2001 qui donne une vision globale et systématique de la profession de criminaliste est conseillée.

2.1.2 L'interprétation de l'indice matériel

Définition

Sur une scène de crime, une trace est produite suivant les principes de divisibilité de la matière et de transfert. Elle est reconnue comme étant une trace susceptible d'être en relation avec l'action délictueuse (pertinence de la trace) lors de l'examen de la scène de crime et est préservée. La trace prend ainsi le statut d'indice. L'essentiel consiste ensuite à déterminer la source de la trace. Lorsqu'une source est suspectée d'être à l'origine de la trace, un échantillon de comparaison est prélevé et comparé à l'indice en obéissant aux principes de classification, d'identification et d'individualisation. A la suite des examens, l'expert est confronté à quatre questions :

Est-ce que des différences inexplicables entre les deux échantillons comparés me convainquent qu'ils proviennent de sources différentes ? (exclusion)

Si la réponse est non alors :

Est-ce que la quantité et la qualité des caractéristiques de classe (propre à ce type d'élément) me convainquent qu'ils peuvent provenir d'une même source au sens large ? (classification)

Si oui

Ai-je un nombre et une qualité de caractéristiques individuelles (propre à cet élément précis) suffisants pour me convaincre qu'ils proviennent de cette même et unique source. (individualisation)

Si oui

Ai-je un nombre et une qualité

- d'informations circonstanciées sur le cas
- de caractéristiques d'action (propre au contact)

pour me convaincre que cette source a eu ce contact avec le receveur. (association)

Les trois premières questions infèrent à une origine commune ou non des échantillons alors que la quatrième infère à un contact entre deux objets. Dans chaque situation décrite par ces questions, l'expert, pour remplir son rôle, doit estimer d'une manière ou d'une autre la force de l'inférence en question.

Ainsi l'interprétation de l'indice matériel est définie comme étant les processus mis en jeu dans l'estimation de la force de cette inférence qu'elle soit sur la source ou sur le contact.

Méthodes d'interprétation

Il existe plusieurs méthodes d'estimation de l'inférence, mais, dans le cadre de ce travail de mémoire nous allons nous intéresser à une seule d'entre elles : l'approche bayésienne. Nous nous intéressons à cette approche puisque, d'une part, elle est mathématique et ne repose pas uniquement sur l'expérience de l'expert et, d'autre part, parce qu'elle est pratiquée et investiguée par nombre d'experts dans le domaine et est préconisée lors des études à l'Institut de Police Scientifique qui seront décrites plus tard dans ce document. (TARONI 2003)

Dans une optique bayésienne, un criminaliste aborde une expertise avec au moins deux hypothèses compétitives en tête, l'une rejoignant la thèse de l'accusation et l'autre celle de la défense, comme dans le jeu d'hypothèses suivant (par convention H1 est toujours l'hypothèse rejoignant l'accusation) :

H1 : la source suspectée est la véritable source de la trace

H2 : une autre source est la véritable source de la trace

Les hypothèses étant posées, le scientifique évalue la force de l'indice à la lumière des deux hypothèses alternatives. Dans cette logique, on utilise ce qu'on appelle des rapports de vraisemblance qui se décrivent de la manière suivante :

$$LR = \frac{P(E|H_1, I)}{P(E|H_2, I)}$$

où :

P : Probabilité

H : Hypothèses

E : Élément de preuve présentant les caractéristique analysées

I : Informations circonstancielle du cas

Comme on peut le déduire de cette formule, une valeur de LR (Likelihood Ratio) supérieure à 1 va soutenir l'hypothèse H1 par rapport à H2 alors qu'une valeur inférieure à 1 soutiendra H2 par rapport à H1, la valeur neutre étant 1.

Lorsqu'on examine une trace, on ne peut pas connaître la probabilité d'une source commune sachant les caractéristiques de la trace que l'on observe, par contre on peut calculer la probabilité d'observer ces caractéristiques en supposant comme vraie une ou l'autre des propositions. Si nous reprenons notre jeu d'hypothèses, en supposant que la trace provient véritablement de cette source, alors la probabilité d'observer des résultats similaires avec la référence est de 1. Par contre, si l'on observe ces concordances sachant que la trace provient d'une autre source, alors cette probabilité est la chance de trouver cette trace au hasard dans une population. Dans ce cas, un calcul de fréquence dans une population est indispensable pour estimer cette probabilité.

Le théorème de Bayes donne un modèle général qui utilise le rapport de vraisemblance pour adapter le degré de certitude sur n'importe quelles propositions grâce à la formule suivante :

$$\underbrace{\frac{P(H_1|I)}{P(H_2|I)}}_{P \text{ a priori}} \times \underbrace{\frac{P(E|H_1, I)}{P(E|H_2, I)}}_{LR} = \underbrace{\frac{P(H_1|E, I)}{P(H_2|E, I)}}_{P \text{ a posteriori}}$$

On remarque dans cette formule que le calcul d'un LR à la lumière d'un nouvel indice multiplie une probabilité a priori pour donner une probabilité a posteriori qui intègre le nouvel indice.

Ainsi, partant de ces principes, la pratique d'interprétation de l'expert consiste essentiellement à déterminer la valeur du numérateur et du dénominateur du LR et par conséquent, les variables qui les influencent. Le problème étant que ces valeurs et ces variables dépendent de la trace et des circonstances du crime, la formulation pratique du LR appliquée à un cas précis aura une formulation propre à ce cas ou à ce type de cas dans ce type de circonstances pour un type d'inférence (inférence sur la source ou sur le contact).

Analyse de l'enseignement de l'interprétation

- Cadre général

Les études académiques en sciences forensiques sont rares en Europe étant donné que l'Institut de Police Scientifique de l'Université de Lausanne est un des seuls à dispenser actuellement des études complètes dans ce domaine.

Ces études sont depuis plusieurs années en pleine expansion. En quelques années, le nombre d'étudiants et de demandes a augmenté de manière drastique passant d'une dizaine à une centaine d'étudiants en première année. En revanche, le nombre d'étudiants qui finissent leurs études semble se stabiliser au alentours d'une vingtaine de personnes par année suite à un processus de sélection à la fin de la première année.

L'interprétation de l'indice matériel est, paradoxalement, un nouveau domaine car jusqu'ici les efforts étaient plutôt portés vers le développement des techniques de traitement des indices que vers la réflexion sur leur signification. Il est enseigné à l'Institut de Police Scientifique d'une manière formelle depuis 1997. C'est depuis, un enseignement devenu essentiel à la formation de généraliste en sciences forensiques. Il se développe chaque année puisqu'il est maintenant enseigné non plus seulement en dernière année, mais sur deux ans et qu'un mastère en techniques d'interprétations se développe dans le cadre des restructurations des études académiques suite aux accords de Bologne.

L'enseignement de l'interprétation est donc un cours théorique sur deux ans qui regroupe une vingtaine d'étudiants. Cependant, il a des retombées sur une multitude de cours pratiques étant donné que l'expert a pour rôle et obligation d'évaluer chaque trace lors d'une expertise.

- Méthodes d'enseignement

L'enseignement de l'interprétation de l'indice matériel se déroule en trois grandes parties. Dans un premier temps, les concepts du rôle de l'expert, d'identification, d'individualisation, etc, sont mis en place. Ensuite, les notions de bases de probabilités et le théorème de Bayes sont introduits. Enfin, on met en pratique la théorie bayésienne avec la résolution de cas

par des d'évaluations de LR dans des scénarios de crime précis. C'est cette dernière partie de l'enseignement que nous allons développer au cours de ce travail.

Comme nous l'avons vu précédemment, la formulation pratique du LR, c'est-à-dire la formulation permettant d'évaluer la probabilité d'observer les caractéristiques de l'indice en regard d'une hypothèse ou d'une autre, qui sert à l'estimation de la force d'inférence, dépend de la trace, des circonstances du cas et du niveau d'inférence voulu (source ou contact). De ce fait, toute la troisième partie de l'enseignement se fait par une déclinaison de scénarios de crime mettant en œuvre des problématiques différentes. Cette méthodologie se retrouve d'ailleurs très fréquemment dans la littérature du domaine proposant des solutions d'évaluation dans un contexte précis.

L'avantage de cette méthode est qu'elle met en place des problématiques de situations de référence propres à chaque type de trace. Par contre l'éventail des scénarios restera toujours infime par rapport à la pléthore de situations possibles dans la réalité.

2.2 Un système expert comme aide à l'enseignement de l'interprétation

2.2.1 Notions générales sur les systèmes experts

Un système de connaissance est un système qui est utilisé pour les connaissances qu'il contient. En général, lorsque cette connaissance provient d'experts hautement qualifiés dans le domaine, il est appelé système expert. La recherche dans ce domaine est fortement liée à l'évolution des progrès en intelligence artificielle (IA) qui est un domaine informatique attaché à la compréhension du phénomène de l'intelligence. C'est son aspect informatique qui distingue ce domaine de la philosophie ou de la psychologie. (STEELS 1991)

En général, un système expert se caractérise par le fait que les connaissances qu'il contient sont séparées des mécanismes qui réalisent le raisonnement.

La représentation interne de la connaissance est alors appelée la base de connaissance alors que les mécanismes de raisonnement sont regroupés dans le moteur d'inférences.

Un système expert peut être comparé au fonctionnement d'un carnet de note interactif. Il est capable d'assimiler des faits tout comme des observations générales, des méthodes pour générer des nouveaux faits, des règles pour la résolution de problèmes ou encore des commentaires. Dans un carnet de note, il est nécessaire de chercher l'information à la main, de la lire et de l'appliquer soi-même alors que le système expert fait ce travail à votre place. (STEELS 1991)

Ainsi des systèmes de connaissances ont été créés dans différents domaines comme l'assistant médical, le manuel actif, le conseiller d'impôt, etc.

