



HAL
open science

**SAFE-NEXT : UNE APPROCHE SYSTEMIQUE
POUR L'EXTRACTION DE CONNAISSANCES DE
DONNEES. Application A La Construction Et A
L'interprétation De Scénarios D'accidents De La Route**

Walid Ben Ahmed

► **To cite this version:**

Walid Ben Ahmed. SAFE-NEXT : UNE APPROCHE SYSTEMIQUE POUR L'EXTRACTION DE CONNAISSANCES DE DONNEES. Application A La Construction Et A L'interprétation De Scénarios D'accidents De La Route. domain_stic.gest. Ecole Centrale Paris, 2005. Français. NNT : . tel-00011540v2

HAL Id: tel-00011540

<https://theses.hal.science/tel-00011540v2>

Submitted on 8 Feb 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**ÉCOLE CENTRALE DES ARTS
ET MANUFACTURES
« ÉCOLE CENTRALE PARIS »**

THÈSE

présentée par

Walid BEN AHMED

pour l'obtention du

GRADE DE DOCTEUR

Spécialité : Génie Industriel

Laboratoire d'accueil : LGI- Laboratoire Génie Industriel

SUJET

**SAFE-NEXT : UNE APPROCHE SYSTEMIQUE POUR L'EXTRACTION DE
CONNAISSANCES DE DONNEES.**
APPLICATION A LA CONSTRUCTION ET A L'INTERPRETATION DE SCENARIOS D'ACCIDENTS
DE LA ROUTE

Soutenue le : 17 janvier 2005

Devant le jury composé de :

Daniel COUTELLIER – Professeur ENSIAME Valenciennes et DG AIP-PRIMECA	Rapporteur/Président
Rose DIENG – Chef du projet ACACIA, INRIA Sophia Antipolis	Rapporteur
Christian TAHON – Maître de conférence, Université de Valenciennes	Rapporteur
Claudia ECKERT – Professeur Engineering Design Center, University of Cambridge	Examineur
Yves PAGE – Adjoint du directeur du LAB PSA Peugeot-Citroën Renault	Examineur
Michel TOLLENAERE – Professeur INPG, Grenoble	Examineur
Mounib MEKHILEF – HDR à l'Université d'Orléans	Directeur de thèse
Michel BIGAND – Maître de Conférence, Ecole Centrale Lille	Co-directeur de thèse

2005 -1

A mes parents

Remerciements

Je tiens à remercier en tout premier lieu Mounib Mekhilef et Michel Bigand qui ont encadré cette thèse dans la continuité de mon stage de DEA. Leurs conseils et idées, tant sur le plan méthodologique que scientifique, ont profondément inspiré et nourri ma réflexion au cours mes travaux. Il m'est difficile d'imaginer que cette thèse aurait pu être menée à bien sans leur direction bienveillante.

Je remercie également Yves page qui m'a encadré au sein du LAB PSA Peugeot-Citroën Renault et sans qui ces travaux de recherche n'auraient jamais vu le jour. Je tiens à lui exprimer ma gratitude pour ses remarques et son esprit critique et rigoureux qui m'ont été infiniment précieux. J'ai pleinement profité de sa grande expérience dans le domaine de l'accidentologie.

Je voudrais remercier le professeur Jean Claude Bocquet, directeur du Laboratoire Génie Industriel à l'École Centrale Paris, de m'avoir accueilli dans son équipe de recherche. J'adresse mes vifs remerciements à Jean-Yves Le Coz (ancien directeur du LAB) et Hervé Guillemot, directeur du LAB, de m'avoir accueilli dans leur laboratoire et permis de réaliser ce travail dans d'excellentes conditions. Je tiens à remercier également Léonard Campione, délégué général du CEESAR, et Thierry Hermite pour l'intérêt qu'il ont porté à mon travail et l'appui qu'il m'ont fourni.

Je souhaite exprimer toute ma gratitude à Richard Driscoll, François Bar, Marie Christine Simon, Sami Nasri, Mustapha Ben Hassine, Sylvain Alleaume, Claire Mautuit et Anne Damville pour leur implication forte dans la réalisation et la validation de mes travaux.

J'adresse mes vifs remerciements aux rapporteurs de ce mémoire, Daniel Coutellier et Rose Dieng pour l'intérêt qu'ils ont porté à mon travail. Merci également à Michel Tollenaere et Claudia Eckert pour avoir accepté de faire partie du jury de cette thèse.

Mes remerciements s'adressent également à Jean-Pierre Bourrey pour ses précieuses recommandations en UML et à Bernard Yannou pour ses conseils et les travaux de collaboration que nous avons menés ensemble.

Je tiens à reconnaître une vive reconnaissance à l'ensemble des personnes du LAB et du CEESAR pour leur soutien durant les quatre années que nous avons passées ensemble. En particulier, je remercie Thierry pour nos discussions interminables, Nathalie et Martine pour leur gentillesse et leur disponibilité, Mohamed, Olivier, Wajih, Clément, Fabien, Emmanuel et tous les autres. Je remercie la famille du Laboratoire Génie Industriel pour son accueil chaleureux. En particulier, Anne, Sylvie, Anne Jubert, Zied, Barth, Leila, Aurélie, Caroline, Ali, Walid, Selma, Yacine, Hamza et les autres.

Abdelbasset et Ahmed, merci d'avoir traversé avec moi ces trois années et de m'avoir accompagné et soutenu jusqu'à la fin de cette thèse.

Enfin, je remercie ma famille, tous mes amis Jihen, Nadra, Ahmed, Nader et tous ceux qui, de loin ou de près, m'ont soutenu.

Résumé

Aujourd'hui, l'informatisation des saisies de données et la puissance des systèmes de collecte conduisent à la construction de grandes Bases de Données (BD). L'exploitation de ces millions de données en accidentologie et dans beaucoup d'autres domaines (e.g. management, marketing, etc.) fait appel à des techniques d'Extraction de Connaissances de Données (ECD). La complexité des données, du domaine d'application et des connaissances recherchées rendent fondamentale l'intégration des connaissances expertes dans le processus d'ECD. Cela nécessite la mise en place d'approches, méthodes et techniques d'identification, de représentation et d'opérationnalisation de ces connaissances.

Dans ce travail de thèse, nous proposons une nouvelle approche, appelée *SAFE-Next* (Systemic Approach For Enhanced kNowledge EXTraction, approche systémique pour l'extraction des connaissances) qui intègre les quatre approches suivantes : La première est appelée *ASMEC* (Approche Systémique de ModElisation des Connaissances). Elle consiste en une méthode de modélisation des connaissances multi-vues et selon une architecture à plusieurs niveaux d'abstraction. La deuxième approche, *AICEF* (Approche d'Incorporation des Connaissances Expertes dans la Fouille de données), propose l'élaboration et l'utilisation de *méta-données multi-vues* comme un moyen pour l'incorporation des connaissances formalisées par *ASMEC* dans le processus d'ECD. La troisième approche, *ASAIC* (Approche Systémique d'Analyse d'Impact de Changement), utilise le modèle de connaissances d'*ASMEC* pour une analyse interactive et multi-vues de l'impact d'un changement sur un système. La quatrième approche, *ASEM* (Approche Systémique d'Evaluation de Modèles), fournit un *modèle général d'évaluation de modèles de connaissances*.

Les fondements épistémologiques et méthodologiques de nos travaux sont respectivement le *constructivisme* et la *systémique* (ou *cybernétique*). En se basant sur ces fondements, nos travaux de recherche ont conduit à des contributions réparties en quatre domaines : En accidentologie, *SAFE-Next* fournit un outil efficace pour l'élaboration des STA permettant une meilleure analyse et compréhension de l'accident. Elle fournit aussi un moyen de capitalisation des connaissances offrant une vision synthétique des différents types de connaissances du domaine de l'accidentologie. En Ingénierie des Connaissances (IC), *SAFE-Next* propose un modèle général multi-vues et multi-niveaux d'abstraction de modélisation des connaissances pour le développement des Systèmes à Base de Connaissances (SBC). Elle permet aussi de guider l'élicitation des connaissances selon un modèle multi-vues. En ECD, *SAFE-Next* propose l'utilisation des métadonnées multi-vues pour l'incorporation des connaissances expertes du domaine dans la première et la dernière phase du processus d'ECD (i.e. préparation des données et interprétation des résultats). En conception de nouveaux systèmes, *SAFE-Next* fournit à travers les STA un moyen de communication entre les accidentologues et les concepteurs des systèmes de sécurité embarqués dans les véhicules. Cette interface entre les deux métiers (i.e. conception et accidentologie) permet la construction de l'espace de conception pour développer et évaluer les systèmes de sécurité. Elle offre aussi un moyen d'analyse de l'impact d'un changement (e.g. introduction d'un nouveau système de sécurité) sur le comportement du système Conducteur-Véhicule-Environnement.

Mots clefs : Ingénierie des Connaissances, Extraction de Connaissances de Données, Modélisation des connaissances, Modélisation des systèmes complexes, Cybernétique, Systémique, Analyse de changement, Conception de nouveaux systèmes, Evaluation de modèles

Abstract

Nowadays, given the automation of data collection, very large databases are constructed. The exploitation of these data in accidentology and several others fields (e.g. marketing, engineering, etc.) requires automatic techniques of Knowledge Discovery in Databases (KDD). Incorporating expert knowledge in the KDD process is fundamental to handle the complexity of data, domain and knowledge. This necessitates the development of approaches, methods and techniques intended to identify, represent and operationalize expert knowledge.

In this dissertation, we propose a new approach, *SAFE-Next* (Systemic Approach For Enhanced kNnowledge EXtraction), which integrates the following four approaches: the first one, *ASMEC* (Approche Systémique de ModElisation des Connaissances), allows knowledge modeling according to multiple viewpoints and granularity levels. The second approach, *AICEF* (Approche d'Incorporation des Connaissances Expertes dans la Fouille de données), uses the *ASMEC* knowledge model to elaborate multi-view *metadata*. It then uses these metadata as a tool for incorporating expert knowledge into the KDD process. The third approach, *ASAIC* (Approche Systémique d'Analyse d'Impact de Changement), uses the *ASMEC* knowledge model to carry out multi-view change impact analysis. The fourth approach, *ASEM* (Approche Systémique d'Evaluation de Modèles), provides an assessment framework for knowledge models.

The epistemological and methodological foundations of our work are constructivism and systemic approach (or cybernetics). Based on these backgrounds, our research contributions concern several disciplines, ranging from Accidentology, Knowledge Engineering, Knowledge Discovery in Databases and Design. In accidentology, *SAFE-Next* provides experts with an efficient tool for knowledge management. It enables the elaboration of multi-view accident scenarios, which are a powerful tool for understanding accident mechanisms in order to develop safety counter-measures. Furthermore, *SAFE-Next* provides a knowledge capitalization tool. In knowledge engineering, *SAFE-Next* supplies, via *ASMEC*, a multi-view knowledge model and thereby allows the integration of different viewpoints stemming from different users. Furthermore, it provides a multi-granularity knowledge model and in that way addresses the difficulty of knowledge identification and formalization. At the same time, *SAFE-Next* permits, via *ASEM*, the evaluation of knowledge models, an issue rarely addressed in literature. In Knowledge Discovery in Database, *SAFE-Next* enables, via *AICEF*, the incorporation of domain knowledge in the data preprocessing step (i.e. the first step in a KDD process) and more specifically in the attribute selection task. Likewise, the multi-view metadata enables the incorporation of domain knowledge in the interpretation step (i.e. the last step in a KDD process). In design, *SAFE-Next* provides safety system developers with an efficient tool to construct the design space. Scenarios enable them to understand complex behaviors and thereby to define solutions and alternatives. *SAFE-Next* also provides, via *ASAIC*, an approach for multi-view change impact analysis. Moreover, it proposes an extension of the change impact analysis to the use process of a given product as well as the evaluation process instead of limiting it to the design process.

Keywords: Knowledge Engineering, Knowledge Discovery in Databases, Knowledge Modeling, Complex System Modeling, Cybernetics, Systemic Approach, Change Impact Analysis, New System Design, Model Assessment.

Table des matières

Introduction	1
Guide de lecture	5
Partie I Présentation de la problématique	7
Présentation de la partie I	8
Chapitre 1. Contexte	9
1.1. Contexte économique-socio-politique	9
1.2. Sécurité routière et image de marque des constructeurs automobile.....	10
1.3. Contexte industriel : LAB (Peugeot-Citroën Renault).....	11
1.4. Les connaissances en entreprise : cruciales mais coûteuses	11
Chapitre 2. Positionnement scientifique et contributions	12
2.1. Positionnement épistémologique.....	12
2.2. Positionnement méthodologique.....	13
2.3. Positionnement par rapport à l'accidentologie routière	14
2.4. Positionnement par rapport à la conception de nouveaux systèmes.....	15
2.5. Positionnement par rapport à l'ingénierie d'analyse d'impact de changement.....	16
2.6. Positionnement par rapport à l'extraction de connaissances de données.....	17
2.7. Positionnement par rapport à l'ingénierie des connaissances	18
2.8. Synthèse du positionnement scientifique	19
Synthèse de la partie I.....	20
Partie II Etat de l'art	21
Présentation de la partie II	22
Chapitre 3. Cybernétique & théorie du système général	24
3.1. Absolutisme versus relativisme	24
3.2. Subjectivisme versus objectivisme.....	25
3.3. Constructivisme versus positivisme.....	26
3.4. L'approche systémique versus l'approche analytique.....	27
3.5. L'accident de la route : un phénomène complexe ou compliqué ?	31
3.6. La cybernétique de second ordre : vers une vision systémique de la modélisation	32
3.7. La systémique : une approche réflexive.....	33
3.8. Synthèse.....	34

Chapitre 4.	Accidentologie	36
4.1.	La sécurité routière	36
4.2.	Les EDA : Etudes Détaillées d'Accidents	37
4.3.	Modélisation de l'accident de la route.....	38
4.4.	Concept de scénario-type d'accidents	44
4.5.	Synthèse et positionnement	49
Chapitre 5.	Modélisation des connaissances en ingénierie des connaissances	50
5.1.	Connaissance : définitions et typologies	50
5.2.	Ingénierie des connaissances : présentation générale.....	53
5.3.	La modélisation des connaissances.....	54
5.4.	La représentation et l'opérationnalisation des modèles de connaissances	62
5.5.	La notion de point de vue en modélisation des connaissances.....	63
5.6.	La validation des connaissances acquises	65
5.7.	Synthèse et positionnement	65
Chapitre 6.	Extraction des connaissances de données (ECD).....	67
6.1.	Processus d'ECD.....	67
6.2.	Complexité et extraction de connaissances de données.....	73
6.3.	Intégration des connaissances du domaine dans le processus ECD	74
6.4.	Synthèse et positionnement	76
	Synthèse de la partie II	77
Partie III	SAFE-Next : Développement d'une approche systémique pour l'extraction des connaissances	79
	Présentation de la partie III.....	80
Chapitre 7.	Elaboration d'un modèle de connaissance en accidentologie : utilisation de CommonKADS	82
7.1.	La connaissance du domaine	83
7.2.	La connaissance d'inférences	88
7.3.	La connaissance de la tâche d'élaboration des STA.....	94
7.4.	Critiques.....	96
7.5.	Synthèse	97

Chapitre 8.	ASMEC : une Approche Systémique de Modélisation	
d'Es Connaissances	98
8.1.	Objectifs d'ASMEC	99
8.2.	Choix d'une stratégie de modélisation.....	99
8.3.	Modélisation dirigée par un méta-modèle général	100
8.4.	Architecture globale d'ASMEC.....	101
8.5.	Spécification du méta-modèle	103
8.6.	Instrumentation du méta-modèle par les modèles.....	105
8.7.	Instrumentation des modèles par les concepts	110
8.8.	Construction des métadonnées : caractériser les concepts par les attributs	113
8.9.	Illustration de l'architecture d'ASMEC.....	121
8.10.	Synthèse.....	121
Chapitre 9.	AICEF : une Approche d'Incorporation des	
Connaissances Expertes dans la Fouille de données	123
9.1.	Utilisation des métadonnées pour la sélection des attributs.....	124
9.2.	Utilisation des métadonnées pour l'interprétation des résultats de classification.....	126
9.3.	Synthèse.....	140
Chapitre 10.	ASAIC : une Approche Systémique d'Analyse	
d'Impact de Changement	142
10.1.	Besoin d'analyse d'impact d'un changement sur le système CVE.....	143
10.2.	Processus d'Ingénierie d'Analyse d'Impact de Changement (IAIC)	143
10.3.	Différentes approches d'IAIC.....	144
10.4.	Elaboration d'un modèle d'analyse d'impact d'un changement sur le système CVE	145
10.5.	Cas d'étude : analyse de l'impact de l'ESP sur le comportement du système CVE.....	148
10.6.	ASAIC : vers une approche globale d'analyse de changement	149
10.7.	Synthèse.....	153
Chapitre 11.	ASEM : une Architecture Systémique d'Evaluation	
de Modèles	154
11.1.	Motivations de l'élaboration d'ASEM.....	155
11.2.	Critères d'évaluation de connaissances	157
11.3.	Critères d'évaluation d'un modèle	159
11.4.	Relation critères d'évaluation d'un modèle /critères d'évaluation de connaissances.....	165
11.5.	Relation intra-critères d'évaluation d'un modèle.....	165
11.6.	Synthèse.....	168

Chapitre 12.	Validation et apports de SAFE-Next	169
12.1.	Démarche de validation.....	169
12.2.	Les apports de SAFE-Next	171
	Synthèse de la partie III	179
Conclusion générale	181	
	Synthèse générale de nos travaux.....	182
	Synthèse des contributions	188
	Discussion & perspectives.....	190
Références bibliographiques.....	193	
Annexe	205	
	A-1. Modèle de connaissances de l'élaboration des STA en accidentologie...	206
	A-2. Enrichir la base EDA en utilisant ASMEC	236
	A-3. Représentation XML de la projection d'un accident sur les méta- données multi-vues.....	237
	A-4. Traduction de l'objectif de l'étude sous forme d'une requête SQL.....	242
Liste des figures	245	
Liste des définitions	248	
Liste des abréviations	249	

Introduction

Le bilan des accidents et des victimes de la route recensés au cours de l'année 2003 en France, présente 5 731 tués et 115 929 blessés, soit un coût total de 24 milliards d'euros. Tel est le bilan des 86 992 accidents dénombrés sur les routes de la France. En 2003, des réductions ont été enregistrées en France par rapport à 2002 : 14,5% pour les accidents, 20,9% pour les tués et 15,9% pour les blessés en 2003. Malgré ces améliorations, le bilan reste lourd faisant de la France le quatrième pays dans l'union européenne ayant le nombre de tués par million d'habitants le plus élevé qui s'estime à 122 en 2002.

Ce bilan fait de la sécurité routière un enjeu social, économique, technologique et politique. L'importance de ces enjeux s'est traduite sous forme de plusieurs types de contre-mesures qui se manifestent à travers des lois, des réglementations, des actions de prévention et des actions de recherche. Pour améliorer la sécurité routière, les constructeurs automobiles français ont créé une structure commune qui est le Laboratoire d'Accidentologie, de Biomécanique et du Comportement Humain (LAB). L'objectif du LAB est d'apporter un retour d'expérience sur l'accident de la route vers les constructeurs PSA Peugeot Citroën et Renault. Ceci est assuré par la construction et l'exploitation de base de données d'accidents et de comportement humain, des études en biomécanique et le développement d'outils et de critères permettant d'accroître la sécurité routière.

La mission du LAB est de construire une base de connaissances en accidentologie exploitable essentiellement par les constructeurs pour l'amélioration de la sécurité routière. Le concept de Scénario Type d'Accidents (STA) est considéré comme l'un des supports de cette connaissance en accidentologie. Ce concept est apparu pour résoudre la difficulté de synthèse dont la problématique est la suivante : « *comment passer de la compréhension de chaque accident, pris individuellement, à une vue plus générale des problèmes de sécurité sur l'échantillon d'accidents étudiés* » [Brenac et Fleury, 1999]. L'utilisation des STA est basée sur le fait qu'a priori les accidents qui présentent des similitudes au niveau de leur déroulement, présentent des similitudes au niveau des contre-mesures. Ce concept est en France dû à l'INRETS¹ qui a, au début des années 1980, développé ce genre d'approche pour caractériser des situations capables de rendre compte synthétiquement d'un certain nombre de configurations. Il est utilisé au LAB comme base de réflexion pour l'évaluation a priori ou a posteriori des Systèmes de Sécurité Embarqués (SSE) dans les véhicules et pour identifier de nouvelles prestations de sécurité [Page, 2002].

L'élaboration de ces STA peut être effectuée d'une manière purement experte en commençant par une analyse en profondeur et cas par cas des accidents [Brenac et Fleury, 1999; Megherbi, 1999]. Après cette analyse, on procède à un regroupement des cas qui présentent des similitudes. Ce regroupement se base sur une comparaison qualitative et holistique faite par l'expert. Cependant, le coût de construction des STA est très élevé étant donné la complexité des données, la variété de leurs sources, le grand nombre d'attributs caractérisant un accident, la complexité des interactions entre ces attributs, le nombre important d'accidents stockés dans les BD, la variété des études (perte de contrôle, freinage) et la multitude de points de vue d'analyse de l'accident.

L'Extraction de Connaissances à partir de Données d'accidents (ECD) (ou la fouille de données) est une autre approche permettant la construction de ces scénarios : puisque les STA sont

¹ Institut National de Recherche sur les Transports et leur Sécurité

basés sur la notion de similarité, alors leur construction peut se faire par une classification automatique des accidents en groupes similaires. Mais, pour les mêmes raisons citées dans le paragraphe précédent, il est fondamental d'intégrer les connaissances expertes du domaine dans le processus d'ECD. En effet, dans la première phase du processus ECD (i.e. préparation des données), les connaissances expertes du domaine sont nécessaires pour traduire l'objectif d'une étude donnée en attributs d'étude. Les attributs utilisés pour étudier la perte de contrôle sont différents de ceux utilisés pour étudier l'effet de l'alcool sur la conduite. De même, lors de la phase d'interprétation des résultats issus de la classification automatique, dernière phase du processus ECD, les connaissances expertes sont cruciales : il ne suffit pas d'obtenir des classes d'accidents, mais il faut permettre leur interprétation pour en extraire des connaissances pertinentes et exploitables.

Ainsi, les deux alternatives, automatique et « au cas par cas », présentent plusieurs limites. D'un côté, la quantité importante des données, leur nature multi-sources et le besoin de construire des STA à la demande (selon un objectif donné) rendent une construction experte très coûteuse. D'un autre côté, la complexité de l'accident, la pluridisciplinarité de l'accidentologie et le besoin de combiner plusieurs points de vue d'analyse rendent une construction automatique de ces STA inadaptée en terme de qualité des résultats. Nous proposons alors une approche qui combine les deux approches experte et automatisée.

En plus des STA, le LAB a besoin de méthodes et outils permettant l'Analyse d'Impact de Changement (AIC) sur le comportement du système Conducteur-Véhicule-Environnement (CVE) et sur le métier de l'accidentologie lui-même. Un changement au niveau de l'un de ces composantes (e.g. nouveau système de sécurité, nouvelle technologie, nouvelle loi) peut affecter dramatiquement le comportement du système CVE et donc la sécurité routière en général. Par exemple, un conducteur avec un véhicule équipé d'un ABS¹ ou d'un ESP² peut se sentir plus en sécurité et donc prendre plus de risques dans sa conduite. Les configurations des accidents ainsi que le besoin en sécurité peuvent donc changer. Ces effets doivent être anticipés pour éviter les effets pervers. Là encore, le besoin d'automatiser une partie du processus d'AIC en intégrant l'expertise s'avère crucial.

L'identification des connaissances expertes nécessaires à l'élaboration des STA et à l'analyse des changements, la formalisation des ces connaissances et leur opérationnalisation (i.e. mise en œuvre) sont donc les principaux défis auxquels nous sommes confrontés. La problématique générale dans laquelle s'intègrent nos travaux est la Gestion des Connaissances ou KM³ : il s'agit de développer de nouvelles connaissances en accidentologie exploitables pour le développement et l'évaluation des systèmes de sécurité.

¹ ABS (Antiblockier system) : système antiblocage des freins. Il permet d'éviter le blocage d'une ou plusieurs roues pendant le freinage.

² ESP (Electronic Stability Program) : Programme de Contrôle de Stabilité. Il redresse le véhicule en cas de sous-virage ou sur-virage pour empêcher la perte de contrôle.

³ KM (Knowledge Management) : c'est le terme le plus utilisé dans la communauté scientifique pour désigner la gestion de connaissances.

Nos travaux sont à cheval sur plusieurs disciplines (cf. Figure 1) : en **Ingénierie des Connaissances (IC)** (cf. section 2.7), nous proposons l'utilisation de l'*approche systémique* de modélisation des systèmes complexes pour l'identification des points de vue du domaine. Cela nous a permis l'élaboration d'une approche que nous appelons *ASMEC (Approche Systémique de ModElisation des Connaissances)* offrant un cadre général pour la modélisation multi-vues et multi-niveaux d'abstraction des connaissances expertes. Nous proposons aussi une méthode d'évaluation des modèles de connaissances que nous appelons *ASEM (Approche Systémique d'Evaluation de Modèles)*. En **extraction de connaissances de données** (cf. section 2.6), nous proposons une approche que nous appelons *AICEF (Approche d'Incorporation des Connaissances Expertes dans la Fouille de données)*. Elle permet l'élaboration et l'utilisation des métadonnées multi-vues pour l'incorporation des connaissances expertes du domaine dans la première et la dernière phase du processus ECD (i.e. préparation des données et interprétation des résultats). En **accidentologie** (cf. section 2.3), nous proposons l'utilisation d'*ASMEC* et d'*AICEF* pour l'élaboration de STA multi-vue. En **conception de nouveaux systèmes** (cf. section 2.4), nous proposons l'utilisation des STA comme un moyen de communication entre les accidentologues et les concepteurs des SSE. Ces STA aident aussi les concepteurs des SSE à construire l'*espace de conception*. En effet, ils permettent de comprendre les mécanismes complexes de production de l'accident et donc de fournir aux concepteurs une base de réflexion pour définir et évaluer des prestations de sécurité. Nous développons aussi une approche que nous appelons *ASAIC (Approche Systémique d'Analyse d'Impact de Changement)* permettant une analyse interactive de l'impact d'un changement sur la sécurité routière. Elle permet essentiellement l'évaluation d'un nouveau système de sécurité. Des pistes sur l'évaluation de l'impact d'un changement sur le processus de conception des véhicules ainsi que sur le métier de l'accidentologie sont proposées aussi.

Nous avons intégré l'ensemble des méthodes développées dans ce mémoire (*ASMEC, AICEF, ASAIC* et *ASEM*) dans une approche que nous appelons **SAFE-Next** (Systemic Approach For Enhanced kNowledge EXTraction) (ou *ASEC : Approche Systémique pour l'Extraction des Connaissances*) (cf. Figure 1).

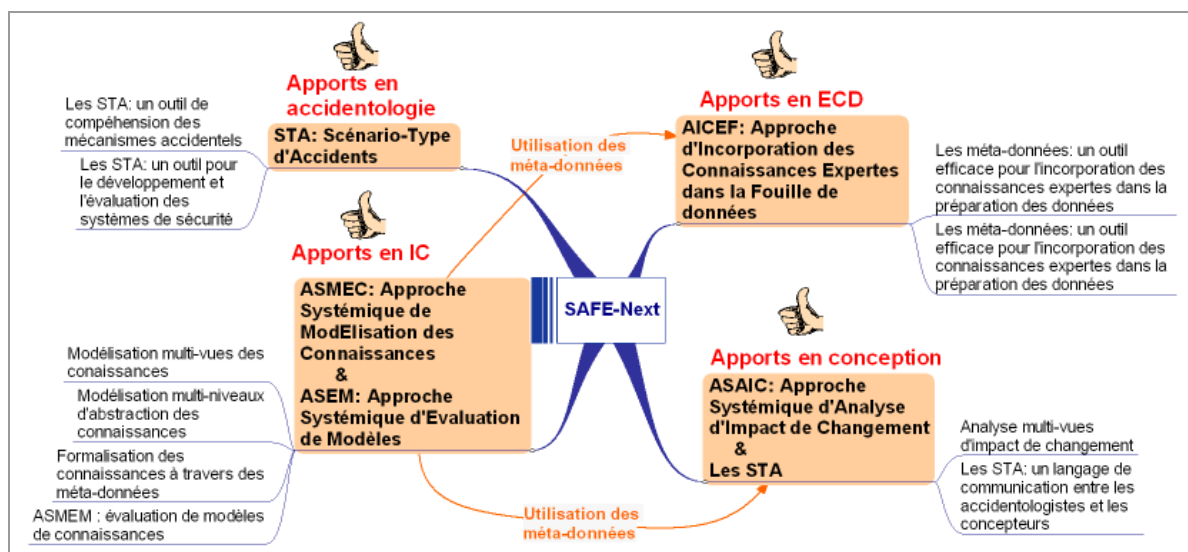


Figure 1 - Une vue domaine de nos contributions : l'apport de SAFE-Next dans les domaines d'accidentologie, de conception, d'ingénierie des connaissances et d'extraction de connaissances de données

Guide de lecture

Ce mémoire de thèse est présenté en quatre parties qui se décomposent en chapitres. Le plan ainsi que les différentes parties sont présentés dans la Figure 1 et détaillés dans les points suivants :

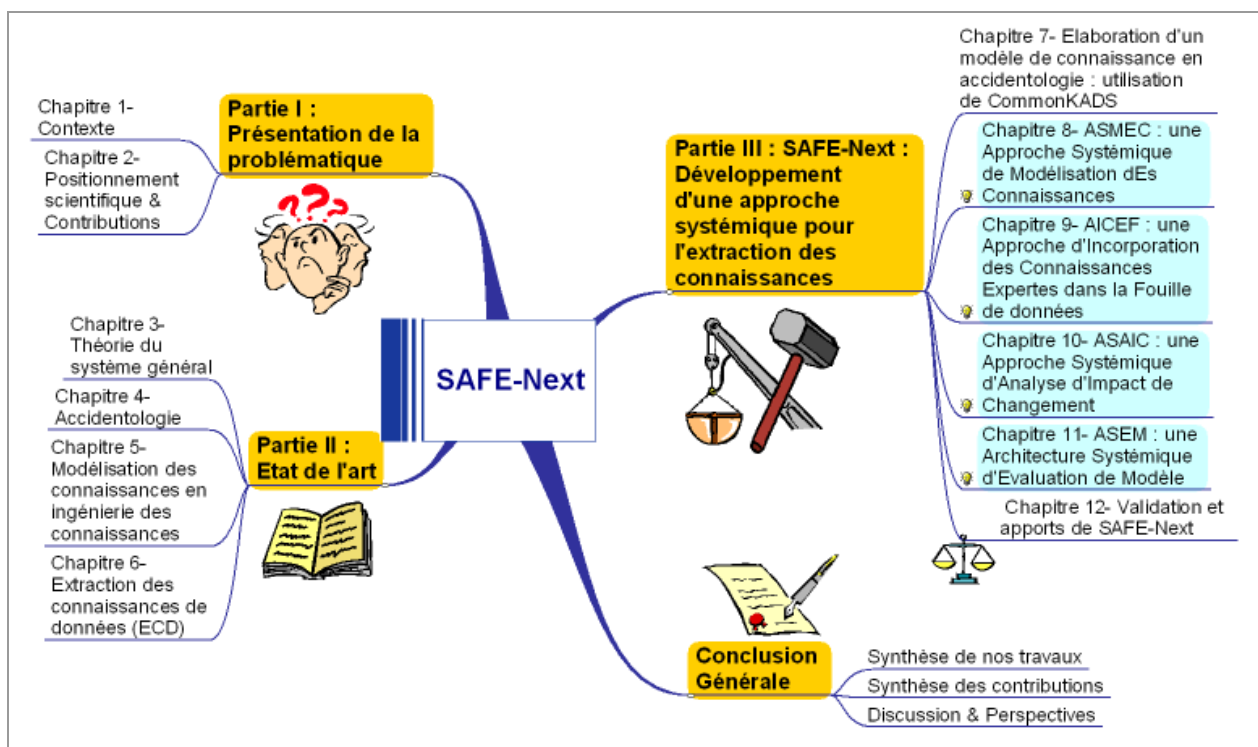


Figure 2- Guide de lecture du mémoire

- 1- La première partie se compose de deux chapitres. Dans le premier nous présentons la problématique industrielle de nos travaux. Dans le deuxième, nous commençons par un positionnement épistémologique et méthodologique de nos travaux. Ensuite, nous positionnons nos contributions par rapport aux différentes disciplines auxquelles nous avons fait appel (i.e. Accidentologie, IC, ECD et conception).

-
- 2- La deuxième partie est consacrée à l'état de l'art. Du Chapitre 3 au Chapitre 6, nous présentons respectivement : des états de l'art sur la théorie du système général que nous utilisons dans nos travaux de modélisation, l'accidentologie, l'élaboration de modèles du domaine en IC et l'intégration des connaissances expertes dans le processus d'ECD.
 - 3- Dans la troisième partie, nous commençons dans le Chapitre 7 par l'application d'une méthode d'IC existant dans la littérature pour voir si elle répond à nos besoins en terme de modélisation des connaissances du domaine. La méthode choisie est CommonKADS. Nous montrons les limites de cette approche par rapport à nos besoins et nous développons dans le Chapitre 8 notre approche ASMEC de modélisation multi-vues de connaissances. Dans le Chapitre 9, nous développons notre deuxième approche AICEF et montrons comment elle utilise les métadonnées pour l'incorporation des connaissances du domaine dans le processus d'ECD. Nous consacrons le Chapitre 10 à la présentation de notre troisième approche ASAIC pour l'évaluation de l'impact de changement en sécurité routière. Etant donné qu'une partie essentielle de nos travaux porte sur l'élaboration de modèles du domaine, nous avons consacré le Chapitre 11 à mener une réflexion sur l'évaluation des ces modèles. Une approche d'évaluation de modèles, ASEM, est alors élaborée. Le dernier chapitre de ce mémoire est consacré à l'évaluation de notre approche globale SAFE-Next. Nous discutons ses différents apports en KM, en accidentologie, en IC, en ECD et en conception de nouveaux produits.
 - 4- La dernière partie conclue ce mémoire et ouvre de nouvelles perspectives.

Partie I

Présentation de la
problématique

Présentation de la partie I

Cette partie est consacrée à l'expression de notre problématique de recherche. Elle se décompose en deux chapitres (cf. Figure I.1) : le Chapitre 1 est dédié à la description du contexte industriel de nos travaux de recherche ainsi qu'aux enjeux sociaux, économiques et politiques. Dans le Chapitre 2, nous commençons par un positionnement épistémologique et méthodologique de nos travaux. Ensuite, nous effectuons un positionnement par rapport aux différentes problématiques en accidentologie, en conception, en extraction de connaissances de données et en ingénierie des connaissances.

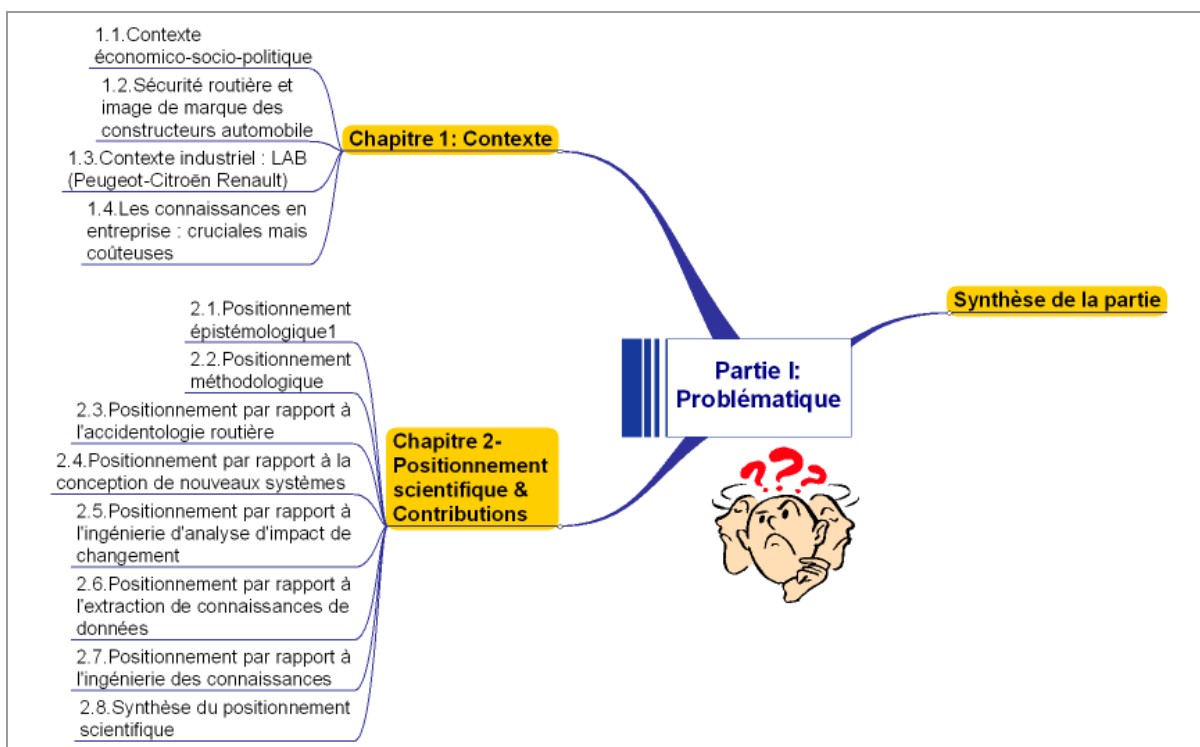


Figure I.1- Partie I : Présentation de la problématique.

Chapitre 1.

Contexte

1.1. Contexte économique-socio-politique

Le bilan annuel de victimes et de dégâts matériels fait de la sécurité routière en France (et dans le monde en général) un enjeu social, économique, technologique et politique :

- Un *enjeu humain* : les accidents de la route causent des souffrances humaines considérables : la famille, les amis et la communauté de chaque victime sont frappés de plein fouet par les conséquences physiques, psychologiques et économiques du décès, des blessures ou de l'invalidité de leur proche. Les rescapés et leurs familles subissent les conséquences souvent longues et douloureuses des traumatismes, des incapacités et de la réadaptation. Il n'est pas rare que la disparition du soutien de famille, les frais d'obsèques, le coût des soins et la perte de revenu consécutive à une incapacité plongent la famille dans la pauvreté [OMS, 2004] ¹;
- Un *enjeu social* : l'insécurité routière affecte la société sur plusieurs plans. En effet, elle se manifeste comme un problème de santé publique puisqu'elle cause la mort et des dégâts dramatiques sur les personnes. L'inconfort, le bruit et la pollution sont considérés aussi comme des défis à surmonter pour assurer la sécurité du public ;
- Un *enjeu économique* : les accidents de la route engendrent aussi des dégâts matériels et humains importants. Dans la plupart des pays, le coût des accidents de la route s'élève à 1% ou 2% du PNB [OMS, 2004]. Le coût tutélaire de la sécurité routière en 2002 est de 1 054 949 euros pour un tué, 158 243 euros pour un blessé grave, 23 209 euros pour un blessé léger et 5802 euros pour un accident matériel. Le coût global s'élève à 24,1 milliards d'euros ;

¹ OMS : Organisation Mondiale de la Santé.

- Un *enjeu technique* : l'enjeu social et l'enjeu économique imposent des contraintes contradictoires. Le compromis recherché est d'assurer la sécurité mais en respectant les contraintes économiques en minimisant les coûts. Le développement de nouvelles techniques et technologies paraît l'un des éléments essentiels pour assurer ce compromis en intégrant les objectifs de sécurité dès les phases de conception des véhicules et infrastructures. D'où l'apparition de nouveaux matériaux, de nouvelles structures et de nouveaux dispositifs de sécurité ;
- Un *enjeu politique* : face à l'enjeu socio-économique, la sécurité routière est devenue un enjeu politique : elle a été présente au débat politique lors des dernières élections présidentielles en France. En 2003, un projet de loi « à finalité préventive » a été voté au parlement pour lutter contre l'insécurité routière. De nouvelles mesures de prévention et de répression ont été prises (e.g. la suppression des « permis blancs », la création de nouvelles peines complémentaires, la mise en place d'une chaîne automatisée contrôle-sanction, la mise en place d'un permis probatoire).

1.2. Sécurité routière et image de marque des constructeurs automobile

Sur tous les marchés, y compris hors automobile, on observe le même mouvement : la marque contribue à donner une valeur ajoutée au produit, d'où le besoin des constructeurs automobiles de travailler sur leur marque même si l'image de leurs produits et leurs performances commerciales sont bonnes. Les leviers pour améliorer l'image de marque sont appelés les *fondamentaux* qui sont des points produit « *disqualifiants* ». En dessous d'un certain niveau (réel ou perçu) sur l'un ou l'autre des fondamentaux, une marque sort tout simplement du marché. La *qualité*, depuis les années 80, la *sécurité* depuis le salon de Francfort 95, et demain le *respect de l'environnement* sont des fondamentaux sur lesquels Renault, par exemple, se veut d'être sans reproche, car ils représentent un dû pour le client.

Les systèmes de sécurité embarqués dans les véhicules contribuent à améliorer l'un des points fondamentaux qui est la sécurité. Les constructeurs se considèrent comme un partenaire des pouvoirs publics partout dans le monde et veulent être un partenaire actif pour améliorer la sécurité routière. PSA Peugeot-Citroën et Renault ont pris dès les années soixante l'initiative d'une recherche approfondie afin d'améliorer la connaissance des mécanismes accidentels et lésionnels et proposer des contre-mesures. Selon Louis Schweitzer, PDG Renault, « *Faire progresser la sécurité des automobiles est pour Renault une priorité depuis plus de trente ans... Près de 100 millions d'euros sont ainsi investis chaque année dans la recherche et le développement afin que de nouveaux équipements issus des technologies les plus avancées viennent renforcer aussi bien la sécurité active que la sécurité passive de l'ensemble de notre gamme* ». Dans ce cadre aussi, des manifestations sont organisées : Le programme *Sécurité pour tous*, concours *Sécurité Sans Frontières*, participation à la *semaine nationale de sécurité routière*, formation interne à la sécurité routière, journées *porte ouverte sécurité routière*, etc.

1.3. Contexte industriel : LAB (Peugeot-Citroën Renault)

Conscients de l'enjeu de la sécurité, les deux constructeurs français, Renault et PSA Peugeot-Citroën, ont créé leurs propres structures internes et des structures communes pour développer des connaissances en accidentologie qui leur permettent d'améliorer la sécurité routière. L'ARP (Association Peugeot-Renault) est la première structure créée en 1969. Ensuite, dans les années 80, le LAB (Laboratoire d'accidentologie et de Biomécanique et d'étude du comportement humain) a été créé. Son objectif est d'apporter un retour d'expérience sur l'accident de la route vers les constructeurs pour développer et évaluer des prestations de sécurité.

Trois activités principales peuvent être distinguées au sein du LAB : (i) *L'accidentologie* pour identifier les situations vécues sur les routes et comprendre les accidents. Cette activité est basée essentiellement sur la construction et l'étude détaillée de bases de données d'accidents (cf. section 4.2). (ii) La *biomécanique* pour créer des connaissances sur le comportement biomécanique des occupants d'un véhicule. (iii) *L'étude de comportement humain* pour construire des connaissances sur le comportement physiologique et psychologique des occupants d'un véhicule [Le Coz et Page, 2003]. A travers ces trois activités, le LAB développe des outils et des critères permettant d'accroître la sécurité routière. Les travaux présentés dans ce mémoire s'intègrent dans l'activité de l'accidentologie. L'objectif est d'utiliser les EDA (Etudes Détaillées d'Accidents (cf. section 4.2)) pour développer des connaissances en accidentologie exploitables par les constructeurs automobiles pour développer et évaluer des prestations de sécurité.

1.4. Les connaissances en entreprise : cruciales mais coûteuses

Dans un contexte industriel à complexité croissante dans un environnement évolutif, les départs en retraite anticipée, les transferts de personnes et les suppressions d'effectifs d'un côté, les besoins de capitaliser les expériences et les savoir-faire pour une réutilisation future d'un autre côté, ont entraîné chez les firmes une prise de conscience de la valeur de leur capital immatériel et notamment de leur capital de connaissances. Ainsi, la connaissance est élément essentiel de la création de valeur. Elle représente, désormais, un capital économique immatériel, une ressource stratégique, un facteur de stabilité et un avantage concurrentiel décisif pour toute organisation. La gestion des connaissances au sein de l'entreprise devient alors un levier important pour améliorer sa performance. Il s'agit de l'identification, préservation, partage, capitalisation et création de nouvelles connaissances.

La *capitalisation* et la *création de nouvelles connaissances* en accidentologie sont fondamentales pour les constructeurs automobile afin de développer des systèmes de sécurité embarqués et améliorer de la sécurité routière en général. Le concept STA (cf. section 4.4), grâce à sa représentation synthétique et compréhensible, est utilisé comme l'un des supports de cette connaissance. Cependant, une grande expertise est nécessaire pour extraire ces connaissances à partir de données complexes et multi-sources. De plus, ces STA dépendent de l'objectif de l'étude et de la perception de l'expert. La combinaison des approches expertes et automatiques semble être une perspective intéressante pour construire ces STA.

Chapitre 2.

Positionnement scientifique et contributions

Dans ce chapitre, nous commençons par un positionnement épistémologique de nos travaux suivi d'un positionnement méthodologique. Ensuite, dans les sections suivantes, nous effectuons un positionnement par rapport aux disciplines aux quelles nous avons fait appel dans nos travaux et dans lesquels se situent nos apports : l'accidentologie, la conception de nouveaux systèmes, l'ingénierie d'analyse d'impact de changement, l'ECD et l'IC.

2.1. Positionnement épistémologique

Comme nous l'avons montré dans l'introduction et au premier chapitre, l'objectif de nos travaux est de développer des approches/méthodes et outils pour le développement de nouvelles connaissances en combinant des approches expertes et des approches automatiques. Il convient donc de se poser des questions relatives à la *nature*, l'*objectif*, les *sources* et la *validité* de ces connaissances. L'*épistémologie* est la branche de la philosophie qui traite ces questions relatives au « QUOI » de la connaissance. Un positionnement épistémologique de nos travaux s'impose donc au début de ce chapitre et c'est l'objectif de cette section.

Le *constructivisme* est le fondement épistémologique des travaux présentés dans ce mémoire. Le constructivisme reconnaît le caractère relatif de la connaissance et sa dépendance de la construction du sens par les individus en se basant sur leurs expériences. Un système dans une perspective constructiviste est défini comme une représentation de la réalité perçue par un certain nombre d'individus dans un contexte donné.

Le constructivisme est opposé au *positivisme* qui insiste sur le caractère absolu, objectif et stable de la connaissance et suppose que la science doit être limitée à ce qui est observable et mesurable. L'épistémologie positiviste est connue aussi sous le nom de « l'empirisme » ou « l'idéalisme platonicien ». Un système est défini dans une perspective positiviste comme un

ensemble d'entités existantes, réelles et stables avec des caractères universels qui peuvent être manipulés et mesurés.

Le constructivisme affirme qu'une modélisation ne dépend pas seulement du système modélisé, sinon chaque système a un modèle unique ce qui n'est pas le cas : pour un seul système, on peut avoir plusieurs modèles. La modélisation dépend de la nature du système à modéliser, de la discipline, de l'objectif de l'étude, du contexte de modélisation, etc. Cette conception multi-vues et contextuelle de la modélisation nous a paru adaptée pour :

- La modélisation de l'accident de la route : en effet la perception du phénomène de l'accident chez un accidentologue, par exemple, n'est pas la même que celle d'un psychologue, ni celle d'un cogniticien, ni celle d'un biomécanicien, ni celle d'un agent de police. Sa modélisation n'est pas unique même lorsqu'il s'agit d'une même discipline. Elle dépend, entre autres, de l'objectif de la modélisation, de l'expérience du modélisateur et de la discipline traitant le phénomène.
- L'intégration des connaissances du domaine dans le processus d'extraction de connaissances de données (ECD) : en effet, l'ECD est un processus interactif et itératif avec une forte implication de l'opérateur [Brachman et Anand, 1996]. L'intégration de l'objectif et du point de vue de l'utilisateur durant les différentes étapes du processus d'ECD est donc fondamentale.

Plus de détails sur le constructivisme, le positivisme ainsi que leurs origines sont présentés en état de l'art, Chapitre 3.

2.2. Positionnement méthodologique

Si l'épistémologie s'intéresse aux questions relatives au « QUOI » de la connaissance, une *méthodologie*, quant à elle, s'intéresse aux *pratiques* (méthodes) utilisées pour acquérir la connaissance. Autrement dit, elle concerne le « COMMENT » produire des connaissances, c'est-à-dire l'ensemble des procédures et règles qui permettent aux chercheurs de produire des connaissances.

Nos travaux se basent essentiellement sur la modélisation de *systèmes (ou objets) complexes* (accident, connaissances, etc.). Notre positionnement méthodologique se fait donc par rapport aux *méthodes de modélisation*. Toutes les épistémologies sont d'accord sur le fait qu'un *modèle* est l'interface entre un sujet et le monde réel pour construire la connaissance [Heylighen, 1993]. Cependant, ces épistémologies ont différentes conceptions de *ce qu'est un système et de la façon de le modéliser*. D'où l'existence de plusieurs *méthodologies* de modélisation.

Comme une grande partie de notre travail porte sur la modélisation de systèmes et de processus complexes (i.e. accident de la route et processus d'ECD), il convient de poser la question suivante : *comment construire ces modèles ?* En d'autres termes : *quelle méthodologie de modélisation choisir ?*

Le Moigne (1999) expose deux logiques de modélisation : une *logique disjonctive* (appelée aujourd'hui formelle) et une *logique conjonctive*. Selon la logique disjonctive l'observateur (e.g. le

modélisateur) est dissocié de *l'observation* (e.g. *la modélisation*) et de *l'objet observé* (e.g. *le système à modéliser*). Ainsi, un modèle est unique, puisqu'il ne dépend que du système modélisé, qui est unique. Le modélisateur ne fait que recevoir les données, et puisque ces données sont réelles, stables et universelles, on peut les représenter par un modèle déterministe. Cette logique prend ses fondements dans *l'épistémologie positiviste* (cf. section 2.1 et 3.3).

En ce qui concerne l'accident de la route et la tâche de conduite en général, il s'agit d'un comportement qui fonctionne et se transforme en même temps dans un environnement et contexte qui sont évolutifs. L'accident ne peut pas non plus être modélisé indépendamment de l'objectif (ou téléologie) de la modélisation. Plusieurs acteurs peuvent être impliqués quand il y a un accident de la route (e.g. un expert véhicule, un expert infrastructure, les concepteurs des systèmes, etc.). Chacun de ces acteurs a une perception de l'accident qui est différente de celle des autres.

Il est donc clair qu'un modèle « fermé » est inadapté pour appréhender la complexité du système CVE et son évolutivité car la complexité implique nécessairement une forme d'imprévisibilité. Il faut des modèles qui font la conjonction entre le système (structure, fonctions, évolutions, etc.), son environnement et les objectifs de la modélisation.

C'est la *logique de modélisation conjonctive* que nous adoptons dans ce travail. Cette logique prend ses fondements dans l'épistémologie constructiviste (cf. section 3.3) et elle a donné naissance à la *méthode systémique* (appelée aussi *systémographie*) pour la modélisation de ces systèmes. L'approche systémique [Le Moigne, 1974; Morin et Le Moigne, 1999] est connue dans la littérature anglo-saxonne sous le nom de la *cybernétique* [Ashby, 1965; Wiener, 1948] ou de la *théorie du système général* [Bertalanffy, 1969]. Une présentation détaillée de cette approche est donnée dans la section 3.4.

2.3. Positionnement par rapport à l'accidentologie routière

L'accidentologie est l'un des leviers pour améliorer la sécurité routière.

Définition 2.1- Accidentologie

Il s'agit des études qui ont pour objectif de comprendre l'étiologie de l'accident et des lésions, afin de proposer des contre-mesures pour diminuer le nombre des accidents de la route et les lésions corporelles conséquentes et les évaluer [Foret-Bruno et al., 2001].

Deux approches de l'accidentologie peuvent être distinguées [Le Coz et Page, 2003; Megherbi, 1999] :

Définition 2.2- Une approche macroscopique ou épidémiologique : la macro-accidentologie

Elle a des fondements statistiques et un aspect évaluatif. Elle s'intéresse à l'accident dans son aspect global telles que les partitions spatio-temporelles des accidents et des risques pour identifier des enjeux et des cibles principales pour mener des actions de prévention. Elle permet également d'évaluer l'efficacité de certaines aides à la conduite en terme d'accidents évités ou de réduction de leur gravité.

Définition 2.3- Une approche microscopique ou clinique : la micro-accidentologie

Elle a des fondements cliniques. Elle se base sur des analyses au cas par cas d'un échantillon limité d'accidents. Elle vise à comprendre et évaluer les mécanismes accidentels. Ces études sont généralement coûteuses puisqu'elles sont menées par des experts qui se basent sur des études détaillées de cas d'accidents.

Dans [Page et al., 2004], les auteurs proposent une approche *méso-accidentologique* : à mi-chemin entre l'analyse au cas par cas (i.e. micro) et l'analyse statistique agrégée (i.e. macro).

Le concept STA est l'un des outils développés et utilisés dans l'approche clinique. Cependant la construction de ces STA requiert une importante expertise du domaine. Ceci rend leur construction par une *approche purement experte* très coûteuse. De même, la complexité de l'accident (cf. section 3.5) et du processus de construction des STA écarte la possibilité d'une *construction purement automatique*. Nos travaux se situent donc dans la micro-accidentologie et proposent une méthodologie qui combine les deux approches : *experte* et *automatisée* pour l'élaboration de STA.

2.4. Positionnement par rapport à la conception de nouveaux systèmes

L'analyse de *l'environnement du fonctionnement* d'un nouveau système est une tâche fondamentale dans son processus de conception. En effet, elle permet la compréhension de son interaction avec son environnement afin de : définir les fonctionnalités du nouveau système, anticiper ses impacts sur son environnement (et inversement) pour éviter des effets pervers (évaluation a priori) et évaluer son efficacité après son développement (évaluation a posteriori).

C'est typiquement le cas quand il s'agit de développer un nouveau système de sécurité embarqué dans les véhicules. Ceci requiert des connaissances sur le processus de conduite en général et en accidentologie en particulier. Ces connaissances permettent de (i) détecter et de comprendre des problèmes de sécurité qui vont être utilisés comme des « entrées » dans le processus de conception ; (ii) anticiper l'impact du nouveau système sur le comportement du système CVE et (iii) évaluer les systèmes qui sont déjà utilisés (e.g. ABS¹, ESP²).

Sans ces connaissances, des méthodes classiques d'aide à la conception comme l'analyse fonctionnelle (externe et interne), la méthode APTE³ et la QFD (Query Function Deployment) ou des méthodes d'aide à l'innovation comme TRIZ (Theory of Inventive Problem Solving) ne peuvent pas être appliquées. Ces méthodes permettent la représentation et la structuration des fonctionnalités du nouveau système, donc elles supposent une connaissance préalable de ces fonctionnalités. Ainsi, elles sont des méthodes de *structuration et représentation de l'espace de conception*. Mais, ce dont on a besoin dans notre cas ce sont des méthodes de *construction de cet espace de conception*.

Ce que nous entendons par « *la construction de l'espace de conception* » des systèmes de sécurité est : la définition de *fonctions de sécurité* (e.g. les contre mesures) à réaliser à travers des solutions techniques ainsi que la spécification des contraintes à respecter lors de la conception des SSE. Cela nécessite

¹ ABS : Antiblockier system : système antiblocage des freins.

² ESP : Electronic Stability Program : Système de Contrôle de Stabilité.

³ APTE : méthode d'analyse fonctionnelle et d'analyse de la valeur.

la compréhension des mécanismes accidentels, afin de proposer les contre-mesures adéquates. Etant donné que les STA forment un outil efficace pour la compréhension de l'accident, ils peuvent alors être utilisés comme un langage compréhensible entre les accidentologues et les concepteurs pour la définition de ces fonctions de sécurité. Nous développons donc un outil d'aide à la construction de l'espace de conception des SSE.

2.5. Positionnement par rapport à l'ingénierie d'analyse d'impact de changement

Dans un environnement qui ne cesse d'évoluer, l'anticipation des changements et de leurs impacts est fondamentale pour qu'un système (e.g. une organisation, une entreprise) survive. La sécurité routière fait intervenir plusieurs systèmes qui sont évolutifs et en interaction continue. Cette évolutivité est l'une des sources de complexité du domaine. L'analyse des changements (au niveau de chacun des systèmes) et de leurs impacts sur les autres systèmes s'avère donc cruciale pour répondre aux besoins en terme de sécurité routière. Nous assimilons le domaine de la sécurité routière à un système constitué lui-même d'autres sous-systèmes. Nous nous limitons pour le besoin de notre étude aux trois sous-systèmes suivants : les *constructeurs automobiles*, l'*accidentologie* (le LAB) et le *système CVE*.

Prenons l'exemple de l'introduction d'un nouveau système de sécurité dans le véhicule. Il s'agit d'un changement par rapport au produit « véhicule ». Ce changement a une influence très probable sur le *comportement du système CVE*. En effet, le comportement d'un conducteur avec un véhicule équipé peut changer (e.g. plus de prise risques dans sa conduite). Ce changement implique de nouvelles contraintes en terme de sécurité dont il faut tenir compte dans le *processus de développement* du véhicule chez les constructeurs automobile. Ce changement peut affecter aussi le métier de l'accidentologie lui-même. Le LAB, par exemple, doit anticiper les besoins et développer de nouvelles méthodes et techniques pour l'évaluation et la validation du système (e.g. des nouvelles variables à définir et collecter, des nouvelles études à mener). Cela peut même affecter l'organisation du LAB (nouvelles ressources mobilisés, nouvelles antennes de collecte, nouvelles organisation de travail, etc.).

L'Ingénierie d'Analyse d'Impact de Changement (IAIC) a été développée pour répondre au besoin des entreprises qui veulent faire face à l'évolutivité de l'environnement dans lequel elles existent. Plusieurs types de changement peuvent affecter un système et à plusieurs niveaux : sa structure, ses fonctionnalités, son processus de développement, etc. [Pikosz et Malmqvist, 1998]. Parmi les techniques existantes dans la littérature et permettant d'analyser l'impact d'un changement, nous citons le *Design Rationale* [Bose, 1998], la représentation orientée objet du produit ou du processus [Gorti et al., 1998], etc. (cf. section 10.3 pour un état de l'art sur les différentes approches).

ASAIC (Approche Systémique d'Analyse d'Impact de Changement) (cf. Chapitre 10) est une approche que nous développons dans SAFE-Next pour l'analyse d'un changement en sécurité routière.

2.6. Positionnement par rapport à l'extraction de connaissances de données

L'accidentologie représente un retour d'expérience sur le phénomène d'accident de la route vers les constructeurs automobiles. La construction de bases de données d'accidents (les EDA) et l'étude de ces bases afin d'en extraire des connaissances sont parmi les moyens assurant ce retour d'expérience.

Définition 2.4- Processus d'Extraction de Connaissances de Données (ECD)

Le processus d'ECD ou processus de fouille de données est défini comme : « le processus d'identification de patrons (patterns) valables, nouveaux, potentiellement utiles et compréhensibles dans les données » [Fayyad et al., 1996]. Dans notre cas, les STA sont les supports de ces patrons.

Plusieurs techniques d'extraction de connaissances existent (e.g. classification automatique, classement, règles d'association). Mais, nous nous intéressons uniquement aux problématiques de classification automatique (e.g. création de groupes d'individus similaires d'un point de vue donné (cf. section 4.4.2)). En effet, un STA est, par définition, un groupe d'accidents présentant des similarités. La construction de ces STA peut donc faire appel à des méthodes de classification.

Le processus d'ECD (cf. section 6.1) est interactif et itératif, impliquant de nombreuses étapes avec des décisions prises par l'utilisateur. Trois étapes principales peuvent être distinguées : préparation de données, application de techniques de data mining (classification dans notre cas) et interprétation des résultats. L'implication de l'utilisateur est plus importante dans la première et la dernière phase du processus d'ECD.

Nous avons identifié très peu de travaux génériques dans la littérature sur la problématique d'incorporation de connaissances expertes dans le processus d'ECD (cf. Chapitre 6). Pour pallier ce déficit, nous développons AICEF : une Approche d'Incorporation des Connaissances Expertes dans la Fouille de données (cf. Chapitre 9).

En ce qui concerne la *préparation des données* (voir état de l'art section 6.2), nous nous intéressons à la sélection des attributs. La problématique est formulée ainsi : *Comment décliner l'objectif d'une étude donnée en attributs d'étude ?* Les variables utilisées pour l'étude d'une perte de contrôle, par exemple, sont différentes de celles utilisées pour l'étude de freinage. Même si l'objectif de l'étude est le même, les variables sélectionnées peuvent être différentes selon le point de vue de celui qui va effectuer l'étude (e.g. psychologue, biomécanicien, mécanicien). L'approche AICEF permet d'aider l'utilisateur dans la première phase d'ECD en lui fournissant une sélection multi-points de vue des attributs permettant d'utiliser les connaissances du domaine.

En ce qui concerne *l'interprétation des classes d'accidents* (voir état de l'art section 6.3.2), la problématique est formulée ainsi : *Comment trouver une description des classes avec un bon niveau d'abstraction, compréhensible et par les accidentologues et par les concepteurs ?* Les connaissances du domaine et donc l'implication de l'utilisateur sont fondamentales dans cette phase aussi. En effet, même si les algorithmes de classification utilisés sont performants, on peut obtenir des résultats difficilement exploitables (i.e. difficilement interprétables et utilisables) par les utilisateurs finaux. De plus, le même résultat peut être interprété différemment selon le point de vue de l'utilisateur.

Pour cela, notre approche AICEF a pour objectif l'identification et la formalisation des connaissances du domaine et leur utilisation dans la phase d'interprétation des résultats.

2.7. Positionnement par rapport à l'ingénierie des connaissances

L'ingénierie des connaissances a été développée dans le domaine de l'Intelligence Artificielle (IA) et elle contient plusieurs champs de recherche en systèmes d'information, modélisation par objet, travail coopératif, Interface Homme/Machine (IHM), Raisonnement à Partir de Cas (RàPC), etc.

Définition 2.5- Ingénierie des connaissances (IC)

C'est l'étude des concepts, méthodes et techniques permettant de modéliser et/ou d'acquérir les connaissances dans des domaines se formalisant a priori, peu ou pas [Charlet et al., 2000].

L'identification, la modélisation, l'explicitation, l'opérationnalisation et l'utilisation des connaissances sont parmi les activités centrales en ingénierie des connaissances. Dans nos travaux nous nous intéressons à l'élaboration de *modèles du domaine*.

Définition 2.6- Modèle du domaine

Il est défini selon CommonKADS [Wielinga et al., 1994] comme « *une structuration des ontologies du domaine (concepts, relations entre les concepts, termes, etc.) selon un point de vue donné* ».

La *notion de point de vue* est donc centrale dans l'élaboration des modèles du domaine. Cependant, pour l'identification de ces points de vue, la majorité des études dans la littérature se base essentiellement sur une étape d'*élicitation* (i.e. collecte de données, entretiens) suivie d'une étape de conceptualisation. Cette *approche ascendante (Data-Driven)* présente des limites que nous détaillons dans les paragraphes 5.3.1.3 et 5.3.3.2 (coûteuse, risque d'incomplétude en terme de points de vue, etc.). D'autres études utilisent des *approches descendantes (Model-Driven)* qui commencent par la sélection, à partir d'une bibliothèque de modèles génériques comme celle de CommonKADS, du modèle le plus proche de l'application. Cette approche présente aussi des limites (e.g. difficulté de trouver un modèle générique) qui ont amené au développement d'approches mixtes.

L'approche que nous développons dans ces travaux, ASMEC (pour Approche Systémique de Modélisation d'Es Connaissances), est mixte (Chapitre 8). De plus, grâce à l'utilisation de l'approche systémique, elle permet à l'utilisateur de développer des *modèles multi-vues*.

Nous nous intéressons aussi à la problématique d'évaluation des modèles élaborés, une problématique très peu traitée en littérature (cf. Chapitre 11). L'*évolutivité* du modèle, son *invariance*, sa *contrôlabilité*, sa *généralité*, sa *ré-utilisabilité*, etc. (voir définitions dans section 11.3) sont des questions dont il faut tenir compte durant le processus de modélisation. Nous développons alors un *modèle d'évaluation de modèles*, que nous intitulons ASEM (cf. Chapitre 11). Dans ce modèle, nous définissons des critères d'évaluation qui sont à utiliser comme des *contraintes de modélisation* durant le processus de modélisation. Ils sont utilisés aussi comme des critères d'évaluation a posteriori une fois le modèle réalisé ou pour choisir entre des modèles existants.

2.8. Synthèse du positionnement scientifique

Dans ce chapitre, nous avons fait un positionnement scientifique par rapport aux différentes disciplines auxquelles nous avons fait appel dans nos travaux. Nous avons choisi le constructivisme comme fondement épistémologique pour notre recherche. Le constructivisme reconnaît le caractère relatif de la connaissance et sa dépendance des individus et du contexte. Il nous a donc paru adapté à notre problématique de construction de nouvelles connaissances dans un contexte multidisciplinaire et évolutif qu'est l'accidentologie.

Sur le plan *méthodologique*, nous avons choisi la systémique. Elle est adaptée à la modélisation des systèmes et des processus complexes tels que l'accident de la route et le raisonnement des experts lors de la résolution d'une tâche, notamment lors de la construction de STA.

Par rapport à *l'accidentologie*, nos travaux se situent dans la micro-accidentologie qui a un fondement clinique pour étudier et comprendre l'accident de la route. Le concept STA est l'un des outils de cette approche clinique.

Par rapport à la *conception de nouveaux systèmes*, nos travaux se positionnent dans la phase de préconception. En effet, le concept scénario-type d'accidents offre un langage compréhensible par l'accidentologue et le concepteur des systèmes de sécurité pour définir des contre-mesures. Il s'agit en d'autres termes d'un outil d'aide à la construction d'un *espace de conception*.

Par rapport à l'ingénierie d'analyse d'impact de changement, nous proposons ASAIC qui permet une analyse multi-vues de cet impact.

Par rapport à *l'ingénierie des connaissances*, nous proposons une nouvelle approche, ASMEC, permettant une modélisation multi-vues des connaissances en combinant des approches ascendantes et des approches descendantes. Nous proposons aussi ASEM, une approche qui permet l'évaluation de modèles de connaissances.

Par rapport aux techniques *d'extraction de connaissances de données*, nous proposons AICEF, une méthodologie qui permet la modélisation des connaissances du domaine et leur incorporation dans le processus d'ECD.

Synthèse de la partie I

L'objectif de nos travaux peut se résumer comme suit : *développer une méthode permettant l'élaboration de connaissances en accidentologie sous forme de scénarios-types d'accidents exploitables pour le développement des systèmes de sécurité embarqués (SSE)*. Cette méthode doit tenir compte des points suivants :

- La complexité de l'accident de la route et donc le besoin d'analyse multi-points de vue nécessitant une grande expertise du domaine ;
- Le rôle fondamental des connaissances expertes pour l'élaboration des STA, ce qui rend difficile l'automatisation de cette construction ;
- Le coût élevé d'une élaboration purement experte, ce qui la rend insatisfaisante ;
- La nature complexe des connaissances à construire ;
- La dépendance des connaissances des experts, du contexte de leur développement et du contexte de leur utilisation ;
- Les STA doivent fournir un langage entre les accidentologues et les concepteurs des SSE facilitant leur communication pour définir de nouvelles prestations de sécurité.

Pour répondre à ces différents points, nous proposons *une méthodologie de construction de STA combinant des approches expertes et automatisées*. Pour ce faire, nous utilisons des approches d'ingénierie des connaissances et d'extraction de connaissances de données pour :

- Développer une approche de modélisation multi-vues des connaissances du domaine et leur formalisation (ASMEC) ;
- Incorporer ces connaissances dans le processus d'ECD (AICEF) ;
- Incorporer ces connaissances dans le processus d'analyse d'impact du changement (ASAIC).

Partie II
Etat de l'art

Présentation de la partie II

Cette partie est consacrée à l'état de l'art concernant les problématiques de nos travaux (cf. Figure II.1) : *modélisation des systèmes complexes* (Chapitre 3), *création de connaissances en accidentologie* (Chapitre 4), *modélisation des connaissances en IC* (Chapitre 5) et *incorporation des connaissances du domaine dans le processus d'ECD* (Chapitre 6). Pour des raisons de clarté, nous avons préféré la présentation de l'état de l'art sur l'ingénierie d'analyse d'impact du changement dans le chapitre consacré au développement d'ASAIC. Notre état de l'art se décompose donc en quatre chapitres :

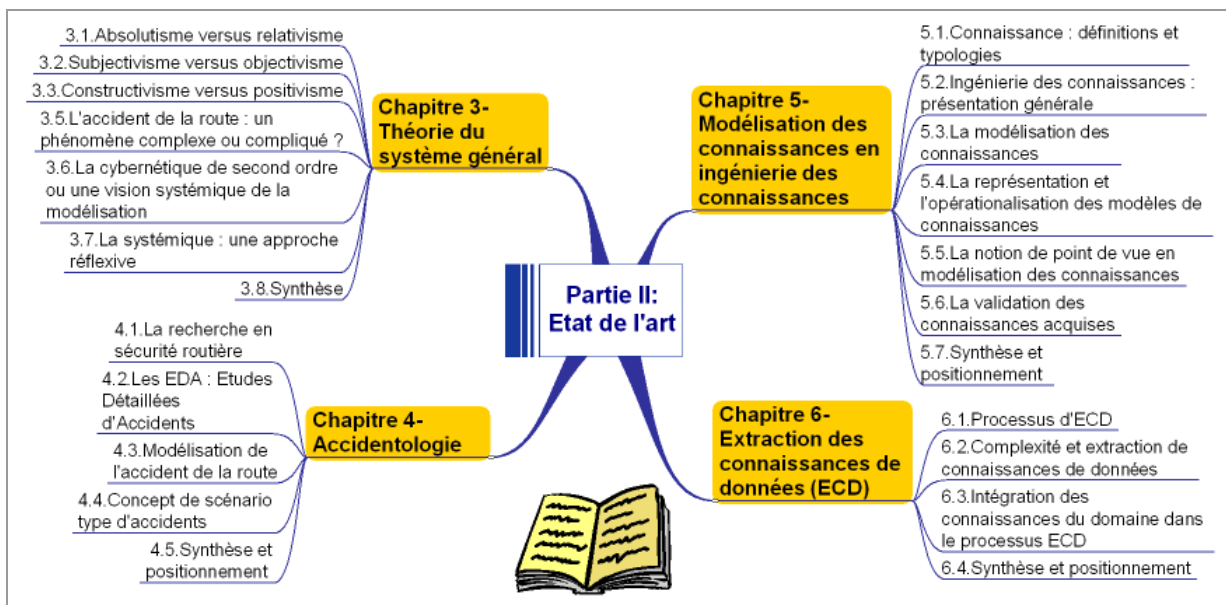


Figure II.1- Partie II : Etat de l'art.

Le Chapitre 3 est consacré à une présentation des origines épistémologiques et philosophiques de l'approche systémique que nous utilisons dans nos travaux de modélisation de systèmes complexes. Il retrace l'historique, selon notre point de vue, de cette approche et les différents axiomes sur lesquels elle est fondée.

Le Chapitre 4 se focalise sur les différentes recherches dans le domaine de l'accidentologie : les approches et modèles d'analyse de l'accident et le concept scénario-type d'accidents ainsi que les différentes approches utilisées pour les construire.

Le Chapitre 5 propose un état de l'art sur la problématique de modélisation des connaissances en ingénierie des connaissances. Cet état de l'art couvre différents aspects relatifs à la connaissance qui sont en rapport avec notre problématique d'incorporation des connaissances dans le processus d'ECD. Définitions des connaissances, typologies des connaissances, différentes approches de modélisation des connaissances, représentation et opérationnalisation des connaissances et évaluation des connaissances sont les différents thèmes abordés dans ce chapitre.

Le Chapitre 6 se focalise sur la problématique d'extraction de connaissances de données. Après une présentation générale du processus d'ECD, nous montrons les sources de complexité de ce processus. Ensuite, nous nous focalisons sur les aspects en rapport avec notre problématique.

Un positionnement de nos contributions par rapport à la littérature est effectué dans chacun de ces chapitres.

Chapitre 3.

Cybernétique & théorie du système général

La *modélisation* peut être perçue comme une méthodologie de construction de connaissances. Selon [Heylighen, 1993], la question fondamentale à laquelle la modélisation doit répondre est : « *Comment un environnement complexe peut être représenté par un modèle qui est nécessairement plus simple et qui permet à un sujet (i.e. un utilisateur) de construire des connaissances pour prédire le comportement de cet environnement ?* ».

Deux courants philosophiques, *l'absolutisme* et le *relativisme* (cf. section 3.1), ont influencé les épistémologies existantes ainsi que les méthodes de modélisation. Cette dualité est présente en économie entre le *courant objectiviste* et le *courant subjectiviste* (cf. section 3.2) et en épistémologie entre le *constructivisme* et le *positivisme* (cf. section 3.3).

Dans nos travaux, nous utilisons la *systemique* (ou la *cybernétique*) (cf. section 3.4) comme méthode de modélisation des connaissances. Elle est issue de l'épistémologie *constructiviste* (cf. section 3.3) influencée par le courant *absolutiste* en philosophie (cf. section 3.1) et le courant *subjectiviste* en économie (cf. section 3.2). Nous commençons par retracer, d'une manière simplifiée, les origines de la systemique, telle que nous les percevons.

3.1. Absolutisme versus relativisme

Pour illustrer la différence entre ces deux courants antagonistes en philosophie, nous avons choisi de présenter la différence de leur perception de la notion de « *valeur* » analysée par [Compte-Sponville, 1998].

La première conception est celle d'Aristote et Platon. C'est une conception *absolutiste*, *objectiviste* et *ontologique* selon laquelle la valeur est intrinsèque à l'être, on parle alors de quelque chose qui vaut absolument, indépendamment du sujet qui la vise et indépendamment du contexte.

Donc, c’est la valeur d’une chose qui génère le désir de cette chose. D’après Aristote, « *Nous désirons une chose parce qu’elle nous semble bonne, plutôt qu’elle ne nous semble bonne parce que nous la désirons* » (lu dans [Compte-Sponville, 1998]). La valeur d’un objet est donc indépendante du contexte. Ce courant a influencé l’objectivisme (cf. section 3.2) en économie et le positivisme en épistémologie (cf. section 3.3).

La deuxième conception c’est celle de Spinoza. Elle est *relativiste, subjectiviste ou anthropologique*. Selon celle-ci, « *l’être est sans valeur intrinsèque comme la valeur est sans être propre ; il n’est de valeur que pour l’homme et par le désir* », écrit Spinoza (lu dans [Compte-Sponville, 1998]). Donc la valeur est soumise au désir qui est un élément du contexte. Un objet peut voir sa valeur changer selon le contexte. D’après Spinoza, « *Nous ne nous efforçons à rien, ne voulons, n’appétons ni ne désirons aucune chose, parce que nous la jugeons bonne ; mais au contraire, nous jugeons qu’une chose est bonne parce que nous nous efforçons vers elle, la voulons, appétons et désirons* ». La valeur d’un objet dépend donc du contexte dans lequel le sujet et l’objet se trouvent. Ce courant a influencé le subjectivisme (cf. section 3.2) en économie et le constructivisme en épistémologie (cf. section 3.3).

La dualité du débat entre les deux perceptions philosophiques s’est étendue à travers un débat entre économistes subjectivistes et objectivistes.

3.2. Subjectivisme versus objectivisme

L’*objectivisme* en économie s’est manifesté dans les travaux des économistes classiques (Smith, Ricardo, Marx...) qui font la différence entre *valeur d’échange* et *valeur d’usage*. Pour eux, « *tout objet n’est en effet fabriqué que s’il est susceptible d’être utile, c’est-à-dire propre à satisfaire un besoin déterminé : c’est précisément ce qui lui confère une valeur d’usage* » [Marx, 1969]¹. Ces valeurs d’usage sont toutes différentes et donc non comparables entre-elles : on ne peut pas échanger des valeurs d’usage contre des valeurs d’usage car l’échange implique comparaison, possibilité de trouver un équivalent entre des produits. Il s’agit donc de trouver ce qui permet de comparer diverses marchandises aux valeurs d’usages différentes, d’où la notion de *valeur d’échange*. Cette dernière permet la comparaison et donc d’échanger les marchandises. Les valeurs d’échange des marchandises doivent être ramenées à quelque chose qui leur est commun. Smith définit cette chose comme la *quantité de travail* nécessaire pour produire la marchandise. Selon lui, *il est naturel que ce qui est ordinairement le produit de deux jours ou de deux heures de travail vaille le double de ce qui est le produit de ce qui est ordinairement le produit d’un jour ou d’une heure de travail* » [Smith, 1776]. Cette valeur est indépendante du contexte.

A partir du début des années 1870, l’approche objective de l’école classique était substituée par l’*approche subjective* qui a caractérisé le courant néoclassique ou la révolution marginaliste. L’approche subjective prend ses racines du concept de la *valeur-utilité* qui a été fondée par Condillac [Condillac, 1776] : la valeur d’un objet dépend de son utilité. Adam Smith a ensuite soulevé le *paradoxe de l’eau et du diamant* pour intégrer la notion de *rareté* dans le jugement de valeur : l’eau a une faible valeur marchande par rapport au diamant malgré qu’elle soit plus utile. Les marginalistes

¹ Edité par Eleanor Marx Aveling, International Publishers, New York, 1969

ont donc une théorie dualiste qui combine la *rareté* et l'*utilité* comme sources de valeur. Contrairement à l'approche objective pour qui la valeur préexiste à l'échange, pour les marginalistes, la valeur d'échange se détermine sur le marché « *l'offre n'est qu'une conséquence de la demande* » [Walras, 1874]. Dès lors que la valeur dépend du désir du consommateur, elle n'est plus stable, ni relative à un temps de travail socialement nécessaire, ni intrinsèque à l'objet et elle n'est plus absolue. Elle est variable, relative et dépend du sujet et du contexte de l'échange. Il s'agit d'une *valeur-désir* au lieu d'une *valeur-travail*. Un même objet peut voir sa valeur augmenter ou diminuer selon que le désir du sujet qui en a besoin augmente ou diminue, c'est donc le contexte qui décide de la valeur. (voir [Ben Ahmed et Yannou, 2003] pour plus de détails)

3.3. Constructivisme versus positivisme

En psychologie, le *relativisme* (cf. section 3.1) et le *subjectivisme* (cf. section 3.2) se sont manifestés à travers le *constructivisme* de Jean Piaget (1896-1980). Il a développé l'*épistémologie génétique* [Piaget, 1970] (génétique ici a le sens de la genèse de structures cognitives et non celui de l'hérédité) appelé aussi *constructivisme génétique*. Il a décrit les différentes étapes par lesquelles passe un enfant en construisant un modèle du monde réel dans des situations de résolution de problème. La thèse centrale de l'épistémologie génétique est que *toute connaissance est le résultat d'une expérience individuelle d'apprentissage qui s'effectue à travers des processus d'accommodation et d'assimilation*. Les modèles psychologiques du développement cognitif, centrés sur des mécanismes individuels, ont été mis en cause ensuite pour laisser la place à une approche psycho-sociale des activités cognitives, le *socio-constructivisme*. Cette approche insiste davantage sur les dimensions sociales dans la formation des compétences et des connaissances. L'idée fondamentale du socio-constructivisme est qu'il est nécessaire de passer d'une psychologie « *binnaire* » (interaction individu-tâche) à une psychologie « *ternaire* » interaction individu-tâche-alter. Le développement ne peut plus être considéré comme indépendant de l'apprentissage, et l'apprentissage ne peut pas être seulement une relation « *privée* » entre un *enfant* et un *objet*.

Le constructivisme a donné naissance à ce qu'on appelle l'*épistémologie constructiviste* pour répondre à des questions relatives à la nature, à l'objectif, aux sources et à la validité des connaissances. L'épistémologie constructiviste *reconnaît le caractère relatif de la connaissance et sa dépendance de la construction du sens par les individus en se basant sur leurs expériences et leurs interactions avec leur environnement (contexte)*. Cette conception est proche du *relativisme* (cf. section 3.1) en philosophie et du *subjectivisme* (cf. section 3.2) en économie.

Un *système* dans une perspective constructiviste est défini comme une représentation de la réalité perçue par un certain nombre d'individus dans un contexte donné. Un *modèle* est donc une représentation de la réalité et il n'est valide que dans un contexte donné. Les *problèmes* ne sont pas indépendants des perceptions des individus qui les traitent. Il existe alors plusieurs *solutions* et la solution optimale est celle qui est acceptable pour la majorité. Les *méthodes de recherche* ainsi que leurs résultats reflètent donc la perception et l'interprétation des chercheurs. Les *sources d'information* doivent être diversifiées pour couvrir la diversité des opinions. En ce qui concerne les *données*, on accorde plus d'importance au processus de leur collecte qu'aux données elles-mêmes (une synthèse de ces principes peut être trouvée dans [Richard, 2003]).

Une interprétation extrême du constructivisme peut conduire au *relativisme absolu* ou ce qu’on appelle *solipsisme* selon quoi tout est subjectif et la réalité n’existe pas. Selon le solipsisme, il n’y a aucune différence entre la perception et l’hallucination, entre le rêve et l’imagination. Le « danger » de ce relativisme extrême est qu’on ne peut pas choisir entre des modèles différents, tous les modèles sont considérés comme ayant la même pertinence quel que soit le contexte.

L’épistémologie constructiviste adoptée dans nos travaux est celle qui reconnaît la relativité des connaissances et de leur dépendance du contexte, mais qui reconnaît en même temps la possibilité d’évaluer cette connaissance. Nous partons de ce principe pour développer, dans le Chapitre 11, un *modèle d’évaluation de modèles* que nous avons appelé ASEM.

Le constructivisme est opposé au *positivisme* qui insiste sur le caractère absolu, objectif et stable de la connaissance et suppose que la science doit être limitée à ce qui est observable et mesurable. L’épistémologie positiviste est connue aussi sous le nom de « l’empirisme » ou « l’idéalisme platonicien » ou « absolutisme » (cf. section 3.1). Un *système* est défini dans une perspective positiviste comme un ensemble d’entités existantes, réelles et stables avec des caractères universels qui peuvent être manipulés et mesurés. Un *modèle* ne dépend donc pas du contexte de modélisation, mais exclusivement du système. Les *problèmes* sont perçus dans cette perspective comme objectifs, indépendamment de la perception des individus. Leurs *solutions* sont apportées par ceux qui sont capables de connaître « l’idéal » qui sont les « experts ». Les *méthodes de recherche* sont orientées vers la collecte de données « réelles » et « objectives ». Le rôle d’un chercheur, qui est nécessairement un expert, est d’identifier le « problème objectif » et de proposer la « solution idéale » (une synthèse de ces principes peut être trouvée dans [Richard, 2003]).

3.4. L’approche systémique versus l’approche analytique

3.4.1 Apparition et évolution de la cybernétique

L’épistémologie constructiviste était à la base de l’apparition d’une nouvelle théorie dans les années quarante qui est la *cybernétique* ou la *science du contrôle et des communications dans l’homme, l’animal et la machine*.

Définition 3.1- Cybernétique

Ce terme est d’origine grecque, kybernetes, qui signifie « Art de gouverner un navire ». Selon [Heylighen et Joslyn, 2001], il est apparu dans l’antiquité avec Platon et réutilisé au 19^{ème} siècle par Ampère pour signifier science de gouvernement. Sa formalisation est attribuée au mathématicien américain Norbert Wiener en 1948 [Wiener, 1948].

La cybernétique distingue les systèmes complexes et les systèmes compliqués, « la complexité n’est pas la complication », nous dit Edgar Morin [Le Moigne, 1999; Morin et Le Moigne, 1999].

Définition 3.2- Système compliqué

C'est un système qui est caractérisé par un comportement qui peut être prévu par l'analyse des interactions entre ses composantes, il est déterministe (e.g. un ordinateur). Les approches analytiques issues de l'épistémologie positiviste (cf. section 2.2 et 3.3) sont adaptées à la modélisation de ce type de système.

Définition 3.3- Systèmes complexes ou système cybernétique

C'est un système non-déterministe dont le comportement ne peut pas être prévisible par analyse disjonctive de ses différents éléments. Un système complexe selon Miller [Miller, 1995] est un système vivant, évolutif et ouvert à son environnement avec lequel il est en interaction continue. C'est donc un système qui fonctionne et se transforme en même temps. Son comportement peut être décrit en terme de feedbacks et boucles récursives [Le Moigne, 1999]. Ces boucles peuvent être négatives (i.e. évoluant vers la stabilité) ou positives (i.e. évoluant vers l'amplification). Ces boucles récursives sont aussi responsables de l'auto-production et l'auto-organisation du système [Morin et Le Moigne, 1999].

L'approche cybernétique ne se focalise pas sur la composition matérielle d'un système, mais insiste les interactions entre les composants. Les éléments d'un système sont en interaction réciproque. L'action d'un élément sur un autre entraîne en retour une réponse du second élément vers le premier. On dit alors que ces deux éléments sont reliés par une boucle de feedback (ou boucle de rétroaction). Les interactions entre les éléments d'un système sont régies selon le principe du holisme.

Définition 3.4- Holisme

« Le tout est supérieur à la somme des parties » : Les interactions entre les éléments d'un système donnent à l'ensemble des propriétés que ne possèdent pas les éléments pris séparément.