2.2.2 Comparaison avec d'autres programmes standards

Une question qui revient systématiquement est : qu'est ce qui diffère un système expert d'autres programmes informatiques standards. Un des premiers points pour répondre à cette

question est qu'il ne faut pas voir une division nette mais plutôt un continuum entre les deux type de programmation. Plus il y a de management de l'information plus on s'approche des techniques traditionnelles avec des accès à des bases de données et des calculs et plus il y a de raisonnement, plus on se dirige vers un système de connaissance.

Nous allons cependant, sans entrer dans les détails, exposer deux différences majeures :

Le système de connaissance réalise un raisonnement. Il sert d'aide non seulement pour des calculs et des recherches d'informations mais aussi pour le raisonnement.

Il est basé sur un modèle d'expert humain du domaine : le principal avantage de cette approche est que le système est capable d'expliquer son fonctionnement d'une manière compréhensible et cohérente pour l'utilisateur.

2.2.3 Fonctionnement d'un système expert

Tous les systèmes experts comprennent au moins trois éléments de base : une base de connaissances, un moteur d'inférence et enfin une interface avec l'utilisateur.

Il existe une multitude de façon de représenter les connaissances et donc de multiples formalismes. Cependant, de manière générale, une base de connaissances est composée d'un ensemble de faits ou de prédicats et de règles d'inférence. Une règle d'inférence est de la forme :

Si... conditions, alors... conséquences

La partie gauche exprimant les conditions d'applicabilité de la règle et la partie droite représentant la conclusion, qui peut être une action à effectuer ou une assertion à ajouter à la base de faits (HOPGOOD 2001).

Le moteur d'inférence est le coeur du système expert ; c'est lui qui, alimenté par la base de connaissances, construit dynamiquement le raisonnement, décidant quelles règles déclencher et dans quel ordre.

Il existe pour les moteurs d'inférences deux grandes familles de raisonnements qui peuvent être utilisés : le chaînage avant et le chaînage arrière.

A cette distinction entre modes de raisonnement, s'ajoutent les stratégies suivantes :

La stratégie irrévocable : les règles appliquées ne sont jamais remises en cause.

La stratégie par tentatives : elle peut remettre en cause l'application d'une règle si ce choix débouche sur un échec.

La stratégie en profondeur : en premier lieu, le moteur explore l'arbre des solutions dans sa profondeur c'est-à-dire en explorant une branche jusqu'au bout.

La stratégie en largeur : d'abord, le moteur explore l'arbre des solutions dans sa largeur et donc en explorant toutes les branches possibles.

Ces stratégies ont toutes un point commun : elles s'intéressent toutes à la globalité de l'arbre de solutions ; on les considère donc comme complètes. Il existe des situations où ce genre de

stratégies n'est pas adapté. Prenons l'exemple d'une partie d'échec ; ces stratégies ont pour effet d'évaluer à chaque coup toutes les possibilités offertes. Il serait plus judicieux d'opter pour une stratégie qui nous guide, sans étudier toutes les possibilités, vers les solutions les plus prometteuses. Les stratégies heuristiques remplissent ce rôle.

Le chaînage avant

Dans le mode de chaînage avant, le moteur d'inférence part des faits pour arriver au but, c'est-à-dire qu'il ne sélectionne que les règles dont les conditions de la partie gauche sont vérifiées, puis applique une de ces règles qui ajoute des faits à la base. Ce processus est réitéré jusqu'à ce qu'il n'y ait plus de règles applicables ou que le but soit atteint. L'efficacité du moteur d'inférence réside dans la pertinence de la décision prise (règle élue) lors de la phase de choix. Le chaînage avant procède par stratégie irrévocable.

Le chaînage arrière

Le système part du but et essaie de remonter aux faits pour le démontrer. Les règles sélectionnées sont celles dont la partie droite (conséquence) correspond au but recherché. Les conditions inconnues (partie gauche) de ces règles deviennent autant de sous-buts à démontrer. Ce processus est répété jusqu'à ce que tous les différents niveaux de buts soient démontrés - le but est alors atteint - ou jusqu'à ce qu'il ne soit plus possible de sélectionner des règles. Dans ce cas, le système peut demander à l'utilisateur de résoudre un ou plusieurs sous-buts (questions, tests) et le processus recommence. L'échec intervient quand le système ne peut plus sélectionner de règle ni poser de question à l'utilisateur. Le chaînage arrière procède en stratégie par tentatives. Le moteur d'inférence opère alors un retour arrière (backtracking) pour remettre en cause l'application d'une règle qui débouche sur un échec et pour essayer une règle écartée précédemment.

2.2.4 Un système expert comme outil d'enseignement

Un système expert ne peut pas être assimilé proprement dit à un enseignement intelligent assisté par ordinateur (EIAO) dans le sens où, il n'a pas véritablement de fonction ou d'implémentation spécifique à l'enseignement. (MENDELSON & DILLENBOURG 1991)

En effet, même s'il existe bien un modèle expert, il ne prend en compte que les protocoles de l'expert : il n'y a donc qu'un seul point de vue contrairement à l'EIAO.

Il n'y a pas de modèle élève qui permet de tenir compte des capacités de l'élève et ainsi d'adapter la difficulté ou l'interaction.

Il n'y a pas non plus de modèle pédagogique car le système ne permet pas véritablement de faire des erreurs étant donné qu'il est amené à proposer des solutions adaptées en fonction de critères spécifiés par l'utilisateur. Par contre il est capable de justifier ces choix et de donner ses critères de décision. Il est aussi capable de donner une aide circonstancielle sur le domaine, de manière plus ou moins interactive.

Le système expert ne doit donc en aucun cas être considéré comme un module d'enseignement à part entière. Par contre, il peut être utilisé comme outil de support à la résolution de problèmes dans le domaine concerné et, grâce à des aides ou des explications spécifiques et

à des justifications de prise de décision, permettre éventuellement à l'utilisateur d'assimiler le raisonnement de référence de l'expert.

De plus, dans l'optique qu'un système expert est en quelque sorte un carnet de note interactif contenant le savoir de l'expert dans le domaine, comme décrit au chapitre 2.2.1, on peut considérer qu'il aide l'apprentissage dans le sens où l'accès à l'information pertinente est facilité par rapport à une recherche bibliographique ou une lecture d'articles.

2.2.5 Perspectives pour l'enseignement de l'interprétation

Dans le chapitre 2.1.2 on a pu remarquer que l'enseignement de l'interprétation est basé pour une bonne partie sur l'étude de cas. En effet, on propose des scénarios de crimes et on tente d'évaluer la force des inférences sur la source ou sur le contact à la lumière d'hypothèses compétitives et d'obtenir ainsi un rapport de vraisemblance entre ces deux hypothèses. On a pu voir aussi que l'évaluation de ces probabilités d'inférences est spécifique à chaque trace et à chaque cas, ce qui empêche une généralisation de la démarche.

Deux problèmes essentiels découlent donc de ce type d'enseignement ; il ne faut pas voir ici une critique de l'enseignement donné mais plutôt une limitation de celui-ci.

D'une part, il n'est possible de traiter que de quelques cas de référence par type de trace, et d'autre part il y a un problème d'adaptabilité du cas de référence à la situation réelle à laquelle l'étudiant est confronté.

Un système expert pourrait éventuellement apporter une solution si une généralisation de la collecte des données pour pouvoir accumuler le plus de scénarios possibles est imaginable. Alors ce système saurait gérer une multitude de cas de référence qui aboutirait à une formulation pratique, dans le sens directement utilisable, du LR comprenant toutes les variables significatives qui entrent en jeu.

On pourrait même voir plus loin en imaginant que ce système ne donne pas seulement une formulation finale correcte du LR en fonction des critères remplis. Il pourrait, en plus, expliquer tout d'abord les critères sur lesquels il s'est basé pour arriver à une telle formulation, donner aussi une explication sur les simplifications apportées à la formulation (variables négligeables) et, enfin, fournir des pistes pour l'estimation des variables significatives et permettre de varier ces paramètres pour voir leur répercussion sur la valeur du LR.

Il serait aussi nécessaire que le système expert donne des informations sur l'utilité des valeurs demandées tout au long de la collecte et qu'il les remette dans le contexte général de la scène de crime. Ces informations sont primordiales pour la compréhension d'un cas et des facteurs qui entrent en jeu.

La mise en place d'une généralisation de la collecte d'informations liées au cas et à la trace permettrait à un système expert de gérer une multitude de cas.

La formulation d'un LR adapté à la situation décrite, l'explicitation des critères qui lui ont permis d'arriver à cette formulation, l'explication sur le développement du LR et de ses simplifications, la direction à prendre pour l'évaluation des variables significatives du LR ainsi que l'information donnée sur l'influence et le contexte général d'une donnée demandée lors de la collecte sont autant de fonctionnalités nécessaires à un système expert pour être

utile à l'enseignement de l'indice matériel.

Chapitre 3

Objectif et problématique de recherche

3.1 Hypothèse et but du travail

L'analyse générale faite dans le chapitre 2 du domaine de l'interprétation de l'indice matériel et des possibilités des systèmes experts a permis de mettre en avant quelques fonctionnalités qu'un système expert doit avoir pour éventuellement être utilisé à des fins d'enseignement dans cette pratique des sciences forensiques.

Ce travail de mémoire s'intéresse donc au développement d'un tel système et à sa mise en place dans un contexte défini. Ces deux pôles d'intérêt distincts reposent sur deux hypothèses principales :

il est possible de réaliser un système expert qui retrace la pratique d'un expert de référence dans le domaine.

l'outil développé et son interface sont utilisables par des criminalistes.