Dans les années soixante-dix, cette vision de la cybernétique plus ou moins mécanique a été critiquée car elle assure qu'un système peut être observé et manipulé indépendamment de son observateur. De nouveaux concepts ont été alors développés : l'autonomie, l'auto-organisation, la cognition et surtout le rôle de l'observateur dans la modélisation du système. L'observateur et l'objet observé ne sont plus alors séparés et le résultat de l'observation dépend de leur interaction [Heylighen et Joslyn, 2001]. Il y a eu reconnaissance du fait que toutes nos connaissances sur les systèmes sont basées sur des représentations simplifiées (i.e. des modèles). Ces dernières ignorent les propriétés du système qui ne sont pas pertinentes pour l'observateur. On trouve là la définition même du constructivisme (cf. section 3.3).

Cette évolution a donné naissance à ce qu'on appelle la *cybernétique du second ordre* [Von Foerster, 1995]. Un système n'est plus considéré comme une entité passive qu'on peut observer et manipuler, mais comme un agent qui interagit avec son environnement et avec un autre agent qu'est l'observateur. L'observateur est aussi perçu comme un système cybernétique (voir Définition 3.3) qui construit un modèle d'un autre système cybernétique. Il s'agit donc d'appliquer la cybernétique sur elle-même ou ce que le fondateur de cette théorie, Heinz Von Foerster appelle de la *cybernétique du second ordre* dans son livre « *Cybernetics of Cybernetics* » [Von Foerster, 1995].

Dans les années cinquante, les fondateurs de la cybernétique se sont associés à l'école de la *théorie du système général* fondée à peu près à la même époque par Ludwig Von Bertalanffy [Bertalanffy, 1969]. Aujourd'hui, les termes de la *cybernétique*, *cybernétique du second ordre* [Von Foerster, 1995], la *théorie du système général*, la *systémique* et la *systémographie* [Le Moigne, 1999; Morin et Le Moigne, 1999] sont utilisés pour désigner, à peu près, la même approche [Heylighen, 2003;

Heylighen et Joslyn, 2001]. Durant le reste de ce mémoire nous utilisons les termes *systémique* et *cybernétique* .

Parmi les concepts centraux de l’approche systémique, nous citons : complexité, auto-organisation, auto-production, autonomie, réseaux, communication, interaction, régulation, processus, circularité, connexionisme, conjonctivisme, adaptation, évolution, boucles de rétroaction (ou feedback), etc. Pour plus de détails sur les définitions de ces concepts, voir [Heylighen et Joslyn, 2001].

Grâce à sa transdisciplinarité, la systémique a été utilisée pour l’étude de l’organisation et comportement des phénomènes complexes indépendamment de leur matière, types ou échelle spatio-temporelle de leur existence. Elle explore les principes communs à tous les systèmes complexes et les modèles qui peuvent être utilisés pour leur description [Heylighen, 2003; Heylighen et Joslyn, 2001]. Elle touche à un niveau conceptuel très abstrait toutes les disciplines traditionnelles telles que les mathématiques (Norbert Wiener 1894-1964, John Von Neumann 1903-1957), la biologie (Ludwig Von Bertalanffy 1901-1972, Humberto Maturana né en 1928), l’information (Claude Shannon 1916-2001), la psychologie (W. Ross Ashby 1903-1972), la physique (Heinz Von Foerster 1911-2002), les sciences sociales (Donald T. Campbell 1916-1996), l’économie (Kenneth E. Boulding 1910-1993), etc.

3.4.2 La systémique et la logique conjonctive

La systémique (ou la cybernétique) est fondée sur *la logique de conjonction* ou ce que Le Moigne appelle *l’axiomatique conjonctive* . Cette axiomatique part du constat que « *l’esprit humain s’avère fort naturellement capable de raisonner en conjoignant au moins aussi aisément qu’en disjoignant* » [Le Moigne, 1999; Le Moigne, 1974]. Le Moigne montre que la conjonction est fondamentale et que la disjonction est secondaire. Il justifie ceci par le fait que les langues soient des conjonctions de *signes* et de *significations* , que toute analyse ou compréhension d’un système commence par une *comparaison* (acte de conjonction) avec des représentations ou des connaissances acquises. La *logique de conjonction* ne sépare pas le signe de sa signification, l’opérateur de ses opérations ou l’action de son objectif et de ses résultats.

Le Moigne (1999) distingue deux types de procédures dans un comportement complexe : la procédure cybernétique et la procédure structuraliste et il présente la systémique comme étant une conjonction de ces deux procédures :

- **La procédure cybernétique** : elle est fondée sur la conjonction des concepts d’ *environnement actif* et de *téléologie* (ou projet). Pour illustrer cette procédure, Le Moigne (1999) donne l’exemple des deux concepts suivants : < *Environnement actif = Pôle Nord* > et < *Téléologie = Survivre* >. La conjonction de l’environnement actif « *pôle nord* » avec la téléologie « *survivre* » conduit au modèle du système < *Survivre, Pôle Nord* > ayant a priori les caractéristiques comportementales d’un ours blanc. On peut ainsi construire un modèle opérationnel de ce système « ours blanc » sans l’avoir jamais vu ;
- **La procédure structuraliste** : elle est fondée sur la conjonction des concepts de *fonctionnement* (synchronique) et de *transformation* (diachronique). Le comportement d’un système se représente inséparablement par les deux composantes : le « *Faire* » (fonction) et

le « *Devenir* » (transformation). Cette procédure est caractérisée par le modèle classique de la forme, qui en fonctionnant se transforme, et en se transformant assure quelques fonctions. L'exemple de la respiration chlorophyllienne des plantes en est une illustration. En effet, on ne peut pas étudier ce comportement sans dommage si on ne fait pas la conjonction du *fonctionnement synchronique* (absorption de gaz carbonique et émission d'oxygène) et des *transformations diachroniques* des plantes (croissance des feuilles).

Définition 3.5- La conjonction systémique

C'est la conjonction des deux procédures cybernétique et structuraliste. Elle propose de tenir pour inséparable le fonctionnement et la transformation d'un phénomène, des environnements actifs dans lesquels ils s'exercent et des projets par rapport auxquels il est identifiable [Le Moigne, 1999].

La systémique est souvent opposée à l'*approche analytique* et sa *logique disjonctive*. Selon Le Moigne (1999), le fondement épistémologique de l'approche analytique est l'épistémologie positiviste (cf. section 3.3). Cette approche peut être résumée à travers les quatre préceptes du « discours de la méthode » proposés par Descartes qui dit « *Ainsi, au lieu de ce grand nombre de préceptes dont la logique est composée, je crus que j'aurais assez des quatre suivants ...* » :

- « *Le premier était de ne recevoir jamais aucune chose pour vraie que je ne la connusse évidemment être telle, c'est-à-dire d'éviter soigneusement la précipitation et la prévention, et de ne comprendre rien de plus en mes jugements que ce qui se présenterait si clairement et si distinctement à mon esprit que je n'eusse aucune occasion de la mettre en doute* ». Dans ce précepte, on trouve deux principes du positivisme (cf. section 3.3) : le premier est celui qui suggère de se limiter à ce qui est *observable, objectif, stable et mesurable*. Le deuxième est celui qui, contrairement à la cybernétique, ne conçoit par la subjectivité de la *modélisation* et sa dépendance de l'*observateur* (i.e. modélisateur) ;
- « *Le second, de deviser chacune des difficultés que j'examinerais en autant de parcelles qu'il se pourrait et qu'il serait requis pour le mieux résoudre* ». On retrouve dans ce précepte l'influence de la logique disjonctive opposée à celle du holisme (voir Définition 3.4) ;
- « *Le troisième de conduire par ordre mes pensées en commençant par les plus simples et les plus aisées à connaître, pour monter peu à peu comme par degrés jusqu'à la connaissance des plus composées, et supposant même de l'ordre entre celles qui ne se précèdent point naturellement les unes des autres* ».
- « *Et le dernier, de faire partout des dénombrements si entiers et des revues si générales que je fusse assuré de ne rien omettre* ». Cela suppose la possibilité d'exhaustivité et de créer un modèle fermé.

D'après Le Moigne (1999), pour résoudre la complication il suffit de la simplifier. Un système compliqué est donc un *système réductible à des modèles explicatifs*, son intelligibilité (explication) se fait par sa simplification. Alors que la complexité « *implique toujours quelque forme d'imprévisibilité, elle ne peut aisément être tenue pour déterministe* ».

3.5. L’accident de la route : un phénomène complexe ou compliqué ?

Le système CVE peut être considéré comme un *système complexe* (ou *cybernétique*) (cf. Définition 3.3, section 3.4) dans le sens où il s’agit d’un ensemble d’éléments en interaction dont le comportement peut être décrit par boucles récursives (feedbacks). En effet, le conducteur joue, avec les systèmes de gestion du trafic et de contrôle, un rôle central de régulateur : il essaie d’adapter, grâce à un processus de traitement d’information, son comportement et celui du véhicule selon la situation dans laquelle se trouve tout le système CVE (cf. section 4.3.3). Les interactions entre les différents éléments du système CVE (i.e. conducteur, véhicule et environnement) consistent en des échanges de matière (e.g. le frottement des pneus sur la chaussée), d’énergie (e.g. l’appui du conducteur sur la pédale de frein) ou d’information (e.g. saisie d’information par le conducteur). Face à ces échanges et grâce au rôle régulateur du conducteur et celui des composantes « *intelligentes* » embarquées dans l’infrastructure ou dans le véhicule, les éléments du système réagissent en changeant d’état ou en modifiant leur action.

On observe dans le comportement du système CVE les deux procédures cybernétique et structuraliste. En effet, on y identifie une conjonction entre un *environnement actif* <trafic, conditions ambiantes, infrastructure, etc.> et une *téléologie* qui consiste en un <déplacement vers une destination tout en assurant la sécurité>. On y identifie aussi une conjonction entre des *fonctionnements* <accélérer, tourner, freiner, etc.> et des *transformations* <fatigue, hypovigilance, situation d’urgence, usure de pneu, etc.>. La régulation du comportement par le conducteur illustre bien le modèle d’un système, qui *en fonctionnant se transforme, et en se transformant assure quelques fonctions*. Tous les concepts centraux de la complexité (cf. section 3.4) peuvent être identifiés dans le comportement d’un système CVE : auto-organisation, auto-production, interaction, régulation, processus, adaptation, évolution, boucles de rétroaction, etc.

L’accident de la route peut être perçu comme un « phénomène » complexe dans le sens où il s’agit d’une défaillance dans le comportement du système complexe CVE. La complexité de l’accident est due à son *imprévisibilité*. Cette dernière caractérise le comportement du conducteur et l’état dans lequel peuvent se trouver d’autres composantes du système telles que les usagers, l’infrastructure, les conditions ambiantes et le trafic. Le conducteur, s’il occupe une position centrale dans la régulation du fonctionnement du système CVE, il n’est pas le seul acteur impliqué dans l’accident : les constructeurs automobile et le pouvoir public sont aussi impliqués en terme de sûreté des véhicules et de l’infrastructure.

L’évolutivité du « phénomène » de l’accident, due à l’évolutivité du triptyque CVE, est aussi l’une des sources de complexité. En effet, un changement au niveau du véhicule (e.g. un nouveau système embarqué), au niveau de l’environnement (e.g. une nouvelle loi) ou au niveau du conducteur (e.g. une nouvelle situation socioprofessionnelle) peut affecter le comportement global du système CVE. De nouvelles configurations d’accidents peuvent donc apparaître.

Tous les éléments cités ci-dessus ont rendu difficile (voire impossible) une analyse déterministe de l’accident de la route. A cela s’ajoute, comme nous l’avons déjà noté, une part de

subjectivité importante due au fait que chacun des experts accidentologues peut avoir sa propre perception du même phénomène. Cette perception varie selon le contexte de l'étude (objectif, outils utilisés, etc.), la discipline de l'expert (ergonomie cognitive, mécanique, psychologie, etc.), sa spécialité et sa tâche (diagnostic, collecte de données, etc.). Même si deux experts partagent la même discipline, la même spécialité et la même tâche, leurs perceptions de l'accident peuvent être différentes. Dans l'étude menée dans le cadre du projet ACACIA¹ (cf. §5.3.3.1 pour plus de détails), on peut vérifier que les représentations de l'accident utilisées par les experts véhicules sont différentes [Dieng et al., 1996].

3.6. La cybernétique de second ordre : vers une vision systémique de la modélisation

Le Moigne (1999) définit la modélisation comme suit :

Définition 3.6- La modélisation

C'est une action d'élaboration et de construction intentionnelle de modèles par composition de symboles. Ces modèles sont susceptibles de rendre intelligible un phénomène perçu complexe et d'amplifier le raisonnement de l'acteur projetant une intervention délibérée au sein du phénomène ; raisonnement visant notamment à anticiper les conséquences de ces projets d'actions. La modélisation est une activité intrinsèque à toute tâche humaine et elle est la base de construction de connaissances [Le Moigne, 1999].

Dans cette définition, Le Moigne (1999) présente la modélisation comme étant un processus « projectif », c'est-à-dire qu'elle est dépendante d'une téléologie. Cela rejoint l'un des concepts centraux du constructivisme (cf. section 3.3) qui affirme la relativité d'un *modèle* et sa dépendance du contexte de son élaboration et des perceptions des individus qui l'élaborent. Cette définition affirme aussi la *récurtivité* d'un processus de modélisation. En effet, on peut observer dans un processus de modélisation les boucles de rétroaction s'établissant entre le *système modélisé* et le *système modélisant* (le i.e. le *modélisateur*). La perception du modélisateur génère des signes et des symboles qui peuvent influencer sa propre perception ce qui génère de nouveaux résultats. Le schéma d'un système, qui en fonctionnant se transforme, et en se transformant assure quelques fonctions, est donc bel et bien présent. Dans cette définition, on trouve les deux procédures cybernétique et structuraliste conjoignant des concepts de fonctionnement (e.g. observer, représenter, simuler) et de transformation (e.g. signes, symboles, schéma), d'environnement actif dans lesquels ils s'exercent la modélisation et des projets par rapport auxquels il est identifiable.

Ainsi, la modélisation d'un système complexe est perçue comme un *processus complexe* avec tous les traits caractéristiques de la complexité que nous avons présentés dans la section 3.4 (i.e. auto-organisation, évolution, boucles de rétroaction, etc.).

La cybernétique, qui est une façon de voir un système complexe et donc de le modéliser, a été appliquée à elle-même. C'est ce qu'a fait Von Foerster (1995) dans son livre « *Cybernetics of*

¹ ACACIA : Acquisition des Connaissances pour l'Assistance à la Conception par Interaction entre Agents. Il s'agit d'une étude qui a été menée dans le cadre de ce projet pour l'élaboration d'une mémoire d'entreprise pour l'analyse de l'accident à l'INRETS.

Cybernetics ». Il a fait correspondre « *un système de modélisation* » à la tâche de modélisation. On parle alors d’un « *système modélisé* » (le phénomène perçu complexe) et d’un « *système modélisant* » (le modélisateur perçu comme agent complexe) et d’un « *contexte de modélisation* ».

Dans la majorité de la littérature sur la cybernétique du second ordre, on parle de l’inséparabilité de *l’observateur (système modélisant)* et de *l’objet observé (système modélisé)*. Le contexte de l’observation est souvent implicite. Nous avons choisi d’explicitier ce concept et de l’intégrer comme un troisième point de vue à tenir inséparable des deux autres lors de la modélisation. Pour résumer, la cybernétique de second ordre consiste à effectuer une *vision cybernétique (ou systémique)* de la *modélisation cybernétique* elle-même. Cela a donné le concept de « *système de modélisation* » qui conjoint le *système modélisé (ou objet observé)*, le *système modélisant (ou observateur)* et le *contexte de modélisation (ou contexte d’observation)*.

3.7. La systémique : une approche réflexive

Une analyse systémique du système CVE (cf. section 3.5) fait apparaître quatre axes qui sont nécessaires pour la représentation et l’analyse d’un système complexe en général. Nous présentons brièvement ces quatre axes et nous utilisons l’accident de la route pour montrer la réflexivité de l’approche systémique :

- **Axe ontologique ou structurel** (*ce qu’est le système*) : le terme ontologie est issu du domaine de la philosophie où il signifie « *explication systématique de l’existence* ». L’axe ontologique représente pour nous les composants du système (e.g. conducteur, infrastructure, trafic, conditions ambiantes et véhicule), les différentes interactions entre ces composants ainsi que leurs taxonomies ;
- **Axe fonctionnel** (*ce que fait le système*) : il représente le processus global de fonctionnement du système complexe Conducteur-Véhicule-Environnement (CVE) ;
- **Axe transformationnel ou génétique** (*comment évolue le système, quels sont les états générés*) : il décrit l’aspect dynamique, évolutionnel et génétique (dans le sens de la genèse et non celui de l’hérédité) du comportement du système CVE lors d’un accident ;
- **Axe téléologique ou motivationnel** (*quels sont l’objectif et la motivation du système*) : la téléologie signifie en philosophie « *l’étude de la finalité* ». Une analyse de l’accident selon l’axe téléologique consiste en une analyse à travers les objectifs du système CVE lors de la conduite en général et lors d’un accident en particulier.

La réflexivité de l’approche systémique réside dans le fait que chacun des axes peut être analysé selon les autres axes et ainsi de suite (cf. Figure 3-1).

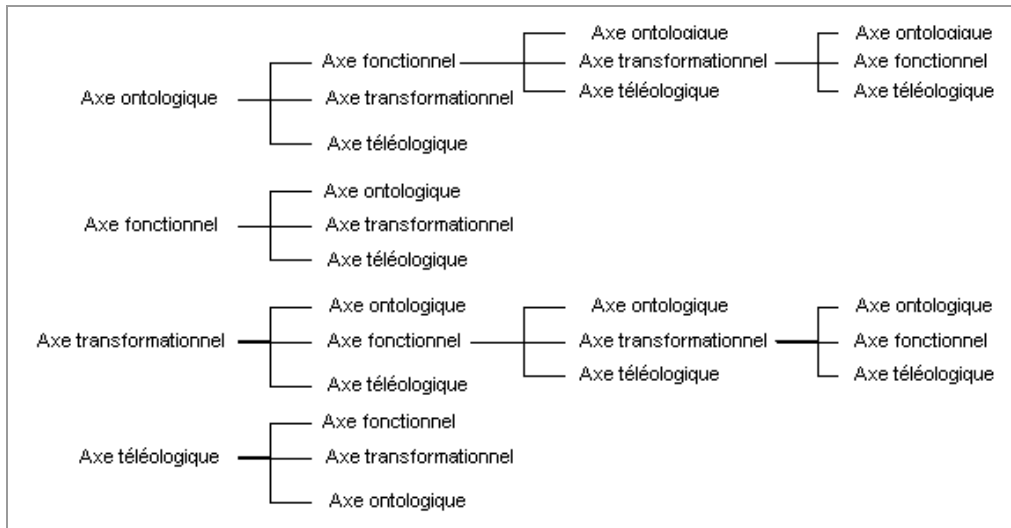


Figure 3-1- Réflexivité de l'approche systémique

Par exemple, le « *conducteur* » est l'un des composants du modèle ontologique que nous pouvons analyser selon les autres axes. Une analyse transformationnelle de ce composant, par exemple, consiste à analyser l'évolution de son état (physique et psychologique) durant les étapes séquentielles¹ d'un accident de la route (cf. section 4.3.2). L'axe transformationnel peut être analysé à son tour selon les autres axes. Si on prend l'une de ces étapes, « *l'urgence* »² par exemple, elle peut être analysée selon l'axe fonctionnel. Cela peut être effectué en utilisant le modèle de traitement d'information (cf. section 4.3.3). Chacune des étapes fonctionnelles peut, à son tour, être analysée selon l'axe téléologique, c'est-à-dire l'objectif du conducteur. On peut refaire la même chose en commençant par d'autres axes et on aboutit à la même analyse.

3.8. Synthèse

L'objectif de ce chapitre est de présenter les fondements philosophiques et épistémologiques de l'*approche systémique* que nous utilisons comme méthode de modélisation dans nos travaux. L'origine philosophique de la systémique (cf. section 3.1) remonte à la perception relativiste du monde que nous avons présentée à travers la perception de la *notion de valeur* chez Spinoza selon qui *la valeur d'un objet est soumise au contexte de son échange*. Cette perception est opposée à la perception *absolutiste* d'Aristote et Platon pour qui *la valeur est intrinsèque à l'objet indépendamment de son contexte d'échange*.

Cette dualité s'est manifestée à travers deux perceptions antagonistes de la *valeur* chez les économistes classiques et néo-classiques (cf. section 3.2) : la première est *objective*, on la trouve chez Smith, Ricardo et Marx qui ont défini la *valeur d'échange* comme étant relative à la *quantité du travail*, donc qui préexiste à l'échange (indépendante de l'échange). La deuxième est *subjective*, on la trouve chez les néoclassiques ou marginalistes tel que Walras, pour qui la valeur d'échange dépend du

¹ conduite normale, rupture, urgence et choc

² l'étape où le conducteur doit entreprendre une manœuvre de récupération pour éviter le choc

contexte car elle se détermine sur le marché « *l’offre n’est qu’une conséquence de la demande* » [Walras, 1874].

Sur le plan *épistémologique* (cf. section 3.3), on distingue l’*épistémologie constructiviste* et l’*épistémologie positiviste*. La première reconnaît le caractère relatif de la connaissance et sa dépendance de la construction du sens par les individus en se basant sur leurs expériences et leurs interactions avec leur environnement (contexte). Cette conception est donc proche du *relativisme* (cf. section 3.1) en philosophie et du *subjectivisme* (cf. section 3.2) en économie. Quant à l’*épistémologie positiviste*, elle insiste sur le caractère absolu, objectif et stable de la connaissance et suppose que la science doit être limitée à ce qui est observable et mesurable. L’*épistémologie positiviste* est connue aussi sous le nom de « *l’empirisme* » ou « *l’idéalisme* » (ou *absolutisme*) platonicien (cf. section 3.1).

Sur le plan de la *méthodologie* (cf. section 3.4), nous avons distingué entre la *systémique* (appelée aussi dans la littérature *cybernétique*, *cybernétique du second ordre* ou *théorie du système général*) et l’*approche analytique*. Les éléments d’un *système cybernétique* (Définition 3.3) sont en interaction réciproque et sont reliés par des boucles de rétroaction. Les interactions entre les éléments donnent à l’ensemble des propriétés que ne possèdent pas les éléments pris séparément, c’est le principe du *holisme* (voir Définition 3.4). Une analyse systémique se base alors sur une logique conjonctive qui prend pour inséparable l’ontologie du système de sa téléologie et le fonctionnement de sa transformation.

En se basant sur la définition de la complexité dans la section 3.4, nous avons montré dans la section 3.5 que le système CVE est un système complexe et que l’accident de la route peut être perçu comme un « phénomène » complexe. La combinaison de plusieurs points de vue d’analyse de l’accident s’avère donc fondamentale pour appréhender cette complexité.

Dans la section 3.7, nous avons expliqué le principe de la *cybernétique de second ordre*, c’est-à-dire appliquer la cybernétique à elle-même. En d’autres termes, il s’agit d’une *vision systémique* de la *modélisation systémique* elle-même. Cela a donné le concept de « *système de modélisation* » qui conjoint un « *système modélisé* » (ou objet observé), un « *système modélisant* » (ou observateur) et un « *contexte de modélisation* » (ou contexte d’observation).

Dans la dernière section de ce chapitre (cf. section 3.6) nous avons utilisé l’exemple de l’accident de la route pour montrer la réflexivité de l’approche systémique. Cette réflexivité réside dans le fait que chacun des axes d’analyse systémique (i.e. ontologique, fonctionnel, transformationnel et téléologique) peut être analysé selon les autres axes et ainsi de suite.

Nous tenons à noter à la fin de cette synthèse que les catégorisations des épistémologies et des méthodologies et les étiquettes en général sont simplificatrices. Von Foerster, le fondateur de la cybernétique du second ordre, a toujours refusé le dogmatisme et les étiquettes. Dans son dernier livre [Von Foerster et Poerksen, 2002], il répond à la question de Bernhard Poerksen s’il est constructiviste : « *Non, non. Je suis viennois. C’est la seule étiquette que j’accepte (...). Je n’adhère à aucune épistémologie* ». Néanmoins, comme le précise Bernhard Poerksen, « *les étiquettes nous permettent de dialoguer sur la base d’une sorte de compréhension préliminaire par laquelle nous pouvons faire l’économie d’avoir à clarifier les notions de base au début de chaque discussion...* ».

Chapitre 4.

Accidentologie

Ce chapitre se décompose en quatre sections dans lesquelles nous présentons respectivement : un aperçu succinct des différentes approches dans le domaine de la sécurité routière, des Etudes Détaillées d'Accident (EDA), des différentes approches et modèles d'analyse de l'accident de la route et le concept scénario-type d'accidents ainsi que les différentes approches utilisées pour les construire.

4.1. La sécurité routière

La sécurité routière « a pour finalité spécifique d'assurer les déplacements routiers sans effets externes indésirables (e.g. sentiment d'insécurité, accidents de la route, blessures, etc.). Elle doit être assurée tant dans l'organisation des déplacements qu'à l'occasion de chaque déplacement » [Le Coz et Page, 2003]. Le besoin en sécurité routière est apparu avec l'apparition des accidents, donc dès l'apparition des véhicules et de leur utilisation. Ce besoin s'est intensifié par l'augmentation de l'insécurité routière qui a atteint son record en France en 1972 par un bilan de 16617 morts. Une action de sécurité s'appuie sur les quatre phases suivantes [Fleury, 1998] : le *diagnostic*, la *définition des modalités de l'action (d'intervention)*, la *mise en œuvre* et l'*évaluation*.

La sécurité peut être traitée sous trois angles [Le Coz et Page, 2003; Perron, 1997]:

Définition 4.1- La sécurité primaire

Elle vise à la réduction des accidents par leur prévention et leur évitement.

Définition 4.2- La sécurité secondaire

Elle vise à réduire les conséquences des accidents (e.g. protection des occupants).

Définition 4.3- La sécurité tertiaire

Elle vise à améliorer le secours aux personnes impliquées et diminuer les séquelles physiques et psychologiques après un accident.

D’autres définitions font la distinction entre la *sécurité active* et la *sécurité passive* :

Définition 4.4- La sécurité active

Elle fait référence aux systèmes nécessitant l’intervention du conducteur (e.g. ceinture de sécurité, ABS, etc.)

Définition 4.5- La sécurité passive

Elle fait référence aux contre-mesures indépendantes de l’intervention du conducteur (airbag, régulateur de vitesse, etc.)

Les différentes « branches » de la sécurité sont regroupées aujourd’hui dans ce qu’on appelle la « *sécurité intégrée* ». Nos travaux de recherche s’intègrent dans les études en sécurité primaire (cf. Définition 4.1) car les STA sont destinés au développement des systèmes d’évitement de l’accident.

4.2. Les EDA : Etudes Détaillées d’Accidents

Les Etudes Détaillées d’Accidents (EDA) représentent l’un des outils qui ont servi à approfondir les connaissances en accidentologie. Elles permettent de détecter des nouveaux problèmes en sécurité routière, formuler de nouvelles hypothèses qui peuvent ensuite être testées expérimentalement, révéler les interactions entre les facteurs d’accidents, informer sur le système de production des accidents et comprendre les mécanismes accidentels pour pouvoir définir et mettre en œuvre des contre-mesures de sécurité [Ferrandez, 1995]. Les précurseurs de ces études approfondies en France sont l’Organisme Nationale de Sécurité Routière (ONSER) puis le département Mécanismes d’Accident (MA) de l’Institut National de Recherche sur les Transports et leur Sécurité (INRETS). Dans le cadre des EDA, MA a développé une méthodologie de recueil de données sur les accidents. Un document de référence dans ce domaine est intitulé : « Etude Détaillée d’Accidents orientée vers la sécurité primaire, méthodologie de recueil et de pré-analyse » [Ferrandez, 1995].

Le LAB a choisi de développer des études détaillées d’accidents sur la scène de l’accident et a confié au CEESAR¹ la réalisation de ces études. Il s’agit d’envoyer une équipe de spécialistes le plus rapidement sur les lieux de l’accident et de rechercher tous les indices pouvant aider à expliquer ce qui s’est passé. Deux équipes du CEESAR étaient chargées de ces études, une à Amiens et l’autre à Evreux. Chaque équipe est composée d’un psychologue, d’un technicien du véhicule et d’un technicien de la route. Les données recueillies portent sur l’impliqué (état civil, permis de conduire, expérience de conduite, état au moment de l’accident, motif de déplacement, manœuvres effectuées, etc.), les véhicules impliqués (identification, caractéristiques techniques, défaillances, équipements, charge, pneus, direction, etc.) et l’environnement (conditions ambiantes, infrastructure, trafic, etc.).

Cette équipe travaille en collaboration avec d’autres intervenants sur les lieux d’accidents (les secours, les policiers et les gendarmes) et avec les intervenants qui peuvent faciliter le déroulement des enquêtes ou fournir des informations autour de l’accident (la Direction Départementale de

¹ CEESAR : Centre Européen d’Etude de Sécurité et Analyse des Risques

l'Équipement pour l'Infrastructure et le Trafic, le Tribunal de Grande Instance pour les autorisations et l'accès aux procès-verbaux d'accidents, les services hospitaliers pour les bilans médicaux et l'autorisation de procéder à des entretiens avec les impliqués).

Une grande partie des données est recueillie sur le site. Les informations complémentaires sont recueillies plus tard par les différents moyens suivants : consultation du procès-verbal à la brigade de gendarmerie ou à la compagnie de CRS, des entretiens avec les impliqués à l'hôpital ou chez eux, des discussions avec les secours et les policiers, des contacts avec les garagistes et des recherches à la DDE. Les bilans médicaux, établissant les lésions des impliqués lors de l'accident, sont obtenus auprès des services hospitaliers.

Une grande partie de l'information recueillie et/ou analysée est codifiée puis stockée dans une base de données pour la rendre accessible plus facilement. L'autre partie est accessible dans le dossier d'accident et dans le logiciel de reconstruction. L'information codée est structurée dans la base de donnée selon quatre axes : l'usager (conducteur, piéton ou occupant), le véhicule, l'infrastructure et les caractéristiques générales de l'accident. Ces données serviront comme sources d'informations pour les études micro ou macro-accidentologiques. Les bases de données contiennent aussi les entretiens avec les impliqués et les tableaux de collisions.

4.3. Modélisation de l'accident de la route

Dès la fin des années cinquante, des études en profondeur ont été menées dans le cadre de la recherche pour tenir compte de la complexité des accidents. Il a fallu créer des modèles de représentation des accidents permettant leur analyse en détail tout en respectant leur aspect dynamique, séquentiel, causal ainsi que les interactions entre ses différents composants (i.e. conducteur, véhicule et environnement). Nous avons choisi dans ce chapitre de ne présenter que les modèles les plus utilisés.

4.3.1 Modèle CVE

Le modèle CVE (Conducteur-Véhicule-Environnement) permet une analyse de l'accident selon les trois points de vue *conducteur*, *véhicule* et *environnement*, ainsi qu'à travers les interactions entre ces trois composantes. Les « composantes » de chacun des trois sous-systèmes sont identifiées et caractérisées par des attributs. L'environnement, par exemple, regroupe {*l'infrastructure*, les *conditions ambiantes*, *l'équipement de la route* et le *trafic*}. Le véhicule regroupe { *l'aide à la conduite*, les *commandes*, la *charge*, les *ouvrants*, le *vitrage*, la *sellerie*, les *pneus*, *l'éclairage/signalisation*, la *suspension/amortissement*, la *direction*, etc.}.

Dans la Figure 4-1, nous présentons une version simplifiée du modèle CVE avec quelques exemples d'interactions entre les trois composantes. Une version plus détaillée de ce modèle est présentée dans la Figure 7-4.

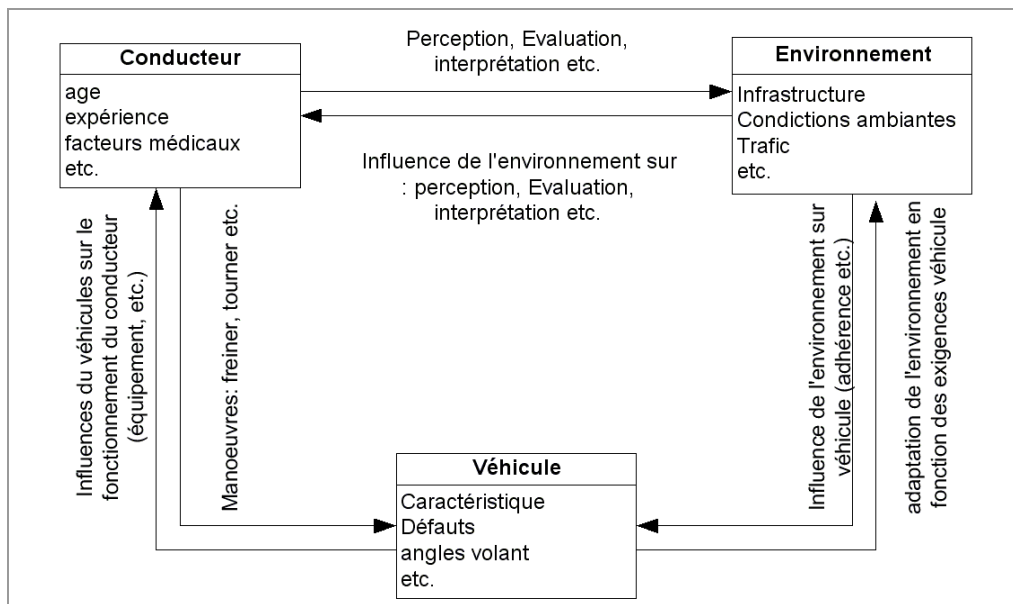


Figure 4-1- Modèle CVE.

4.3.2 Modèle séquentiel

L’approche séquentielle est parmi les approches les plus utilisées pour l’analyse de l’accident car elle représente en même temps l’aspect causal, dynamique et spatio-temporel de l’accident. Parmi les premiers modèles étudiant ces aspects, nous citons celui de Baker (1960) qui représente l’accident à travers une séquence des phases suivantes : *Préparation du déplacement* (fixation d’un objectif de déplacement), *la stratégie de conduite* (choix du mode de déplacement et planification de l’itinéraire), *la tactique d’évitement devant une situation dangereuse* (choix d’une manœuvre d’évitement pour éviter la situation dangereuse), *l’événement dommageable* (occurrence du choc dû à l’échec de la tactique d’évitement choisie) [Baker, 1960].

Nous avons identifié un autre modèle, celui de l’accident de « *véhicule seul* » (*Single Vehicle Accident*) élaboré en 1977 à l’université de Miami pour la NHTSA [Kurucz et al., 1977]. Il s’agit d’un modèle séquentiel pré-accidentel qui distingue les cinq phases suivantes : *manœuvre pré-accidentelle*, *tâche* (but du conducteur), *traitement de l’information*, *action conducteur*, *réaction du conducteur* [Kurucz, 1977].

Dans ces deux modèles, on trouve des aspects relatifs à l’*objectif* du conducteur (e.g. tâche) et à son *fonctionnement* (e.g. traitement d’information, action). Ces deux aspects appartiennent à d’autres types de modèles que nous présenterons dans les paragraphes 4.3.4 et 4.3.3.

Nous avons donc choisi de présenter le troisième modèle que nous avons identifié dans la littérature et qui a été développé par l’INRETS. Ce choix a été encouragé aussi par le fait qu’un modèle semblable est utilisé par le LAB dans les EDA.

Définition 4.6- Modèle d'analyse séquentielle de l'accident

Ce modèle décrit le déroulement de l'accident à travers les quatre phases (ou situations) suivantes [Brenac, 1997] (cf. Figure 4-2):

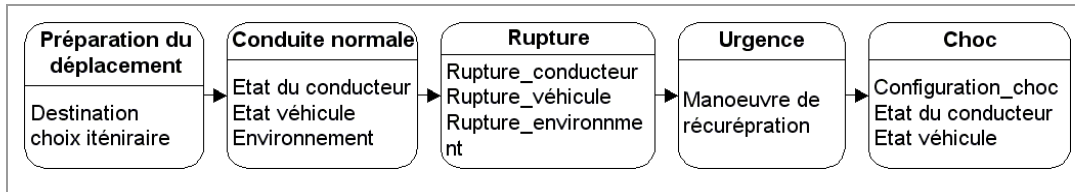


Figure 4-2- Modèle séquentiel de l'INRETS [Brenac, 1997].

Préparation du déplacement

Elle correspond au choix de l'itinéraire, la stratégie de conduite, etc. ;

Une situation de conduite normale

Elle correspond au comportement « normal » ou « stable » du système CVE. Elle est caractérisée par l'objectif du conducteur, sa stratégie de conduite et son mode de déplacement. Cette situation intègre aussi le cadre spatio-temporel de la fonction de conduite avant l'apparition du facteur initiateur de l'accident ;

La situation d'accident ou situation de rupture

Elle est instantanée et courte en durée. Il s'agit d'une rupture qui se produit par rapport à la situation précédente. Elle est caractérisée par un événement (manœuvre inattendue d'un autre usager, configuration inattendue de l'infrastructure, etc.) qui fait basculer le système CVE vers une situation d'urgence ;

La situation d'urgence

C'est une situation dégradée dans laquelle le conducteur doit mettre en œuvre une tentative de récupération sous forme d'une manœuvre d'urgence pour faire face aux contraintes spatiales et temporelles de la situation ;

La situation de choc

Elle marque l'échec des manœuvres d'évitement entreprises dans la situation d'urgence. Elle englobe le choc, son équivalent (chute, renversement, etc.) et leurs conséquences.

La Figure 4-2 donne une illustration graphique du modèle séquentiel.

4.3.3 Modèle de fonctionnement de l'opérateur humain

Le conducteur occupe une position centrale dans la régulation du fonctionnement du système CVE dont il fait partie. Des études ont porté donc sur le conducteur, sur ses caractéristiques, son comportement d'une part en situation normale de conduite et d'autre part en situation d'accident [Van Elslande et Alberton, 1997], [Leplat, 1985; Reason, 1993]. Elles se sont focalisées sur des études du processus de traitement de l'information par l'opérateur humain (cf. §4.3.3.1) et sur les mécanismes de production de la défaillance (cf. §4.3.3.2).

4.3.3.1 Traitement d'informations par l'opérateur humain

Plusieurs modèles ont été réalisés dans le cadre de recherche sur le fonctionnement cognitif de l'homme pour expliquer son mode de traitement d'information et son mode de prise décision. On distingue trois conceptions :

- (i) *L’homme : un système de traitement de l’information à capacité limitée.* Ce courant met en évidence le fait que l’homme ait une charge mentale maximale admissible et donc sa capacité de traitement de l’information est limitée [Leplat, 1985] ;
- (ii) *L’homme : un système de traitement séquentiel et algorithmique de l’information.* Ce courant repose sur les travaux de Newell pour la résolution de problème qui formalisent le raisonnement menant à la solution d’un problème sous forme d’un algorithme optimal [Van Elslande et Alberton, 1997]. La limite de cette conception réside dans le caractère algorithmique supposé recouvrir le traitement mis en œuvre en négligeant par exemple les heuristiques et l’apprentissage [Leplat, 1985] ;
- (iii) *L’homme : un système de capitalisation.* D’après ce courant, l’homme construit, par apprentissage, une base de connaissances composée par des modèles mentaux de représentation des systèmes et de leurs réactions [Nicolet et al., 1989; Warner, 1981]. Ces modèles sont réactualisés au cours de l’expérience ce qui leur confère un aspect dynamique. La prise de décision est plus ou moins automatique selon le degré d’apprentissage et de familiarité de l’opérateur avec la situation confrontée [Leplat, 1985; Reason, 1993].

4.3.3.2 Production de l’erreur humaine

Plusieurs modèles ont été élaborés pour comprendre les mécanismes de production de l’erreur humaine en général et celle du conducteur en particulier. La différence entre ces modèles réside dans la définition de la notion d’erreur. Ces modèles sont si nombreux dans la littérature qu’il est difficile de les classer. Néanmoins, nous proposons de les présenter suivant quatre axes :

- (i) *Explication de l’erreur selon le modèle de traitement de l’information adopté :* trois conceptions de l’erreur humaine peuvent être distinguées dans cet axe : l’erreur en tant que *dépassement des ressources* [Neboit, 1995], l’erreur en tant que *défaillance d’une étape* de déroulement du processus de traitement de l’information [Van Elslande et Alberton, 1997] et l’erreur en tant que *distorsion* entre représentation et monde réel ;
- (ii) *Explication de l’erreur selon le niveau de l’activité :* Reason (1993) et Leplat (1990) différencient les erreurs selon qu’elles seraient faites au niveau de la planification (*niveau stratégique*), au niveau de la mémorisation d’une séquence d’action (*niveau tactique*), ou au niveau de l’exécution (*niveau opérationnel*) ;
- (iii) *Explication de l’erreur selon la norme de référence :* généralement lorsqu’on parle d’erreur c’est par rapport à une norme ou une référence. Mais, cette norme, comme l’indiquent Cellier (1990) et Van Elslande et al. (1997), est multi-référentiels. L’erreur est, du *point de vue l’expert*, l’écart entre une *tâche prescrite* et la *tâche effective* définie par l’opérateur. Du *point de vue l’opérateur*, l’erreur est définie comme l’écart entre ce qu’il a voulu faire (la *tâche*) et ce qu’il a fait (*l’activité*). Si on prend comme norme de référence *l’activité* telle qu’elle se déroule normalement, l’erreur est l’écart entre cette *activité habituelle* et *l’activité exécutée* par l’opérateur ;
- (iv) *Explication de l’erreur selon sa source* [Perron, 1997] : on distingue une *erreur de conception* : elle est due à une mauvaise étude du besoin (la sécurité dans notre cas) ou du savoir-faire. Donc, c’est un écart entre le besoin réel (la sécurité) et les fonctions envisagées par le

concepteur. Une *erreur de réalisation* : elle est due à un écart entre les fonctions envisagées et spécifiées par le concepteur et leurs réalisations effectives. Ce type d'erreurs englobe les erreurs de fabrication et les erreurs de montage. Une *erreur d'utilisation* : c'est l'erreur commise par l'utilisateur du système (e.g. conducteur, piéton). Elle est due à un écart entre ce que le système offre comme fonctionnalités et ce que l'utilisateur a utilisé pour faire face à une situation. On peut identifier aussi, du point de vue de l'opérateur, des *erreurs internes* (relatives à l'opérateur lui-même) et des *erreurs externes* (relatives à son environnement). En suivant le même raisonnement, Reason (1993) distingue deux types d'erreurs. Les *erreurs actives* : elles correspondent à l'activité des opérateurs impliqués directement dans la régulation du système (exemple : le conducteur). Les *erreurs latentes* : elles correspondent aux activités des opérateurs éloignés de l'interface de contrôle et d'utilisation direct du système (e.g. le concepteur du système). D'autres études distinguent facteurs humains et facteurs contextuels. Les *facteurs humains d'erreur* peuvent être expliqués à travers la nature de cognition humaine et son mode fonctionnement (exemple : mode de traitement de l'information) et à travers la nature même de l'opérateur humain (état physique, psychologique, etc.). Les *facteurs contextuels de l'erreur* sont relatifs aux conditions dans lesquelles se trouve l'opérateur. Ils génèrent des situations non favorables [De Keyser, 1985] ce qui rend impossible l'exécution de la tâche. Les caractéristiques organisationnelles contraignantes (charges de travail) font partie de cette catégorie de facteurs contextuels dans la mesure où elles minimisent l'efficacité de l'opérateur [Leplat, 1985].