Ces deux hypothèses nécessitent, pour être confirmées, un certain nombre d'hypothèses intermédiaires qui guident la démarche de travail. En effet, pour rendre possible le développement de ce système, il est nécessaire notamment de pouvoir :

modéliser la connaissance et la pratique de l'expert et ainsi avoir une représentation de la connaissance afin de la rendre utilisable.

transformer ce modèle en règles et faits pour rendre cette connaissance compréhensible à un système expert.

généraliser des cas afin de limiter le plus possible le nombre de règles du système.

Il est important de remarquer que même si ce travail tient compte des fonctionnalités pédagogiques que doit posséder ce système expert, il ne s'intéresse pas directement aux retombées de l'outil sur l'enseignement du domaine. En effet, l'idée est ici plutôt de convaincre du potentiel des systèmes experts pour une utilisation pratique dans le contexte réel de l'enseignement de l'interprétation de l'indice matériel que d'en étudier l'efficacité.

Le but de ce travail est donc de donner des pistes pour le développement d'un système expert qui répond aux critères énoncés auparavant afin d' être éventuellement utilisable à l'enseignement de l'interprétation en réalisant **un prototype**. Ce prototype devra être suffisamment élaboré pour permettre de déterminer des perspectives d'avenir pour ce type d'outil dans ce domaine.

3.2 Méthodologie

L'approche méthodologique de ce travail se divise en cinq étapes principales :

La première consiste à dresser un état de la situation dans ce domaine. Le but est ici de savoir ce qui se fait et ce qui a été fait comme système expert pour l'interprétation de l'indice et éventuellement pour son enseignement. Cette étude est fondamentale car elle peut permettre d'orienter la recherche en fonction de ce qui est déjà connu. (chapitre 4.1)

Une deuxième étape est d'étudier les différentes possibilités qu'offre la recherche dans le domaine des système de connaissance pour être en mesure de faire des choix sur le formalisme à utiliser (chapitre 4.2)

La troisième étape, la modélisation de la connaissance, est essentielle au développement d'un système expert. Cette structuration de la connaissance est faite par une analyse des facteurs impliqués et des rapports qu'ils ont entre eux dans le contexte général de la pratique des sciences forensiques symbolisée par le paradigme introduit dans le chapitre 2.1.1. (chapitre 4.3)

La quatrième étape est la transformation du modèle de connaissance en un formalisme compris par la technologie utilisée et le développement du prototype en fonction des options retenues lors des recherches précédentes. La stratégie est ici de procéder par des petits tests afin de choisir une technologie précise.

Pour le développement proprement dit, par analogie, la stratégie choisie est, dans un premier temps, une stratégie en profondeur comme décrite dans le fonctionnement des systèmes experts (chapitre 2.2.3). Elle consiste à progresser dans l'arbre de résolution en suivant la branche jusqu'au bout. Comme chaque branche est un scénario ou un groupe de scénarios, cette approche permet de développer un *système expert* complet pour ce cas précis. L'avantage est que, au final, le prototype fonctionne entièrement sur au moins un scénario ce que ne garantit pas une approche moins spécifique. Par contre, dans un deuxième temps, une stratégie en largeur (chapitre 2.2.3) est adoptée car elle permet, jusqu'à un certain point, une généralisation de la démarche. En effet, cette approche permet de trouver des critères de sélection généraux qui caractérisent des groupes de cas. (chapitre 4.4)

Enfin la dernière étape consiste à tester le prototype. Il s'agit ici de mettre en place une expérimentation qui offre la possibilité d'étudier de manière générale s'il est utilisable par un criminaliste. Le but est d'obtenir des informations sur la qualité et la clarté de l'interface ainsi que sur la facilité de prise en main du programme. Ces tests ne sont

pas orientés en vue d'évaluer l'utilité d'un aspect du logiciel ou de son efficacité pour l'enseignement. (chapitre 4.5)

Chapitre 4

Développement du prototype

4.1 Etat de la situation

Le chapitre 2.1.2 a montré que le domaine concerné était relativement récent et qu'il était en plein développement. Deux axes de recherche en rapport avec ce travail se distinguent comme aide à l'interprétation liés à cette grande effervescence :

La mise en place et l'utilisation de réseaux bayesiens

Le développement de systèmes experts.

Il est à remarquer que ces deux outils ne s'utilisent pas forcément dans le même contexte.

4.1.1 Les réseaux bayesiens

Le réseau bayésien est une méthode graphique, mathématique et statistique qui aide à gérer l'incertitude. Appliqué aux sciences forensiques, il est utilisé comme méthode de raisonnement qui assiste le criminaliste à comprendre toutes les dépendances qui peuvent exister entre les différents aspects de l'indice et qui participent au processus de prise de décision.

Graphiquement, un réseau bayésien est une succession acyclique de nœuds, qui représentent des états incertains de variables, reliés par des flèches qui sont les relations de cause à effet entre les variables. A chaque variable est associée une table contenant les probabilités d'un nombre fini d'états mutuellement exclusifs. Si un nœud ne possède pas de flèche entrante, alors la table contiendra des probabilités non conditionnelles. Si par contre, il en reçoit, il devient le nœud enfant d'un ou plusieurs parents, et la table associée à ce nœud contiendra des probabilités conditionnelles dépendantes des variables des nœuds parents (TARONI, BIEDERMANN, GARBOLINO & AITKEN 2003). Il existe trois types de connexion de nœuds : en série, convergente et divergente qui sont illustrées dans la figure 4.1.

Dans ce travail, nous n'allons pas détailler les techniques de propagation des probabilités à l'intérieur du réseau qui permettent le processus de prise de décision. Plusieurs ouvrages sont disponibles dans ce domaine. (CASTILLO, GUTIÉRREZ & HADI 1997) (G.COWELL, DAWID, LAURITZEN & SPIEGELHALTER 1999)

Le but de la mise en place de ces réseaux bayesiens est d'avoir d'une part une représentation

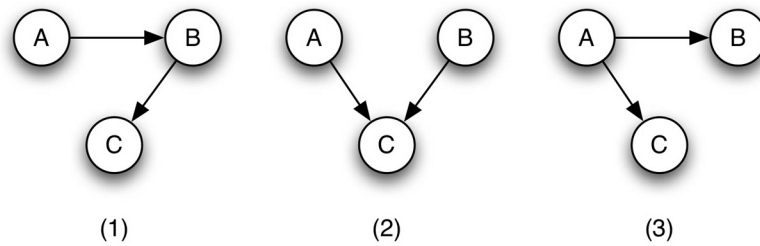


FIG. 4.1: Les différents type de connections dans un réseaux bayésien : (1) en série, (2) convergente et (3) divergente

graphique du problème et, d'autre part, un outil qui permette une représentation dynamique de la prise de décision en fonction de n'importe quelle situation possible découlant du scénario représenté. Il faut souligner que cette prise de décision se fait en fonction des valeurs de probabilité entrées dans les tableaux de chaque variable. Ces probabilités ne reposent pas toujours sur des valeurs réelles qui sont disponibles, mais doivent souvent être estimées par une méthode ou par une autre. Ainsi, il faut considérer la décision, c'est-à-dire la probabilité d'une proposition par rapport à d'autres, prise par le réseau, comme une décision issue d'un contexte très précis qui ne correspond pas totalement à la réalité.

L'application de ces réseaux bayésiens est réalisée dans le cadre de plusieurs types de trace comme l'interprétation des traces d'accélération dans les incendies, du verre, des fibres, des traces biologiques, etc.

4.1.2 Les systèmes experts

Plusieurs projets de grande envergure ont été développés ces dernières années. Deux de ces projets sont particulièrement intéressants et vont dans le sens de cette recherche. Le premier est le projet Cage2000 qui s'intéresse à l'interprétation de l'indice verre développé à l'Institut de Police Scientifique (HICKS 2003) et le second est un projet né d'une collaboration du Forensic Science Service (FSS) avec l'Université de Warwick : FINDS (PUSH, CHAMPOD, EVETT, SMITH, WYNN & BATES 2003).

Cage2000

Il s'agit d'un servlet, une application Java qui fonctionne coté serveur, avec une interface web qui permet l'évaluation de l'indice verre grâce à un calcul de LR dépendant des informations fournies à l'application. Ainsi, dans ce programme, une procédure de collecte d'information a été mise en place permettant, grâce à un jeu de règles, d'adapter le calcul du LR en fonction de la situation décrite et voulue par l'utilisateur. Le système met également en place une étape de pré-évaluation de la trace où l'expert doit entrer les informations qu'il pense devoir trouver a priori en fonction des circonstances connues du crime. Cette pré-évaluation fait partie de la pratique de l'expert comme décrite par Evett (EVETT, COOK, JACKSON, JONES & LAMBERT 1998). Elle permet notamment à l'expert d'estimer, avant de commencer les analyses, la force de l'inférence que la trace est susceptible de donner et

dans quel contexte, et donc de savoir si les analyses valent la peine d'être toutes réalisées ou pas. Comme ce programme est aussi capable de garder en mémoire des bibliothèques de cas dans une base de données, une fonction de recherche de cas similaires dans ces bibliothèques est proposée à l'utilisateur afin de l'aider à estimer la valeur des variables qui entrent en jeu dans le calcul du LR adapté à ce cas.

Ce programme présente donc deux voire trois des critères que nous avons définis au chapitre 2.2.5 permettant à un système expert d'être utilisé à des fins pédagogiques soit :

La mise en place d'une généralisation de la collecte d'informations liées au cas et à la trace.

La formulation d'un LR adapté à la situation décrite. Même si la formule n'est pas explicitement décrite dans Cage2000.