Le modèle utilisé au LAB pour l'analyse du *fonctionnement/dysfonctionnement du conducteur* est similaire, quoique moins complet, à celui développé par [Van Elslande et Alberton, 1997].

Il identifie les étapes suivantes :

Situation de pré-accident

Elle caractérise l'état du système CVE avant le processus de conduite ;

Etape perceptive

Elle est caractérisée par les fonctions de recherche et de détection des informations ;

Etape de diagnostic

Elle est caractérisée par une fonction de compréhension qui concerne l'interprétation des données détectées ;

Etape de pronostic

Elle est caractérisée par une fonction d'anticipation relative à l'évolution attendue d'une situation déjà identifiée. Elle est caractérisée aussi par une fonction de prévision relative aux attentes développées sur la rencontre d'un événement non encore présent dans la scène visuelle ;

Etape décisionnelle

Elle fait référence à la prise de décision ;

Etape motrice (ou action)

Elle correspond à la fonction d'exécution des actions décidées lors de la prise de décision.

Ce modèle, comme le précisent ses auteurs [Van Elslande et Alberton, 1997], n'est pas un modèle de fonctionnement du conducteur (qui n'est pas forcément linéaire) au sens strict, mais plutôt une décomposition des étapes engagées qui doit servir à identifier le moment où une défaillance est intervenue. Nous donnons une illustration graphique de ce modèle dans la Figure 4-3.

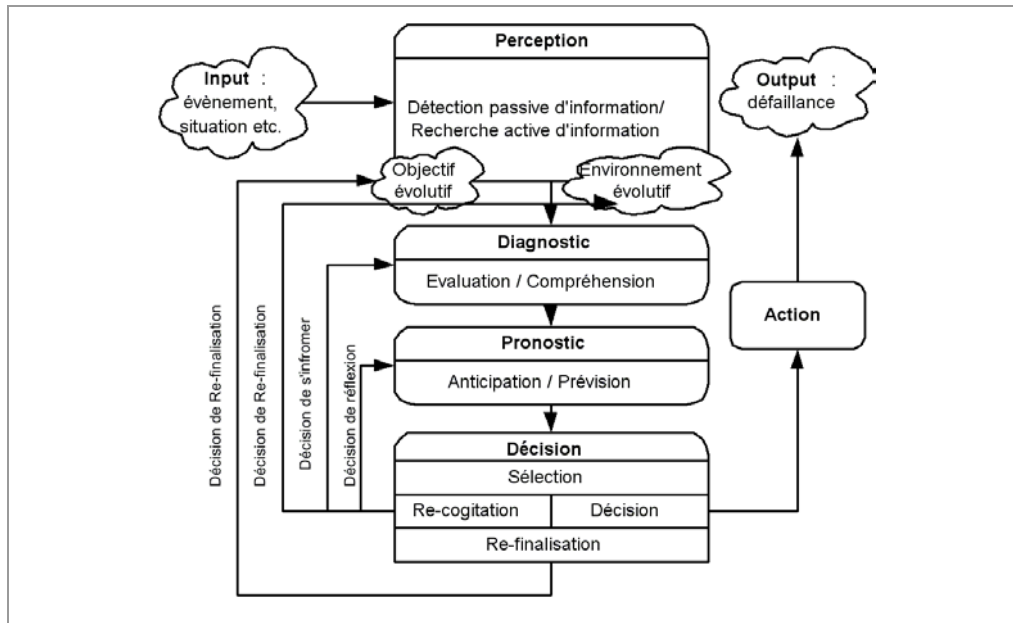


Figure 4-3- Modèle de traitement de l'information par le conducteur (adapté de [Van Elslande et Alberton, 1997]).

4.3.4 Modèle de la tâche de conduite

Leplat et Hoc distinguent les notions de *tâche* et d'*activité* [Leplat et Hoc, 1992] :

Définition 4.7- Tâche

Elle est définie comme un but donné dans des conditions déterminées.

Ces conditions sont relatives aux états à parcourir avant d'atteindre le but fixé [Perron, 1997]. Il s'agit de conditions technologiques, organisationnelles, physiques et socio-économiques. La tâche peut être décomposée en sous-tâches, chacune définissant ce qu'il y a à faire dans une situation donnée [Malaterre et Fontaine, 1992].

Définition 4.8- Activité

Elle est définie comme ce que le sujet met en œuvre pour exécuter la tâche.

L'activité regroupe l'ensemble de conduites (observables ou non, adéquates ou non) mises en jeu pour exécuter la tâche [ISDF, 1994]. On distingue aussi l'*activité réelle* et l'*activité habituelle* [Cellier, 1990]. L'activité réelle est l'activité effectivement mise en œuvre par l'opérateur pour réaliser la tâche qu'il a définie. L'activité habituelle est ce qu'on met habituellement en œuvre pour réaliser la tâche prescrite.

Exemple :

- Tâche définie : dépasser un véhicule.
- Activité effective : se mettre à droite et accélérer.
- Activité habituelle : se mettre à gauche et accélérer.

Le *modèle de la tâche* est un modèle qui représente le processus de conduite à travers trois tâches principales : la *navigation*, le *guidage* et le *contrôle*.

Navigation

Elle est caractérisée par les activités dont l'objectif est la collecte d'informations (e.g. lire la signalisation, chercher son chemin, etc.).

Guidage

Cette tâche est caractérisée par les activités entreprises par le conducteur pour le positionnement longitudinal et transversal par rapport à l'infrastructure (e.g. sortie de voie ou sortie de route) et par rapport au trafic (e.g. conflit avec un autre usager).

Contrôle

Cette tâche est caractérisée par les activités entreprises par le conducteur pour le contrôle dynamique du véhicule en longitudinal (e.g. accélération, freinage) et en latéral (e.g. action volant).

Ce modèle est utilisé pour classifier les systèmes d'assistance en distinguant [Le Coz et Page, 2003; Perron, 1997] : des aides d'ordre *stratégique* (e.g. planification de l'itinéraire et la navigation), *tactique* (e.g. choix et réalisation de la manœuvre de conduite adaptée à la situation rencontrée) et *opérationnel* (e.g. contrôler la trajectoire du véhicule).

La combinaison des différents modèle, i.e. le modèle CVE (cf. section 4.3.1), le modèle production de l'erreur humaine (cf. section 4.3.3), le modèle séquentiel (cf. section 4.3.2) et le modèle de la tâche (cf. section 4.3.4) permet de construire une approche étiologique de l'accident pour comprendre ses mécanismes.

4.4. Concept de scénario-type d'accidents

Le concept « *scénario* » est une technique de description qui a attiré l'attention de plusieurs praticiens et chercheurs dans de nombreux domaines [Leite et al., 2000]. En économie et en finance, par exemple, les scénarios sont utilisés par des institutions financières pour anticiper le comportement des marchés. En politiques et géopolitique, ils sont élaborés en se basant sur les contextes socio-politiques pour anticiper des changements, des risques de guerre. En analyse de risques dans les projets, dans des centrales nucléaires ou chimiques, ils sont utilisés pour anticiper ces risques et préparer des plan d'actions [Scheringer et al., 2001]. Cependant, les publications sur les méthodes de construction de ces scénarios dans ces domaines reste relativement limitées.

Le domaine des systèmes d'information est l'un des domaines où on trouve le plus de travaux sur le développement et l'utilisation des techniques de scénario. Une synthèse de ces travaux peut être trouvée dans [Jarke et al., 1998]. Les scénarios sont utilisés comme un moyen pour comprendre le domaine dans lequel le logiciel va être développé afin de répondre au mieux au cahier des charges [Kazman et al., 1996; Potts, 1995; Sutcliffe, 1998; Van Helvert et Fowler, 2003]. Ils sont considérés comme une façon efficace de capturer les besoins des utilisateurs dans leur contexte, ce qui est essentiel pour la conception du système et son intégration dans l'environnement de travail [Caroll, 1995,1998; Gandon et Dieng].

Un scénario combine généralement une *description informelle* (i.e. une scène de la vie réelle) et une *description formelle* (i.e. un modèle abstrait) [Haumer et al., 1998]. Sa description est généralement effectuée sous forme textuelle, story-board (suite de schémas d'écrans), schémas (e.g. arborescence), matrices fonctionnelles, etc. Sa construction se base sur des approches ascendantes

(partir de l’analyse des données terrain et construire un scénario) ou descendantes (construire un scénario en se basant sur un modèle donné et ensuite le valider sur les données). Un excellent état de l’art sur les différentes méthodes de construction des scénarios dans le domaine de conception des systèmes d’information peut être trouvé dans [Leite *et al.*, 2000].

Nous nous intéressons ici uniquement à l’élaboration et l’utilisation des scénarios dans le domaine de l’accidentologie. Sa définition dans ce domaine est assez proche de celle dans le domaine des systèmes d’information.

Définition 4.9- Scénario-Type d’Accidents (STA)

Il est défini comme un déroulement prototypique (prototype de déroulement) correspondant à un groupe d’accidents présentant des similarités d’ensemble du point de vue de l’enchaînement des faits et des relations de causalité dans les différentes phases [Brenac et Fleury, 1999].

En France, le concept STA a été développé à l’INRETS où des travaux ont montré que son utilisation est particulièrement efficace dans des tâches de diagnostic [Brenac et Megherbi, 1996; Fleury *et al.*, 1991; Megherbi, 1999; Van Elslande et Alberton, 1997]. En effet, la réalisation d’actions de prévention en sécurité routière se base, entre autres, sur une analyse fine des accidents pour identifier les facteurs générateurs des accidents. La deuxième étape consiste à rechercher des solutions aux problèmes identifiés lors de la première phase. Comme nous l’avons déjà noté, les accidents qui présentent des similarités au niveau de leur déroulement, présentent a priori des similarités au niveau des contre mesures. Dans la littérature anglo-saxon, on parle plutôt de « groupes d’accidents » (accident clusters) ou « configurations d’accidents » [Cameron, 1997; Chovan *et al.*, 1994; Najm *et al.*, 2001; Sohn et Lee, 2003; Sohn et Shin, 2001].

Au LAB, le concept STA est utilisé comme un outil de *synthèse des catégories d’accidents*, de *compréhension des mécanismes accidentels* et de *préparation de l’évaluation des systèmes de sécurité*. [Bar et Page, 2002; Kassagi, 2003; Page *et al.*, 2004]. Nos travaux ont pour objectif de fournir des méthodes et des outils efficaces pour la construction de ces STA.

Nous considérons les STA comme un support de connaissances d’experts : ils sont issus de travail basé sur des connaissances expertes. Il s’agit aussi d’un outil de capitalisation de connaissances et de retour d’expérience. Un STA, lorsqu’il est mis sous une forme suffisamment simple, peut constituer un support de communication sur les accidents entre les experts et les non-experts ou dans le cadre d’actions de formation ou d’information [Brenac et Fleury, 1999].

Dans les paragraphes suivants, nous présentons un aperçu sur les approches et méthodes existantes dans la littérature pour la construction des scénarios d’accidents. Nous avons distingué deux approches principales : une *approche experte* (cf. section 4.4.1) et une *approche automatisée* (cf. section 4.4.2).

4.4.1 Elaboration experte des STA

Plusieurs « configurations » (ou taxonomies) d’accidents de la route existent dans la littérature. Ces configurations sont le résultat de travaux cliniques basés essentiellement sur l’expertise. Parmi les méthodes les plus formalisées et expliquées sur la problématique de construction experte des STA, on trouve celle développée dans les travaux de [Brenac et Fleury,

1999; Brenac et Megherbi, 1996; Megherbi, 1999]. Elle se base sur une analyse en profondeur et cas par cas des accidents en utilisant la méthode séquentielle d'analyse d'accident (cf. 4.3.2). L'investigation porte ensuite, pour chaque phase, sur les processus fonctionnels et les facteurs de causalité qui ont déterminé ces faits ou événements [Brenac et Fleury, 1999].

Après cette analyse, on procède à un regroupement des cas qui présentent des similitudes au niveau de l'enchaînement des phénomènes du point de vue événementiel, fonctionnel et causal. Ce regroupement se base sur une **comparaison qualitative et holistique. Il n'y a pas de comparaison formelle. La comparaison fait l'objet d'un jugement global** [Brenac et Fleury, 1999] qui permet d'identifier les traits principaux des accidents, les plus fréquents et de construire un scénario prototypique de déroulement. Un exemple est présenté dans la Figure 4-4.

<p><u>Condition générale et situation de conduite :</u> Un usager local plutôt jeune (3 cas < 25 ans) circule un jour de semaine vers 7h - 8 h ou 16h - 18h</p> <p><u>Situation d'accident (rupture) :</u> Suite au détournement de l'attention de l'utilisateur « A », pour effectuer une tâche concurrente relative à la conduite (contrôle à gauche pour dépasser, recherche directionnelle, etc.), le véhicule rattrape un usager « B » circulant sur la même file.</p> <p><u>Situation d'urgence :</u> L'utilisateur « A » ne réagit pas (2 cas) ou réagit tardivement en freinant (2 cas)</p> <p><u>Situation de choc :</u> L'utilisateur « A » percute l'arrière de l'utilisateur « B »</p> <p><u>Facteurs accidentogènes :</u></p> <ul style="list-style-type: none"> • Faible expérience de la conduite (d'où difficulté de partage de l'attention entre différentes tâches relatives à la conduite) (2 cas) • Problème de signalisation en amont d'une sortie <p><u>Autres éléments explicatifs :</u> circulation dense.</p> <p><u>Autres aspects :</u></p> <ul style="list-style-type: none"> • Détournement de l'attention de l'utilisateur lié à une tâche concurrente relative à la conduite • Ralentissement de la circulation • Dans deux cas, il est probable que l'utilisateur « B » circule à faible allure ou ralentit

Figure 4-4- Exemple de scénario d'accidents présenté selon le modèle séquentiel [Brenac et Fleury, 1999].

D'autres modèles de l'accident peuvent être utilisés pour l'élaboration « manuelle » de STA. Dans [Van Elslande et Alberton, 1997], par exemple, on met l'accent sur la composante humaine en définissant *des scénarios de génération de l'erreur chez l'opérateur*. Un scénario, dans ce cas, regroupe les accidents qui présentent une similitude au niveau des défaillances fonctionnelles.

L'élaboration des STA en se basant sur une approche purement experte présente certaines limites :

- L'examen au cas par cas est coûteux en terme de temps et d'expertise. Ceci est dû au nombre important d'accidents nécessaire à l'analyse, au nombre important d'attributs caractérisant chaque accident et à la variété des types des données (données structurées, entretiens, reconstructions et photos) ;
- Comme nous l'avons noté au début de la section 4.4, un scénario combine généralement un modèle abstrait et des scènes réelles. En accidentologie, plusieurs modèles peuvent être utilisés pour la construction des scénarios. Par exemple, c'est le modèle séquentiel qui a été utilisé dans [Brenac et Fleury, 1999], alors que c'est le modèle de traitement d'information

qui a été utilisé dans [Van Elslande et Alberton, 1997]. Chacun de ces modèles permet de mettre l’accent sur l’un des aspects de l’accident. La combinaison de ces différents aspects est souvent nécessaire pour fournir une analyse multi-vues de l’accident afin d’appréhender sa complexité. Cependant, étant donné le coût élevé de la construction « manuelle » des STA, leur élaboration selon les différents modèles devient difficilement réalisable ;

- Les STA peuvent varier selon l’objectif de l’étude. A cause de leur coût élevé, une élaboration à la demande est quasi impossible ;
- En partant des mêmes cas d’accidents, il est possible que des experts différents aboutissent à des représentations différentes et non partageables. Il est possible aussi, qu’un même expert fournisse des résultats différents s’il effectue deux fois le même travail.

4.4.2 Elaboration automatique des STA

Nous avons identifié deux types de techniques de construction automatique de STA : les techniques d’analyse discriminante (cf. Définition 4.10) et les techniques de classification automatique (cf. Définition 4.11).

Définition 4.10- Analyse discriminante (classification en anglais)

Il s’agit d’une famille de techniques qui couvrent deux aspects : le premier aspect, descriptif et explicatif, consiste à déterminer les caractères discriminants d’une population répartie en groupes, le second est décisionnel et aide à affecter un nouvel individu à un groupe. En analyse discriminante, les classes sont prédéfinies, mais elles sont élaborées (automatiquement ou par expertise) en fonction d’une seule variable, appelée variable cible. L’objectif est de trouver des descriptions discriminantes pour chacune des classes en fonction d’autres variables, appelées variables explicatives (d’où le premier aspect descriptif et explicatif de la méthode). Après la caractérisation des classes, les nouveaux individus peuvent être affectés aux classes existantes (d’où l’aspect décisionnel de la méthode). Les méthodes d’analyse discriminante remontent aux travaux de [Ficher, 1936].

Définition 4.11- Classification automatique (clustering en anglais)

Il s’agit d’une famille de techniques destinées à la création de classes homogènes de n individus mesurés par d variables. Les classes ne sont donc pas prédéfinies comme dans le cas des techniques d’analyse discriminante. De plus, plusieurs variables sont utilisées simultanément pour effectuer la création de classes. L’objectif d’une classification automatique est de minimiser les distances intra-groupes et d’augmenter les distances intergroupes. Les premiers concepts et préalables fondamentaux sont apparus dans les années soixante. Parmi les premiers manuels publiés, on cite [Anderberg, 1973; Benzécri, 1973; Lerman, 1970; Sneath et Sokal, 1973] (Il faut noter que le terme classification en anglais signifie analyse discriminante en français).

La construction des STA en utilisant des techniques d’analyse discriminante (cf. Définition 4.10) consiste à définir des groupes d’accidents en se basant uniquement sur un seul attribut (i.e. *variable cible*). Dans [Sohn et Lee, 2003; Sohn et Shin, 2001], par exemple, la variable cible qui a été sélectionnée était « la sévérité de l’accident ». Trois classes ont été donc définies : « mort et blessure grave », « blessure légère » et « dégât matériel ». 22 variables explicatives ont été utilisées. Trois techniques de segmentation ont été appliquées et comparées : les réseaux de neurones, les arbres de décision et la régression logistique. Dans [Najm *et al.*, 2001], [Chovan *et al.*, 1994] et [Ragland et Zabyshny, 2003], c’est l’attribut « manœuvre du conducteur à l’intersection » qui a été choisi comme variable cible. Des configurations en fonction de cette variable ont été alors définies avant d’appliquer des techniques d’analyse discriminantes pour trouver une caractérisation de chacune de ces configurations.

L'inconvénient majeur de la construction de STA en utilisant l'analyse discriminante réside dans le fait que non seulement on part d'une classification prédéfinie, mais en plus cette classification est basée sur une seule variable qu'est la *variable cible*. Etant donnée la complexité de l'accident de la route (cf. section 3.5), son analyse nécessite de combiner plusieurs variables. Kononov et Janson (2002), par exemple, montrent qu'il est impossible de construire des connaissances en accidentologie pour développer des contre-mesures si on ne combine pas systématiquement la fréquence des accidents et leur sévérité à un grand nombre de facteurs. Chen et Jovanis (2002) montrent que les approches statistiques classiques atteignent leurs limites dès que le nombre de variables devient très important, d'où l'importance des méthodes de data mining ou d'ECD telles que les méthodes de classification automatique. Cameron (1997) montre l'efficacité de ces méthodes, dites de clustering aussi, pour comprendre les mécanismes accidentels et définir des contre-mesures. Plusieurs études récentes soutiennent cet avis [Bar et Page, 2002; Filip et al., 2004; Geurts et al., 2003; Page, 2002; Page et al., 2004].

Dans l'élaboration de STA en utilisant des techniques de classification, il n'y a pas de classes prédéfinies. Il s'agit de partir d'un ensemble de données contenant n individus (accidents) caractérisés par p attributs et de les regrouper en classes d'accidents qui sont similaires. La majorité des travaux que nous avons identifiés sur cette problématique dans le domaine de l'accidentologie se focalise sur les problèmes algorithmiques (e.g. complexité combinatoire et qualité statistique des classes) [Filip et al., 2004; Geurts et al., 2003]. Pour nous positionner par rapport à ces travaux, nous nous focalisons sur d'autres aspects qui sont relatifs à *l'incorporation des connaissances expertes dans le processus de classification* (cf. notre positionnement à la section 2.6).

Dans [Nasri, 2003], nous avons appliqué plusieurs méthodes de classification (cf. section 6.1.2) sur un même échantillon d'accidents afin de comparer leurs résultats. Nous nous sommes rendu compte qu'une bonne classification au sens statistique, n'est pas systématiquement un gage d'interprétabilité des classes. En effet, le fait que les *distances intra-groupes* soient minimales et que les *distances inter-groupes* soient maximales, n'est pas suffisant pour que les classes soient interprétables et donc utilisables par l'expert.

En se basant sur cette étude, nous avons pu distinguer deux notions différentes : la *validation des classes* et *l'interprétation des classes*. Une classification est qualifiée de « *valide* » si les classes découvertes correspondent à des groupements statistiquement homogènes et séparés. Une classification est qualifiée d'« *interprétable* » si on arrive à associer la signification appropriée à chaque classe. Autrement dit, les classes ont une signification pour l'utilisateur et elles sont utilisables pour l'objectif pour lequel elles ont été créées.

Nous avons pu comparer la validité des classifications issues des différents algorithmes utilisés en utilisant des critères statistiques comme les distances intra et inter-groupes. Cependant, il nous était très difficile d'évaluer l'interprétabilité des classifications. Ceci est dû au fait que l'évaluation de l'interprétabilité nécessite des connaissances expertes. D'où le besoin d'identifier ces connaissances expertes et de les incorporer dans le processus de classification.

4.4.3 Combiner des approches expertes et automatisées pour élaborer des STA

L’approche automatisée pour l’élaboration des STA consiste en l’utilisation des techniques de classification automatique (cf. section 6.1.2). Combiner cette approche avec l’approche experte consiste à incorporer les connaissances expertes tout au long du processus de classification. Ceci est important notamment lors de la phase de sélection des attributs qui vont servir pour le calcul des distances et lors de la phase de validation et d’interprétation des classes. L’utilisation des différents modèles de l’accident (cf. section 4.3), par exemple, peut être une piste pour incorporer les connaissances du domaine dans le processus d’ECD.

Mais, dès qu’il s’agit de manipuler des connaissances expertes (les identifier, formaliser et opérationnaliser), la tâche devient rapidement complexe. Cette complexité est due aux différents aspects des connaissances expertes relatifs aux notions de multi-points de vue, subjectivité, difficulté d’identification (aspect implicite), difficulté de formalisation (risque de perte d’information) et difficulté d’utilisation. D’où notre recours à des méthodes de modélisation des systèmes complexes pour modéliser ces connaissances.

Dans la section 6.3, nous présentons un état de l’art sur la problématique d’incorporation des connaissances du domaine dans le processus d’ECD. Nous montrons le manque de techniques efficaces permettant cette tâche et nous consacrons le Chapitre 9 à proposer une nouvelle technique se basant sur l’ingénierie des connaissances.

4.5. Synthèse et positionnement

Dans ce chapitre (Chapitre 4), nous avons commencé par présenter quelques approches en sécurité routière (sécurité primaire, secondaire, active, passive, etc.)(cf. section 4.1). Nous avons présenté ensuite les modèles d’analyse de l’accident de la route les plus connus (cf. 4.3). Dans la section 4.3, nous avons détaillé la méthode d’Etude Détaillée d’Accidents (EDA) développée et utilisée par l’INRETS pour la collecte et l’analyse des données d’accident sur le site de l’accident. Nous avons présenté aussi les différents types de données disponibles dans les EDA. La section 4.4 a été dédiée à la définition du concept scénario-type d’accidents (STA) ainsi que les différentes approches utilisées pour les construire. Nous avons présenté l’approche experte d’élaboration des STA appliquée dans la littérature ainsi que ses limites. Nous avons ensuite exploré les possibilités et les difficultés d’une élaboration automatisée des STA. A la fin, nous avons proposé une nouvelle approche qui combine des approches automatisées d’ECD, telle que la classification, et des approches expertes en incorporant les connaissances expertes dans le processus d’ECD.

Puisque cette proposition nécessite une connaissance du processus d’ECD ainsi que des techniques de modélisation et d’opérationnalisation des connaissances, nous consacrons les deux prochains chapitres à un état de l’art sur chacun de ces domaines.

Chapitre 5.

Modélisation des connaissances en ingénierie des connaissances

5.1. Connaissance : définitions et typologies

5.1.1 Définitions de la connaissance

Plusieurs définitions peuvent être trouvées dans la littérature pour définir la *connaissance*. En utilisant le triangle sémiotique, Jean-Louis Ermine utilise la conjonction S^3 : *Signe*, *Signifié*, et *Signifiant* pour définir ce qu'il appelle le *patrimoine de connaissances* [Ermine et al., 1996]. Il s'agit d'une conception de la connaissance comme un objet qui s'observe selon trois axes : *syntactique* (*signe, forme*), *sémantique* (*signifié, le sens*) et *pragmatique* (*signifiant, le contexte*). Cette vue selon le triangle sémiotique permet d'éviter « l'amalgame » entre *connaissance* et *information*. En effet, la connaissance ne peut se résumer à l'information, celle-ci n'est qu'un point de vue, une projection sur l'axe syntaxique qui ne peut s'y soustraire [Cantzler, 1996]. Dans une autre définition donnée par Ballay (1997), la connaissance n'est plus un objet passif et statique qu'on peut manipuler, mais elle devient un « agent » qui a un rôle : *la connaissance associe des événements et des expériences qui, évoqués par des signaux externes, provoquent des processus de résolution de problèmes* [Ballay, 1997]. Elle peut être considérée comme le produit de trois ressources : *notre expérience intime, les informations dont nous disposons et les savoirs que nous avons acquis* [Ballay, 1999]. Quant à Dunham (1996), pour définir la notion de connaissance, il s'est basé sur les travaux des biologistes Humberto Maturana et Francisco Varela [Maturana et Varela, 2002] qui ont développé une théorie de la cognition dans laquelle « *Connaître c'est faire* » ou « *connaître c'est agir* ». A partir de ces travaux, il affirme qu'il ne faut pas se contenter d'assimiler la notion de connaissance aux notions de compréhension et d'information et omettre ce qui est essentiel et qui est le fait de *produire des résultats* [Dunham, 1996].

Notre choix de ces quatre définitions [Ballay, 1997; Ballay, 1999; Cantzler, 1996] [Dunham, 1996]) qui bien qu’elles portent sur le même concept paraissent différentes, n’est pas un simple hasard. Il ne s’agit pas non plus de choisir LA définition de la connaissance parce qu’il n’y en a pas. Nous les avons choisies parce qu’elles nous paraissent complémentaires. En effet, chacune d’entre elles porte sur l’un des aspects systémiques de la connaissance : *ontologique* (ce qu’est la connaissance), *fonctionnel* (ce qu’elle fait), *génétique* (génétique ici a le sens de la genèse non celui de l’hérédité) ou *transformationnel* (d’où elle vient et qu’est ce qu’elle devient) et *téléologique* (le pourquoi) (voir les axiomes de la systémique dans §3.4). En se basant sur ces quatre aspects, nous donnons la définition suivante :

Définition 5.1- La connaissance

Il ne s’agit de donner « LA » définition de la connaissance. Nous proposons une vision systémique de la connaissance. Les trois niveaux de formalisation (*forme, sens et contexte*) de [Cantzler, 1996] insistent sur l’aspect ontologique de la connaissance. La conception de la connaissance qui « *associe des événements et des expériences qui, évoqués par des signaux externes, provoquent des processus de résolution de problèmes* » [Ballay, 1997] insiste sur l’aspect fonctionnel de la connaissance. La conception de la connaissance comme « *le produit de trois ressources : expérience, informations et savoirs acquis et le fait qu’elle évolue* » [Ballay, 1999] insiste sur son aspect génétique. La conception « *connaître c’est faire pour produire des résultats* » [Dunham, 1996] insiste quant à elle sur l’aspect téléologique de la connaissance.

5.1.2 Typologies de connaissances

En sociologie, D.Vinck, cité dans [Verzat, 2000], a présenté différents types de connaissances dans les organisations:

- *Faits et règles formelles* : correspondent aux règles explicites du type loi scientifique. C’est la seule connaissance qui puisse être véritablement dite déclarative, c’est à dire transmissible quel que soit le contexte et indépendante des buts et intentions des acteurs ;
- *Heuristiques* : règles empiriques explicites et pratiques standardisées, qui apparaissent éventuellement (mais rarement) dans les manuels ;
- *Savoir-faire manuels et perceptifs* : ils sont appris par imitation et intériorisés sans passer par la conscience (par exemple savoir-faire du vélo) ;
- *Savoir-faire culturels* : ils correspondent à la capacité à reconnaître un objet dans un environnement confus et à lire et comprendre les informations grâce à une expérience commune.

En ergonomie cognitive, Rasmussen (1983) identifie trois types de connaissances à travers lesquels il explique le fonctionnement cognitif de l’opérateur humain en prise de décision [Rasmussen, 1983] :

- *Connaissances* : face à une nouvelle situation, les informations sont identifiées comme des *symboles* à partir desquels l’opérateur humain réalise des inférences en mettant en œuvre un modèle mental du système auquel il est confronté. Un comportement basé sur les connaissances est un *comportement cognitif* ;

- **Règles** : les informations sont identifiées comme des *signes* et il s'agit de mettre en œuvre des procédures d'exécution stéréotypées antérieurement acquises, sans choix complexe entre plusieurs alternatives. Il s'agit d'un *comportement procédural* ;
- **Habilité** (« skills ») : l'information est saisie comme un signal qui déclenche une activité sensori-motrice automatisée acquise par la pratique. Il s'agit d'une *activité automatisée quasiment inconsciente*.

En intelligence artificielle, on distingue différents types de connaissances qui présentent certains points en commun avec les typologies présentées ci-dessus :

- **Connaissance procédurale** : elle représente comment un problème est résolu. Elle indique comment réaliser une tâche ;
- **Connaissance déclarative** : elle décrit ce qui est connu du problème (des énoncés décrivant les objets, les concepts, etc.) ;
- **Méta Connaissance** : elle décrit la connaissance sur la connaissance. Par exemple, quelle connaissance est adaptée à un moment donné pour résoudre une tâche donnée ;
- **Connaissance heuristique** : appelée aussi connaissance de surface. C'est une connaissance empirique acquise par un expert au fil de son expérience passée, elle sert à guider le raisonnement ;
- **Connaissance structurée** : décrit un modèle mental de l'expert sous forme de structure (concepts, sous-concepts, objets, etc.).

Newell distingue la *connaissance* et sa *représentation* (les symboles) [Newell, 1982] et il affirme qu'il « existe un niveau du système informatique au-dessus de celui des symboles et qui est caractérisé par la connaissance comme médium et le principe de rationalité comme loi de comportement » [Newell, 1982]. Newell appelle ce niveau le « *knowledge level* ». Dans la littérature francophone, ce niveau est appelé *niveau cognitif* ou *niveau de connaissance*. A ce *niveau cognitif*, un agent est considéré comme ayant un *ensemble d'objectifs*, un *corpus de connaissances* et un *ensemble d'actions*. Son comportement est prédit par le *principe de rationalité* : « le système sélectionne une action s'il possède la connaissance qu'un de ses buts peut être atteint au moyen de l'exécution de l'action » [Newell, 1982].

Le « *knowledge level* » de Newell tient donc compte des aspects de *résolution de problèmes* ainsi que de la *connaissance* mise en œuvre pour cette résolution. Il dissocie la *connaissance* de sa *représentation* et de son *implémentation*. C'est ainsi que la plupart des méthodes de génie cognitif sont orientées sur deux axes : celui dit de la *connaissance statique*, qui modélise les objets et les concepts du domaine considéré et celui dit de la *connaissance dynamique* qui modélise les stratégies d'utilisation de cette connaissance statique destinées à résoudre un ou plusieurs problèmes déterminés [Ermine et al., 1996].

En IC, Clancey (1983) était le premier à montrer l'importance de la modélisation au *niveau connaissance* pour le développement des Systèmes à Base de Connaissances (SBC) (cf. Définition 5). Dans une extension de la notion du « *knowledge level* » de Newell et dans un objectif de la rendre plus pratique en IC, Wielinga et al. (1994) proposent dans leur méthode CommonKADS (cf. §5.3.2) [Wielinga et al., 1994] une modélisation des connaissances selon trois points de vue : les

connaissances nécessaires (*domaine*), l’objectif à réaliser (*tâche*) et les actions nécessaires pour le réaliser (*inférence*) (cf. Chapitre 7 pour un exemple d’application de CommonKADS).

5.2. Ingénierie des connaissances : présentation générale

L’IC a été développée dans le domaine de l’Intelligence Artificielle (IA) et elle contient plusieurs champs de recherche en systèmes d’information, modélisation par objet, travail coopératif, Interface Homme/Machine (IHM), Raisonnement à Partir de Cas (RàPC), etc.

Edward A. Feigenbaum (1977) définit l’activité de l’IC comme « the art of building complex computer programs that represent and reason with knowledge of the world ».

Définition 5.2- Les méthodes d'ingénierie des connaissances

Elles « offrent en premier lieu des **concepts, méthodes et techniques** permettant d'**acquérir** et de **modéliser** les **connaissances** à partir de **sources complexes**. En second lieu, elles offrent des outils permettant d'**opérationnaliser** et de **valider** les **modèles qualitatifs** conçus lors de l'étape précédente en conservant une « **traçabilité** » entre les modèles au **niveau des connaissances** et les modèles pour l'**implémentation**. En troisième lieu, dans une démarche de conception spirale, le processus de validation doit être conçu pour prendre en compte les **retours des experts** et des **utilisateurs** afin d'**améliorer** les modèles qualitatifs élaborés dans la première étape et de progresser dans l'analyse du problème et la définition des méthodes élaborées pour le résoudre » [Charlet et al., 2000].

Selon [Studer et al., 1998], au début des années 1980, l’IC a été perçue comme un *processus de transfert* des connaissances qui consiste à *transférer et transformer des expertises* de résolution de problèmes en des programmes pour construire des Systèmes Experts (SE). Depuis les années 1990, l’IC est perçue comme un *processus de modélisation*. Ceci a été animé par une nouvelle perception plus constructiviste de la connaissance. On ne parle plus alors de *l'extraction des connaissances des têtes des experts* et de leur formalisation à l’aide de *règles de production*, mais de la *modélisation des connaissances expertes selon différents points de vue* [Charlet et al., 2000; Le Roux, 1994]. Cette évolution s’est traduite par l’apparition de ce qu’on appelle aujourd’hui les *Systèmes à Base de Connaissances* (SBC) ou *SE de seconde génération* [David et al., 1993] pour remplacer ce qu’on appelait SE. L’objectif d’un SBC n’est plus la simulation du processus cognitif de l’expert dans sa tâche de résolution de problème, mais de construire un modèle qui fournit les mêmes résultats que celui de l’expert. L’acquisition des connaissances n’est plus donc perçue comme un *processus de transformation de connaissances* sous forme d’un programme compréhensible par la machine, mais comme un *processus de construction d’un modèle* qui tient compte de la *subjectivité* et du *caractère relatif et implicite* des connaissances [Studer et al., 1998].

Définition 5.3- Système à Base de Connaissances (SBC)

Un SBC est un système sémiotique de manipulation d’inscription symbolique, dont le fonctionnement informatique doit permettre à un utilisateur d’interpréter et de comprendre le système dans le cadre de son activité et de ses usages, dans les termes du domaine [Charlet et al., 2000]. C’est donc un système informatique qui utilise des modèles de connaissances et des mécanismes de raisonnement afin de résoudre des problèmes qui nécessitent une grande expertise [Charlet et al., 2000].

En se basant sur la définition des méthodes d’IC (Définition 5.2), nous pouvons distinguer trois activités principales en IC : la *modélisation et l’acquisition des connaissances*, l’*opérationnalisation des modèles de connaissances*, et la *validation des méthodes, modèles et outils conçus et opérationnalisés*.

5.3. La modélisation des connaissances

La modélisation des connaissances est une activité centrale en ingénierie des connaissances [Charlet *et al.*, 2000]. La complexité de cette activité est due à la complexité de la tâche de *modélisation* (cf. §3.4) en général et à la nature complexe de la « *connaissance* » (cf. §5.1.1 et §5.1.2).

Définition 5.4- La modélisation

Est une action d'élaboration et de construction intentionnelle, par composition de symboles, de modèles susceptibles de rendre intelligible un phénomène perçu complexe, et d'amplifier le raisonnement de l'acteur projetant une intervention délibérée au sein du phénomène ; raisonnement visant notamment à anticiper les conséquences de ces projets d'actions. La modélisation est une activité intrinsèque à toute tâche humaine et elle est la base de construction de connaissances [Le Moigne, 1999].

Cette définition issue de la systémique (cf. §3.4) identifie trois objets dans une activité de modélisation : *l'objet à modéliser* (ou *l'objet observé*) qui est dans notre cas à complexité du processus de modélisation des connaissances peut être donc analysée au niveau de ces trois objets.

Les concepts de *modélisation* (Définition 5.4) et de *connaissance* (Définition 5.1) étant définis, nous utilisons notre vision systémique de la connaissance pour définir la *modélisation des connaissances* :

Définition 5.5- Modélisation des connaissances

C'est un processus de construction intentionnelle (résoudre des problèmes), à partir d'expérience (savoir-faire), informations et savoir, de modèles susceptibles de rendre intelligible (ayant un sens dans un contexte donné) cette connaissance, et d'amplifier le raisonnement de l'acteur projetant une intervention délibérée sur elle pour résoudre des problèmes.

Cette définition fait apparaître les deux composantes classiques d'un modèle *conceptuel de connaissances* qui sont présentés dans la définition suivante :

Définition 5.6- Modèle conceptuel des connaissances (MC)

Le MC d'un SBC est une description abstraite des connaissances indépendamment de l'implémentation. Il consiste en deux composantes : le modèle conceptuel de raisonnement, appelé aussi savoir-faire : c'est une description abstraite du raisonnement de l'expert lors de la résolution de problème. Le modèle conceptuel du domaine, appelé aussi savoir : il s'agit d'une représentation et structuration des connaissances du domaine étudié à travers des concepts et des relations entre ces concepts.

Le MC peut être utilisé pour guider l'acquisition des connaissances [Aussenac-Gilles *et al.*, 1992]. Les aspects relatifs à l'implémentation d'un SBC sont traités dans le *modèle opérationnel*.

Définition 5.7- Modèle opérationnel des connaissances

L'opérationnalisation d'un modèle consiste à traduire les connaissances identifiées grâce au MC dans des structures de données et la mise en place de mécanismes de mise en œuvre de ces connaissances. Le modèle opérationnel respecte souvent le « principe de correspondance structurelle » qui préconise que sa structure reflète celle du MC. Il permet ainsi de tester et d'affiner la modélisation.

Les travaux et les apports présentés dans ce mémoire s'inscrivent dans la famille de méthodes de construction du modèle conceptuel de connaissances. Nous focalisons donc notre état de l'art sur les méthodes de modélisation des connaissances. Un bref aperçu sur les langages de représentation et d'opérationnalisation des modèles de connaissances est présenté dans la section 5.4.

Plusieurs méthodes et techniques sont utilisées pour la modélisation des connaissances. Ces méthodes peuvent être regroupées en trois types d'approches : *approche ascendante*, *approche*

descendante et *approche mixte*. Nous présentons uniquement quelques exemples de chaque catégorie, sans chercher l’exhaustivité. Nous avons choisi des méthodes qui sont très connues dans la littérature et qui nous permettent de mieux positionner nos travaux.

5.3.1 Approche ascendante de modélisation

L’*approche ascendante (Bottom-Up)* ou *dirigée par les données (data-driven)* se base sur une étape d’élicitation des connaissances des experts (entretiens, analyse de documents, analyse des tâches, etc.) suivie d’une étape de conceptualisation. La méthode KOD (Knowledge Oriented Design) [Vogel, 1989] et la méthode MIKE (Model-based and Incremental Knowledge Engineering) [Angele et al., 1996] sont des exemples typiques de cette famille.

5.3.1.1 La méthode KOD (Knowledge Oriented Design)

C’est une méthode d’aide à l’extraction et à la modélisation des connaissances qui a été développée par Vogel [Vogel, 1989]. Elle consiste à élaborer trois modèles :

- Le **modèle pratique** : permet d’identifier les éléments de l’expertise. Ce modèle permet d’identifier et d’extraire plusieurs types de manifestations verbales recueillies en entretien avec l’expert : (1) les *taxèmes* : objets du monde physique, (2) les *actèmes* : des descriptions de changement d’état causé par un sujet sur un destinataire à l’aide de ressources et (3) les *schémèmes* : des schémas d’interprétation (déduction, induction) qui se manifestent par des contraintes et des inférences ;
- Le **modèle cognitif** : permet de structurer et valider les connaissances acquises. L’expert utilise : (i) des *taxinomies* : représentations hiérarchiques des entités du monde physique (e.g. les concepts), (ii) les *actinomies* : structuration des schémas mentaux de l’expert, permettant de répondre à des situations déjà connues, et (iii) les *schémas d’interprétation* : des aides à l’interprétation du réel. Ce sont des ensembles de règles d’inférence associés correspondant aux stéréotypes de raisonnement de l’expert, qui se déclenchent selon les observations du domaine ;
- Le **modèle informatique** : permet la mise en forme des connaissances dans un programme informatique. Il est constitué d’objets, de méthodes et de règles.

5.3.1.2 La méthode MIKE (Model-based and Incremental Knowledge Engineering)

Cette méthode a été développée par [Angele et al., 1996]. Elle recouvre toutes les étapes de construction d’un SBC que nous présentons ci-dessous et qui sont effectuées dans un processus en spirale :

- **Elicitation** : c’est la première étape de MIKE et elle consiste à faire des interviews avec les experts pour recueillir des connaissances sur le domaine et la tâche de résolutions du problème. Ces connaissances sont stockées dans des *protocoles de connaissances* ;
- **Interprétation** : elle consiste à structurer les connaissances recueillies dans l’étape précédente selon le modèle expert de CommonKADS (domaine, tâche et inférences) (cf.

5.3.2) dans un langage semi-formel. Cette étape aboutit à la construction du *modèle structurel* ;

- **Formalisation/opérationnalisation** : le modèle structurel est représenté dans le langage formel KARL (cf. section 5.4) ce qui donne le *modèle KARL* ;
- **Conception** : le modèle KARL, issu des étapes d'acquisition ci-dessus, représente tous les besoins fonctionnels du SBC. L'étape de conception permet de considérer les besoins non-fonctionnels du SBC tels que l'efficacité, la maintenabilité, etc. Cette étape donne le *modèle de conception* ;
- **Implémentation** : il s'agit d'implémenter le modèle de conception dans un environnement logiciel.