La direction à prendre pour l'évaluation des variables significatives du LR grâce à un accès aux cas contenus dans la base. Il s'agit bien d'une aide à l'estimation, mais par contre elle n'informe pas du tout sur les justifications de ces valeurs. C'est pourquoi, la fonction de direction n'est pas tout à fait remplie.

Il présente, en plus, une étape de pré-évaluation qui correspond à la pratique de l'expert en sciences forensiques.

FINDS

FINDS est une application locale qui permet d'évaluer la force d'une inférence dans deux scénarios prédéfinis. L'un se rapportant à l'indice verre (cambriolage avec bris de vitre) et l'autre à l'indice fibre (vol avec cagoule). Ce programme est un projet de grande envergure qui a été conçu dès le départ pour la formation des experts travaillant au FSS. Il permet à l'utilisateur de faire varier tous les paramètres impliqués dans les deux scénarios et propose une approche continue et non plus discrète du problème. Tout comme Cage, il réalise une pré-évaluation de la force de l'inférence en fonction des caractéristiques que l'expert estime devoir trouver dans ce type de circonstances.

Le logiciel fonctionne avec un réseau bayésien propre à chacun des scénarios comme moteur d'inférence du système, si bien qu'il est prévu à l'avenir, d'augmenter le nombre de ces réseaux, donc le nombre de scénarios et ainsi pouvoir proposer son utilisation sur des cas réels.

Par contre, ce logiciel ne fournit aucune aide ni aucune explication sur les variables ou sur leurs interactions. De ce fait, il est possible de voir les répercussions d'un facteur sur la valeur du LR mais pas forcément d'en comprendre les causes.

Ce programme, lui aussi, présente donc deux voire trois propriétés désirées malgré le fait qu'il était prévu dès le départ pour la formation :

La mise en place d'une généralisation de la collecte d'informations liées au cas et à la trace.

La formulation d'un LR adapté à la situation décrite mais tout comme Cage2000 il ne

l'explique pas.

La direction à prendre pour l'évaluation des variables significatives du LR, mais avec un système différent que celui de Cage2000. FINDS permet de faire varier les paramètres et d'observer les changements en direct par des représentations graphiques et numériques. Par ce biais on peut éventuellement se rendre compte si la valeur de la variable prend la bonne direction ou pas. Par contre, tout comme Cage2000, il ne donne aucune information sur les critères qui influencent ces valeurs.

Cette première approche de l'utilisation de système expert en sciences forensiques est plutôt encourageante étant donné que plusieurs institutions se sont penchées sur le sujet et ont développé des prototypes. La problématique développée pour la réalisation de ces deux projets va servir de base de réflexion pour construire notre prototype.

4.2 Choix des techniques

Avant de nous pencher sur les choix réalisés dans le cadre de ce travail de recherche, commençons par fixer certaines caractéristiques que doit posséder cette technologie en fonction de nos attentes.

En considérant les buts fixés dans le chapitre 3.1, on peut dès le départ exclure la technologie des réseaux bayesiens. En effet, l'utilisation de ces réseaux permet bel et bien d'avoir une vision d'ensemble de la problématique et des interactions en jeu, mais ne permet pas, à eux seuls, de les expliquer. Il faudrait pour cela réaliser un logiciel dans l'esprit de FINDS qui utilise ces réseaux comme moteur d'inférences, mais cette entreprise est trop ambitieuse dans le cadre de ce travail. Par contre, il serait intéressant d'utiliser dans les explications la représentation graphique de ces réseaux car elle permet une remise en contexte rapide et efficace des variables et de leurs interactions.

Dans cette perspective, le choix s'est donc porté sur un système à base de règles. Il faut préciser ici que l'idée du projet est d'identifier et d'utiliser une technique prête à l'emploi qui dans ce cas est un moteur d'inférence déjà existant et de s'intéresser plutôt au développement de la base de connaissance. Ainsi, un premier critère de choix est que la technologie soit suffisamment simple pour permettre un apprentissage rapide. Elle doit aussi être disponible gratuitement puisqu'il s'agit de faire un prototype. Il faut que celle-ci soit assez perfectionnée pour permettre d'intégrer toutes les fonctionnalités désirées. Enfin, une interface conviviale doit pouvoir être réalisée pour que le programme soit utilisable par un usager moyen.

Avec ces critères et suite aux tests décrits dans la méthodologie de travail (chapitre 3.2), le choix s'est porté sur un moteur d'inférences appelé CLIPS. Ce moteur n'ayant qu'une interface texte avec l'utilisateur, un autre logiciel a été choisi pour créer une interface plus conviviale : WXCLIPS. Le problème de ce choix est que ces deux programmes ont leur langage propre et que toute l'interface se réalise à la main.

Il ne faut surtout pas voir cette solution comme le choix idéal, mais plutôt comme une solution appropriée pour atteindre le but fixé dans ce travail, soit réaliser un prototype.

4.3 Modélisation de la connaissance

4.3.1 Limitation

Lors de ce travail, étant donné que la démarche et la pratique du criminaliste varient d'un type de trace à un autre, il a été nécessaire de limiter le cadre de cette recherche à un seul type de trace. Comme les systèmes étudiés dans le chapitre 4.1 traitent essentiellement de l'indice verre. Dès lors, l'idée de considérer un type de trace ayant une problématique proche et tout aussi intéressante permettrait de bénéficier de la recherche déjà effectuée. C'est pour ces raisons ce travail se rapporte à l'analyse des traces *fibres*.

4.3.2 Modélisation des principes

Dans le chapitre 2.1.1, les six principes de base formant le Paradigme des sciences forensiques ont été définis de manière générale. Ces principes seront maintenant étudiés appliqués à l'indice fibre. Par fibre on entend ici tout élément filamenteux allongé, d'origine naturelle ou non, constitutif d'un fil, d'une feuille de papier, d'une toison animale, etc.

La divisibilité : La question qu'il faut se poser ici est quels sont les facteurs qui influencent ce processus si la résultante est une fibre. On peut instinctivement imaginer par exemple que si le support est un pull, il y a la taille de la maille, l'épaisseur du fil, la qualité du tissage du fil, etc. On remarque qu'on pourrait trouver une multitude de causes qui vont faire varier le facteur de divisibilité du support. Dans ce cas, chaque support a un taux de divisibilité qui lui est propre à un moment donné. En effet on peut imaginer qu'un pull neuf aura tendance à perdre plus de fibres que le même pull après une année. Comme l'on ne maîtrise pas tous les facteurs, la stratégie est de regrouper toutes ces variables en une que l'on nomme : *la qualité du donneur* et qu'on qualifie avec une échelle verbale qui va d'excellente à très mauvaise.

Le transfert : Le transfert est le nombre de fibres qui passe d'une source à une cible. Il dépend lui aussi de plusieurs variables. Tout d'abord il est nécessaire de distinguer trois états du transfert : l'état au moment de l'action, transfert à T0, l'état au moment du prélèvement, transfert à T1, et l'état après le prélèvement, soit les caractéristiques réelles de la trace. En se basant sur le modèle proposé par Curran (CURRAN, TRIGGS, BUCKLETON, WALSH & HICKS 2003) concernant le transfert et la persistance des fragments de verre, on a :

- A T0, essentiellement trois variables interviennent, la première est bien sûr la qualité du donneur définie ci-dessus mais il y a aussi la force ou le type de contact et le *taux d'accroche* de la cible.
- Entre T0 et T1, il faut considérer les variables de persistance des fibres transférées à la cible, le temps, le taux de rétention de la cible, l'activité de la cible ou l'exposition de la cible à un type d'activité. En général, on ne fait pas de distinction entre le taux d'accroche et de la rétention de la cible qui sont pourtant deux phénomènes bien différents.

- A T1 enfin, il y a surtout l'efficacité de la méthode de prélèvement qui est importante avec la rétention de la cible.

Tous ces facteurs de transfert et de persistance sont résumés dans la figure 4.2.

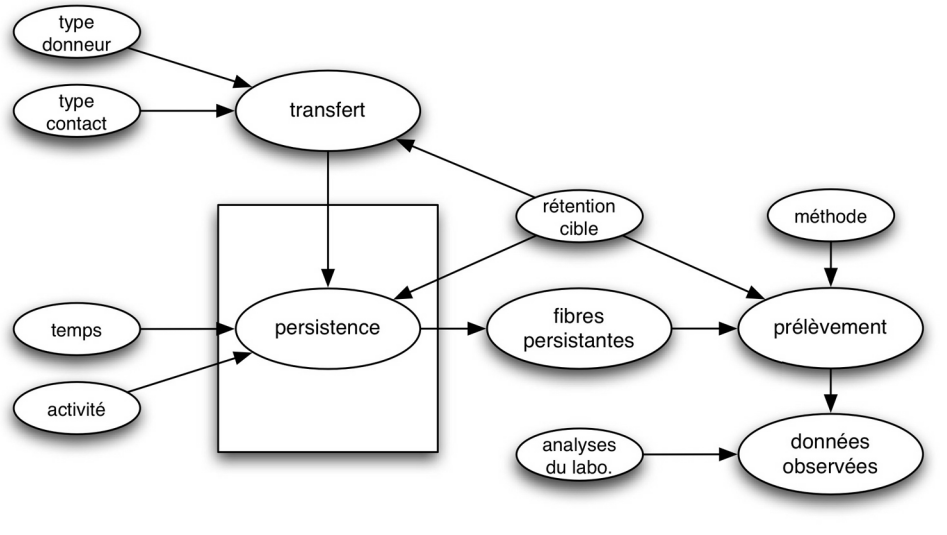


FIG. 4.2: Modèle graphique pour le transfert et la persistance des fibres sur une cible

La classification, l'identification et l'individualisation ont comme but d'inférer que l'indice provient de la source. Ces trois processus interviennent à tour de rôle dans la pratique du criminaliste. Appliquée aux fibres, la démarche peut se résumer comme dans la figure 4.3. (MARGOT 2003).

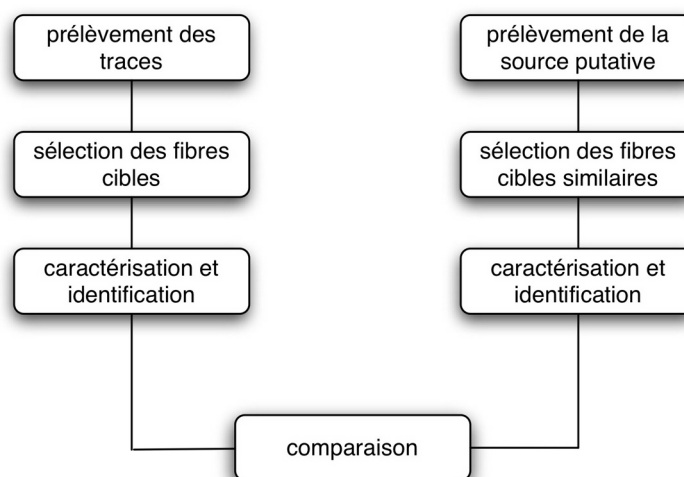


FIG. 4.3: Démarche du criminaliste dans le processus d'individualisation des fibres

La classification et l'identification sont des processus qui s'intéressent soit uniquement

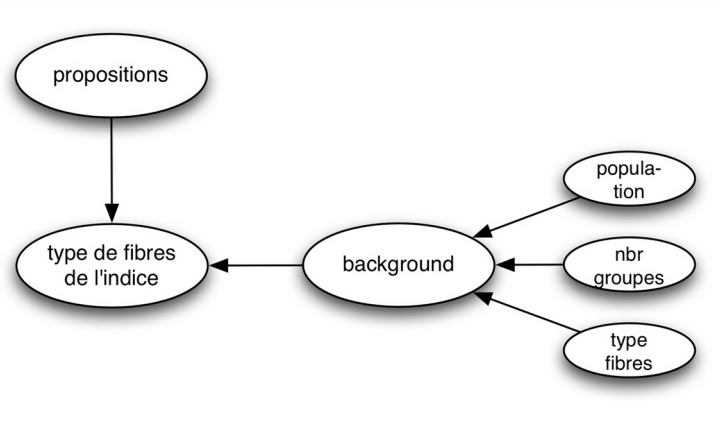


FIG. 4.4: Réseau bayésien de la connaissance dans le contexte d'une inférence à la source

à l'indice, soit à la source putative. Il faut simplement caractériser les propriétés physicochimiques de ceux-ci, indépendamment l'un de l'autre. Par contre, l'individualisation est un processus de comparaison entre ces deux objets.

Ainsi les examens concernés par ces trois processus sont uniquement des examens liés aux propriétés physicochimiques des fibres en cause. Dans cette étape, les propriétés et les variables qui influencent le transfert sont complètement ignorées. Le calcul du LR à ce niveau d'inférence se réalise sur un jeu d'hypothèse du type :

H1 : les fibres proviennent de cet objet

H2 : les fibres ne proviennent pas de cet objet

Dans ce calcul interviennent donc les notions de fréquence d'apparition dans une population. La fréquence dépendant des propriétés physicochimiques de l'indice et du niveau de caractérisation de ceux-ci dans une population qui, elle, est fixée par les données circonstancielles du cas et par le choix de l'hypothèse alternative H2 qui reflète la position de la défense. Ce processus est représenté par un réseau bayésien dans la figure 4.4 ci-dessus.

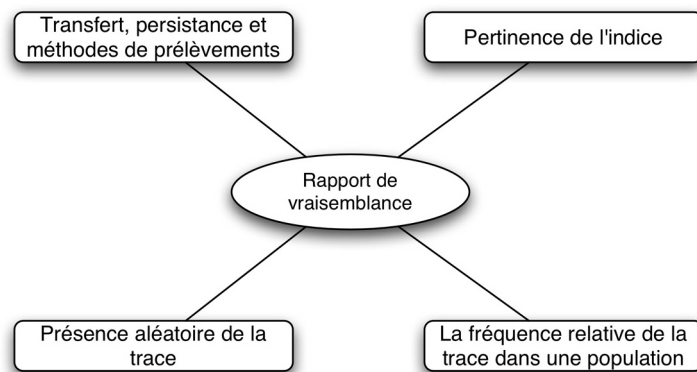


FIG. 4.5: Paramètres qui influence l'évaluation du rapport de vraisemblance

L'association est, par contre, l'étape qui s'intéresse aux propriétés du transfert dans un contexte de crime. Cette étape est très importante car c'est ce stade d'inférence que doit tenter d'atteindre le criminaliste pour remplir son rôle. Dans ce contexte, il est nécessaire de s'interroger sur la pertinence de la trace : la trace est-elle en rapport avec le crime ou pas? Champod et Taroni décrivent dans le schéma de la figure 4.5, en se basant sur des recherches de Evett, les paramètres principaux qui influencent la valeur du LR au niveau de l'association. (CHAMPOD & TARONI 1999)

Dans un scénario de fibres, il est donc nécessaire de poser deux jeux d'hypothèses qui, selon les circonstances, ne sont pas forcément indépendants (ROUX 1997). Un jeu concernant la pertinence et l'autre relatif au contact :

- P1 : Les fibres ont été transférées durant la commission du délit
- P2 : Les fibres proviennent du bruit de fond
- H1 : Cet objet a eu un contact avec la cible
- H2 : Cet objet n'a pas eu un contact avec la cible

Les hypothèses proposées ci-dessus sont des hypothèses générales qui peuvent être spécifiées par les circonstances du cas et par la stratégie de la défense.

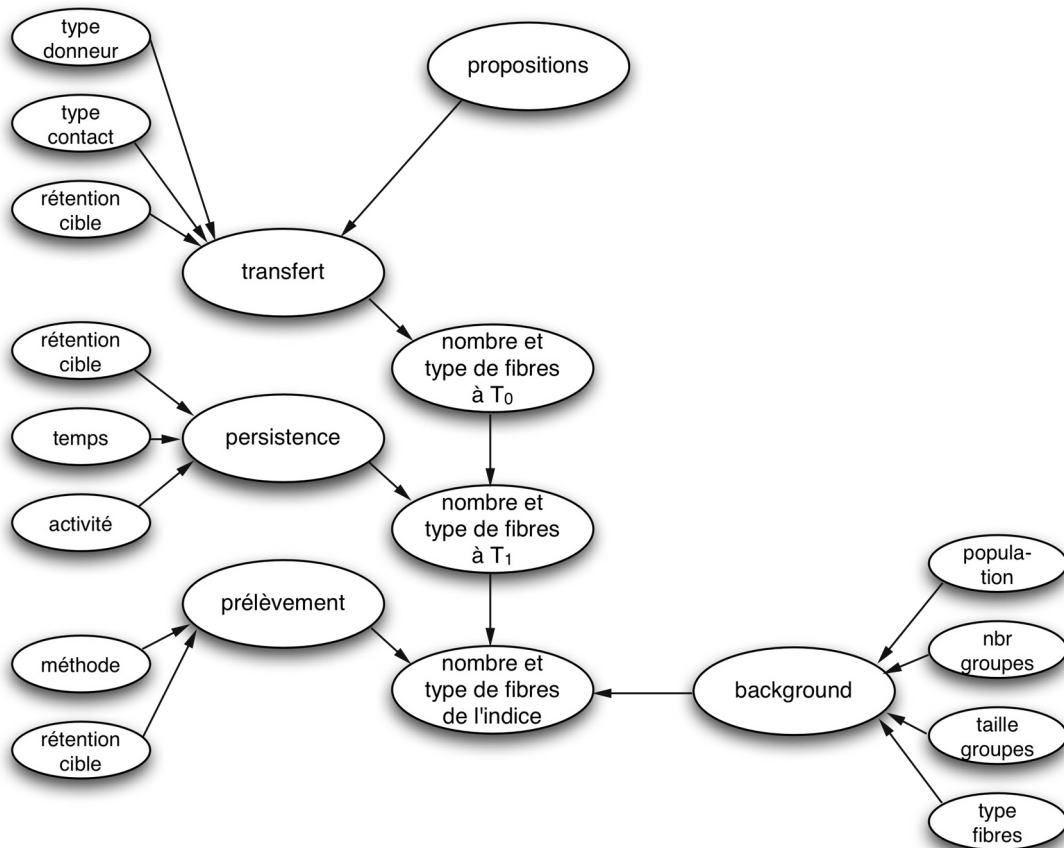


FIG. 4.6: Réseau bayésien de la connaissance dans le contexte d'une inférence au contact de deux objets

Le calcul du LR est dans ce cas plus complexe. Il faut considérer la globalité des variables de transfert et de persistance. Une représentation globale de ce processus d'association est représentée dans le réseau bayésien de la figure 4.6.

La reconstruction qui s'intéresse à l'historique et au déroulement de l'action criminelle sort du cadre de ce travail et ne sera pas abordée. En effet, ce processus est essentiellement utile pour vérifier des allégations ou pour l'analyse de crimes sériels (MARGOT 2003).

4.4 Développement du logiciel

La modélisation graphique en réseaux bayésiens des paramètres et de leurs relations intervenant dans un scénario général d'un cas fibre, permet d'identifier toutes les variables utilisées et aussi leur contexte d'utilisation.