5.3.1.3 Avantages et limites de l'approche ascendante

L'un des avantages de l'approche ascendante est qu'elle laisse la liberté aux experts d'exprimer leurs perceptions ainsi que leurs tâches sans les contraindre [Bigand et al., 1999]. De plus, les modèles élaborés correspondent aux points de vue existants. Cette approche était largement utilisée pour l'élaboration des premières générations des systèmes experts. Aujourd'hui, elle reste particulièrement adaptée à des domaines où il n'existe pas de modèles de connaissances prédéfinis et où il manque de formalisation. L'exploration d'un nouveau domaine en est un exemple.

Cependant, l'approche ascendante présente un certain nombre d'inconvénients [Dieng et al., 1996] : le coût élevé en terme de temps et d'expertise dans le processus d'élicitation ainsi que dans le processus de validation aussi bien pour les experts du domaine que pour les ingénieurs de connaissances. La seconde difficulté de cette approche est relative à la gestion des conflits de perception entre les experts aussi bien sur le plan opérationnel (e.g. différentes façons d'effectuer une tâche de diagnostic de l'accident) que sur le plan conceptuel (e.g. différents modèles de l'accident utilisés, différents modes de raisonnement). Les modèles issus de cette approche souffrent aussi de manque d'abstraction, de généralité et sont donc difficilement réutilisables [Duribreux-Cocquebert et Houriez, 2000].

5.3.2 Approche descendante de modélisation

La deuxième famille de méthodes d'ingénierie des connaissances est *descendante (Top-Down)*, appelée aussi *approche dirigée par les modèles (Model-Driven)*. L'élaboration de modèles de connaissances dans cette approche consiste à trouver des modèles génériques préexistants (dans des bibliothèques, par exemple) et les adapter au domaine et à l'application concernés. Deux sous-approches peuvent être distinguées dans cette famille [Wielinga et al., 1994]: l'approche *sélectionner-et-modifier* et l'approche *modélisation-par-composition à partir d'éléments de bibliothèque*.

L'approche sélectionner-et-modifier

La méthode KADS-I [Schreiber et al., 1993] est un exemple typique de cette approche. Il s'agit de sélectionner, à partir d'une bibliothèque, un modèle complet (appelé modèle d'interprétation) et de l'adapter à l'application concernée. L'avantage de cette approche réside

dans la rapidité de l’élaboration du modèle. Cependant, le gros inconvénient réside dans la difficulté de trouver un modèle suffisamment générique pour pouvoir l’adapter.

L’approche modélisation-par-composition à partir d’éléments de bibliothèque

La méthode CommonKADS [Wielinga *et al.*, 1994] est l’exemple le plus connu de cette approche. Pour acquérir les connaissances, cette approche propose l’utilisation non seulement de modèles génériques [Wielinga *et al.*, 1994] préexistant, connus aussi sous le nom de modèles squelettiques [Motta *et al.*, 1990], mais aussi de *composantes génériques* de ces modèles. L’avantage de cette approche par rapport à l’approche *sélectionner-et-modifier* vient du niveau de granularité des composantes préexistantes dans la bibliothèque qui est suffisamment fin pour assurer une adaptation plus facile de ces composantes.

Parmi les approches ascendantes, nous avons choisi de présenter CommonKADS [Schreiber *et al.*, 1994a; Schreiber *et al.*, 1994c] et MKSM [Ermine *et al.*, 1996].

5.3.2.1 La méthode CommonKADS

CommonKADS [Schreiber *et al.*, 1994a; Schreiber *et al.*, 1994c] est la méthode de référence en IC pour la modélisation et le développement de systèmes à base de connaissances. Elle a été développée sur la base de la méthode KADS-I [Schreiber *et al.*, 1993] dans le contexte du Programme Européen ESPRIT. Pour élaborer un SBC, l’approche CommonKADS [Schreiber *et al.*, 1994c] propose une *phase de spécification* développant cinq modèles (organisation, tâche, agent, communication et expert) et une *phase de conception* pour le développement d’un modèle de conception :

- **Modèle de l’organisation** : il permet l’analyse des facettes majeures de l’organisation (structure, processus, personnel et ressources) afin de déterminer la faisabilité des solutions SBC et d’évaluer leur impact sur l’organisation ;
- **Modèle des tâches** : il permet d’identifier et d’analyser les tâches globales de l’organisation (les entrées et les sorties, les pré-conditions et les critères de performance, ainsi que les ressources et compétences nécessaires dans l’organisation) ;
- **Modèle d’agents** : il décrit les caractéristiques des agents qui peuvent être soit des humains, soit des systèmes d’information, soit d’autres entités capables d’effectuer une tâche. Les agents sont les exécuteurs des tâches. Le modèle décrit leurs compétences, leur autorité et leurs contraintes ainsi que les liens de communication ;
- **Modèle de communication** : il permet de modéliser le système de communication entre les agents dans lequel le SBC doit évoluer ;
- **Modèle conceptuel d’expertise** : il permet la représentation des connaissances de résolution de problème dans une application donnée ainsi que les connaissances du domaine nécessaires à cette résolution. Ce modèle représente trois types de connaissances : *connaissances du domaine*, *connaissances de la tâche* et *connaissances d’inférences*. Les *connaissances du domaine* sont décrites à l’aide du *schéma du domaine* (les concepts et les relations entre les concepts) et la *base de connaissances* (règles de résolution de problème) (cf. section 7.1). Les *connaissances d’inférences* sont décrites à l’aide des *rôles de connaissances* et

des *structures d'inférence* (cf. section 7.2). Les *connaissances de la tâche* sont décrites à l'aide du modèle de la tâche et méthode de la tâche (cf. section 7.3). Toutes ces notions sont définies avec des exemples dans le Chapitre 7;

- **Modèle de conception** : permet la spécification technique du système basé sur les cinq autres modèles en matière d'architecture, de plate-forme d'implantation, de modules logiciels, de modes de représentations et de mécanismes de calculs nécessaires pour mettre en application les fonctions spécifiées dans le modèle de connaissances et de communications.

5.3.2.2 La méthode MKSM

MKSM (Methodology for Knowledge System Management), appelée également MASK (Method for Analysing and Structuring Knowledge), est une méthode de gestion des connaissances qui a été élaborée au sein du CEA (Commissariat à l'Énergie Atomique) [Ermine *et al.*, 1996]. D'après son auteur J-L. Ermine, MKSM « se présente comme une méthode d'analyse des systèmes de connaissances, et a pour but de rendre ces systèmes intelligibles à ceux qui en sont les acteurs, afin qu'ils puissent mettre eux-mêmes en place leur propre système de connaissances » [Ermine *et al.*, 1996]. Elle n'aboutit donc pas forcément à la conception d'un système informatique. Mais, elle offre, comme dans le cas de CommonKADS, tout un cadre général de modélisation du *patrimoine de connaissances* dans une entreprise.

La méthode MKSM se base sur un *modèle général* de la connaissance au sens de la systémique. Ce modèle général se base dans un premier niveau sur le *triangle sémiotique* dont les trois axes correspondent aux trois points de vue suivants [Ermine *et al.*, 1996]:

- **Points de vue syntaxique de la connaissance** : il concerne *l'information*. Cette dernière représente la partie visible de la connaissance : la forme sous laquelle se traduit la connaissance, le code qu'elle utilise pour prendre forme, etc. ;
- **Point de vue sémantique de la connaissance** : il concerne la *signification* de l'information. C'est ce point de vue qui s'occupe du fond par rapport à la forme de la connaissance, de la structure qu'elle utilise pour prendre sens ;
- **Point de vue pragmatique** : il concerne le *contexte* dans lequel le sens de l'information se met en place. C'est donc ce point de vue qui s'occupe du système, de l'environnement que la connaissance utilise pour se mettre en contexte.

Le deuxième niveau du modèle général dans MKSM consiste à analyser chacun des points de vue du premier niveau selon le triangle systémique composé des axes : *ontologique* (structure), *fonctionnel* (fonction) et *génétique* (évolution) que nous avons définis dans la section 3.7. Ainsi, le modèle général offre neuf points de vue d'analyse de la connaissance.

En se basant sur ce modèle général, Ermine [Ermine *et al.*, 1996] propose un « cycle de modélisation » à cinq phases :

1. **Phase 1** : définir et délimiter le système sur lequel on désire travailler. Cela fournit une première cartographie des connaissances. Une représentation du système selon le triptyque

- système opérant/système d’information/système de décision* avec les flux cognitifs peut être utilisé ;
2. **Phase 2** : élaborer le *modèle du domaine* qui décrit l’ensemble des processus qui se déroulent dans le système étudié. Chacun des processus est décrit par un modèle *Source-Cible-Flux-Champ* (SCFC). Des liens de causalité et d’enchaînement entre les différents processus permettant de décrire des scénarios de processus ;
 3. **Phase 3** : élaborer le *modèle d’activité*. Il s’agit d’une analyse de l’activité du système qui produit ou utilise les connaissances, et cherche à replacer les connaissances du domaine dans le cadre d’une utilisation opérationnelle ;
 4. **Phase 4** : élaborer le modèle des concepts qui représente l’aspect « statique » de la connaissance. Un modèle des concepts dans MKSM est très proche du *schéma du domaine* dans CommonKADS (cf. section 7.1.1) ;
 5. **Phase 5** : élaborer le *modèle des tâches* qui décrit la *connaissance dynamique*. La connaissance dynamique est vue sous deux angles. Le premier est celui de la résolution de problèmes, il correspond au modèle de la connaissance de la tâche dans CommonKADS (cf. section 7.3). Le deuxième aspect de la connaissance dynamique est celui de la manipulation de la connaissance statique et il correspond au modèle de la connaissance des inférences dans CommonKADS (cf. section 7.27.3).

Les phases 2 et 3 permettent de décrire les connaissances selon le *point de vue du contexte* à travers la construction des modèles du domaine et d’activité. Les phases 4 et 5 permettent de modéliser les connaissances de *point de vue sémantique*. L’avantage de MKSM par rapport à CommonKADS réside dans le fait qu’elle met plus le point sur les processus notamment à travers le modèle *Source-Cible-Flux-Champ*. Cependant, elle est moins orientée conception de SBC, il n’y a pas de modèle de conception, par exemple, comme dans CommonKADS. De plus, les différences entre le modèle de la tâche et le modèle d’activité et entre le modèle du domaine et celui des concepts n’est pas très net à part le fait qu’ils décrivent la même chose à des niveaux de granularité différents. Le contexte dans CommonKADS est mieux formalisé à travers les modèles de *communication*, *d’organisation* et *d’agent* (voir ci-dessus pour les définitions).

5.3.2.3 Avantages et limites de l’approche descendante

L’intérêt majeur des approches descendantes, comme CommonKADS et MKSM, par rapport à l’approche *dirigée par les données* est qu’elles offrent un support pour guider l’acquisition des connaissances. L’autre intérêt réside dans la généralité et la réutilisabilité des modèles qui en résultent. Cependant, il est nécessaire de disposer d’une large bibliothèque de modèles et/ou de composantes génériques. De plus, l’adaptation d’un modèle existant à une application spécifique est généralement une tâche difficile. D’autres limites spécifiques à la méthode CommonKADS peuvent être trouvées dans [Bruaux, 2001] et [Garrido De Ceita, 2002].

5.3.3 Approche mixte de modélisation

Des approches mixtes ont été proposées dans [Duribreux-Cocquebert et Houriez, 2000] et [Aussenac-Gilles *et al.*, 1992] pour bénéficier des avantages des différentes approches présentées dans les paragraphes §5.3.1 et §5.3.2. Il s’agit de construire des modèles de connaissances en combinant des étapes issues des deux approches précédentes : 1) dans la première étape ascendante, il s’agit de recueillir des connaissances sur le domaine et la tâche de résolution de problème à partir des données sans qu’aucun modèle ne soit a priori choisi. 2) La deuxième étape est descendante et elle consiste à choisir, en tenant compte des connaissances préalablement recueillies, un schéma du modèle conceptuel. 3) La troisième étape consiste à instancier le schéma par les données collectées pour construire le MC. 4) La dernière étape consiste en l’opérationnalisation du MC (i.e. le rendre exécutable) en construisant le *modèle opérationnel* [Aussenac-Gilles *et al.*, 1992]. Parmi les exemples d’application de l’approche mixte dans le domaine de l’accidentologie, nous citons une étude menée dans le projet ACACIA (Acquisition des Connaissances pour l’Assistance à la Conception par Interaction entre Agents) et que nous présentons dans le paragraphe suivant.

5.3.3.1 Le projet ACACIA

ACACIA est un projet pluridisciplinaire au sein de l’INRIA¹ qui vise à développer des aides méthodologiques et logicielles pour l’acquisition, la modélisation et la capitalisation des connaissances pour la construction, la gestion et la diffusion de mémoire d’entreprise. Dans le domaine de l’accidentologie, une étude a été menée dans le cadre de ce projet en collaboration avec l’INRETS. Son objectif était l’élaboration d’une mémoire d’entreprise pour l’analyse de l’accident. Huit ingénieurs de connaissances et sept experts en accidentologie ont participé à cette étude qui a consisté en quatre étapes principales [Dieng *et al.*, 1996] :

- 1) **Elicitation des connaissances.** Le protocole utilisé a consisté en une collection des documents à l’INRETS (rapport, article, notes, etc.), des entretiens non directifs avec chacun des experts pour décrire son travail (ses tâches, ses activités, etc.), résolution individuelle de l’analyse d’un cas d’accident et une résolution collective de l’analyse d’un cas d’accident ;
- 2) **Analyse des connaissances élicitées et élaboration de modèles de connaissances.** Elaboration de modèles de connaissances : un modèle par expert, un modèle par type d’expert et des modèles théoriques. Nous avons remarqué que certains modèles portent sur le processus cognitif des experts lors de la résolution de la tâche d’analyse d’accidents. D’autres, sont propres à l’accident et utilisé par les experts pour l’analyse de l’accident. Exemple : *le modèle CVE* (cf. section 4.3.1), *le modèle de traitement d’information* (cf. section 4.3.3) et *le modèle séquentiel* (cf. section 4.3.2). Pour les modèles relatif à la tâche d’analyse de l’accident, des *modèles génériques de diagnostic*, de *planification*, de *conception*, etc. ont été sélectionnés et adaptés de la bibliothèque de CommonKADS [Breuker et Van de Velde, 1994]. Un *modèle expert* et un *modèle d’agent* ont été alors construits (cf. les définitions à la section 5.3.2.1) ;

¹ INRIA : Institut National de Recherche en Intelligence Artificielle.

- 3) **Validation des modèles de connaissances obtenus.** Deux types de validation ont été conduites. La première validation des modèles a été effectuée entre les ingénieurs de connaissances. La deuxième a été effectuée avec les accidentologues : les modèles ont été présentés et expliqués aux accidentologues qui ont donné leur avis. Leurs remarques ont été ensuite intégrées pour améliorer les modèles ;
- 4) **Construction d’une mémoire d’entreprise.** L’ensemble des modèles obtenus ainsi que les liens entre ces différents modèles forment une mémoire d’entreprise.

5.3.3.2 Le besoin de modèles pour guider l’élucidation des connaissances

Dans ce paragraphe, nous discutons des limites des approches de modélisation de connaissances (ascendantes, descendantes ou mixtes) concernant l’étape d’élucidation de connaissances. Dans le paragraphe 5.3.1.3, nous avons présenté les difficultés relatives à cette étape et nous avons vu dans le paragraphe 5.3.2.3 comment les approches descendantes permettent de résoudre une partie de ces limites en utilisant des modèles génériques pour guider la phase d’élucidation et d’acquisition des connaissances en général. Nous montrons ci-dessous que pour certains types de connaissances, notamment le *schéma du domaine* (cf. §5.3.2.1), nous n’avons pas trouvé de méthode permettant de guider leur élucidation.

Notre remarque concerne essentiellement le *modèle expert* (cf. §5.3.2.1). Comme on peut le remarquer, dans le modèle expert on distingue des connaissances relative à la tâche de résolution du problème (les règles, les inférence et les tâches) et des connaissances propres au domaine, indépendantes de la tâche, mais utilisées par cette tâche (e.g. les concepts du schéma du domaine). Prenons l’exemple du modèle CVE (cf. section 4.3.1) : ce modèle est utilisé par les accidentologues pour élaborer des STA (cf. section 4.4.1), mais il est indépendant du processus d’élaboration des STA (i.e. sélection des attributs, classification, interprétation).

Ce que nous avons constaté dans les approches de modélisation de connaissances actuelles est l’existence de modèles génériques (e.g. de diagnostic, de planification, etc.) pour guider la construction de modèles de résolution de tâches dans un domaine donné (e.g. analyse de l’accident, élaboration des STA, etc.). Cependant, nous avons noté l’absence de modèles pour guider l’élaboration et/ou l’identification de modèles du domaine tels que le modèle CVE ou le modèle séquentiel. La construction et/ou l’identification de ces modèles reste essentiellement basée sur l’élucidation sans qu’elle soit guidée par un modèle donné. L’étude menée dans le cadre d’ACACIA en est une bonne illustration (cf. §5.3.3.1).

A cause de l’absence de méthodes pour guider l’élucidation des connaissances du domaine, nous avons constaté un risque d’incomplétude des représentations obtenues. En effet, comme la construction du modèle est basée sur l’élucidation, si l’un des experts ayant un point de vue particulier ne participe pas à la phase d’élucidation, ce point de vue risque de ne pas apparaître dans le modèle final. Prenons l’exemple du modèle CVE : si parmi les experts qui ont participé dans ACACIA, par exemple, aucun n’utilise ce modèle pour l’analyse de l’accident, ce modèle n’aurait pas été identifié. Les modèles élaborés sont donc très dépendants des experts qui ont participé à leur élaboration. De plus, si dans l’entreprise, où les modèles sont à construire, il manque des

compétences ou des points de vue existant ailleurs et qui sont importants pour la tâche de résolution de problème, ces points de vue ne peuvent pas être représentés dans les modèles élaborés par élicitation. Nous revenons à ce point avec des exemples dans le paragraphe 8.6.2.3.

Pour nous situer par rapport aux approches existantes de modélisation de connaissances, nous proposons une approche mixte, mais qui permet de guider la phase d'élicitation des connaissances du domaine par un méta-modèle (cf. ASMEC, Chapitre 8).

5.4. La représentation et l'opérationnalisation des modèles de connaissances

D'après [Charlet *et al.*, 2000], l'évolution des langages de représentation et d'opérationnalisation des connaissances a suivi celle de l'IC (cf. section 5.3). En effet, au début des années 80, avec la perspective de « transfert des connaissances », la représentation des connaissances se basait essentiellement sur les langages de *règles de production*. Ensuite, avec la deuxième génération des SE [David *et al.*, 1993] et la perspective de « construction coopérative de modèles de connaissances », de nouvelles méthodes d'IC comme KADS-I et CommonKADS [Schreiber *et al.*, 1994a; Schreiber *et al.*, 1994c; Wielinga *et al.*, 1994] (cf. §5.3.2.1), distinguaient deux types de connaissances : « la connaissance du domaine » et « la connaissance de la tâche ». Les règles de production atteignent alors leurs limites pour pouvoir représenter et spécifier ces deux types de connaissances. De nouveaux langages plus déclaratifs sont apparus pour remédier à ce problème. Ceci a été accompagné par une forte inspiration du domaine du génie logiciel en définissant le cycle de développement d'un SBC distinguant entre le « modèle conceptuel » et le « modèle opérationnel » (cf. Définition 5.6 et Définition 5.7).

Dans [Charlet *et al.*, 2000], on distingue des *langages de modélisation* reposant sur une notation informelle des connaissances et des *langages de représentation*, reposant sur une syntaxe définie par des règles formelles. Une autre distinction est donnée dans [Fensel et Van Harmelen, 1994] entre des *langages de représentation* (dits aussi de modélisation ou de formalisation), des *langages d'opérationnalisation* et des *langages de représentation et d'opérationnalisation*. Nous avons choisi de faire la distinction la plus courante : celle qui distingue des *langages de modélisation* (ou de représentation ou de formalisation) utilisés, en amont, pour guider l'acquisition des connaissances lors de la construction du MC et des *langages d'opérationnalisation* utilisés, en aval, pour l'élaboration du modèle opérationnel. Dans cette section, nous présentons quelques exemples de chaque catégorie sans entrer dans les détails. Pour une comparaison détaillée des différents langages, voir [Fensel et Van Harmelen, 1994; Kassel *et al.*, 2000; Trichet, 1998].

Langages de représentation des connaissances

Les *langages de représentation* proposent des constructions permettant d'exprimer les différentes connaissances à un niveau abstrait, indépendant de l'implémentation. Des langages de modélisation « classiques » comme SADT [Marca et McGowan, 1987], UML [Booch *et al.*, 1999] et les Graphes Conceptuels (GC) [Sowa, 1984] peuvent être utilisés pour la modélisation et la représentation des connaissances. (cf. [Dieng, 1997; Dieng *et al.*, 1996] pour l'utilisation des GC et [Bruaux *et al.*, 2003] pour l'utilisation de SADT).

D’autres langages plus spécialisés sont utilisés pour la modélisation et la représentation des connaissances. Certains sont non-formels, comme MONA [Aussenac-Gilles et Matta, 1994], CML (Conceptual Modeling Langage) [Schreiber et al., 1994b] qui offrent une première spécification d’un MC. D’autres sont formels comme (ML)² [Van Harmelen et Balder, 1992] et FORKADS [Wetter, 1990]. Ils offrent un niveau de formalisation plus fin afin de réduire les ambiguïtés demeurant dans un modèle informel et faciliter le passage au modèle opérationnel [Kassel *et al.*, 2000]. Dans nos travaux, nous utilisons, dans le Chapitre 7, le langage CML [Schreiber *et al.*, 1994b] pour la représentation des connaissances du domaine. Une présentation détaillée de ce langage est donc présentée dans le Chapitre 7.

Langages d’opérationnalisation des connaissances

Trichet [Trichet, 1998] note que peu de travaux d’acquisition des connaissances portent sur l’opérationnalisation du MC et que la plupart portent sur la phase de modélisation : élaboration de méthodes, modèles et outils pour construire et raffiner des modèles conceptuels. L’auteur souligne l’intérêt d’utilisation de langage d’opérationnalisation pour le raffinement et la maintenance et du MC et du SBC.

Les langages de programmation « classiques », comme C, C++, LISP, etc., peuvent être utilisés pour opérationnaliser des modèles conceptuels de connaissances. Cependant, de nombreux travaux ont montré l’intérêt d’utiliser des langages spécialisés, dits langages d’opérationnalisation [Tchounikine et al., 2000; Trichet, 1998]. Par exemple, le principe de correspondance structurelle entre le MC et son implémentation (i.e. le modèle opérationnel) permet de maintenir en même temps le SBC et le MC. En effet, toute modification du MC peut facilement être reportée dans le système et inversement.

Parmi les langages spécialisés, nous citons KARL [Fensel et al., 1991], OMOS [Linster, 1992], MODEL-K [Karbach et VoB, 1993], ZOLA [Istennes, 1997; Tchounikine *et al.*, 2000], LISA [Jacobdelouis et Krivine, 1995], DEF-* [Kassel *et al.*, 2000], OCML [Motta, 1998], etc. Pour une comparaison détaillées de ces différents langages, le lecteur peut se reporter aux travaux de [Trichet, 1998] et [Fensel et Van Harmelen, 1994].

5.5. La notion de point de vue en modélisation des connaissances

Plusieurs travaux ont été menés sur la notion de *point de vue* [Carré et Dekker, 1990; Labidi, 1995; Marcaillou, 1995; Ribière, 1999]. C’est une notion polysémique dont la définition change selon le domaine (conception, base de données, science cognitives, etc.). Un excellent état de l’art sur les différentes définitions de cette notion ainsi que des comparaisons entre elles sont présentées dans [Ribière, 1999].

Dans un objectif de catégoriser ces définitions, Ribière (1999) distingue deux notions : la notion de **perspective** et la notion **d’opinion**.

- **Perspective** : il s’agit de point de vue qui fait référence à « *un placement par rapport à l’objet* ». on entend ici un placement au sens cognitif. En d’autres termes, il s’agit « *d’une position conceptuelle par rapport à un objet, cette position servant à en donner une description particulière* »

[Rivière, 1999]. Cette description est une abstraction partielle de l'objet permettant d'accéder à un sous-ensemble d'informations [Marcaillou, 1995]. En se basant sur ces définitions, les sous-systèmes d'un système de modélisation, définis par la *cybernétique de second ordre* (cf. section 3.6), sont des points de vue et ils appartiennent à la catégorie « perspective ». En effet, *l'objet observé*, *l'observateur* et le *contexte d'observation* sont des abstractions d'un système de modélisation et des façons de voir ce système afin de ne prélever que certaines informations parmi tout un ensemble. Les axes systémiques que nous avons présentés dans la section 3.7 sont, eux aussi, des points de vue appartenant à la catégorie « perspective ». L'analyse d'un système selon les axes *ontologique*, *fonctionnel*, *transformationnel* et *téléologique* n'est autre qu'un mécanisme d'accès à un sous-ensemble d'informations sur le système ;

- **L'opinion** : « *par comparaison avec la perspective, l'opinion fait appel à bien plus qu'un placement conceptuel, il s'agit d'un avis donné ou d'une évaluation sur un objet* » [Rivière, 1999]. La notion d'opinion est souvent en relation étroite avec une personne (e.g. un expert) et elle tient compte de son expérience, ses connaissances, ses tâches, etc. Prenons l'exemple des trois points de vue de la cybernétique < *objet observé*, *observateur*, *contexte d'observation* >, ils forment une sorte de patron (pattern en anglais) qui peut être utilisé par plusieurs personnes pour effectuer une analyse. Les résultats, qui peuvent être perçus comme des instanciations de ce patron, dépendent de ces personnes (i.e. de leurs expériences, de leurs disciplines, etc.), il s'agit donc d'opinions. Ainsi, l'instrumentation des points de vue systémique par des modèles, l'instrumentation de ces modèles par des concepts et la caractérisation des concepts par des attributs correspondent à des opinions (cf. ASMEC, Chapitre 8).

Pour nous positionner par rapport aux travaux sur la notion de point de vue, nous n'envisageons pas d'élaborer un *système de gestion de point de vue* comme dans les travaux de [Carré et Dekker, 1990] et [Marcaillou, 1995]. Nous ne cherchons pas non plus à montrer l'utilisation d'une technique ou d'un formalisme de représentation tels que UML ou les Graphes Conceptuels (GC) pour la représentation multi-points de vue d'un objet comme dans [Rivière, 1999]. Par rapport aux travaux sur les applications multi-expertes ou multi-disciplinaires, notre recherche porte sur les *applications inter-disciplinaires*. En effet, [Labidi, 1995] distingue deux classes d'applications multi-disciplinaires : (i) les applications inter-disciplinaires où les liens entre les différentes disciplines est fort et demandent un vrai travail coopératif et (ii) les applications pluri-disciplinaires où il y a peu d'interactions entre les disciplines et où la multi-expertise consiste en une juxtaposition des résultats des différents spécialistes mono-disciplinaire. L'analyse d'un cas d'accident dans les EDA ne consiste pas à juxtaposer des résultats issus de différents experts, mais elle est le résultat de discussion entre les différents experts pour reconstruire le scénario du déroulement le plus proche de la réalité.

Dans ce mémoire, nous proposons de résoudre le problème de conflit de points de vue dans un contexte multi-experts en développons des points de vue (au sens de perspectives) suffisamment générales et génériques (cf. ASMEC, section 8.4) pour pouvoir intégrer les différents points de vue existants (au sens d'opinions). Pour atteindre cet objectif, nous utilisons l'approche systémique pour la modélisation des connaissances dans un domaine donné.

5.6. La validation des connaissances acquises

La problématique d’évaluation des modèles de connaissance en IC est, à notre connaissance, très rarement étudiée dans la littérature. Ceci est peut être dû à la complexité de la tâche d’évaluation en général qui est accentuée par la complexité de l’objet à évaluer (i.e. la connaissance). L’un des rares travaux d’évaluation des connaissances en IC est celui de [Wielinga *et al.*, 1994] où cinq critères ont été définis :

La complétude

Il s’agit de vérifier si tous les éléments de la connaissance ont été recueillis et formalisés : les concepts, les relations entre les concepts, les tâches, les inférences, les rôles de connaissances, etc.

L’exactitude

Il s’agit de vérifier si les différents modèles de CommonKADS ainsi que les composantes de ces modèles sont interprétées correctement. En ce qui concerne les inférences, par exemple, il s’agit de vérifier si les définitions des inférences correspondent à celles existantes dans la bibliothèque des inférences. Vérifier si les entrées et les sorties de chacune des inférences sont correctes, etc. En ce qui concerne les concepts, il s’agit de vérifier si les relations hiérarchiques entre les concepts sont correctes.

Non-contradiction

Des contradictions peuvent apparaître au niveau de l’ordre des tâches, des entrées/sorties au niveau des inférences et des relations hiérarchiques entre les concepts. Il s’agit donc de détecter et de résoudre ces contradictions.

Non-ambiguïté

Il faut éviter les ambiguïtés dans les noms des concepts, tâches et inférences. Deux éléments différents doivent avoir des noms différents.

Cohérence

D’autres aspects de la cohérence, comme ceux relatif à la terminologie, doivent être définis et vérifiés.

Ces critères définis dans [Wielinga *et al.*, 1994] nous ont paru importants, mais insuffisants pour l’évaluation des connaissances acquises. En effet, nous avons identifié d’autres aspects dont l’évaluation nous a paru nécessaire (cf. section Chapitre 10). De plus, nous avons constaté que ces critères portent sur des « objets » différents qui sont : la *connaissance*, sa *représentation* et son *implémentation*. Certains critères, comme l’*exactitude*, portent sur le contenu de la connaissance, alors que la *non-ambiguïté* porte sur la représentation. Nous consacrons une section de ce mémoire à la problématique d’évaluation des connaissances et des modèles de connaissances (cf. section Chapitre 10).

5.7. Synthèse et positionnement

Ce chapitre a été consacré à un état de l’art qui couvre plusieurs aspects de la connaissance : définitions, typologies, modélisation, représentation, opérationnalisation et évaluation.

Après avoir présenté quelques définitions de la littérature qui nous ont paru pertinentes (cf. section 5.1), nous avons proposé une définition systémique de la connaissance selon les axes suivant : ontologique (*ce qu’elle est*), fonctionnel (*ce qu’elle fait*), transformationnel (*ce qu’elle devient*) et téléologique (*ce qu’on fait avec*). Ensuite, nous avons présenté quelques typologies de la connaissance en sociologie, en ergonomie, en IA et IC.

En ce qui concerne la *modélisation des connaissances* (cf. section 5.3), nous avons comparé des *méthodes ascendantes* ou *dirigées-par-les-données* comme KOD [Vogel, 1989] et MIKE [Angele *et al.*, 1996] et des méthodes descendantes ou dirigées-par-les-modèles comme CommonKADS [Schreiber *et al.*, 1994a] et MKSM [Ermine *et al.*, 1996]. L'avantage de l'approche ascendante est qu'elle laisse la liberté aux experts d'exprimer leurs perceptions ainsi que leurs tâches sans les contraindre. Les modèles élaborés correspondent donc aux points de vue existants. Cependant, elle est généralement coûteuse en terme de temps et d'expertise et les modèles sont dépendants des experts donc difficilement généralisables. L'approche descendante quant à elle offre un support pour guider l'acquisition des connaissances et ses modèles sont réutilisables. Cependant, il est nécessaire de disposer d'une large bibliothèque de modèles et/ou de composantes génériques. De plus, l'adaptation d'un modèle existant à une application spécifique est généralement une tâche difficile. Dans ce mémoire, nous utilisons une approche mixte (cf. section 5.3.3).

En ce qui concerne la représentation et l'opérationnalisation des connaissances (cf. section 5.4), nous avons noté l'évolution des langages de règles vers de nouveaux langages capables représenter les différents types de connaissances (i.e. connaissances du domaine et connaissances de la tâche). Nous avons présenté brièvement un aperçu sur les différents langages existants.

En ce qui concerne la validation des connaissances (cf. section 5.5), nous avons noté qu'il s'agit d'une problématique très rarement étudiée en IC. De plus, des critères comme ceux proposés dans [Wielinga *et al.*, 1994] portent dans leur évaluation sur des concepts différents : la *connaissance*, sa *représentation* et son *implémentation*.

Chapitre 6.

Extraction des connaissances de données (ECD)

6.1. Processus d’ECD

Aujourd’hui avec l’informatisation des saisies de données (utilisation des codes à barres, informatisation des transactions, etc.) et la puissance des systèmes de collecte de ces données (satellites, ordinateurs, etc.), des grandes BD (Bases de Données) sont construites et ne cessent de s’agrandir. L’exploitation de ces millions de données en management, en administration, en médecine, en géologie, en biologie et dans beaucoup d’autres domaines fait appel à des techniques d’extraction de connaissances de données.

Trois étapes principales peuvent être distinguées dans un processus d’ECD : *préparation de données*, *application de techniques de data mining* et *interprétation des résultats* (voir pour plus de détails [Brachman et Anand, 1996; Fayyad *et al.*, 1996 ; Han et Kamber, 2000]).

6.1.1 Préparation des données

La qualité des résultats d’un processus d’ECB dépend en grande partie de la qualité des données utilisées, d’où l’importance de l’étape de préparation de ces données [Fayyad *et al.*, 1996]. D’après [Famili *et al.*, 1997], le prétraitement de données consiste en toute action effectuée sur les données avant l’application d’une technique de data mining (e.g. la classification automatique, le classement et les règles d’association). C’est essentiellement une transformation T qui transforme les vecteurs de données brutes X_{ik} en un jeu de nouveaux vecteurs de données Y_{ij} plus utiles que X_{ik} , éliminant au moins un problème des X_{ik} et préservant l’information. En effet, les données initiales peuvent être incomplètes, bruitées, aberrantes et incohérentes [Famili *et al.*, 1997 ; Han et Kamber, 2000 ; Pyle, 1999]. Un processus de pré-traitement et de préparation de données peut être décrit selon les étapes suivantes [Famili *et al.*, 1997; Han et Kamber, 2000] :

- La *nettoyage de données* : cette tâche consiste en la détection et la suppression des erreurs, du bruit et de l'incohérence de données pour améliorer leur qualité [Rahm et Do, 2000] ;
- L'*intégration de données* : cette tâche combine des données de sources multiples, détecte et résout des conflits de valeurs [Calvanese et al., 2001] ;
- La *transformation de données* : cette tâche consiste en l'extraction/construction de nouvelles variables (features) pour fournir une nouvelle représentation des données adéquate à l'application, au domaine et à l'objectif de l'étude [Hellerstein, 1997]. Plusieurs méthodes telles que l'agrégation, la généralisation, la normalisation permettent l'extraction de ces variables [Han et Kamber, 2000] ;
- La *réduction de données* : le coût de mesure et l'exactitude des résultats d'une technique de data-mining sont les deux raisons principales pour maintenir le nombre de données le plus faible possible [Hellerstein, 1997]. Mais cette réduction peut causer une perte d'information. Il s'agit donc de trouver un compromis. Parmi les stratégies de réduction de données, nous citons l'agrégation par les cubes de données, la compression de données, la discrétisation et la génération de hiérarchie [Han et Kamber, 2000].

6.1.2 Data mining

Les concepts de data mining et d'ECD sont parfois confondus et considérés comme synonymes. Mais, la définition la plus partagée du concept de data mining le voit comme un élément essentiel intervenant dans le processus ECD. L'équivalent de ce terme en français est « *fouille de données* ». Le data mining fait appel à un lot de méthodes issues de la statistique, de l'analyse des données, de la reconnaissance des formes, de l'intelligence artificielle ou de l'apprentissage automatique.

Parmi les méthodes les plus connues du data mining, nous citons les *règles d'association* [Agrawal et al., 1993], les méthodes de *classement* (ou d'apprentissage supervisé ou de classification supervisée) (cf. Définition 4.10) et les méthodes de *classification automatique* (ou d'apprentissage non-supervisée ou de classification non-supervisée). Dans [Jouini et al., 2003], nous avons effectué un état de l'art complet sur ces différentes approches.

Comme nous l'avons déjà noté dans la section 4.4.2, la construction de STA est un problème de classification et non de classement. D'où le choix de focaliser sur les méthodes de classification automatique. D'autres travaux de synthèse sur les méthodes de classification peuvent être trouvés dans [Everitt, 1993; Han et Kamber, 2000; Jain et Dubes, 1988; Jain et al., 1999] [Ben Ahmed *et al.*, 2002]. Des comparaisons des algorithmes de classification automatique sont élaborées dans [Halkidi et al., 2001; Zaït et Messatfa, 1997].

Définition 6.1- Classification automatique : formulation mathématique

Une formulation mathématique de la définition de la **classification automatique** peut être la suivante : Soient $D = \{1, 2, \dots, n\}$ un ensemble de n objets. L'objet $N^{\circ}i$ est désigné par $O_i = [x_{i1}, x_{i2}, \dots, x_{id}]$ ($i : 1 \dots n$), il est caractérisé par d attributs a_j dans un espace d -dimensions ($j : 1 \dots d$). X_{ij} représente la valeur de l'attribut a_j pour l'objet O_i . L'objectif d'une classification automatique est de partitionner D en k (k est un entier positif) groupes C_α similaires, $\alpha : 1 \dots k$.

$D = C_1 \cup \dots \cup C_k \cup C_{\text{outliers}}$, $C_i \cap C_j = \Phi$ pour $1 < i, j < k$. La recherche de ces classes est généralement guidée par la maximisation d'une fonction de similarité intra-classes et la maximisation d'une fonction de dissimilarité interclasses.

Dans [Ben Ahmed et al., 2002], nous distinguons des méthodes de classification hiérarchiques (ascendantes et descendantes), des méthodes de classification par partitionnement (agrégation autour de centres mobiles, K-means, K-modes, K-prototypes, nuées dynamiques, K-medoïdes), méthodes de classification à base de densité, méthodes de classification à base de maillage et méthodes de classification basée sur la co-occurrence.

Nous avons fait le choix, que nous justifions ci-dessous, de combiner un algorithme de classification par partitionnement et un algorithme de classification hiérarchique ascendante.

6.1.2.1 Classification par partitionnement

Les méthodes de partitionnement (partitional methods) sont connues dans la littérature française sous le nom de *méthodes non-hiérarchiques*. Il y a deux variantes dans cette famille : les algorithmes de classification autour de centres mobiles (algorithme de Forgy [Forgy, 1965], *k-means* [MacQueen, 1967], *nuées dynamiques* [Diday, 1971], *k-modes* [Huang, 1997b], *k-prototypes* [Huang, 1997a]) et les algorithmes de classification autour d'objets représentatifs (PAM (Partitioning Around Medoids) [Kaufman et Rousseeuw, 1990], CLARA (Clustering LARge Applications) [Kaufman et Rousseeuw, 1990], CLARANS (Clustering Large Applications based on RANdomized Search) [Ng et Han, 1994]).

La présentation des différents algorithmes de classification dépasse largement le cadre de ce mémoire (cf. [Ben Ahmed et al., 2002] pour une présentation détaillée). Nous présentons ici uniquement l'algorithme que nous avons utilisé, *k-means* [MacQueen, 1967] qui est le plus utilisé pour la classification par partitionnement. Il est fondé sur les quatre étapes présentées dans la Figure 6-1 et est illustré dans la Figure 6-2 :

1. Sélection k centres moyens initiaux ;
2. Calcul de la dissimilarité entre un objet et les centres moyens des groupes ; affecter l'objet au groupe dont le centre est le plus proche selon la distance d . Mettre à jour les noyaux du groupe après chaque affectation ;
3. Une fois tous les objets répartis dans les groupes, réévaluer la dissimilarité d'objets contre les centres moyens actuels. Si un objet est tel que son centre le plus proche appartient à un autre groupe, réattribuer l'objet à ce groupe et mettre à jour les noyaux des deux groupes correspondants ;
4. Répéter 3 jusqu'à ce que l'algorithme converge, c'est-à-dire jusqu'à ce qu'aucun objet ne change de position

Figure 6-1- Algorithme *k-means*.

La différence entre les algorithmes de classification autour de centres mobiles et les algorithmes de classification autour d'objets représentatifs se situe au niveau de la définition des centres des classes.

Dans la première famille le centre est un objet fictif, alors que dans la deuxième famille on utilise l'individu le plus centré du groupe pour jouer le rôle du centre du groupe.

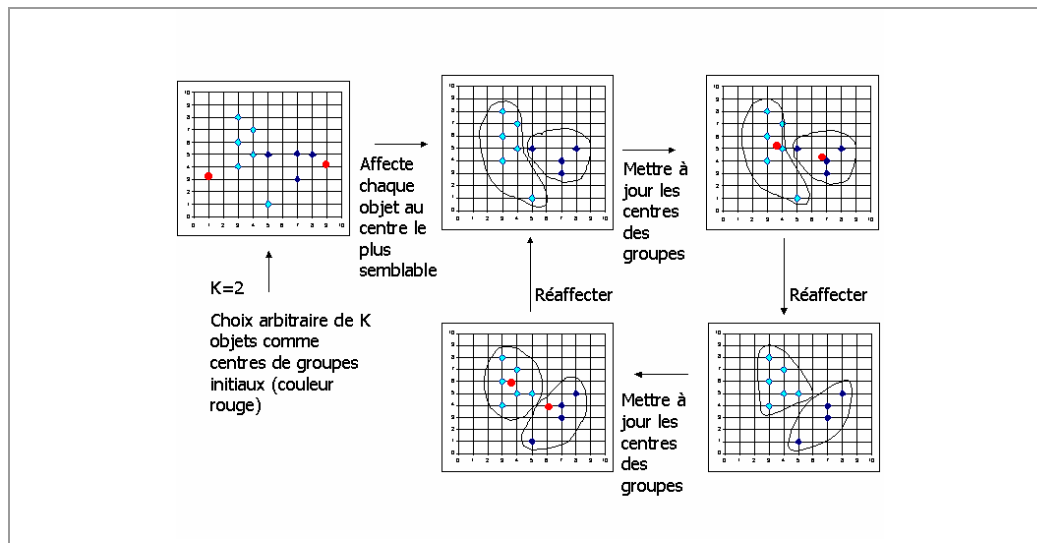


Figure 6-2- Algorithme *k-means*.

L'algorithme *k-means* présente certaines limites [Ball et Hall, 1967; Huang, 1997b] : il est restreint aux données numériques parce qu'il est basé sur le calcul de moyenne pour la détermination des centres des groupes. Donc, *k-means* ne permet pas de traiter des données caractérisées par des *attributs catégoriels* (cf. Définition 6.2).