Il est donc possible de réaliser une procédure de collecte de données qui permette de demander toutes les informations nécessaires. Même si toutes les variables ne sont pas forcément nécessaires dans tous les cas, elles sont malgré tout demandées durant cette procédure. La justification de cette stratégie réside dans le rôle de l'expert qui implique que tout doit être mis en œuvre pour se prononcer au niveau du contact. Ce n'est qu'en constatant les informations nécessaires manquantes qu'il pourra se contenter de l'inférence sur la source. Lors de la collecte, le programme permet d'accéder par un bouton d'aide à de l'information sur

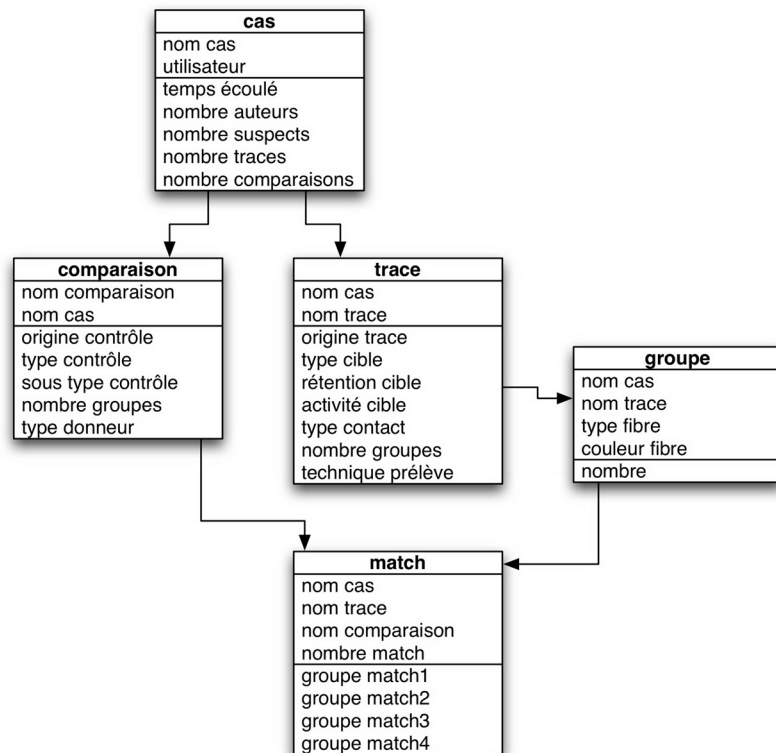


FIG. 4.7: schéma relationnel de la prise d'information du système et de son stockage

chaque valeur demandée qui remet la variable en question dans son contexte général. La collecte d'information se déroule en quatre étapes successives modélisées dans le schéma relationnel de la figure 4.7.

La collecte d'informations générales terminée il est alors possible de guider l'utilisateur en fonction des informations rentrées vers le niveau d'inférence le plus adapté à sa situation. La décision du système est employée ici uniquement comme guide ; la décision finale revenant à l'utilisateur. La décision du système doit être expliquée pour qu'elle puisse être comprise par l'utilisateur. Le processus de prise de décision imaginé pour ce projet est schématisé dans la figure 4.8.

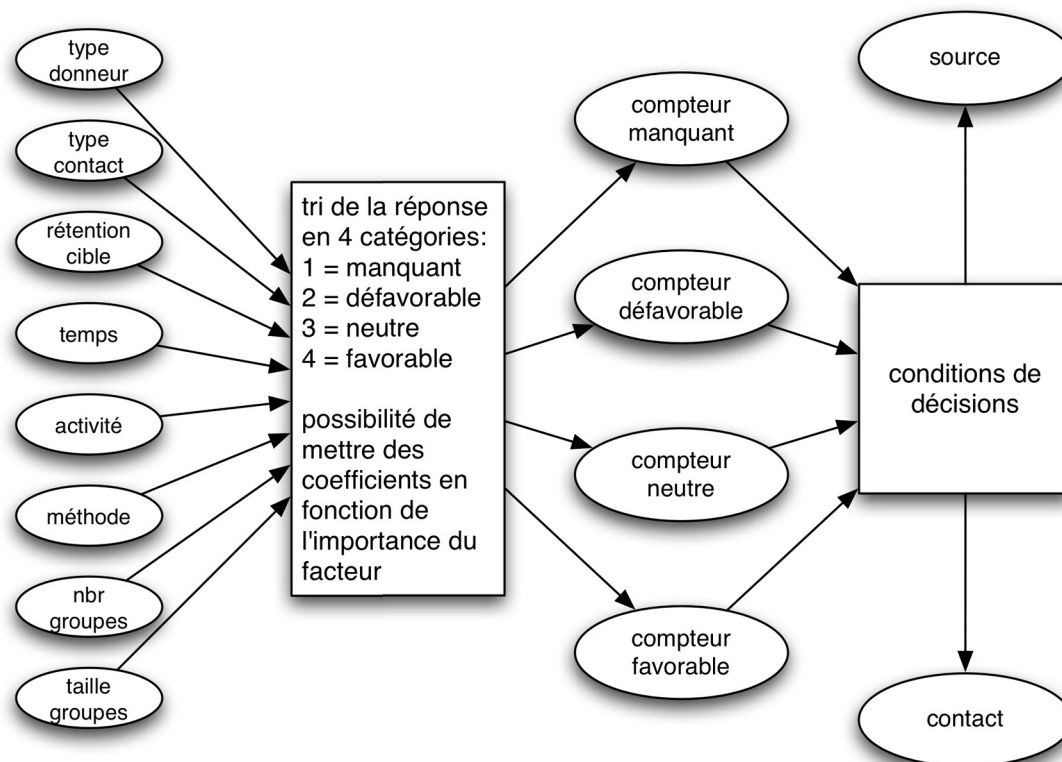


FIG. 4.8: Modèle de la prise de décision en fonction des données entrées par l'utilisateur

Ce processus ne représente pas la réalité et celui-ci est tout à fait critiquable, mais il répond parfaitement au but de ce projet qui est de faire un prototype permettant de guider l'utilisateur et d'explicitier les critères de ses choix. Ainsi, malgré la simplification de ce système de décision, il peut tout de même être très utile dans un contexte d'enseignement.

Le niveau d'inférence étant choisi, le but du système est de formuler un LR correct selon la littérature sur le sujet et suivant les circonstances du cas. Si le choix s'est orienté au niveau de la source, le système a toutes les informations nécessaires (figure 4.9). Par contre, si le choix est au niveau du le contact, la situation est beaucoup plus complexe et il y a encore quelques informations nécessaires. Ces informations sont décrites dans l'arbre de la figure 4.10.

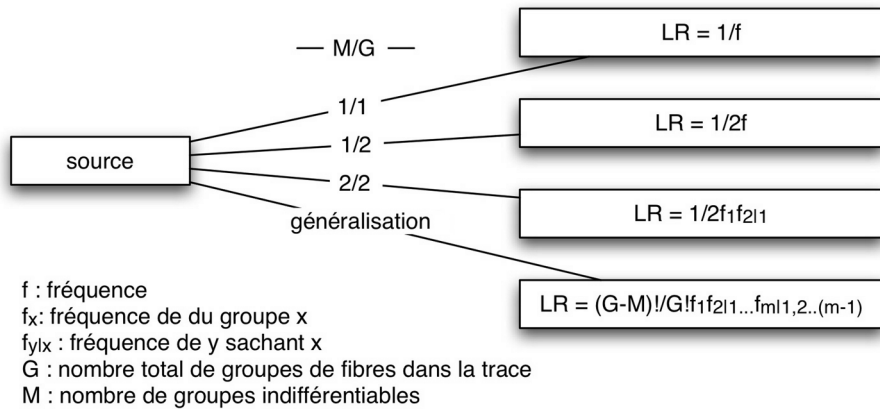


FIG. 4.9: Arbre des possibilités de formulation du LR au niveau de la source en fonction des informations données par l'utilisateur

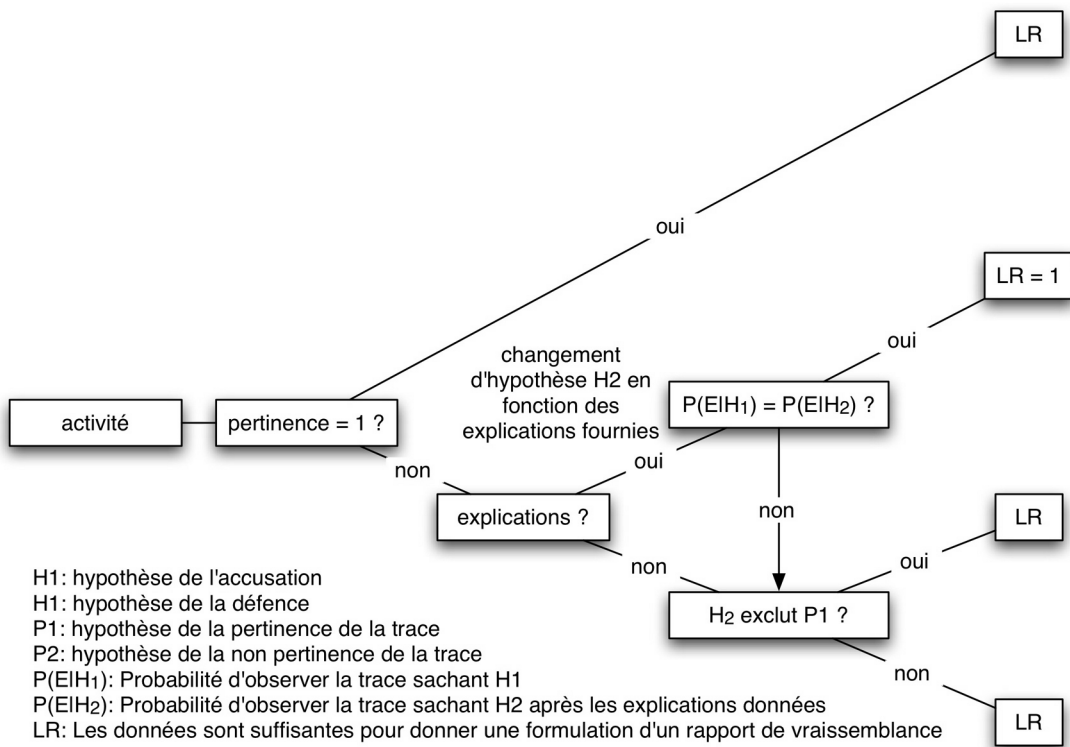


FIG. 4.10: Arbre des possibilités de formulation du LR au niveau du contact. Plusieurs informations sont alors nécessaires pour pouvoir formuler un LR. L'arbre se termine lorsque le système a toutes les informations requises

Le but de ce prototype n'est pas d'implémenter toutes les possibilités de ces arbres mais de montrer qu'il serait possible de le faire et de donner des pistes pour savoir comment le réaliser.

4.5 Test pratique

Le test effectué dans le cadre de ce mémoire consiste à confronter 5 utilisateurs avec le système en leur proposant un scénario de cas précis. Le scénario proposé était le suivant :

un prélèvement constitué de 67 fibres de laine bleue et de 54 fibres de laine rouge est réalisé sur le siège conducteur d'une voiture volée. La présence de ces fibres ne s'explique pas par les habits du propriétaire. Suite aux analyses d'identification et de comparaison de l'indice avec la veste d'un suspect arrêté pour d'autres raisons, les fibres de laine bleue ne sont pas différenciées de celles de la veste constituée d'un seul type de fibres. Calculer la valeur du LR pour une inférence au niveau de la source sachant que :

p	f
générale	0.015
vestes bleues	0.2
vestes en laine	0.24
vestes en laine bleue	0.76
vestes rouges et bleues	0.27

Où :

p : la population de référence

f : la fréquence d'apparition de ce type de fibres (selon les caractéristiques observées) dans une population de référence

Le but de ce test est, pour l'utilisateur, de réussir à entrer les valeurs correspondant au scénario dans le système, tout en découvrant éventuellement les aides qui sont proposées et de calculer un LR pour le cas en question. L'exercice est considéré comme réussi si la valeur du LR obtenu par l'utilisateur correspond à celle attendue.

Une fois ce petit exercice terminé, les impressions des utilisateurs sont recueillies oralement au cours d'une petite discussion sur l'interface, les difficultés, etc.

Chapitre 5

Résultats et discussion

5.1 Le prototype

Le prototype développé dans ce mémoire fonctionne sur une plateforme PC. Les possibilités offertes par le langage de wxClips ont été utilisées pour générer une interface de programme tout à fait comparable à une application standard comme le montre l'image de la figure 5.1.

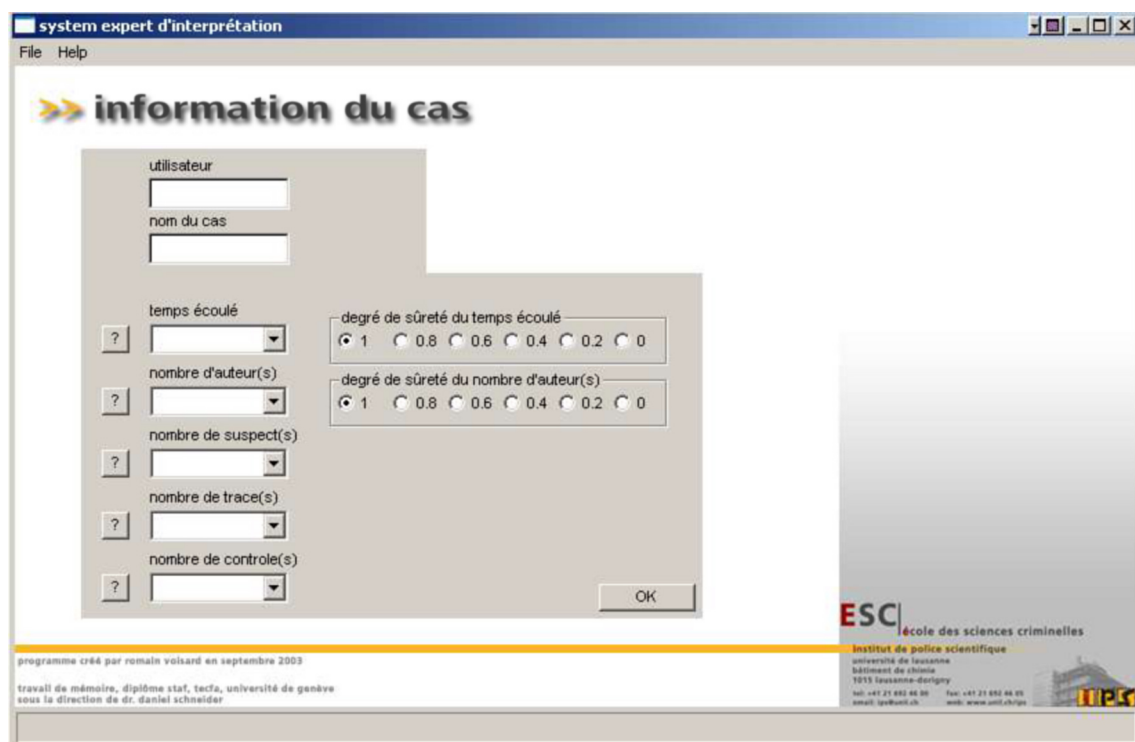


FIG. 5.1: Illustration du prototype, module de collecte des données au niveau du cas

Le module de collecte des informations se déroule donc en cinq étapes, comme montré dans le schéma relationnel 4.7. Pour des raisons de cohérence de l'application, les informations sont transmises au système par des menus déroulants. Les possibilités offertes par les options de chaque menu doivent permettre de couvrir le plus grand nombre de cas mais en même temps limiter ce choix à des situations qui sont possibles.

De plus, des procédures de contrôle sont mises en place pour vérifier que certains champs obligatoires sont remplis et renvoie un message d'erreur dans le cas contraire. Ces procédures de vérification sont activées lors de la validation de toute la page d'interactions par l'utilisateur.

Tout au long du programme, l'utilisateur peut accéder à de l'information en activant les boutons d'aide. Ceux-ci ouvrent une nouvelle fenêtre dans laquelle se trouvent des explications relatives au bouton activé. Ces informations, issues de la littérature du domaine, remettent le paramètre dans son contexte et expliquent son importance dans le processus d'interprétation. Ces aides faisaient défaut dans les deux autres systèmes étudiés au chapitre 4.1.2

La navigation choisie pour ce système est une navigation par nœud dans une arborescence. Les nœuds sont des faits qui sont chargés au départ par le système. Ensuite, en fonction de la valeur d'un autre fait appelé navigation, le système active la règle et déclenche ainsi l'interaction liée au nœud ayant le même nom que le fait navigation. Ainsi, potentiellement, il est possible de faire des retours en arrière ou des sauts dans l'arborescence. Cette fonctionnalité n'a pas été mise en place dans ce prototype si bien qu'à la validation de chaque interaction il n'est pas possible de revenir en arrière.

Toutes les informations liées à chaque cas sont stockées dans des fichiers de texte de manière structurée comme dans le schéma relationnel 4.7. Etant donné que le cas peut être identifié par son nom, l'unicité de ce nom étant garanti par la stratégie d'organisation de l'institution, le système peut à tout moment rechercher des informations dans cette bibliothèque de cas. Les procédures de recherche d'informations dans ces fichiers ont été implémentés pour retrouver des valeurs et ainsi ne pas devoir systématiquement les stocker dans des variables. Par contre, un module de recherche de cas comme le propose Cage2000, n'a pas été réalisé, mais est cependant tout à fait imaginable.

La cohérence du système a été un facteur important lors du développement du prototype. En effet, les menus déroulant s'adaptent en fonction des cas possibles suivant les informations fournies par l'utilisateur. Par exemple, si une source putative possède deux groupes de fibres et la trace trois, il ne sera pas possible d'avoir les trois groupes de la trace qui correspondent. Dans ce cas le système proposera donc dans le menu deux groupes au maximum. Plusieurs contrôles de ce type ont du être implantés au système pour éviter le plus possible la création de cas incohérents. Par contre la mise en place de ces procédures implique une explosion non négligeable du temps de développement.

La réalisation de la prise de décision comme décrite dans la figure 4.8 permet de conseiller sur le choix du niveau d'inférence. Elle renseigne aussi sur les critères de décision. Le choix de ce niveau est actuellement la seule véritable prise de décision implantée dans le système. En effet, c'est la seule partie du programme qui évalue plusieurs critères en fonctions de règles modifiables.

Enfin, le système est doté d'un module qui permet de proposer des formulations de LR, de les expliquer et de calculer ces rapports en fonction de paramètres déterminés par l'utilisateur.

Les estimations de ces paramètres sont guidées par des informations données par le système. Une évolution possible du système pourrait être qu'il réalise lui-même l'estimation des variables en jeu en fonctions de données d'expérimentations implantées et de réaliser ainsi un système expert jusqu'au bout du processus. Il y a deux problèmes majeurs à cette démarche, la première est qu'elle complexifie énormément le système et donc augmente les coûts de production. La deuxième est que le nombre de situations auxquels peut-être confronté le criminaliste n'est pas limité. Cet effet de nombre porte n'ont pas forcément sur la formulation du rapport de vraisemblance puisqu'une même formulation peut-être utilisé à plusieurs cas dont les circonstances sont différentes. Par contre c'est essentiellement dans l'estimation des variables que ce nombre de possibilité intervient. De ce fait, pour ce prototype, la stratégie a été de ne pas estimer ces variables dépendantes de circonstances précises dans le système, mais plutôt de donner des informations pour que l'utilisateur soit capable de les estimer de manière adaptée.