Définition 6.2- Attribut catégoriel

C'est un attribut dont la seule structure significative est la distinction des valeurs entre elles (exemple : la dichotomie homme-femme). Les variables qualitatives, symboliques, nominales et ordinales sont des variables catégorielles. Une variable numérique peut être transformée en une variable catégorielle.

La deuxième limite réside dans le fait qu'il faut spécifier le nombre de groupes (k) à l'entrée de l'algorithme. La troisième limite réside dans le fait que le résultat de la classification dépend du choix des centres initiaux. En effet, si on applique deux fois *k-means* sur les mêmes données en changeant les centres initiaux, les classes dans les deux partitions peuvent être différentes. Les individus qui sont affectés à une même classe dans les deux partitions forment ce qu'on appelle un « *groupement stable* » (cf. Définition 6.3).

Définition 6.3- Groupement stable (ou forme forte)

*Pour obtenir les groupements stables, il s'agit d'effectuer dans un premier temps plusieurs partitions (2 dans la pratique) par la méthode *k-means* en changeant chaque fois aléatoirement les centres des groupes. Ensuite, il s'agit de trouver les sous-ensembles d'individus qui ont toujours été affectés à une même classe dans chacune des partitions. Ces sous-ensembles forment les groupements stables.*

6.1.2.2 Classification hiérarchique

Les algorithmes hiérarchiques créent une décomposition hiérarchique de l'ensemble X de données. Le résultat est un arbre hiérarchique appelé dendrogramme (cf. Figure 6-3). C'est un arbre qui divise itérativement X en des sous-ensembles plus petits jusqu'à ce que chaque sous-ensemble consiste en un seul objet. Dans une telle hiérarchie, chaque nœud de l'arbre représente une classe de X .

Le dendrogramme peut être créé à partir des feuilles (i.e. les individus) jusqu'à la racine (l'ensemble des individus) en fusionnant des groupes à chaque étape : c'est l'*approche agglomérative* ou *ascendante*. Il peut aussi être créé de la racine jusqu'aux feuilles en bas divisant des groupes à chaque étape : c'est l'*approche divisive* ou *descendante*. Contrairement aux algorithmes de partitionnement (cf. §6.1.2.2), les algorithmes hiérarchiques n'ont pas besoin du nombre de groupe k comme entrée. Cependant, une *condition d'arrêt* doit être définie indiquant quand le processus de division ou d'agglomération devrait être arrêté. Un exemple de condition d'arrêt dans l'approche agglomérative est la distance critique D_{min} entre tous les groupes de la partition.

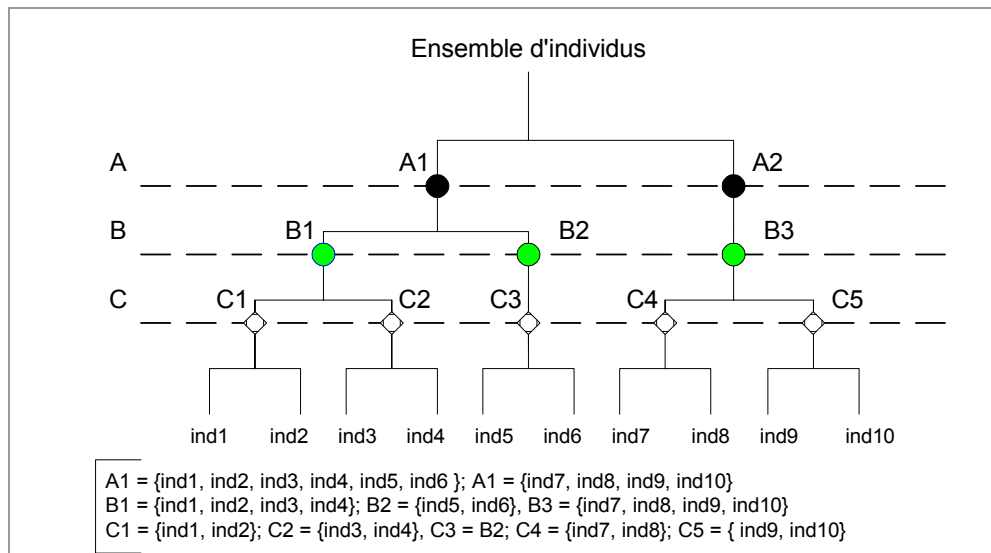


Figure 6-3- Dendrogramme

Dans la Figure 6-3, la classe B1 est appelée une agrégation des deux sous-classes C1 et C2. Plusieurs méthodes existent pour le calcul de la distance entre les deux sous-classes C1 et C2. Ces distances sont appelées *critères d'agrégation*. A chaque critère correspond un algorithme de classification hiérarchique. Une étude bibliographique approfondie des différents algorithmes (cf. [Ben Ahmed *et al.*, 2002]) nous a permis de sélectionner l'algorithme AGNES (Agglomerative NESTing) développé par [Kaufman et Rousseeuw, 1990]. Cet algorithme peut être décrit à travers les trois étapes présentées dans la Figure 6-4.

1. Traiter chaque individu comme une classe. Calculer la matrice de proximité contenant les distances entre chaque paire d'individus;
2. Chercher la paire de classes les plus similaires en utilisant la matrice de proximité. Fusionner les deux classes en une seule. Mettre à jour la matrice de proximité ;
3. Si tous les individus sont dans une seule classe ou le *critère d'arrêt* est atteint, stop. Sinon, retour à l'étape (2);

Figure 6-4- Algorithme AGNES

Les limites de la classification hiérarchique résident dans le fait qu'elle fournisse un grand nombre de partitions possibles (chaque niveau du dendrogramme correspond à une partition) parmi lesquelles on doit choisir une. Il n'est souvent pas aisé de choisir la coupure pertinente. D'autre part, le dendrogramme n'est pas un arbre optimal puisque la partition construite à un niveau donné dépend de la partition obtenue à l'étape précédente.

6.1.2.3 Classification mixte

Comme nous l'avons noté à la fin des paragraphes 6.1.2.1 et 6.1.2.2, les deux types de méthodes de classification (i.e. par partitionnement et hiérarchique) présentent un certains nombre d'inconvénients que nous résumons ci-dessous :

- Pour la classification par partitionnement : elle est restreinte aux données numériques, nécessité de déterminer le nombre de classes à obtenir et dépendance des résultats du choix des centres initiaux ;
- Pour la classification hiérarchique : grand nombre de partitions possibles et résultat non optimal ;

Pour remédier à ces limites, nous commençons par transformer les données catégorielles en données numériques en appliquant une *Analyse des Correspondances Multiple (ACM)*. L'ACM est une technique de description de données catégorielles. C'est en quelque sorte l'équivalent de l'Analyse en Composante Principale (ACP) pour des attributs numériques (cf. [Lebart et al., 1997] ou [Saporta, 1990] pour plus de détails sur ces méthodes). L'ACM nous a permis d'obtenir les descriptions des individus (i.e. les accidents) sur les axes factoriels. Ces descriptions sont des données numériques sur lesquelles nous pouvons appliquer une classification. Nous avons ensuite effectué une utilisation conjointe des deux méthodes de classification : il s'agit d'une *classification mixte* [Wong, 1982]. Elle consiste à appliquer k-means plusieurs fois sur les données issus de l'ACM pour obtenir les *groupements stables* (cf. Définition 6.3). Chaque groupement stable représente dans la suite de la méthode un « vecteur individu ». L'ensemble de ces groupements stables est alors soumis à une classification hiérarchique. La partition finale est obtenue en coupant le dendrogramme (cf. Figure 6-3) par une droite située à un niveau tel que les branches qui en résultent soient nettement séparées. Une excellente explication de la méthode avec des cas d'application peuvent être trouvés dans [Nakache et Confais, 2000].

6.1.3 Interprétation des résultats

Une étude bibliographique approfondie (cf. [Lamiri et al., 2004]) nous a permis de constater, sans surprise, qu'il est quasiment impossible de trouver des travaux génériques sur des méthodes d'interprétation des résultats issus de data mining indépendamment de la technique utilisée (i.e. classification, classement, etc.). Pour cela, nous avons choisi d'orienter notre recherche vers les travaux sur l'interprétation des résultats issus de classification automatique, la technique que nous utilisons pour élaborer des scénarios d'accidents.

Comme nous l'avons déjà noté dans la section 6.1.2, l'objectif d'une classification est l'organisation des éléments d'un ensemble en classes homogènes, a priori inconnues, interprétables et sans tenir compte de l'algorithme utilisé. Nous rappelons que deux tâches sont généralement effectuées à la suite d'une classification automatique [Manco et al., 2004] :

- *Validation des classes* : une classification est qualifiée de « valide » si les classes découvertes correspondent à des groupements statistiquement homogènes et séparés ;

- *Interprétation des classes* : une classification est « interprétable » si on arrive à associer la signification appropriée à chaque classe. Autrement dit, les classes ont une signification pour l'utilisateur et elles sont donc utilisables pour l'objectif pour lequel elles ont été créées.

A travers notre recherche bibliographique, nous avons constaté, là encore, que la majorité des travaux développe des critères statistiques qui sont orientés *validation des classes* plutôt qu'*interprétation des classes*.

Trois types de critères sont généralement utilisés pour la validation des résultats d'une classification automatique [Manco *et al.*, 2004] :

- *Des critères externes* : les résultats sont évalués objectivement par rapport à une partition pré-spécifiée. Exemple : le taux d'erreur (error rate) de l'opération de classification (i.e. le nombre des cas mal classifiés), indice de Rand, coefficient de Jaccard, indice de Huberts, etc. (voir [Manco *et al.*, 2004] et [Halkidi *et al.*, 2002] [Lamiri *et al.*, 2004] pour une description détaillée),
- *Des critères internes* : évaluation des résultats en fonction des quantités calculables à partir des données disponibles. Exemple : similarité intra-classe (i.e. compacité) et dissimilarité inter-classes (séparation), coefficient de corrélation, etc. [Halkidi *et al.*, 2002],
- *Des critères relatifs* : l'évaluation s'effectue en comparant plusieurs résultats issus d'un même algorithme mais avec des paramètres différents. Exemple : calcul des critères de compacité et de séparation en modifiant certains paramètres de la classification comme le nombre de classes, les poids des variables, etc. Plus de détails sont donnés dans [Halkidi *et al.*, 2001].

En ce qui concerne la tâche d'*interprétation des classes*, les techniques les plus utilisées sont des techniques d'exploration visuelles combinées avec les statistiques descriptives. La visualisation des résultats est généralement interactive ce qui permet d'incorporer les connaissances expertes dans la tâche d'interprétation (cf. section 6.3.2).

6.2. Complexité et extraction de connaissances de données

Etant donné le caractère itératif et interactif du processus ECD [Brachman et Anand, 1996], une forte implication de l'expert est requise. L'utilisation et l'intégration des connaissances expertes du domaine dans ce processus sont très importantes pour découvrir de nouvelles connaissances cachées interprétables et utilisables [Palmeri et Blalock, 2000]. Cette forte implication des experts est l'une des sources de la complexité du processus d'ECD. En effet, les feedbacks et les boucles récursives qui caractérisent un système complexe (cf. section 3.4) peuvent être observés durant le processus d'ECD. Chacune des phases du processus est dépendante des phases en amont et elle conditionne les phases en aval. Les fonctions effectuées durant le processus (e.g. nettoyage de données, sélection des attributs, application de data mining, etc.) génèrent des transformations (transformation de données, génération de résultats, etc.) pouvant conduire l'utilisateur à modifier les fonctions déjà effectuées. Ainsi, de nouvelles fonctions pourraient être définies ou redéfinies. Le processus est donc en perpétuelle interaction avec son environnement (utilisateur, contexte d'application, etc.). La forte implication de l'opérateur humain génère une certaine imprévisibilité aussi bien au niveau de la façon de faire qu'au niveau des résultats et leur interprétation.

L'imprévisibilité, source de complexité, peut être aperçue à travers la « non-stabilité » des résultats. En effet, l'application d'un processus d'ECD par des personnes différentes et/ou dans des conditions différentes en utilisant les mêmes données et ayant les mêmes objectifs peut ne pas mener aux mêmes résultats. Nous citons ci-dessous d'autres sources de complexité du processus ECD :

- *La nature des données utilisées* : les différentes sources (i.e. différents domaines) des données, les différentes formes de données (textuelles, images, vidéo), le nombre de données, ainsi que leurs différents niveaux de granularité sont parmi les éléments qui peuvent augmenter la non-répétitivité ou non-reproductibilité des résultats de leur analyse, donc la complexité du processus d'ECD ;
- *La spécificité du domaine ou du phénomène étudié* : quand le phénomène étudié est complexe (cf. définition à la section 3.4), plusieurs points de vue issus de différents domaines et approches sont nécessaires pour appréhender cette complexité. C'est généralement le cas pour tous les systèmes où l'opérateur humain est impliqué. L'accident de la route en est un exemple parfait. Combiner différents points de vue dans un processus d'ECD rend plus complexe ce dernier à cause des difficultés relatives à la gestion de conflits entre les experts (différence de raisonnement, différence de terminologies, etc.) ;
- *La nature des tâches effectuées* durant les différentes phases du processus d'ECD : si ces tâches se basent sur l'expertise dans leur application ou l'interprétation de leurs résultats, le processus d'ECD devient plus complexe. Certains algorithmes exigent une certaine forme de données ce qui nécessite de créer des objets intermédiaires rendant compliqué le processus ;
- *L'interdépendance des étapes d'ECD* : plus les étapes sont dépendantes, plus les feedbacks et la récursivité sont intenses et plus l'implication de l'utilisateur est accrue ;
- *L'objectif de l'étude et la nature des connaissances recherchées* : pour une étude donnée d'extraction de connaissances, il faut trouver un support adéquat pour ces connaissances qui réduit au maximum la perte d'information et qui permet leur exploitation. Pour les connaissances en accidentologie, générer des STA est une tâche plus complexe que celle de la génération de simples chiffres statistiques.

6.3. Intégration des connaissances du domaine dans le processus ECD

6.3.1 Intégration des connaissances du domaine en préparation de données

Chacune des tâches de préparation de données est effectuée selon : *l'objectif de l'étude*, *la nature des données* (i.e. textes, images, vidéo), *le type de techniques de data-mining* à utiliser, *le domaine de l'étude*, etc. Etant donné qu'une forte implication des experts est nécessaire, il est important de développer des méthodes permettant d'intégrer d'une manière efficace leurs connaissances dans ces différentes tâches. Cela permet d'intégrer l'objectif de l'étude (i.e. le traduire en attributs d'étude) et de réduire le nombre d'attributs en minimisant la perte d'information. Plusieurs techniques

statistiques sont utilisées dans la phase de préparation de données (e.g. filtrage, modélisation du bruit, clustering, régressions, etc.), mais nous nous intéressons ici uniquement aux techniques permettant l’intégration de l’expertise.

Parmi les méthodes que nous avons identifiées dans la littérature et qui permettent l’intégration des connaissances dans la préparation des données, citons le concept de hiérarchie. En effet, la représentation hiérarchique experte de données permet de représenter les relations entre les données [Dhar et Tuzhilin, 1993; Han et Fu, 1996]. L’utilisation de ces hiérarchies, résultats de l’expertise, permet en quelque sorte l’intégration de connaissances expertes. Une autre manière permettant l’utilisation des connaissances prédéfinies pour la préparation des données est le re-engineering de la base de données [McClellan et al., 2000]. Elle propose le remplacement des données manquantes et conflictuelles par des ensembles d’attributs du domaine définis par les experts. Les techniques de visualisation des données [Ahlberg et Wistrand, 1995; Keim et Kriegel, 1996] représentent aussi un outil pour l’intégration des connaissances du domaine : elles permettent, par exemple, aux experts la sélection et la transformation des attributs dans la phase de prétraitement des données.

L’approche AICEF que nous développons dans le Chapitre 9, propose l’utilisation des modèles du domaine comme un moyen pour utiliser les connaissances expertes dans la préparation des données. Une interface entre les modèles du domaine et les données a été élaborée pour assister l’utilisateur dans la phase de préparation de données et essentiellement de sélection d’attributs.

6.3.2 Intégration des connaissances du domaine dans l’interprétation des résultats

Les techniques les plus connues pour l’incorporation des connaissances expertes dans la tâche d’interprétation des résultats sont celles de visualisation interactive de ces résultats. Nous en citons ci-dessous quelques exemples :

- La représentation des données multidimensionnelles par un *diagramme plan* [Zhang et al., 2002]. L’idée de base de cette méthode est de remplacer les axes orthogonaux, représentant les attributs, par des lignes verticales parallèles. Les valeurs maximales et minimales des attributs seront affectées, respectivement, aux points extrêmes hauts et bas de ces lignes verticales. Pour représenter un point à N dimensions, il faut tracer N lignes parallèles où chacune d’entre elles représente une dimension donnée. En localisant la valeur de chaque attribut sur la ligne qui lui est associée, et en joignant ces points par N-1 segments on obtient finalement la représentation de notre individu. La structure résultante d’une représentation en coordonnées parallèles permet l’interprétation de plusieurs relations existantes entre les données ;
- Le *graphique matriciel* (ou multiplot) [Grinstein et al., 2001] est l’un des outils de visualisation les plus utilisés en data mining. En multiplot, N dimensions sont projetées sur des graphiques de dispersion. Chaque paire de dimensions est associée à deux de ces graphiques représentant la relation entre les deux dimensions en question. Il permet l’interprétation des classes, ainsi que l’identification des tendances et des corrélations ;

- D'autres techniques existent dans la littérature. Exemple : les *techniques orientées icône*, la *géométrie hyperbolique*, *l'approche surfacique*, etc. (voir [Sachinopoulou, 2001] pour la description de ces techniques).

Les techniques que nous venons de présenter, et qui sont essentiellement des techniques de visualisation sont indépendantes des connaissances du domaine. Prenons l'exemple de l'accident de la route, ces techniques ne permettent pas l'interprétation d'une classe selon, par exemple, le modèle CVE ou le modèle séquentiel. Pour pallier cette limite, nous permettons dans AICEF (cf. Chapitre 9) une interprétation multi-vues des classes en utilisant des méta-données qui formalisent les connaissances expertes.

6.4. Synthèse et positionnement

Ce chapitre a été consacré à un état de l'art sur le processus d'extraction des connaissances de données et plus particulièrement à la problématique qui nous intéresse, i.e. l'incorporation des connaissances du domaine dans la première et dernière phase du processus d'ECD.

Nous avons commencé par une présentation générale des différentes phases d'un processus d'ECD (i.e. *la préparation des données*, *l'application d'une technique de data mining* et *l'interprétation des résultats*) ainsi que les techniques que nous avons utilisées dans nos travaux (cf. section 6.1). Ensuite nous avons montré les sources de complexité du processus ECD telles que la nature des données utilisées, la spécificité du domaine, la nature des connaissances recherchées et l'interdépendance des étapes du processus (cf. section 6.2). Ces éléments font de la connaissance experte et des connaissances du domaine en général un levier crucial pour contourner la complexité du processus d'ECD. Cet objectif peut être atteint en utilisant par exemple, des points de vue du domaine pour sélectionner des attributs dans la phase de préparation des données. Cependant, nous n'avons pas trouvé dans la littérature de techniques génériques permettant d'atteindre cet objectif. La majorité des techniques se base sur des critères statistiques indépendants du domaine (cf. section 6.3.1). De même, nous avons constaté que la majorité des méthodes d'interprétation des classes est orientée *validation des classes* en se basant sur des critères statistiques et non *interprétation de classes* (cf. section 6.3.2). L'intégration des connaissances expertes dans cette phase se limite aux méthodes de visualisation des résultats.

Synthèse de la partie II

Cette partie a été consacrée à l'état de l'art concernant les problématiques de nos travaux et elle se décompose en six chapitres.

L'objectif du Chapitre 3 était de remonter aux origines philosophiques et épistémologiques de l'*approche systémique* que nous utilisons comme méthode de modélisation dans les travaux présentés dans ce mémoire. L'approche systémique est issue de l'épistémologie *constructiviste* qu'on oppose souvent à l'épistémologie *positiviste*. Nous avons vu que la dualité entre les deux épistémologies est la même en philosophie entre le courant relativiste (cf. section 3.1) et le courant absolutiste ou platonicien (cf. section 3.1). C'est aussi la même dualité en économie entre le courant objectiviste (cf. section 3.2) et le courant subjectiviste (cf. section 3.2). Nous avons présenté les principes de la systémique comme méthode de modélisation (cf. section 3.4) et nous avons distingué entre la *systémique* (ou la *cybernétique*) et l'approche analytique. Nous avons défini ce qu'est un système complexe (ou cybernétique) (cf. Définition 3.3) et ce qui le caractérise (principe de *holisme*, *boucles de rétroaction*, etc.). En se basant sur la définition de la complexité dans la section 3.4, nous avons montré dans la section 3.5 que l'accident de la route est un « phénomène » complexe. Nous avons mis l'accent sur la nécessité de combiner plusieurs points de vue d'analyse de l'accident pour appréhender cette complexité. Dans la section 3.6, nous avons expliqué le principe de la *cybernétique de second ordre* qui consiste à appliquer la vision systémique sur la tâche de modélisation. Cette vision est à l'origine du concept de « système de modélisation » qui conjoint un « système modélisé » (ou objet observé), un « système modélisant » (ou observateur) et un « contexte de modélisation » (ou contexte d'observation). Dans la dernière section de ce chapitre (cf. section 3.7), nous avons utilisé l'exemple de l'accident de la route pour montrer la réflexivité (cf. la définition à la section 3.7) de l'approche systémique.

Le Chapitre 4 présente les principales approches en sécurité routière : sécurité primaire, secondaire, active et passive (cf. section 4.1). Dans la section 4.2, nous avons détaillé la méthode d'Etude Détaillée d'Accidents (EDA) développée et utilisée par l'INRETS pour la collecte et

l'analyse des données d'accident sur scène. Dans la section 4.3, nous avons présenté les modèles d'analyse de l'accident de la route les plus connus. Nous avons présenté aussi les différents types de données disponibles dans les aussi La section 4.4 a été dédiée à la définition du concept scénario-type d'accidents (STA). Nous avons présenté l'approche experte d'élaboration des STA appliquée dans la littérature. Nous avons ensuite exploré les possibilités et les difficultés d'une élaboration automatisée des STA. A la fin, nous avons proposé une nouvelle approche qui combine des approches automatisées d'ECD, telle que la classification (cf. section 4.4.2), et des approches expertes. Il s'agit d'une approche qui permet d'incorporer les connaissances du domaine dans le processus d'ECD. Notre proposition nécessite une connaissance du processus d'ECD ainsi que des techniques de modélisation et d'opérationnalisation des connaissances. Nous avons consacré alors le Chapitre 5 et le Chapitre 6 à un état de l'art sur chacun de ces domaines.

Le Chapitre 5 propose alors un état de l'art sur la problématique de modélisation des connaissances en ingénierie des connaissances. Cet état de l'art couvre différents aspects relatifs à la connaissance et qui sont en rapport avec notre problématique d'incorporation des connaissances dans le processus d'ECD. Définitions des connaissances, typologies des connaissances, différentes approches de modélisation des connaissances, représentation et opérationnalisation des connaissances et évaluation des connaissances sont les différents thèmes abordés dans ce chapitre.

Le Chapitre 6 se focalise sur la problématique d'extraction de connaissances de données. Après une présentation générale du processus d'ECD, nous avons montré les sources de complexité de ce processus. Ensuite, nous nous sommes focalisés sur les aspects en rapport avec notre problématique d'incorporation des connaissances du domaine dans le processus d'ECD. Nous avons montré les limites des méthodes existantes dans la littérature par rapport à cet objectif et nous avons proposé l'utilisation des métadonnées. Dans la partie suivante (Partie III), nous allons montrer comment élaborer ces métadonnées et comment les utiliser.

Partie III

SAFE-Next : Développement d'une approche systémique pour l'extraction des connaissances

Présentation de la partie III

Pour montrer l’apport de notre approche SAFE-Next, nous avons effectué une étude pour vérifier si des méthodes d’IC existantes dans la littérature, comme CommonKADS [Schreiber *et al.*, 1994a], permettent de modéliser des connaissances expertes selon différents points de vue et d’incorporer ces connaissances dans un processus d’ECD. Dans le premier chapitre de cette partie (Chapitre 7), nous considérons l’accidentologie comme domaine d’application et nous appliquons la méthode CommonKADS pour développer un modèle de connaissances pour la construction des STA. A travers ce chapitre, nous montrons les limites de CommonKADS pour formaliser certains types de connaissances et pour élaborer des modèles multi-points de vue. Les autres chapitres de cette partie sont consacrés au développement de SAFE-Next qui intègre quatre approches dont chacune fait l’objet d’un chapitre (cf. Figure III-1) :

Le Chapitre 8 est consacré au développement de notre première approche dans SAFE-Next que nous appelons ASMEC (Approche Systémique de ModElisation des Connaissances). Elle consiste en une méthode de modélisation multi-points de vue des connaissances.

Le Chapitre 9 est consacré au développement de notre deuxième approche dans SAFE-Next que nous appelons AICEF (Approche d’Incorporation des Connaissances Expertes dans la Fouille de données). Nous proposons l’élaboration et l’utilisation de *méta-données* comme un moyen pour l’incorporation des connaissances formalisées par ASMEC dans le processus d’ECD.

Le Chapitre 10 est consacré au développement de notre troisième approche dans SAFE-Next que nous appelons ASAIC (Approche Systémique d’Analyse d’Impact de Changement). Cette approche permet une analyse interactive de l’impact d’un changement sur un système existant. Nous considérons comme cas d’application l’analyse de l’impact de l’introduction d’un nouveau système de sécurité sur le comportement du système CVE.

Le Chapitre 11 est consacré au développement de notre quatrième approche dans SAFE-Next que nous appelons ASEM (Approche Systémique d’Evaluation de Modèle). Elle fournit un *modèle général d’évaluation de modèles de connaissances*.

Le Chapitre 12 est dédiée à la validation des différentes approches développées dans SAFE-Next.

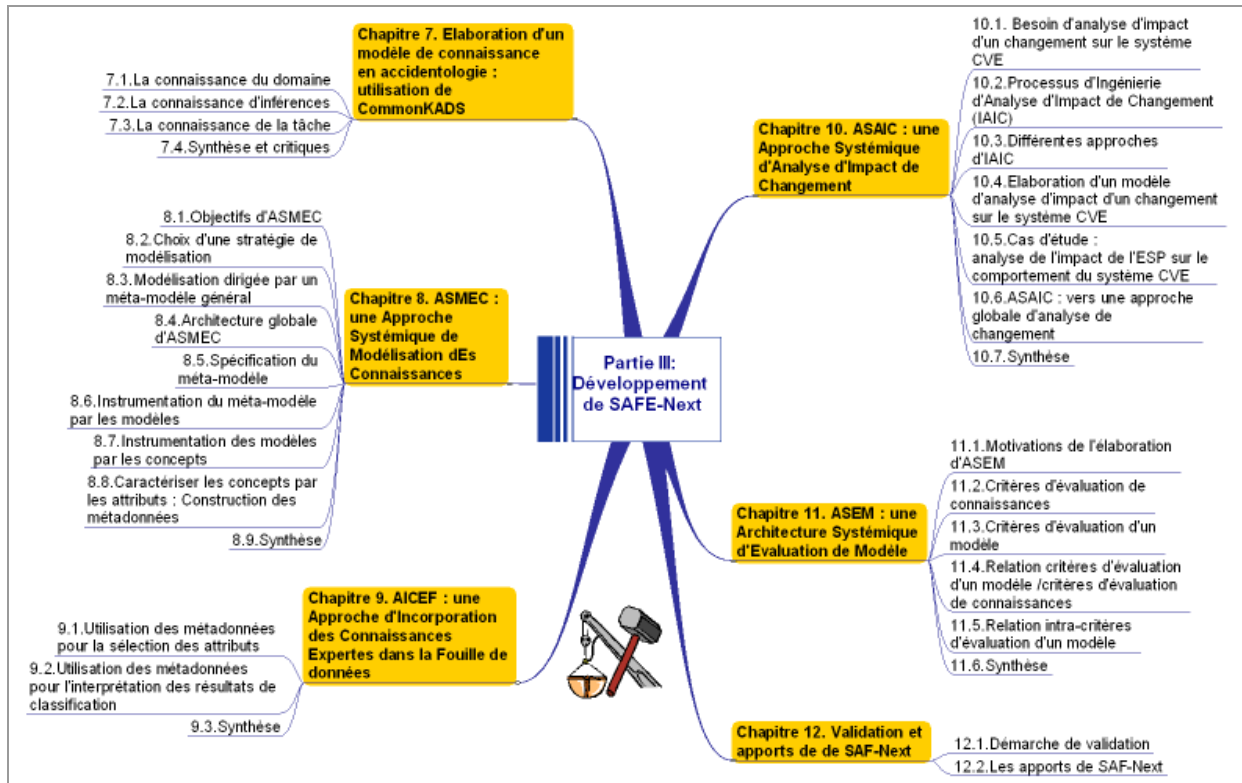


Figure III-1- Partie III : Développement de SAFE-Next.

Chapitre 7.

Elaboration d’un modèle de connaissances pour la construction de STA : utilisation de CommonKADS

Dans ce chapitre, nous appliquons la méthode CommonKADS [Schreiber *et al.*, 1994a] pour l’élaboration d’un modèle de connaissances pour l’élaboration de STA. Notre choix de cette méthode a été basé sur les raisons suivantes :

- Résultat d’un projet européen, CommonKADS est la méthode d’IC qui a mobilisé le plus d’énergie et d’investissement pour son élaboration ;
- Preuve de son efficacité, cette méthode est largement appliquée dans le monde notamment dans la communauté européenne ;
- Les modèles de CommonKADS contiennent les principes généraux que nous trouvons dans la plupart des autres méthodes d’IC.

Comme nous l’avons déjà présenté dans la section 5.3.2.1, trois types de connaissances sont à identifier et formaliser dans CommonKADS : la *connaissance du domaine* (cf. section 7.1), la *connaissance d’inférences* (cf. section 7.2) et la *connaissance de la tâche* (cf. section 7.3).

Dans la dernière section de ce chapitre, section 7.4, nous montrons les limites de CommonKADS pour la modélisation de certains types de connaissances dont nous avons besoin pour l’élaboration des STA. Cela nous permettra de montrer l’apport de notre approche de modélisation des connaissances, ASMEC, que nous présentons dans le Chapitre 8.

La Figure 7-1 illustre les différents types de connaissances que nous allons formaliser à l'aide de CommonKADS.

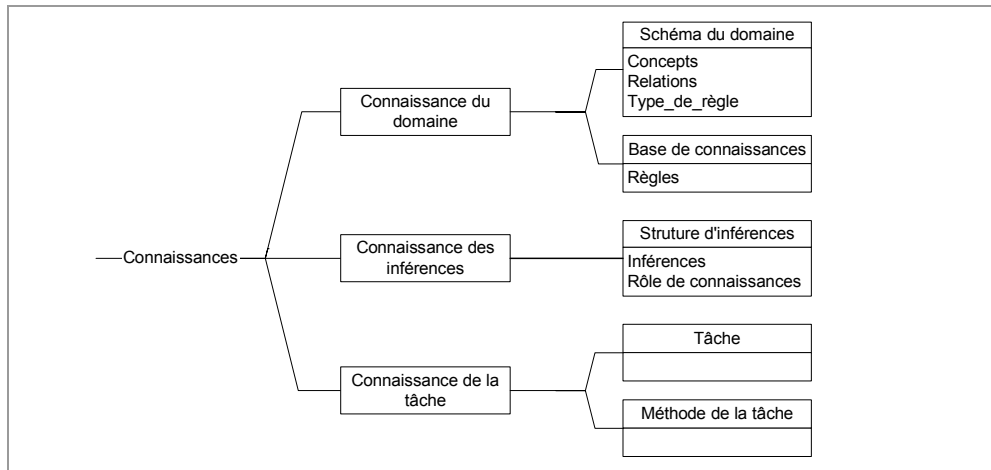


Figure 7-1- Différents types de connaissances formalisés à l'aide de CommonKADS.

7.1. La connaissance du domaine

La *connaissance du domaine* contient les *connaissances statiques* spécifiques à un domaine d'application. Elle est constituée de deux éléments : (i) le *schéma du domaine* qui contient les *concepts* et les *relations* entre ces concepts et (ii) la *base des connaissances* qui contient les *instances des connaissances* sur lesquelles le raisonnement devra s'effectuer.

7.1.1 Le schéma du domaine

7.1.1.1 Les concepts

Définition 7.1- Un concept

Il décrit un ensemble d'objets ou d'instances qui existent dans le domaine d'application et qui ont des caractéristiques similaires [Schreiber et al., 1994c]. Il correspond à la notion de classe dans la modélisation orientée-objet, à la différence que les fonctions ne sont pas incluses dans la définition d'une classe.

Nous utilisons KADS22, un outil disponible en ligne¹ sur le site de CommonKADS, pour la représentation graphique du schéma du domaine. Le formalisme UML est utilisé pour décrire graphiquement les éléments du schéma du domaine. Le langage CML2 (Conceptual Modeling Language), dont la syntaxe est décrite dans [Anjewierden et Schreiber, 2004], est utilisé pour décrire les éléments du modèle de connaissances.

Nous avons identifié trois *concepts* principaux en accidentologie : « *l'impliqué* », le « *véhicule* » et « *l'environnement* ». Chacun de ces concepts est composé de sous-concepts. Par exemple, le concept « *impliqué* » est composé des sous-concepts : « *conducteur* », « *occupant* », « *piéton* », « *blessé* ». Chaque concept est décrit par des *attributs* qui peuvent prendre des *valeurs* de type entier, réel ou texte. La Figure 7-2 présente une représentation du concept « *impliqué* » en CML2.

¹ <http://www.swi.psy.uva.nl/projects/kads22/index.html>


```

Concept impliqué;
  description:
    "les personnes impliquées dans l'accident";
  super-type-of: conducteur, piéton, occupant, blessé;
  disjoint: yes;
  complete: yes;
  attributes:
    Nombre_tués: natural;
    Nombre_blessés_graves: natural;
    Nombre_blessés_légers: natural;
    Nombre_indemnes: natural;
End concept impliqué;
    
```

Figure 7-2- Représentation du concept impliqué en CML.

7.1.1.2 Les relations

Plusieurs types de relations peuvent exister entre les concepts. Parmi celles-ci, nous citons la relation de *généralisation* ou *d'héritage*. Par exemple, le concept « *impliqué* » dans la Figure 7-2 est une super-classe (*super-type-of*) des concepts « *conducteur* », « *occupant* », « *piéton* », « *blessé* ». Une flèche triangulaire permet la représentation graphique de cette relation (cf. Figure 7-4).

L'*agrégation* est un deuxième type de relation entre les concepts. Elle est spécifiée à l'intérieur de la définition des concepts par la syntaxe : *has-parts*. Elle permet de modéliser des agrégations physiques et conceptuelles. Le concept « *véhicule* », par exemple, est une *agrégation conceptuelle* des concepts « *véh_sécur_primaire* » et « *véh_sécur_secondaire* » qui décrivent le véhicule, respectivement, selon un point de vue sécurité primaire et secondaire. Alors que le concept « *environnement* » une *agrégation physique* des concepts « *trafic* », « *conditions_ambiantes* », « *équipement_route* » et « *infrastructure* ». Une flèche sous forme d'un diamant permet la représentation graphique de cette relation.

Les liens entre les différents concepts peuvent être aussi définis par une construction de type *relation*. Cette relation peut elle-même avoir des attributs. « *Est-proprétaire* » est un exemple de ce type de relation : un « *conducteur* » « *est_proprétaire* » d'un ou de plusieurs « *véhicules* » et un

```

Binary-relation est_proprétaire;
  description:
    "Description du rapport du conducteur au véhicule";
  argument-1: conducteur;
    cardinality: 1 +;
  argument-2: véhicule;
    cardinality: 1 +;
  attributes:
    Ancienneté_achat: string;
    Type_achat: {Achat_neuf, Achat_occasion};
    Kilométrage_depuis_achat: string;
End binary-relation est_proprétaire;
    
```

Figure 7-3- Représentation d'une construction de type relation en CML.

véhicule appartient à un ou plusieurs conducteurs (cf. Figure 7-3).

Pour la description du reste des constructions du schéma du domaine (e.g. *type-de-règles*), voir l'annexe A-1.1.1.

première et la dernière tâche du processus d'élaboration des STA sont les tâches qui sollicitent le plus d'implication des experts. Nous avons donc identifié et formalisé deux bases de connaissances :

« connaissances_préparation_données » et « connaissances_interprétation_classe ».

7.1.2.1 Base de connaissances pour la préparation des données

Dans la base de connaissances relative à la préparation des données nous avons identifié des connaissances relatives à l'abstraction des attributs (e.g. « age » <2 → « jeune »), à leur évaluation (cf. Figure 7-5) et à leur modification (e.g. si « nombre-modalités » est élevé → fusionner les modalités).

```

attribut.nbre_inconnue = 'acceptable'           OR
attribut.nbre_inconnue = 'moyen'              AND
attribut.nombre_modalités = 'acceptable'      OR
attribut.nombre_modalités = 'acceptable'      AND
attribut.importance_générale = 'Très important' OR
attribut.importance_générale = 'important'
implique
attribut.pertinence étude = 'attribut pertinent';
    
```

Figure 7-5- Exemple de règles d'évaluation de la pertinence d'un attribut.

La préparation des données consiste aussi à choisir des accidents qui sont « pertinents » pour l'étude. Comme dans le cas de la sélection des attributs, il s'agit d'abstraire les descriptions des accidents, d'évaluer leur pertinence et d'effectuer des modifications sur leurs descriptions si besoin il y a (e.g. remplacer des valeurs manquantes par des valeurs moyennes).

7.1.2.2 Base de connaissances pour l'interprétation des classes

En ce qui concerne la phase d'interprétation, il s'agit de connaissances permettant d'élaborer une description « exploitable » des classes. Ces dernières sont décrites à l'aide de modalités et à chacune de ces modalités sont attachés plusieurs critères statistiques. La Figure 7-6 présente le début d'un tableau qui décrit une classe. Ce tableau représente une sortie classique d'un logiciel de datamining¹.

¹ Le logiciel utilisé est SPAD. On peut obtenir des sortie semblable en utilisant d'autres logiciel de datamining tels que STATISTICA, SPLUS, SPSS, etc.

Libellés des variables	Modalités caractéristiques	% de la modalité dans l'échantillon	% de la modalité dans la classe	% de la classe dans la modalité	Valeur-Test
LocChoc	Hors chaussee	26.64	96.72	30.89	12.26
typeChoc	Tonneau renversement	21.76	78.69	30.77	9.92
Obp	Obp=sol	18.97	68.85	30.88	8.91
vehiprior2	Vehicule seul	29.15	72.13	21.05	7.16
sitace	prob contrôle ve 21#	32.50	73.77	19.31	6.79
critini	Guidance infrastr 5#	15.62	44.26	24.11	5.51
evini	gene exterieur/d 14#	5.72	22.95	34.15	4.68
manident	Section courante 17#	24.83	49.18	16.85	4.19
typace	typace=pilotabilite	55.51	80.33	12.31	4.09
atm	atm=Clair/normal	55.79	80.33	12.25	4.04
surf	surf=Sec	62.62	85.25	11.58	3.89
typlieu	typlieu=H-Agg sur RD	47.98	70.49	12.50	3.58
fondef	fondef=Action	9.07	22.95	21.54	3.30
manident	Changeement file 16#	6.14	18.03	25.00	3.26
typdef	Realisation inco 20#	33.61	52.46	13.28	3.04
mask	mask=Pas de masque	65.13	81.97	10.71	2.86
critini	Perte contrôle tr 8#	17.85	32.79	15.63	2.83
mecdef	mecdef=Panique	5.72	14.75	21.95	2.57
mecdef	Activite annexe 27#	7.67	16.39	18.18	2.23
evini	Droque medicamen 11#	2.51	8.20	27.78	2.21
age	age=25 40	38.35	40.98	9.09	0.31
typdef	Non realisation	17.29	21.31	10.48	0.71

Figure 7-6- Description d'une classe à l'aide des modalités (Sortie du logiciel SPAD d'une classification des accidents)

Les quatre dernières colonnes du tableau ci-dessus représentent les différents critères utilisés pour l'interprétation d'une classe :

- **Global** (3^{ème} colonne du tableau) : le pourcentage de la modalité dans l'échantillon global. Par exemple, on peut lire dans la première ligne du tableau de la Figure 7-6, que 26,64% des accidents dans l'échantillon global vérifient la modalité « localisation de choc = Hors chaussée » ;
- **MOD/CLAS** (4^{ème} colonne du tableau) : le pourcentage des individus décrits par une modalité donnée dans une classe donnée. Par exemple, 96,72% des accidents dans la classe vérifient la modalité « Localisation de choc = Hors chaussée » (Figure 7-6) ;
- **CLASS/MOD** (5^{ème} colonne du tableau) : le pourcentage d'une classe dans une modalité. Par exemple, les individus dans la classe qui vérifient la modalité « localisation de choc = Hors chaussée » représentent 30,89 % de l'ensemble des individus vérifiant cette modalité dans l'échantillon global ;
- **Valeur-Test** (6^{ème} colonne du tableau) : c'est un critère développé par [Lebart et al., 1997]. Il permet de détecter si une modalité est sur ou sous-représentée dans une classe par rapport à l'échantillon global. Prenons le même exemple de la modalité « localisation de choc = Hors chaussée » (Figure 7-6) : on remarque que cette modalité est sur-représentée dans la classe par rapport à l'échantillon global ($MOD/CLAS=96,72\% \gg Global=26,64\%$). Généralement, quand la Valeur-Test est supérieure à deux, la modalité correspondante est sur-représentée. Quand la valeur-Test est inférieure à -2, la modalité correspondante est sous-représentée.

Nous avons formalisé des règles de connaissances expertes manipulant ces trois paramètres lors de la caractérisation des classes dans la phase d'interprétation. Nous avons tout d'abord défini deux notions de pertinence pour les modalités caractérisant une classe : le *pouvoir discriminant* et le *pouvoir descriptif* d'une modalité :

- **Modalité discriminante** : une modalité est dite discriminante si elle est sur-représentée (i.e. $Valeur-Test > 2$) ou sous-représentée (i.e. $Valeur-Test < -2$). Elle permet de *discriminer* (différencier) la classe correspondante par rapport aux autres classe ;
- **Modalité caractérisante** : une modalité est dite caractérisante si le pourcentage MOD/CLASS est supérieur à un seuil fixé par l'accidentologue (le pourcentage généralement utilisé est 30%). Elle permet de caractériser la classe correspondante.

En se basant sur ces deux notions, nous avons pu définir quatre types de modalités :

- **Modalité discriminante et caractérisante** : ($Valeur-Test > 2$ ou < -2) et ($MOD/CLASS > 30\%$). Exemple : la modalité « *localisation de choc = Hors chaussée* » ;
- **Modalité discriminante mais non-caractérisante** : ($Valeur-Test > 2$ ou < -2) et ($MOD/CLASS < 30\%$). Exemple : la modalité « *mecdef=Panique* » (mecdef : mécanisme de défaillance) ;
- **Modalité non-discriminante mais caractérisante** : ($-2 < Valeur-Test < 2$) et ($MOD/CLASS > 30\%$). Exemple : la modalité « *age=25_40* » ;
- **Modalité non-discriminante et non-caractérisante** : ($-2 < Valeur-Test < 2$) et ($MOD/CLASS < 30\%$). Exemple : la modalité « *typdef= Non_realisation* » (typdef : type de défaillance).

7.2. La connaissance d'inférences

La *connaissance d'inférences*, deuxième type de connaissances dans CommonKADS, décrit les inférences de base pour résoudre un problème donné en utilisant les connaissances du domaine. Deux éléments constituent la connaissance d'inférences : les *inférences* et les *rôles de connaissance* [Schreiber *et al.*, 1994c].

7.2.1 Les inférences et les rôles de connaissances

Définition 7.2- Inférences

Elles forment le plus bas niveau des unités de traitement de l'information nécessaires dans un processus de raisonnement. Elles opèrent sur des données d'entrée, et ont la capacité de produire d'autres informations en sortie.

Définition 7.3- Rôles de connaissance

Ils forment les entrées et les sorties d'une inférence. On les appelle rôles de connaissance car ils représentent les rôles que jouent les concepts du domaine par rapport aux inférences. Un rôle peut faire référence à un ou plusieurs concepts du domaine. De plus, les noms des rôles doivent être suffisamment abstraits pour favoriser leur réutilisation dans d'autres applications.

Dans la Figure 7-7, « opérationnaliser » est une inférence. « Besoin », « contraintes », « préférences » et « modèles_multivoques_domaine » sont des rôles de connaissances. Selon cette figure, l'inférence « opérationnaliser » permet de traduire un « besoin » en terme de « contraintes » et « préférences » en utilisant des connaissances du domaine « modèles_multivoques_domaine ». Dans le cas de sélection des attributs d'étude, par exemple, les « contraintes » définissent des restrictions dans le choix des attributs pour en sélectionner ceux qui sont a priori pertinents. Etant donné qu'il peut y avoir plusieurs attributs pertinents, les « préférences » permettent de choisir les plus pertinents pour l'étude. Lors de la phase de sélection des attributs, nous avons constaté que l'expert fait appel

explicitement ou implicitement à des modèles existants tels que le modèle CVE (cf. section 4.3.1), le modèle séquentiel (cf. section 4.3.2), le modèle de traitement d'information (cf. section 4.3.3), le modèle de la tâche (cf. section 4.3.4), etc. Le rôle « *modèles_multivues_domaine* » représente en réalité une conjonction de ces différents modèles.

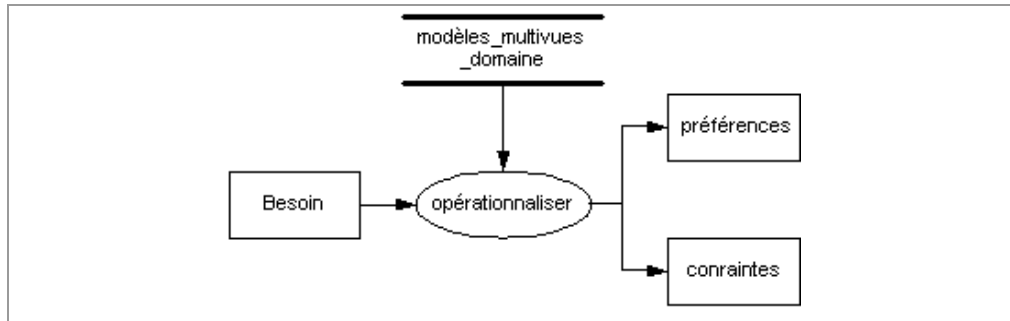


Figure 7-7- Représentation graphique de l'inférence « opérationnaliser ».

On distingue deux types de rôles de connaissance : les *rôles dynamiques* et les *rôles statiques*. Les *rôles dynamiques* sont les entrées et les sorties (durant le temps d'exécution) des inférences. C'est le cas des rôles « *besoin* », « *contraintes* » et « *préférences* » (cf. Figure 7-7). Les *rôles statiques*, quant à eux, sont des ressources stables, qui représentent des connaissances du domaine utilisées pour exécuter une inférence. Ils spécifient la connaissance du domaine dont on a besoin pour réaliser l'inférence. Par exemple, l'inférence « *opérationnaliser* » utilise les « *modèles_multivues_domaine* » pour traduire un « *besoin* » en « *contraintes* » et « *préférences* ». L'un des objectifs de notre approche ASMEC (cf. Chapitre 8) est la modélisation multi-vues de ce type de connaissances statiques.

7.2.2 La structure d'inférences

L'ensemble des inférences et des rôles de connaissances ainsi que les relations entre eux sont organisés dans une *structure d'inférence*.

Définition 7.4- Structure d'inférence

Elle fournit un mécanisme d'abstraction sur les détails du processus de raisonnement. Elle constitue une trace du raisonnement dans la résolution de la tâche.

Pour l'identification et la structuration des inférences nous avons procédé par une approche de *modélisation-par-composition* à partir d'éléments de la bibliothèque (cf. 5.3.2) de CommonKADS [Schreiber *et al.*, 1994c]. Nous avons constaté que durant la première et la dernière tâche d'élaboration des STA (i.e. préparation des données et interprétation des résultats), il s'agit essentiellement d'un problème d'évaluation. En effet, dans la tâche de préparation de données, il s'agit d'évaluer les attributs et les accidents pour en sélectionner les plus pertinents pour la construction de l'échantillon de l'étude. De même, dans la tâche d'interprétation des résultats, il s'agit d'évaluer les classes pour en sélectionner les plus pertinentes pour la construction des STA. Nous avons donc choisi le *modèle d'évaluation (Model Assessment)* présenté dans la bibliothèque CommonKADS [Schreiber *et al.*, 1994c] (Figure 7-8).

Les conventions graphiques dans une structure d'inférence sont les suivantes : Des *ovales* représentent les inférences. Des *rectangles* représentent les rôles dynamiques de la connaissance. Des

traits épais horizontaux indiquent les rôles statiques. Des flèches relient les rôles et les inférences. Des flèches avec un rond noir à leur extrémité signifient que le rôle dynamique s'interprète comme un ensemble d'objets.

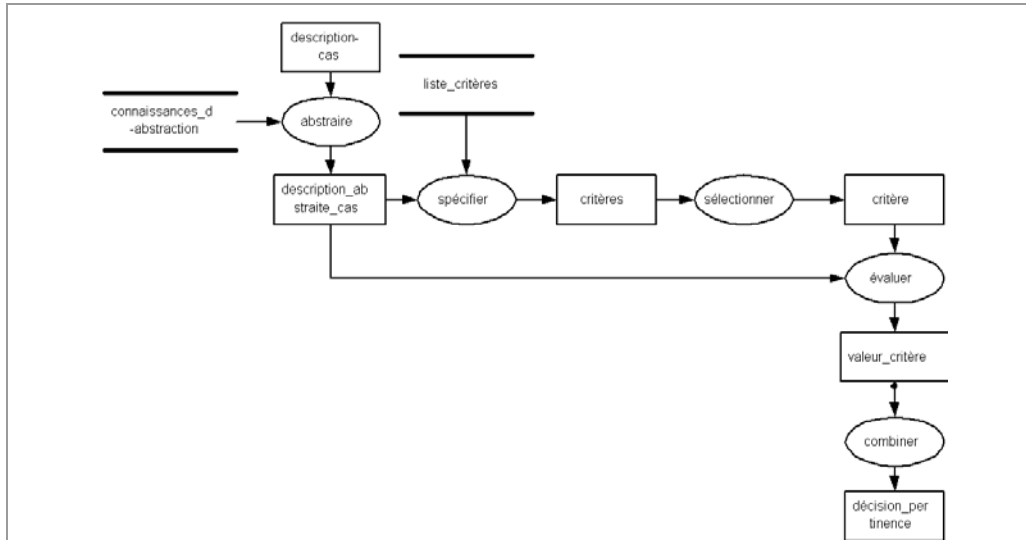


Figure 7-8- Structure d'inférence générique pour l'évaluation d'un objet (e.g. attribut, accident, classe, etc.).

Nous avons aussi constaté que durant l'évaluation, les experts sont amenés à faire des modifications sur différents niveaux : au niveau des attributs (e.g. regrouper des modalités, remplacer des valeurs manquantes, supprimer un attribut), au niveau des accidents (e.g. remplacer certaines valeurs, supprimer un accident) ou au niveau des classes (e.g. rechoisir les modalités discriminantes, rechoisir les modalités descriptives). Nous avons donc sélectionné et adapté certaines composantes du modèle de conception de configurations (« configuration Design ») décrit aussi dans la bibliothèque de CommonKADS. La Figure 7-9 présente uniquement la partie qui nous intéresse du modèle de conception de configurations (cf. [Schreiber et al., 1994a] pour plus de détails sur le modèle complet). Cette figure présente la structure d'inférence générique que nous avons composée et utilisée pour l'évaluation et la modification des attributs, des accidents et des classes.

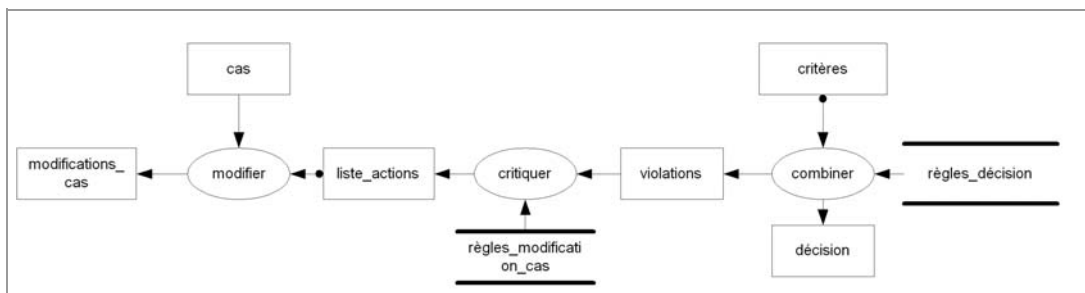


Figure 7-9- Structure d'inférence générique pour la modification d'un objet (e.g. description d'un attribut)

Pour la clarté de la présentation, nous avons choisi de présenter séparément les structures d'inférences relatives aux trois tâches du processus d'élaboration des STA (i.e. préparation des données, classification automatique et interprétation des classes).

7.2.2.1 La structure d'inférence de préparation de données

La tâche de préparation des données se compose de deux sous-tâches : *sélection des attributs pertinents* et *sélection des accidents pertinents*. Nous avons choisi de présenter ici uniquement la structure d'inférence relative à la sélection des attributs (cf. Figure 7-10). La structure d'inférence de sélections des accidents est présentée en annexe (cf. A-1.2.1.2.2).

La sélection des attributs d'étude (cf. Figure 7-10) commence par « opérationnaliser » (i.e. traduire) le « besoin » de l'étude sous forme de « contraintes » de sélection. Ensuite, l'inférence « sélectionner1 » permet de trouver les attributs qui satisfont les contraintes. A la sortie de l'inférence « calculer1 », on obtient des « descriptions » statistiques des attributs (e.g. histogrammes). A travers l'inférence « sélectionner11 », l'expert sélectionne un par un les attributs et fait appel à son expérience (« connaissances d'abstraction ») pour « abstraire » les descriptions des attributs. Par exemple, « age < 25 » implique « jeune conducteur ». Après l'abstraction des attributs, l'expert spécifie (inférence « spécifier1 »), un ensemble de critères à partir d'une « liste de critères » d'évaluation (e.g. nombre de modalités par attribut, nombre de valeurs manquantes, etc.). Il sélectionne un par un les critères (inférence « sélectionner-12 ») et il effectue une évaluation selon chacun de ces critères (inférence « évaluer1 »). Ensuite, il « combine » les différentes valeurs des critères pour « décider » de la « pertinence » de l'attribut en question. A ce niveau de la structure d'inférence, commence la sous-structure relative à la « modification » que nous avons présentée dans la Figure 7-9.

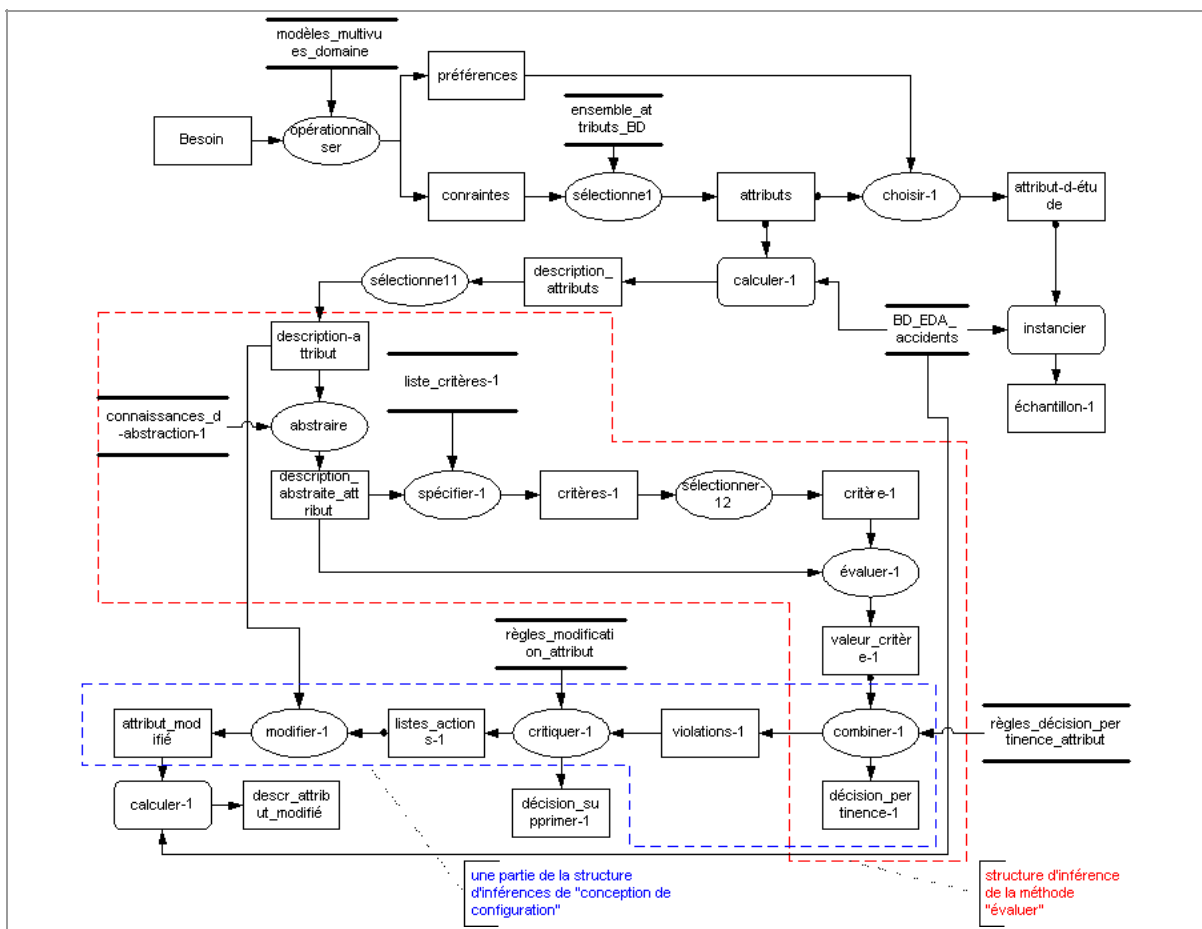


Figure 7-10- Structure d'inférence de sélection des attributs pertinents : combinaison de la structure d'inférence « d'évaluation » et d'une partie de la structure d'inférences de « conception de configuration ».

Le résultat de l'inférence « *combiner1* » peut être « *décision-pertience1* » (l'attribut est pertinent) ou « *violation-1* », c'est-à-dire un certain nombre de critères n'est pas satisfait (e.g. beaucoup de valeurs manquantes). Dans le deuxième cas, l'expert définit une « *liste d'actions* » pour remédier au problème (e.g. compléter les valeurs manquantes). Des « *modifications* » sont alors effectuées. Le même processus d'évaluation est ensuite appliqué sur l' « *attribut modifié* » jusqu'à où une décision soit prise : *sélectionner l'attribut* ou le *rejeter*.

7.2.2.2 La structure d'inférence de classification des accidents

Une fois l'échantillon d'étude préparé et donc les attributs et les accidents pertinents pour l'étude sélectionnés, la classification des accidents se fait automatiquement par un logiciel de data mining¹. L'intervention de l'expert se fait au niveau de la sélection des paramètres de classification tels que le nombre de classes, le nombre d'itérations, les attributs actifs² et les attributs supplémentaires³.

La technique de classification que nous avons appliquée est celle de la *classification mixte*, décrite dans le paragraphe §6.1.2.3 : il s'agit d'appliquer une *Analyse des Correspondances Multiple (ACM)* (cf. §6.1.2.2) suivie d'une *classification par partitionnement (k-means)* (cf. §6.1.2.1), suivie d'une *classification hiérarchique* (cf. §6.1.2.2).

Des logiciels, comme SPAD¹ ou SODAS⁴ (qui intègrent l'approche mixte de classification), offrent des interfaces appelées *filières* permettant de capitaliser la méthode utilisée. La Figure 7-11 illustre la filière que nous avons composée et utilisée pour la classification mixte (cf. 6.1.2.3).

¹ Le logiciel de data mining utilisé au LAB est SPAD qui est commercialisé par la société CISIA.

² ceux qui vont être utilisés dans le calcul des distances entre les accidents

³ Ceux qui vont être utilisés, avec les attributs actifs, pour caractériser les classes sans participer au calcul des distances.

¹ SPAD est logiciel de data mining commercialisé par la société DECISIA :
http://www.cisia.com/SPAD_Presentation.html.

⁴ Logiciel d'analyse de données symbolique, développé au laboratoire CEREMADE à l'université de Dauphine,
<http://www.ceremade.dauphine.fr/~touati/sodas-pagegarde.htm>.

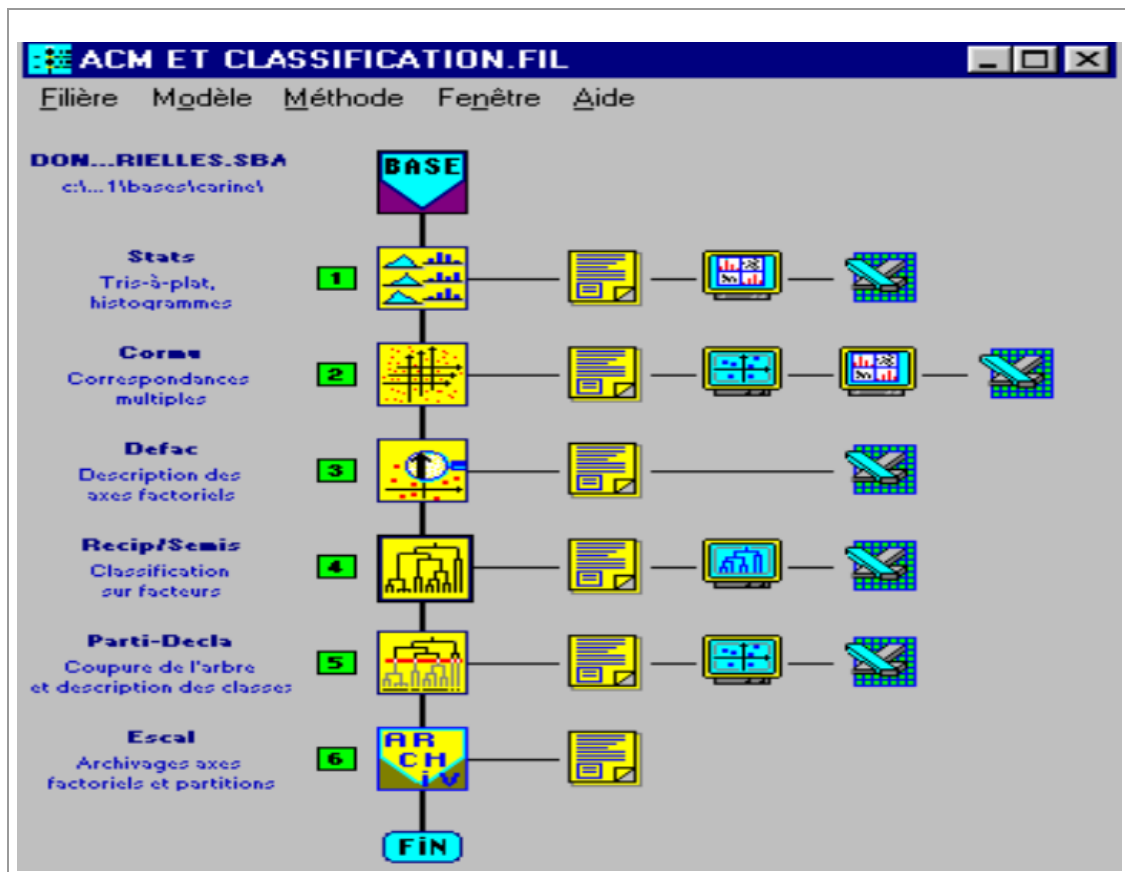


Figure 7-11- Illustrant de la méthode de classification mixte (Filière SPAD).

Ci-dessous quelques commentaires sur les différentes méthodes utilisées dans la filière présentée dans la Figure 7-11 :

- La méthode « Stats » permet d'effectuer des analyses statistiques des données à traiter pour aider l'utilisateur à sélectionner les attributs et les individus ;
- La méthode « Cormu » permet d'appliquer une ACM afin de transformer les données catégorielles en données numériques et la sélection des axes factoriels à utiliser dans la classification ;
- La méthode « Defac » permet la description des axes factoriels retenus ;
- La méthode « Recip/Semis » permet d'effectuer, à partir des composantes factorielles issues de l'ACM, une classification mixte (cf. §6.1.2.3) : application de k-means, obtention des groupements stables et application d'une classification hiérarchique sur ces groupements ;
- La méthode « Parti-Decla » permet de choisir visuellement le niveau hiérarchique adéquat pour la coupeure du dendrogramme afin d'obtenir la partition finale ;
- La méthode « Escal » permet d'archiver les résultats.

D'excellents exemples d'application avec des explications sur le paramétrage de chacune de ces méthodes peuvent être trouvés dans [Nakache et Confais, 2000].

7.2.2.3 La structure d’inférence d’interprétation des classes

Comme dans le cas de la sélection des attributs, l’interprétation des classes se fait à travers une structure d’inférence combinant l’évaluation et la modification. La lecture de cette structure s’effectue donc de la même manière que celle que nous avons présentée dans le paragraphe 7.2.2.1 (cf. Annexe A-1.2.3.3).

7.3. La connaissance de la tâche d’élaboration des STA

La *connaissance de la tâche*, troisième type de connaissances dans CommonKADS, permet de définir l’*application* (la tâche), ses *objectifs* et les *moyens pour les réaliser*. Les tâches peuvent être composées de sous-tâches, qui à leur tour peuvent être composées de tâches terminales indécomposables (i.e. inférences). Les connaissances de la tâche spécifient aussi le *contrôle* (ordre d’exécution, composition, etc.) autour des sous-tâches et inférences qui la composent. Dans notre cas, la tâche correspond à l’élaboration des STA.

7.3.1 La tâche d’élaboration des STA

Une tâche est une fonction de raisonnement complexe, indépendante de la connaissance du domaine. Cette indépendance vis-à-vis du domaine améliore son caractère générique ce qui permet de réutiliser son modèle dans différents domaines d’application. Dans le langage CML2 [Anjewierden et Schreiber, 2004], une tâche est caractérisée par un *objectif* et des *rôles* (entrées et sorties).

Pour illustrer par un exemple en CML2, nous présentons dans la Figure 7-12 uniquement la tâche « *sélectionner_attribut_d-étude* ». La description de l’ensemble des tâches dans tout le processus d’élaboration des STA est donnée dans l’annexe (cf. Annexe 1. A-1.3).

```

Task sélectionner_attribut_d-étude;
  domain-name: sélectionner_attribut_d-étude;
  Goal:
    "Sélectionner les attributs pertinents et construire l'échantillon duquel on va
     sélectionner les accidents pertinents pour construire l'échantillon d'étude";
  Roles:
    Input:
      description_attributs: "l'ensemble des attributs décrits par des paramètres
        statistiques permettant leur évaluation";
      'règles_abstraction_attributs': "Règles utilisées pour l'abstraction des
        attributs";
      'règles_décision_pertinence_attribut': "Règles expertes pour juger la
        pertinence des accidents";
      'règles_modification_attribut': "Règles expertes permettant de déterminer les
        actions adéquates à effectuer pour améliorer la pertinence d'un attribut (e.g.
        regrouper mdalités, remplacer valeurs inconnues, etc.) ou le supprimer";
      BD_EDA_accidents: "il s'agit de la base de données d'accident (EDA) contenant
        des tables";
    Output:
      échantillon-1: "échantillon d'accidents décrit par les attributs pertinents
        pour l'étude";
End task sélectionner_attribut_d-étude;
    
```

Figure 7-12- Tâche sélectionner_attribut_d-étude

On peut lire dans la Figure 7-12 que les entrées de la tâche de sélection des attributs sont « *la description des attributs* », les « *règles d’abstraction des attributs* », les « *règles de décision sur la pertinence des attributs* », les « *règles de modification des attributs non pertinents* » et les « *BD EDA* ». La sortie de cette tâche est « *l’échantillon d’étude* ».

7.3.2 La méthode de la tâche

La description de la décomposition d'une tâche en sous-tâche ainsi que le contrôle d'exécution des sous-tâches sont décrits par la *méthode de la tâche*.

Dans le langage CML2, une méthode de tâche est caractérisée par quatre rubriques. La première, « RÉALISE », indique la tâche concernée par la méthode. La deuxième, « DECOMPOSITION », décrit la stratégie employée pour résoudre la tâche, à travers une décomposition en sous-fonctions qui correspondent à une autre tâche, et/ou à une inférence. La troisième, « RÔLES INTERMÉDIAIRES », indique des rôles additionnels de la tâche, utilisés pour stocker temporairement les résultats du raisonnement. La dernière, « STRUCTURE DU CONTRÔLE » dicte l'ordre d'exécution des sous-tâches. C'est une structure de contrôle qui se lit comme un petit programme. La description en CML2 de l'ensemble des méthodes de toutes les tâches ainsi qu'une représentation graphique de ces tâches est donnée dans l'annexe (cf. Annexe 1. A-1.3).

7.4. Critiques

Revenons maintenant à notre problématique de base qui est l'identification et l'intégration des *connaissances du domaine* dans le processus d'ECD. Grâce à la structure d'inférence (cf. section 7.2), nous avons constaté que les connaissances du domaine que nous cherchons à identifier et incorporer dans le processus d'ECD correspondent aux *rôles de connaissances statiques*. Mais, nous avons constaté aussi que, bien que les *rôles de connaissances statiques* aient la même fonction (i.e. spécifier des connaissances du domaine utilisées pour réaliser une inférence), deux types de ces rôles peuvent être distingués :

- **Des rôles propres à la tâche d'élaboration des STA.** Exemple : le rôle « *connaissances_d-abstraction* » dont la représentation graphique est donnée dans la Figure 7-13. Les instances de connaissances relatives à ce rôle ont été formalisées dans la base de connaissances (cf. Figure 7-14). Le rôle de ces connaissances par rapport aux inférences est défini dans la structure d'inférence (cf. Figure 7-10) ;

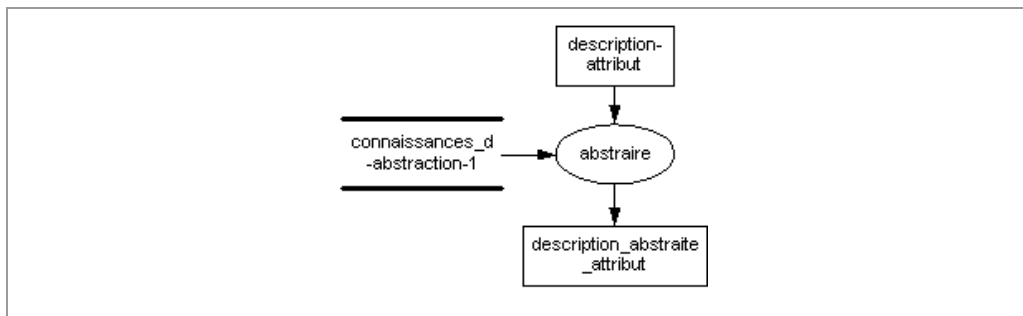


Figure 7-13- Utilisation du rôle de connaissance d'abstraction des attributs dans la structure d'inférence.

```

Attribut_d-étude.modalité.Valeur_test > 2
attribut_d-étude.modalité.Valeur_test <-2
attribut_d-étude.modalité.poids_mod_dans_class > 30
a-pour-abstraction
attribut_d-étude.modalité.pertinence = 'discriminante et
Caractérisante';
  
```

Figure 7-14- L'une des règles abstraction des attributs dans la base de connaissances.

Généralement, il existe des modèles génériques pour l'élicitation et la modélisation de ce type de connaissances relatives à la méthode de résolution du problème. La bibliothèque de CommonKADS [Breuker et Van de Velde, 1994], par exemple, offre un certain nombre de modèles génériques de diagnostic, d'évaluation, etc.

- **Des rôles propres au phénomène étudié** (i.e. accident) **et qui sont plus ou moins indépendants de la tâche** d'élaboration des STA. Exemple : le rôle « *modèles_multivues_domaine* » (cf. Figure 7-15).

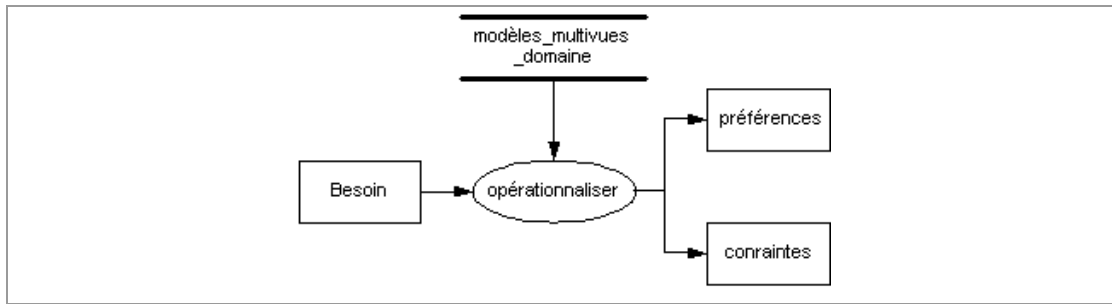


Figure 7-15- Utilisation du rôle de connaissances statique « modèles_multivues_domaine » par l'inférence « opérationnaliser ».

Ce rôle correspond en réalité à des représentations de l'accident auxquelles l'expert fait appel, explicitement ou implicitement, lors de l'élaboration des STA. Parmi ces modèles, on trouve le *modèle CVE*, le *modèle séquentiel*, le *modèle de traitement d'information* et le *modèle de la tâche* qui peuvent être utilisés pour sélectionner les attributs d'étude ou pour interpréter les classes. Ces modèles sont donc bien indépendants de la tâche d'élaboration des STA car ils sont des modèles généraux en accidentologie. De plus, ils sont difficilement, voire non formalisables à travers des règles dans la base des connaissances et c'est pour cela que nous les désignons par « boîtes noires ». CommonKADS permet d'identifier ce type de connaissances statiques à travers la structure d'inférence (cf. Figure 7-15), mais ne fournit pas de modèles guidant leur élicitation.

7.5. Synthèse

Ce chapitre a été consacré à l'élaboration d'un modèle de connaissances pour la classification d'accidents en utilisant la méthode CommonKADS [Schreiber *et al.*, 1994c]. Trois types de connaissances ont été spécifiés : la *connaissance du domaine*, la *connaissance d'inférences* et la *connaissance de la tâche*. Dans la section 7.1, nous avons élaboré un modèle de la *connaissance du domaine* qui est composée de deux éléments : le *schéma du domaine* et la *base de connaissances*. Le *schéma du domaine* contient les *concepts*, les *relations* entre ces concepts et les *types-de-règles*. La *base des connaissances* contient les *instances des connaissances* sur lesquelles le raisonnement devra s'effectuer.

Dans la section 7.2, nous avons élaboré un modèle de la *connaissance d'inférences* qui comporte deux éléments : les *inférences* et les *rôles de connaissances*. Les inférences sont les unités de traitement de l'information nécessaires au processus de raisonnement lors de l'élaboration des STA. Les rôles de connaissances représentent les rôles que jouent les concepts du domaine par rapport aux inférences. Deux types de rôles de connaissances ont été distingués : les *rôles dynamiques* et les *rôles statiques*. Les *rôles dynamiques* sont les entrées et les sorties (durant le temps d'exécution) des inférences. Les *rôles statiques* sont des ressources stables, qui spécifient la connaissance du domaine dont on a besoin pour réaliser l'inférence. La section 7.3 a été consacrée à l'élaboration de la *connaissance de la tâche*. Elle nous a permis la décomposition de la tâche d'élaboration des STA en sous-fonctions ainsi que la représentation de la structure de contrôle d'exécution de ces sous-fonctions. La dernière section (7.4) a été consacrée aux limites de CommonKADS auxquelles nous remédions dans notre approche SAFE-Next que nous développons dans le chapitre suivant.

Chapitre 8.

ASMEC : une Approche Systémique de Modélisation dEs Connaissances

ASMEC est la première approche que nous développons dans SAFE-Next. Son objectif est de *modéliser les connaissances du domaine selon différents points de vue et niveaux d'abstraction et de les opérationnaliser sous forme de méta-données* (cf. Figure 8-1). Ces méta-données sont utilisées par la deuxième approche de SAFE-Next, AICEF, pour l'incorporation des connaissances expertes dans le processus d'ECD (cf. Chapitre 9). Elles sont aussi utilisées par la troisième approche, ASAIC, pour l'analyse d'impact de changement (cf. Chapitre 10).

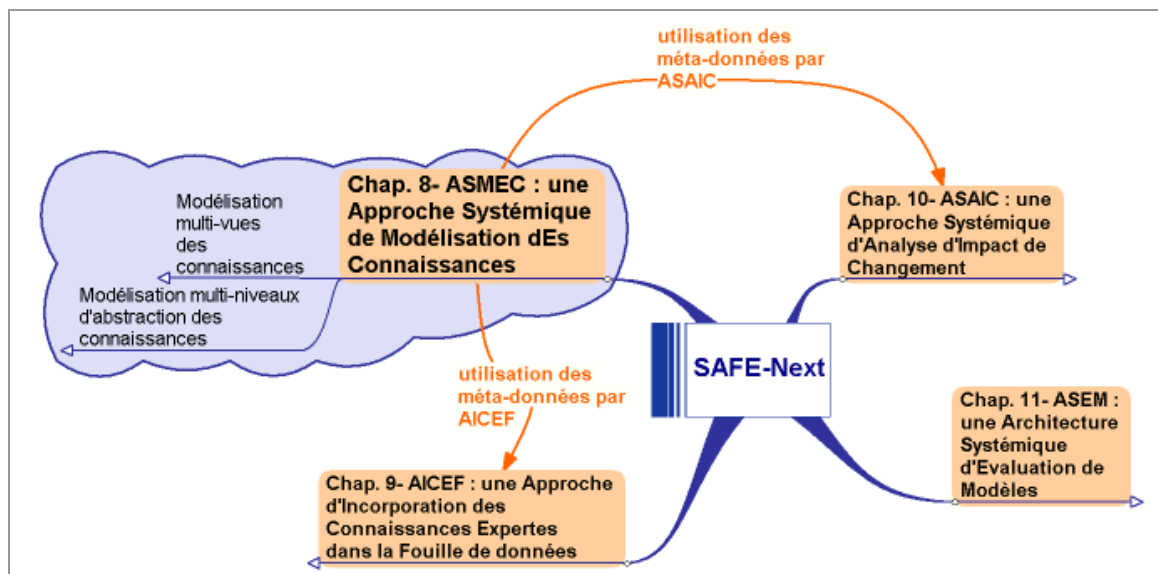


Figure 8-1- ASMEC : approche systémique de modélisation multi-vues et multi-niveaux de granularité de connaissances (première approche dans SAFE-Next)

8.1. Objectifs d'ASMEC

L'objectif d'ASMEC est de pallier deux limites de CommonKADS que nous résumons à travers les deux points suivants :

- Dans CommonKADS, certains rôles de connaissances statiques indépendantes de la MRP comme « *modèles_multivues_domaine* » (cf. section 7.4) ne peuvent pas être exprimés sous forme de règles et restent sous forme de « boîtes noires ». L'élicitation est le seul moyen pour formaliser ce type de connaissances. Cependant, s'il existe aujourd'hui des modèles génériques pour guider l'élicitation des connaissances relatives à la MRP (e.g. bibliothèque CommonKADS), ce n'est pas le cas pour les connaissances indépendantes de la MRP. Dans le paragraphe 5.3.3.2, nous avons montré l'importance de disposer d'un modèle général pour guider l'élicitation des connaissances et éviter certains risques (e.g. incomplétude) : c'est l'une des vocations du méta-modèle d'ASMEC (cf. section 8.4) ;
- Le schéma du domaine dans CommonKADS mélange des concepts portant sur différents aspects systémiques (ontologique, fonctionnel, transformationnel et téléologique) (cf. §7.1.1.3). Nous montrons dans le Chapitre 9 l'intérêt de distinguer les différents types de concepts notamment pour les tâches de sélection des attributs et d'interprétation des classes d'accidents. Nous montrons que la distinction entre ces différents types de concepts permet d'élaborer un schéma multi-vues du domaine, l'un des points faibles dans CommonKADS. Les auteurs de CommonKADS sont conscients de cette limite lorsqu'ils préviennent les lecteurs dans leur livre « *Knowledge Engineering and Management: The CommonKADS methodology* » [Schreiber *et al.*, 1994c] en écrivant : « *Having multiple experts is a considerable risk factor. In the context of this book we cannot go into details about multi-expert situation ...* ».

L'objectif principal de notre approche ASMEC est donc de fournir un outil efficace pour guider l'élicitation de connaissances statiques indépendantes de la MRP et pour la construction d'un *schéma du domaine multi-vues et multi-niveaux de granularité*. Mais, nous montrons également qu'ASMEC permet d'intégrer et de structurer les différents modèles de CommonKADS (cf. §8.6.2.1 et §8.6.2.2).

8.2. Choix d'une stratégie de modélisation

Etant donnée que les rôles de connaissances statiques type « boîte noire » correspondent à des modèles (ou schémas) implicites ou explicites utilisés par les experts, la question qui se pose est la suivante : *Comment identifier et/ou développer ces modèles ? Faut-il se baser sur l'élicitation des connaissances avec les experts participant à l'étude ou faut-il plutôt effectuer un travail de modélisation de l'accident lui-même ?* La différence est que dans le premier cas, il s'agit de modéliser la perception de l'accident par les experts participant à la phase d'élicitation. Cette perception peut être influencée par la spécialité et les objectifs de chacun de ces experts. Alors que dans le deuxième cas, il s'agit de modéliser l'accident avec une autre perspective : l'identification et l'élaboration de modèles de l'accident pour les utiliser comme support d'incorporation de connaissances du domaine dans le processus d'ECD.

Dans le paragraphes 5.3.1.3, nous avons présenté les limites d'une approche *dirigée par les données* (i.e. ascendante) pour la modélisation des connaissances (e.g. coût, gestion de conflits, etc.). Dans le paragraphe 5.3.2.3, nous avons montré que les modèles génériques utilisés dans les approches descendantes permettent de résoudre en partie le problème. Cependant, ces modèles peuvent guider l'élicitation des connaissances relatives à la tâche de résolution de problème, mais pas celles indépendantes de cette tâche comme le rôle « *modèles_multivoies_domaine* ». Dans la section suivante (section 8.3), nous proposons une approche dirigée par un méta-modèle pour modéliser ce type de connaissances et guider leur élicitation.

8.3. Modélisation dirigée par un méta-modèle général

Cette stratégie de modélisation consiste à modéliser l'accident en partant d'un méta-modèle conceptuel comme le *système général* au sens de la systémique (cf. section 3.6). Nous nous inspirons des étapes de modélisation en trois phases proposée par Le Moigne (1999) : le *cadrage du modèle*, le *développement du modèle* et la *l'opérationnalisation du modèle* (le terme utilisé par Le Moigne est l'interprétation du modèle) (Figure 8-2).

1. Le *cadrage du modèle* : cette phase est constituée de deux étapes. Il s'agit de construire dans la première un *méta-modèle* qui doit être suffisamment *général* et *stable* pour que l'on puisse rendre compte de tous les types de complexité du phénomène (ou système) étudié. Le Moigne (1999) appelle ce *méta-modèle* le *système général*. Le méta-modèle que nous proposons est une conjonction d'un ensemble de *points de vue génériques* (ou perspectives (cf. section 5.5)) qui sont fondés sur les différents axiomes de la systémique présentés dans les sections 3.4 et 3.6. Il s'agit ensuite de construire le(s) *modèle(s)* du phénomène étudié par *isomorphie* avec le méta-modèle lui-même (Figure 8-2.a). Autrement dit, il s'agit de développer des modèles selon les différents points de vue du méta-modèle. Le modèle séquentiel, par exemple, est un modèle qui instrumente le point de vue transformationnel. Le modèle CVE est un modèle qui instrumente le point de vue ontologique ;
2. Le *développement du modèle* : il s'agit d'instrumenter (documenter selon les termes de Le Moigne) le(s) modèle(s) par une correspondance *homomorphique* avec les traits perçus du phénomène modélisé (Figure 8-2.b). Pour le faire, nous proposons d'instrumenter (composer) chacun des modèles par des *concepts* du domaine. Par exemple, « *la conduite normale* », « *la rupture* », « *l'urgence* » et le « *choc* » sont quatre concepts qui instrumentent le modèle séquentiel. De même, les concepts « *conducteur* », « *environnement* » et « *véhicule* » instrumentent le modèle CVE ;
3. L'*opérationnalisation du modèle* : il s'agit de rendre les modèles opérationnels pour pouvoir les utiliser et les interpréter pour en tirer les connaissances recherchées.

L'avantage de cette stratégie de modélisation réside dans le fait que le méta-modèle, grâce à sa généralité, permet l'intégration des différents points de vue existants des experts. De plus, il garantit, grâce à la systémique, une certaine complétude des représentations élaborées. Il réduit ainsi les risques d'omettre un point de vue nécessaire à l'identification des connaissances sur le phénomène (cf. §8.6.2.3). Cette vision systémique permet non seulement d'organiser et de structurer

les concepts existants dans le domaine selon les différents axes systémiques, mais aussi de guider les experts pour en créer de nouveaux (cf. la fin de la section 8.8.5).

Nous utilisons cette approche de modélisation pour développer ASMEC. Elle se base sur tous les fondements du constructivisme et de la systémique présentés dans le Chapitre 3 et sur la stratégie de modélisation dirigée par un méta-modèle général que nous venons de présenter et que nous illustrons dans la Figure 8-2.