A ce stade de la recherche, le module est implémenté seulement pour quelques scénarios essentiellement au niveau de la source.

Le prototype développé dans ce mémoire se décompose essentiellement en trois parties. Un module de collecte d'informations qui permet de couvrir un large éventail de scénarios possibles. Un module de décisions qui donne un conseil sur le choix du niveau d'inférence, et enfin, un module de propositions et de calcul de LR qui permet, avec une formulation de LR qui correspondant au cas, de calculer la force de l'inférence en fonction des paramètres évalués par l'utilisateur.

Ce projet regroupe une trentaine de fichiers de code dont six sont voués au stockage de toutes les données récoltées par le système. Les autres fichiers contiennent au fur et à mesure de l'avancement dans l'arborescence les règles, faits, variables et fonctions nécessaires au nœud en question. Le prototype contient, au total, environ soixante règles et une centaine de fonctions qui gèrent plus de cent-cinquante variables et faits. Les fonctions, qui ont un rôle prédominant dans l'élaboration de l'interface utilisateur ainsi que dans la cohérence du programme, ont la particularité d'être spécialement longues étant donné qu'une bonne partie d'entre elles comportent une à plusieurs centaines de lignes de code chacune.

Ainsi, la mise en œuvre d'un tel programme est très longue, d'autant plus pour un débutant dans le domaine. Cette constatation n'est pas nouvelle puisque le projet MYCIN qui est un système expert utilisé pour le diagnostic d'infections bactériologiques et leur traitement, a nécessité 10 ans de développement au début des années septante (STEELS 1991). Une issue pour pallier un tel investissement en temps est de mettre sur pied une équipe de développement tout comme dans le projet MYCIN. En effet, il est très difficile d'endosser tous les rôles que nécessite le développement d'un système expert. Un travail d'équipe, comme dans le cas des deux exemples Cage2000 et FINDS, permettrait d'acquérir beaucoup plus de recul sur les options de développement. De plus, l'apport d'un développeur permettrait un formalisme de développement plus ambitieux que dans le cadre de ce travail de mémoire. Il serait alors possible, à l'image de Cage2000, de songer à une application web plus ambitieuse.

5.2 Test utilisateur

Le test a montré une prise en main relativement facile du système étant donné que les cinq participants ont trouvé une valeur de LR correcte pour le cas donné. Ces résultats sont à considérer avec beaucoup de prudence. En effet, d'une part, le scénario proposé était très simple du fait que seul le niveau de l'inférence sur la source était implémenté au moment du test.

La collecte des données n'a pas posé de problème pour les utilisateurs. Il faut aussi nuancer ce résultat car les cinq utilisateurs emploient quotidiennement des ressources informatiques et sont donc très familiers avec l'utilisation de menus déroulants ou d'interface de ce genre. Par contre en ce qui concerne l'interface, des commentaires ont été formulés sur le fait que des problèmes d'affichage surviennent lors de la régénération de la fenêtre. La cause de ce problème n'a pas été trouvée, mais semble venir de petits problèmes de compatibilité entre le système et l'application.

5.3 Perspectives

Les résultats obtenus dans ce mémoire montrent que la technologie des systèmes experts est intéressante pour le domaine de l'interprétation et que des projets ont déjà été développés dans ce domaine. La voie montrée par Cage2000 et son interface web permettrait d'élargir la perspective d'utilisation d'un tel système à des fins pédagogiques et ainsi d'améliorer le rapport entre le coût de développement et la population cible. Cette ambition tomberait à point nommé dans le contexte actuel de développement d'un mastère en interprétation dont une partie pourrait très bien se faire à distance.

Plutôt que de se pencher uniquement sur les systèmes experts, il serait préférable pour l'enseignement de l'interprétation de songer à intégrer cet outil dans une véritable application pédagogique tenant compte des difficultés de l'utilisateur, qui teste les connaissances, etc. Cette plateforme d'enseignement utiliserait la technologie des systèmes experts, par exemple dans la résolution de cas, et permettrait ainsi à l'apprenant de construire aussi ses propres scénarios. Le logiciel serait donc utilisé dans deux contextes différents, l'un essentiellement pour apprendre le domaine et l'autre plutôt pour le pratiquer.

Chapitre 6

Conclusion

Ce projet s'intéresse à étudier la faisabilité et les possibilités que pourraient offrir les systèmes experts à l'enseignement de l'indice matériel. Dans une première étape, des critères nécessaires à l'utilisation d'un système expert comme aide à l'enseignement ont été identifiés. Ainsi le système doit être capable de gérer une multitude de cas grâce à une collecte d'informations généralisée, de formuler un LR adapté, d'explicitier les critères utilisés lors de la décision, et de donner des informations sur les paramètres en jeu et sur leur évaluation.

Ensuite l'hypothèse qu'un système répondant aux conditions spécifiées est réalisable dans le domaine des sciences forensiques est avancée. Dans ce sens, le prototype proposé dans ce travail répond véritablement au but fixé. En effet même si l'efficacité des fonctionnalités mises en place dans ce prototype n'a pas été encore évaluée, elles ont pu être réalisées. Ce projet confirme donc l'hypothèse qu'il est possible de développer ce type de programme pour le domaine de l'interprétation de l'indice.

Enfin, l'hypothèse que ce prototype et son interface sont utilisables par un criminaliste est aussi confirmée dans ce travail par les tests d'utilisateurs grâce à l'accent mis sur la création d'une interface de logiciel classique.

Le prototype réalisé dans ce travail, associé aux travaux déjà accompli dans ce domaine, permet d'entrevoir un intérêt pratique des systèmes experts dans le domaine de l'interprétation. Cependant, la mise en place d'un tel système est une procédure longue et coûteuse qui nécessite la mise en place d'une équipe de développement. Une des solutions d'avenir afin d'augmenter les retombées du produit, serait le développement d'une application web et son intégration dans le cursus d'un mastère dans le domaine, réalisé à distance.

Bibliographie

- CASTILLO, E., GUTIÉRREZ, J. M. & HADI, A. 1997, *Expert Systems and Probabilistic Network Models*, Springer, New York.
- CHAMPOD, C. & TARONI, F. 1999, The bayesian approach, in J. ROBERTSON & M. GRIEVE, eds, 'Forensic Examination of Fibres', 2 ed. edn, Taylor & Francis, London, pp. 379–398.
- CURRAN, J. M., TRIGGS, C. M., BUCKLETON, J. S., WALSH, K. A. J. & HICKS, T. 2003, 'Assessing transfert probabilities in a bayesian interpretation of forensic glass evidence'. Soumis à publication.
- DEFOREST, P., LEE, S. & GAENSSLEN, R. 1983, *Forensic Science : An Introduction to Criminalistics*, Mc Graw Hill, New York.
- EVETT, I. W., COOK, R., JACKSON, G., JONES, P. J. & LAMBERT, J. A. 1998, 'A model for case assesment and interpretation', *Science & Justice* **38**(3), 151–156.
- G.COWELL, R., DAWID, A. P., LAURITZEN, S. L. & SPIEGELHALTER, D. J. 1999, *Probabilistic Networks and Expert Systems*, Springer, New York.
- HICKS, T. 2003, De l'Interprétation des Fragments de Verre en Sciences Forensiques, PhD thesis, Institut de Police Scientifique, Université de Lausanne. Soumis à publication.
- HOPGOOD, A. A. 2001, *Intelligent Systems for Enginneers and Scientists*, CRC Press, Boca Raton.
- INMAN, K. & RUDIN, N. 2001, *Principles and Practice of Criminalistics : The Profession of forensic science*, CRC Press, Boca Raton.
- KIRK, P. L. 1983, 'The ontogeny of criminalistics', *J. Criminal Law Criminol. Police Sci.* **54**, 235–238.
- LOCARD, E. 1920, *L'Enquete Criminelle et les Méthodes Scientifiques*, Flammarion, Paris.
- MARGOT, P. 2003, *Cours général de sciences forensique*, Institut de Police Scientifique, Université de Lausanne, Lausanne.
- MENDELSON, P. & DILLENBOURG, P. 1991, Le développement de l'enseignement intelligent assisté par ordinateur, in 'Conférence donnée à la réunion de l'Association de Psychologie de Langue Française', Rome.
- OSTERBURG, J. W. 1968, *The Crime Laboratory; Case Studies of Scientific Criminal Investigation*, Indiana University Press, Bloomington.
- PUSH, R. O., CHAMPOD, C., EVETT, I. W., SMITH, J., WYNN, H. P. & BATES, R. A. 2003, 'A bayesian decision support system in forensic science'. Soumis à publication.

- ROUX, C. 1997, La Valeur Indicielle des Fibres Textiles Découverte sur un Siège de Voiture : Problèmes et solutions, PhD thesis, Institut de Police Scientifique, Université de Lausanne.
- SAFERSTEIN, R. 1998, *Criminalistics : An Introduction to Forensic Science*, 6th ed. edn, Prentice-Hall, Englewood Cliffs, NJ.
- STEELS, L. 1991, 'Knowledge systems', Draft published for review purpose, University of Brussels.
- TARONI, F. 2003, *Cours Interprétation de la preuve*, Institut de Police Scientifique, Université de Lausanne, Lausanne.
- TARONI, F., BIEDERMANN, A., GARBOLINO, P. & AITKEN, C. G. C. 2003, 'A general approach to bayesian networks for the interpretation of evidence'. Soumis à publication.