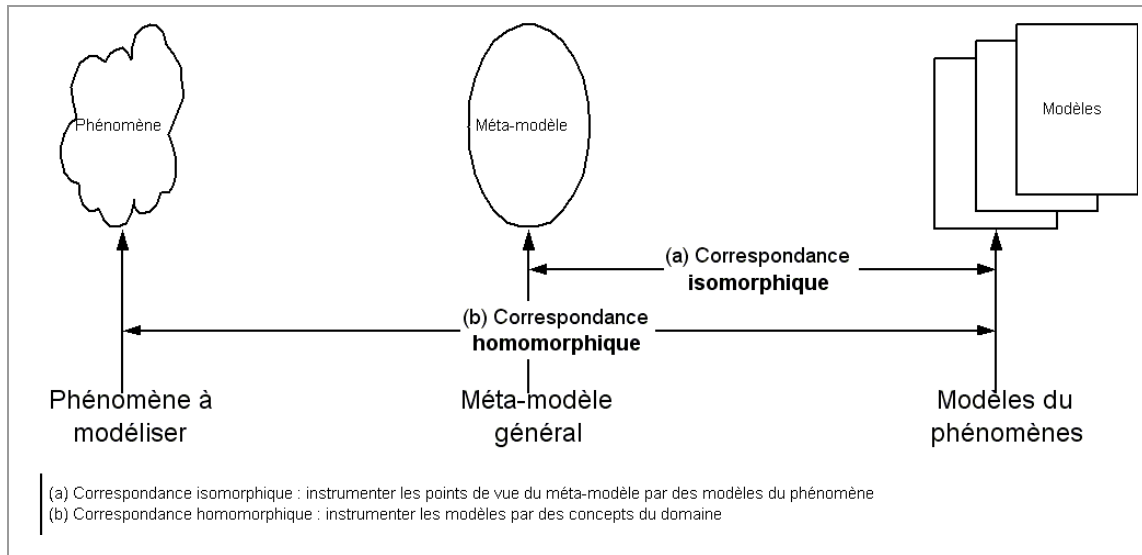


Figure 8-2- Modélisation dirigée par un méta-modèle général (d'après [Le Moigne, 1999]).

8.4. Architecture globale d'ASMEC

En se basant sur la stratégie de modélisation dirigée par un méta-modèle (cf. section 8.3), nous proposons une *architecture de modélisation à plusieurs niveaux d'abstraction* :

1. **Niveau 1 : méta-modèle.** Le méta-modèle doit permettre l'intégration des différents points de vue nécessaires à la construction des connaissances recherchées. Pour le construire, nous proposons les points de vue de la cybernétique de second ordre (i.e. *objet observé*, *observateur* et *contexte d'observation*) (cf. section 3.6) et les axes systémiques (i.e. *ontologique*, *fonctionnel*, *transformationnel* et *téléologique*) (cf. section 3.7). Ces points de vue et ces axes vérifient le critère de généralité. Le méta-modèle sera développé dans la section suivante (cf. 8.5) ;
2. **Niveau 2 : modèle.** Un point de vue au niveau méta-modèle précédent est instrumenté à l'aide de plusieurs modèles. Ces modèles peuvent être préexistants dans le cas d'une approche descendante ou ils sont développés à partir des données dans une approche ascendante. Par exemple, le *modèle séquentiel* permet d'instrumenter l'*axe transformationnel* du point de vue *objet observé* (i.e. l'accident). Ce niveau sera développé dans la section 8.6 ;
3. **Niveau 3 : concept.** Chacun des modèles au niveau précédent est instrumenté par plusieurs concepts du domaine. Exemple : le modèle séquentiel est instrumenté par les concepts « *conduite normale* », « *rupture* », « *urgence* » et « *choc* ». Ce niveau sera développé dans la section 8.7 ;

8.5. Spécification du méta-modèle

Pour la spécification du *méta-modèle général* dans ASMEC qui consiste à l'instrumenter par différents points de vue génériques, nous proposons deux niveaux de points de vue reliés grâce au principe de réflexivité de la systémique (cf. section 3.7) :

Le premier se base sur le principe de la *cybernétique de second ordre* [Von Foerster, 1995] (cf. section 3.6). Il s'agit d'une vision systémique de l'application dont il est question dans l'étude, i.e. élaboration de STA. ASMEC assimile cette application à un système qu'elle analyse selon les trois points de vue suivants : *l'observateur*, *l'objet observé* et le *contexte d'observation* (cf. Figure 8-4) :

- **Point de vue « objet observé »** : l'objet observé correspond au phénomène ou système étudié dans l'application. Dans notre cas, l'objet observé correspond à l'accident. Ce point de vue concerne les *connaissances statiques du domaine* qui sont indépendantes de l'application. Autrement dit, il correspond aux connaissances en accidentologie qui sont indépendantes de la tâche d'élaboration des STA ;
- **Point de vue « observateur »** (ou **point de vue application**) : il intègre *l'objectif de l'observation* et le *processus d'observation*. Ce point de vue concerne tout ce qui est *propre à l'application* et donc à la *résolution de problème* et les *connaissances mises en œuvre pour cette résolution*. Dans notre application, « *l'expert* » représente *l'observateur*, « *élaborer des STA* » représente *l'objectif de l'observation* et la « *méthode de résolution du problème (MRP) d'élaboration des STA* » représente *le processus d'observation* ; non seulement les aspects de résolution du problème sont abordés, mais aussi la connaissance mise en œuvre pour cette résolution est modélisée à part entière.
- **Point de vue « contexte de l'observation »** : le contexte de l'observation est celui de la résolution du problème en question. Il intègre l'environnement dans lequel se déroule la modélisation. Dans notre cas, le contexte de l'observation est la conception de systèmes de sécurité.

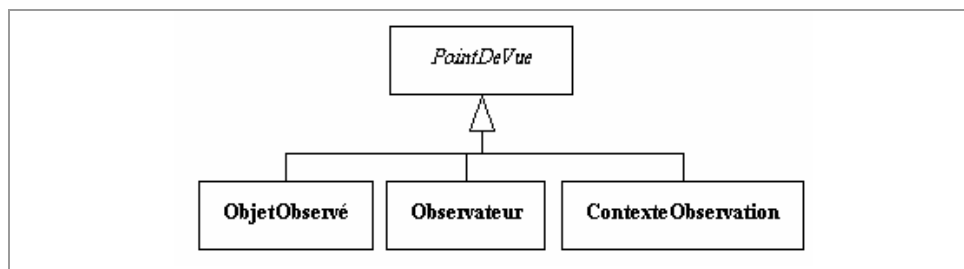


Figure 8-4- Modèle UML du premier niveau du méta-modèle.

Le deuxième niveau de points de vue se base sur la *réflexivité de la systémique* (cf. section 3.7) en faisant correspondre un *système* à chacun des points de vue du niveau précédent (i.e. objet observé, observateur et contexte) et ensuite en analysant chacun de ces « *systèmes points de vue* » selon les quatre axes systémiques suivants (cf. Figure 8-5 et Figure 8-6) :

- *Axe ontologique* : ce qu’est le système,
- *Axe fonctionnel* : ce que fait le système,
- *Axe transformationnel et génétique* : il décrit l’aspect dynamique, évolutionnel et génétique (dans le sens de la genèse et non celui de l’hérédité) du système,
- *Axe téléologique* : quel est l’objectif du système ?

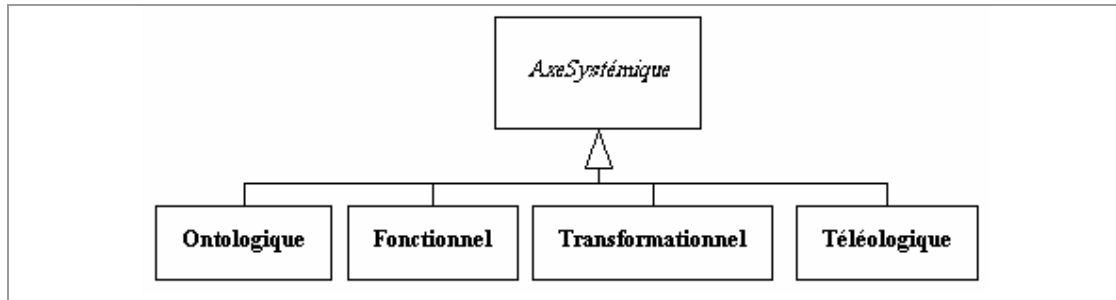


Figure 8-5- Modèle UML du deuxième niveau du méta-modèle.

Avec ces deux niveaux, le méta-modèle présente une combinaison de points de vue (3x4) (cf. Figure 8-6). Le nombre de ces derniers peut augmenter grâce à la réflexivité de l’approche systémique (cf. section 3.7) car chacun des axes systémiques peut, à son tour, être analysé en fonction des autres axes. Mais, dans notre cas, l’application des deux niveaux ci-dessus nous a montré qu’ils sont suffisants pour représenter les différentes connaissances que nous avons pu identifier. Nous nous limitons donc à ces deux niveaux. Cela donne un niveau acceptable de complexité à l’approche et donc assure la facilité de sa compréhension et de son utilisation.

Une représentation UML du méta-modèle est présentée dans la Figure 8-6.

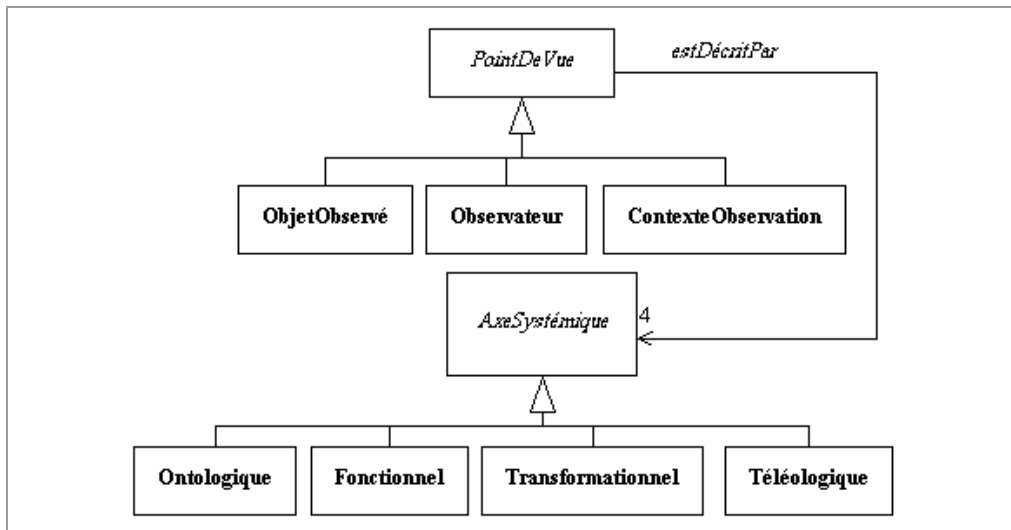


Figure 8-6- Modèle UML du méta-modèle dans ASMEC : chacun des points de vue du premier niveau est analysé selon les quatre axes systémiques

8.6. Instrumentation du méta-modèle par les modèles

Il s'agit dans cette section d'identifier des modèles existants et/ou de construire de nouveaux modèles par *isomorphie* avec les points de vue du méta-modèle présentés dans la section précédente. Concrètement, il s'agit de développer des modèles selon les différents axes systémiques pour chacun des points de vue du premier niveau du méta-modèle (cf. Figure 8-6 et Figure 8-7).

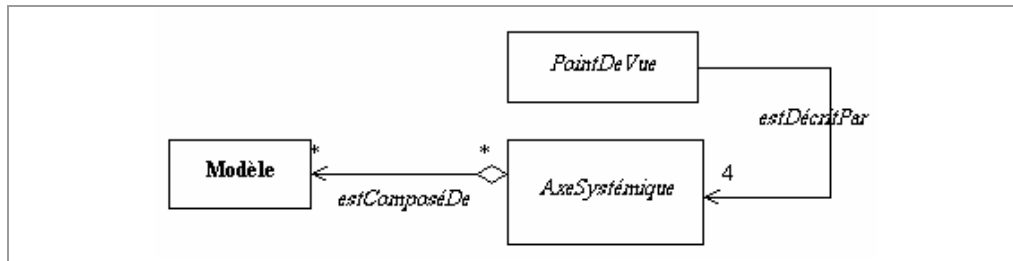


Figure 8-7- Instrumentation des axes systémiques par des modèles.

8.6.1 Comment choisir /construire un modèle pour l'instrumentation d'un axe ?

Plusieurs remarques sont à noter concernant le choix ou la construction d'un modèle pour instrumenter un point de vue :

- Un axe au niveau d'un seul point de vue peut contenir plusieurs modèles. Par exemple, les modèles structurels du « conducteur », du « véhicule » et de « l'environnement » appartiennent tous au même axe ontologique du point de vue « objet observé » ;
- Un modèle ne fait pas obligatoirement partie d'un axe unique. Il peut être composé de concepts appartenant à des axes différents dans un même point de vue. Par exemple, le modèle du « Single Vehicle Accident » [Kurucz et al., 1977] (cf. section 4.3.2) décrit l'accident à l'aide de cinq concepts suivant : « manœuvre pré-accidentelle », « tâche du conducteur », « traitement de l'information », « action conducteur » et « réaction du conducteur ». Dans ce modèle, on trouve des concepts relatifs à la *téléologie* du conducteur (e.g. tâche) et à son *fonctionnement* (e.g. traitement d'information, action). De plus ces concepts sont présentés selon un ordre séquentiel appartenant à l'axe transformationnel. L'apport de la vision systémique dans ce cas réside dans le fait qu'elle permette de distinguer les différents aspects systémiques des concepts dans le modèle. Cela peut être bénéfique pour la compréhension et donc l'utilisation du modèle ;
- Un modèle peut être le résultat de la conjonction de concepts issus d'axes différents appartenant à des points de vue différents. Exemple : le modèle de « représentation de l'accident chez un concepteur » est un modèle pertinent pour notre application. Nous pouvons l'utiliser pour projeter les STA dans l'espace de représentation du concepteur pour que ces STA soient compréhensibles par ce dernier et donc mieux utilisables. Ce modèle porte sur des axes du point de vue objet observé puisqu'il concerne l'objet « accident ». Mais, ces aspects sont projetés aussi sur des axes du point de vue de l'agent « concepteur » qui appartient au point de vue contexte de l'observation ;

- Grâce à la réflexivité de la systémique (cf. section 3.7), un modèle au niveau d'un axe (ou combinant plusieurs axes) peut être considéré comme un sous-système qu'on peut projeter, à son tour, sur les autres axes systémiques. Prenons l'exemple du modèle « *conducteur* » affecté à l'axe ontologique du *point de vue objet observé*. Il s'agit d'un *sous-système* que nous pouvons projeter, par exemple, sur l'axe fonctionnel ce qui nous donne le *modèle de traitement d'information*. Ce dernier est assimilable aussi à un sous-système qu'on peut projeter sur les autres axes systémiques ;
- Dans le même ordre d'idées que la remarque précédente, un modèle peut être considéré, à son tour, comme un point de vue. Le modèle « *conducteur* » est finalement un point de vue sur lequel on peut projeter d'autres modèles. Une projection du « *modèle séquentiel* » selon le point de vue « *conducteur* » donne une description séquentielle du comportement du conducteur ;
- C'est finalement le *point de vue observateur* qui joue un rôle central dans le choix ou la construction d'un modèle dans ASMEC. Il intervient au niveau de l'identification d'abord des *points de vue pertinents* dans le premier niveau du méta-modèle (i.e. objet, observateur ou contexte), ensuite des *axes pertinents* au sein de chacun des points de vue et en fin pour choisir ou construire des *modèles pertinents* au sein de chacun des axes, ensuite des concepts pertinents au sein d'un modèle et enfin les attributs pertinents caractérisant un concept. La notion de pertinence est bien sûr définie par rapport au contexte de l'étude (objectif, etc.) ;
- Les modèles instrumentant un point de vue donné sont choisis en tenant compte des autres points de vue.

Dans le paragraphe suivant, nous présentons les différents modèles que nous avons choisis et construits pour instrumenter notre méta-modèle.

8.6.2 Les modèles choisis pour l'instrumentation des différents points de vue

8.6.2.1 Instrumentation du point de vue observateur

Etant donné que ce point de vue concerne tout ce qui est *propre à la Méthode de Résolution du Problème (MRP)*, plusieurs composantes du *modèle conceptuel d'expertise* de CommonKADS (cf. section 5.3.2) peuvent être utilisées pour instrumenter les différents axes de ce point de vue. Une grande partie de ce point de vue a été développée dans le Chapitre 7 lors de l'élaboration d'un modèle de connaissance en utilisant CommonKADS. Notre objectif dans ce paragraphe est de montrer que les axes systémiques que nous proposons dans le méta-modèle restent génériques pour modéliser les connaissances relatives à la MRP. Autrement dit, nous montrons que les modèles de CommonKADS¹ peuvent être utilisés pour instrumenter les axes de ce point de vue :

- *Axe ontologique* : cet axe concerne les modèles des objets manipulés (e.g. concepts du domaine) et des objets de manipulation (e.g. outils logiciels) utilisés par l'expert et qui sont

¹ les modèles de l'organisation, de tâches, d'agents, de communication, d'expertise et de conception (cf. § 5.3.2.1))

propres à la MRP. Au niveau des concepts, cet axe concerne tous ceux qui sont relatifs aux connaissances dynamiques. Par exemple, le concept « *classe* » fait partie de cet axe car il est propre à la tâche d'élaboration des STA. Ceci n'est pas le cas du concept « *accident* » qui est indépendant de cette tâche. Au niveau des attributs, nous avons développé deux nouveaux modèles selon l'axe ontologique : le « *modèle de la pertinence générale des attributs* » et le « *modèle de fiabilité des attributs* ». Il s'agit dans le premier d'affecter, par les experts, un niveau d'importance (*très important, important, plutôt important, peu important et pas important*) pour chaque attribut. L'affectation d'un poids à un attribut est effectuée selon sa contribution à la compréhension du déroulement de l'accident d'une façon générale (car la notion d'importance est relative à l'objectif de l'étude). Ceci est d'une valeur certaine pour un utilisateur pour la sélection des attributs en tenant compte de leur importance générale. En ce qui concerne le « *modèle de fiabilité des attributs* », il s'agit d'affecter un degré de fiabilité à chacun des attributs (*fiable, peu fiable et pas fiable*). Les données issues de mesures (e.g. la température) sont généralement fiables. Les données issues des déclarations du conducteur (e.g. sa vitesse et ses actions) sont peu fiables. Ceci est important pour un utilisateur pour affecter, par exemple, des poids aux attributs dans une opération de data mining ;

- *Axe transformationnel/génétique* : une structure d'inférences dans CommonKADS comme celle que nous avons développée dans la section 7.2.2 peut être perçue comme un processus de transformation des entrées et des rôles de connaissances pour aboutir au résultat recherché. Elle exprime donc l'aspect transformationnel et génétique : les différentes étapes de résolution du problème décrites par les inférences illustrent l'aspect transformationnel alors que les rôles de connaissances intermédiaires illustrent l'aspect génétique. Les règles de la base de connaissances (cf. section 7.1.2) qui permettent la manipulation des rôles de connaissances dynamiques font partie aussi de cet axe.
- *Axe fonctionnel* : Les *méthodes de tâches* comme celles que nous avons développées dans le paragraphe §7.3.2 permettent la description, à travers les structures de contrôle d'exécution des inférences, le fonctionnement du raisonnement de l'expert durant le processus d'élaboration des STA. Elles peuvent donc être utilisées pour instrumenter cet axe ;
- *Axe téléologique* : il correspond à l'objectif de l'expert. Les *tâches* dans le modèle des *connaissances de tâche* (cf. §7.3.1) illustrent cet axe car elles permettent de décrire et décomposer les objectifs de l'expert durant le processus d'élaboration des STA.

8.6.2.2 Instrumentation du point de vue contexte de l'observation

Comme dans le paragraphe précédent, nous montrons dans ce paragraphe que les axes systémiques restent génériques pour modéliser les *connaissances relatives au contexte* : nous montrons qu'ils intègrent les différents modèles proposés par CommonKADS :

- *Axe ontologique* : le *modèle d'agents*, qui décrit les caractéristiques des agents impliqués (humains, systèmes d'information et autres entités), est un exemple de modèle pouvant instrumenter l'axe ontologique du contexte. Le *modèle de l'organisation* (i.e. structure, processus, personnel et ressources) appartient aussi à cet axe ;

- *Axe fonctionnel* : le modèle de communication représentant le système d'interaction entre les agents est un exemple d'instrumentation de l'axe fonctionnel du contexte ;
- *Axe transformationnel et génétique* : le modèle de l'organisation permet aussi l'instrumentation de cet axe. En effet, il est utilisé pour l'analyse des facettes majeures de l'organisation (structure, processus, personnel et ressources) afin de déterminer la faisabilité des solutions SBC et d'évaluer leur impact sur l'organisation ;
- *Axe téléologique* : le modèle des tâches qui permet d'identifier et d'analyser les tâches globales de l'organisation est un exemple de modèle permettant l'instrumentation de l'axe téléologique du contexte ;

Quant à notre application, la tâche principale exprimée par l'expert est de « *construire des STA utilisables par le concepteur pour développer et évaluer des prestations de sécurité* ». Cette tâche nous a aidé à identifier les modèles pertinents pour le *point de vue contexte*. L'unique point de vue pertinent pour notre application est celui de l'agent « *concepteur* » et c'est précisément l'axe téléologique de ce sous-système qui nous intéresse. Cet axe traduit les objectifs du concepteur derrière l'utilisation des STA. Ceci nous a amenés à réfléchir à un moyen permettant de relier les objectifs du concepteur à la tâche d'élaboration des STA dont il est l'utilisateur final. La solution que nous proposons consiste à développer des modèles de représentation de l'accident selon le point de vue du concepteur. Ces modèles vont nous servir d'espace de projection des STA pour qu'ils soient utilisables par le concepteur et un moyen de communication avec les accidentologues. Il s'agit de deux modèles : le « *modèle véhicule orienté sécurité* » et le « *modèle systèmes de sécurité* » que nous décrivons ci-dessous :

- *Modèle véhicule orienté sécurité* : il ne s'agit pas de présenter un modèle complet du véhicule, mais uniquement une décomposition selon le point de vue « *sécurité* ». Autrement dit, ce modèle présente les composantes du véhicule qui sont impliqués dans l'analyse d'un accident. Il a pour objectif de permettre la « *lecture* » d'un accident à travers le véhicule. Cette perception est a priori plus proche du concepteur de systèmes de sécurité que celle, par exemple, qui analyse l'accident à travers le modèle de traitement d'information. Nous nous sommes basés sur le livre de codage d'informations dans la base de données EDA [CEESAR, 2000] pour définir ce modèle. Deux concepts principaux le composent : « *véhicule sécurité primaire* » et « *véhicule sécurité secondaire* » qui sont composés à leur tour de sous-concepts tels que « *défauts* », « *performance* », « *commandes* », etc. (dans la section 8.7, nous présentons tous les concepts caractérisant ce modèle) ;
- *Modèle systèmes de sécurité* : étant donné que l'objectif des STA est de développer des systèmes de sécurité, nous avons proposé une lecture de l'accident selon les familles des systèmes de sécurité existantes. Ce modèle a pour objectif d'identifier rapidement les fonctions de sécurité qui pourraient être concernées dans un accident ou un groupe d'accidents. L'analyse de l'accident selon ce point de vue permet de guider la réflexion sur de nouvelles prestations de sécurité ou sur une évaluation et/ou éventuelle amélioration des prestations existantes. Nous avons identifié deux grandes catégories de systèmes de sécurité : des *systèmes d'information et d'alerte* et des *systèmes d'assistance et d'intervention* [Ben Hassine, 2004]. Chacune de ces deux familles est décomposée en sous-catégories (cf. §8.6.2.3).

8.6.2.3 Instrumentation du point de vue objet observé

A travers les deux paragraphes précédents (§8.6.2.1 et §8.6.2.2), nous avons montré que tous les modèles fournis par CommonKADS permettent d'instrumenter les différents axes systémiques des deux points de vue « *observateur* » et « *contexte d'observation* ». Cependant, en ce qui concerne le troisième point de vue (i.e. « *objet observé* »), les modèles que nous allons identifier par ASMEC n'ont pas été identifiés lorsque nous avons utilisé CommonKADS dans le Chapitre 7.

L'instrumentation du *point de vue* « *objet observé* » tient compte des deux autres points de vue. En effet, le *point de vue observateur* nous a permis d'identifier les représentations de l'accident, donc de l'objet observé, utilisées par l'expert lors de l'élaboration des STA. Quant au *point de vue contexte*, il nous a permis de développer les deux nouveaux modèles « *véhicule orienté sécurité* » et « *systèmes de sécurité* » que nous avons décrits à la fin du paragraphe précédent. Les différents modèles que nous avons choisi pour instrumenter le *point de vue* « *objet observé* » sont distribués en fonctions des axes systémiques comme suit :

- *Axe ontologique* :
 - Le « *modèle CVE* » (cf. section 4.3.1),
 - Les modèles « *véhicule orienté sécurité* » et « *systèmes de sécurité* » identifiés grâce au *point de vue contexte de l'observation* (cf. la fin du §8.6.2.2) ;
- *Axe transformationnel/génétiq*ue : le « *modèle séquentiel* » (cf. section 4.3.2) ;
- *Axe fonctionnel* : le « *modèle de traitement de l'information* » (cf. section 4.3.3) ;
- *Axe téléologique* : le « *modèle de la tâche de conduite* » (cf. section 4.3.4)

Dans le projet ACACIA [Dieng *et al.*, 1996] (cf. §5.3.3.1), le « *modèle systèmes de sécurité* » et « *modèle véhicule orienté sécurité* » n'ont pas été identifiés. Nous expliquons cela par le fait que la tâche élicitée concernait le « *diagnostic* » et non pas « *l'élaboration des STA* » et que les acteurs (ceux qui ont participé à l'étude) ne sont pas les mêmes et n'avaient pas les mêmes objectifs. Cela montre que le développement du *point de vue* « *objet observé* » dans ASMEC tient en compte des points de vue « *observateur* » et « *contexte d'observation* ». La Figure 8-8 illustre le fait que le développement des modèles du *point de vue* « *objet observé* » tient compte des points de vue « *observateur* » et « *contexte de l'observation* ».

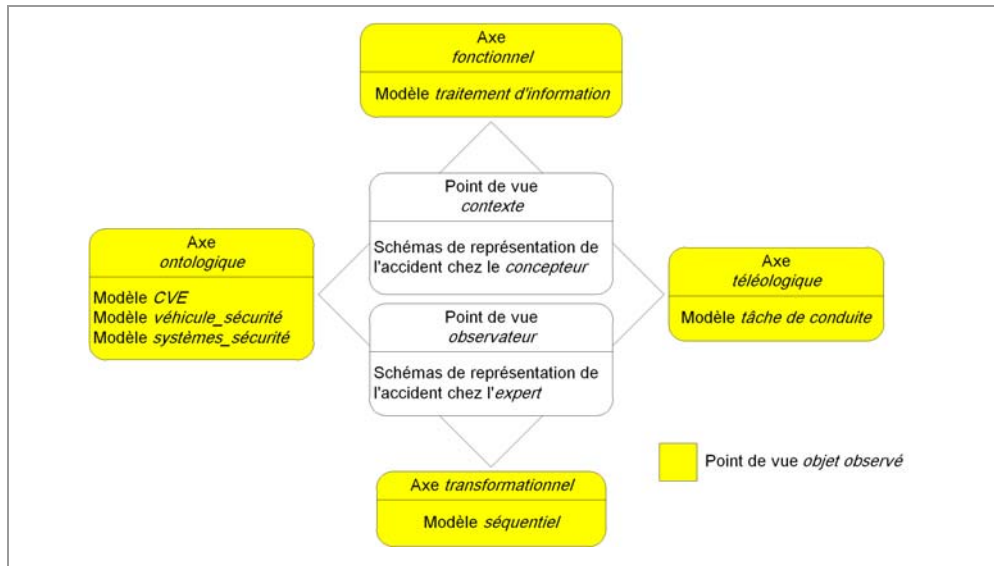


Figure 8-8- Développement du point de vue « objet observé » en tenant compte des points de vue « observateur » et « contexte d'observation ».

Dans ACACIA, la bibliothèque de CommonKADS a été utilisée pour construire les modèles de la MRP de diagnostic (i.e. les connaissances dynamiques). Cependant, l'élaboration des représentations de l'accident chez les experts (i.e. les connaissances statiques) a été basée essentiellement sur l'élicitation. Elle n'a pas été guidée par un méta-modèle comme nous le proposons. Il suffit donc qu'un expert ayant un point de vue différent ne participe pas à l'étude ou qu'un point de vue ne soit pas utilisé dans l'entreprise pour que ce point de vue ne soit pas identifié. Le modèle « tâche de la conduite » (cf. section 4.3.4), par exemple, n'a pas été identifié dans ACACIA sachant qu'il est crucial pour le diagnostic¹. Dans ASMEC, étant donné qu'il faut instrumenter systématiquement tous les axes systémiques, le modèle « tâche de la conduite » a été identifié grâce à l'axe téléologique.

Les connaissances dont nous avons besoin pour les incorporer dans le processus ECD sont celles relatives au *point de vue objet observé*. Nous nous limitons donc, dans les sections suivantes, au développement de ce point de vue en utilisant l'architecture d'ASMEC. Nous passons dans la section au développement du troisième niveau de granularité dans ASMEC, celui de l'instrumentation des modèles par des concepts du domaine.

8.7. Instrumentation des modèles par les concepts

Les modèles choisis pour instrumenter le méta-modèle (cf. section 8.6.2) sont à leur tour instrumentés par des concepts du domaine (cf. Figure 8-9). Ceci est effectué par *homomorphie* aux traits de l'objet observé (cf. Figure 8-10), i.e. l'accident de la route. Les concepts qui ont été choisis tiennent compte des deux autres points de vue (i.e. *observateur* (expert participant à l'étude) et *contexte de l'étude*).

¹ Il permet de comprendre l'erreur comme étant un écart entre ce que le conducteur a voulu faire (sa téléologie) et ce qu'il a fait.

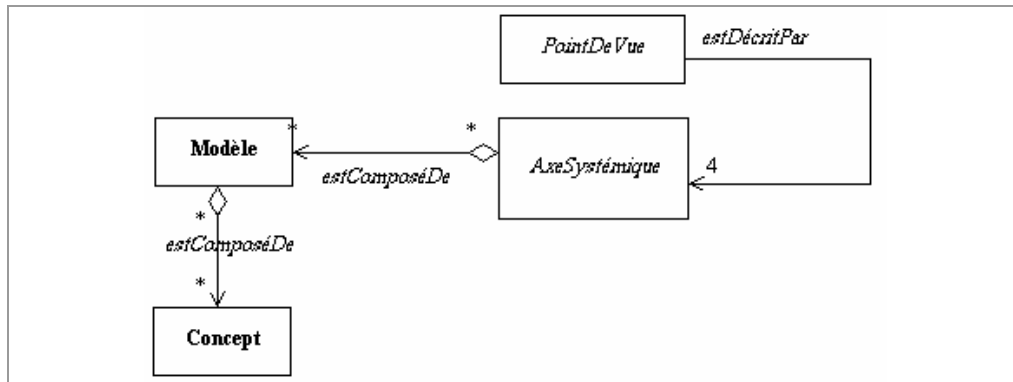


Figure 8-9- Instrumentation des modèles par des concepts du domaine.

Nous présentons ci-dessous les différents modèles ainsi que les concepts qui les constituent selon le patron suivant : $\langle \text{Nom_modèle}, \text{Point(s) de vue impliqué(s)} \rangle = \{\text{concepts}\}$.

Ce patron est appelé « patron de point de vue » qui permet de montrer comment construire les différents modèles en les instrumentant avec les concepts. Les modèles finaux, que nous présentons ci-dessous, sont donc une instantiation de ce patron.

- **Vue ontologique de l'accident :**

- **<Modèle CVE, Accidentologue>** = {Conducteur, Véhicule, Environnement, interaction Cond_Veh, interaction Cond_Envir, interaction Veh_Envir, CVE, Piéton, Impliqué, interaction Cond_Piét, interaction Veh_Piét, interaction Envir_Piét};
- **<Modèle systèmes de sécurité, Concepteur>** = {syst. d'information sur le conducteur, syst. d'information sur le véhicule, syst. d'information sur l'infrastructure, syst. d'information sur le trafic, système de contrôle longitudinal, système de contrôle latéral, système de guidage longitudinale, système de guidage latéral},
- **<Modèle véhicule orienté sécurité, Concepteur>** = {défauts de direction, défauts de suspension, défaut d'éclairage, défauts de pneu, charge au moment de l'accident, spécificité de conception, performances, aides à la conduite, commandes, ouvrant, vitrage, sellerie, caractéristiques sécurité secondaire}.

- **Vue fonctionnelle de l'accident :**

- **<Modèle traitement de l'information, Accidentologue>** = {perception, diagnostic, pronostic, décision, action, général};

- **Vue transformationnelle de l'accident :**

- **<Modèle séquentiel, Accidentologue>** = {conduite normale long-terme, conduite normale court-terme, rupture, urgence, choc};

- **Vue téléologique de l'accident :**

- **<Modèle de la tâche de conduite, Accidentologue>** = {Navigation, Contrôle longitudinal, Contrôle latéral, Guidage longitudinal, Guidage latéral};

La Figure 8-10 reprend tous les points de vue instrumentant le méta-modèle d'ASMEC, les différents modèles instrumentant ces points de vue, les concepts instrumentant les modèles ainsi que les attributs caractérisant les concepts. Dans la section suivante, nous passons au quatrième niveau de granularité dans l'architecture l'ASMEC, celui de la construction de méta-données ou la caractérisation des concepts par des attributs.

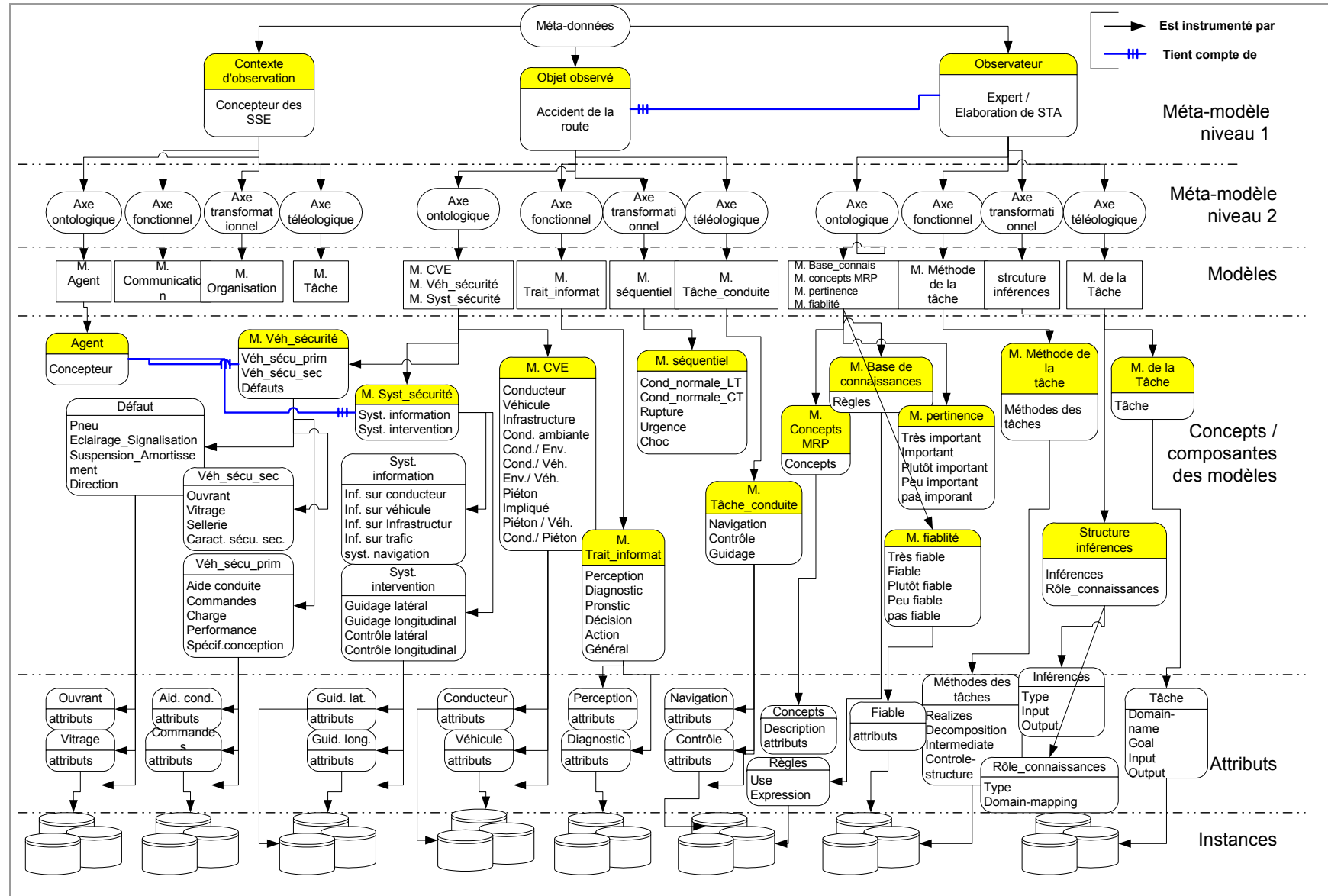


Figure 8-10- Instrumentation des modèles par les concepts et les concepts par les attributs.

8.8. Construction des métadonnées : caractériser les concepts par les attributs

Chacun des concepts est caractérisé par des attributs.

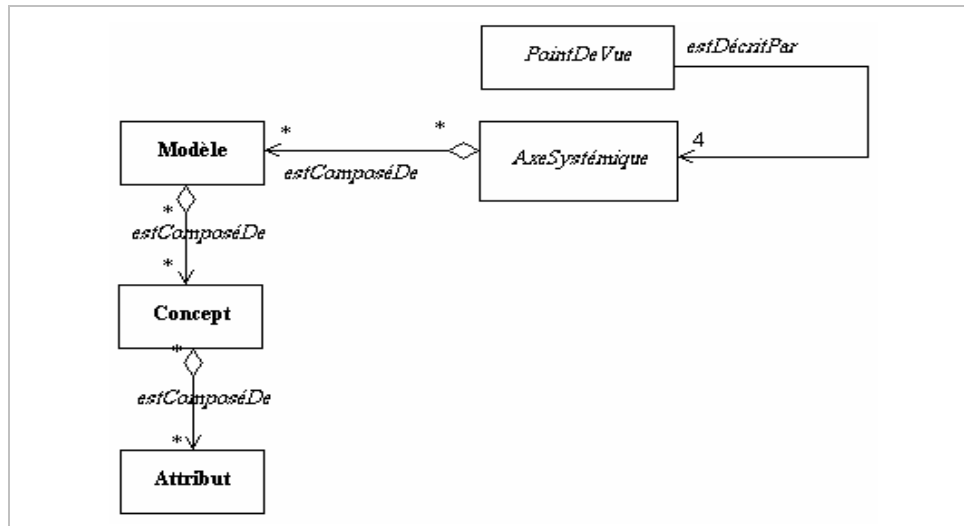


Figure 8-11- Caractérisation des concepts par des attributs.

Le concept « conducteur » du modèle CVE, par exemple, est caractérisé par les attributs {*age, expérience de conduite, nombre d'heure de sommeil, etc.*}. Les 945 attributs caractérisant un accident dans la BD EDA ont été utilisés pour caractériser les concepts. De nouveaux attributs ont été aussi définis par les experts pour cet objectif. La caractérisation des concepts par des attributs peut être perçue comme la création de « données sur les données », donc des méta-données. La méthode de construction de ces méta-données est détaillée dans la section 8.8.2.

Pour la représentation et l'opérationnalisation de ces méta-données, nous utilisons le langage XML (eXtensible Markup Language). Nous commençons dans la section suivante par un aperçu général de ce langage.

8.8.1 XML (eXtensible Markup Language) : présentation générale

Le XML est un langage de balises comme le Html (HyperText Markup Language) mais il est extensible. Le point commun entre le HTML et le XML est qu'ils sont issus tous deux du même langage, soit le SGML (Standardized Generalised Markup Language). Ils sont donc, tous deux, des langages de balises (Markup Language).

Le HTML et le XML sont différents en de très nombreux points, principalement (plusieurs cours sur le XML existent sur le Web, voir par exemple [Van Lancker, 2004]) :

- Le XML décrit, structure et échange des données tandis que le Html ne fait qu'afficher des données ;

- Alors qu'en HTML les balises sont prédéfinies et donc figées, en XML les balises ne sont pas prédéfinies. Le XML est extensible et permet de créer ses propres balises en fonction des données traitées.

Comme le XML n'utilise pas des balises prédéfinies, le navigateur ne « comprend » pas les balises du XML et ne sait pas comment afficher un document XML. Pour le faire, il est nécessaire d'avoir un mécanisme pour décrire comment le document pourrait être affiché. Le XSL pour eXtensible Stylesheet Language est un langage de feuille de style permettant non seulement d'afficher XML, mais aussi (nous présentons dans l'annexe A-3.1 un exemple d'utilisation de XML et XSL) :

- De sélectionner une partie des éléments XML ;
- De trier des éléments XML ;
- De filtrer des éléments XML en fonction de certains critères ;
- De choisir des éléments ;
- De retenir des éléments par des tests conditionnels.

8.8.2 Méthode de construction des méta-données

Les métadonnées ou les « données sur les données » vont nous permettre la formalisation et l'opérationnalisation des *connaissances statiques du domaine* et leur incorporation dans le processus d'ECD. Notre approche pour la construction des métadonnées consiste concrètement à : prendre, un par un, les attributs caractérisant un accident dans les BD EDA et les assigner aux différents concepts des différents modèles déjà choisis pour instrumenter le méta-modèle. Autrement dit, il s'agit de conjointre le niveau 3 (celui des concepts) et le niveau 5 (celui des attributs) d'ASMEC (cf. section 8.4). De nouveaux attributs ont été définis pour enrichir la caractérisation de concepts existants (cf. section 8.8.5) ou pour la caractérisation de nouveaux concepts comme ceux du modèle « *système de sécurité* » (cf. section 8.6.2.3). Ce travail a été effectué par les six experts participant à l'étude (2 équipes de 3 personnes). Avec ces experts, nous avons identifié 101 attributs comme attributs administratifs. Le reste (846) a été projeté sur les différents modèles.

Pour faciliter la compréhension de cette partie, nous avons choisi de travailler sur un exemple concret d'accident (une instance). Nous présentons ci-dessous une brève description textuelle de cet accident.

```
Accident n°503 : Peugeot 206 /Renault Express. L'accident a lieu le vendredi 20
septembre 2003, vers 13h40, au niveau du rond point à l'intersection de la RN 25 et
de la RN 29, hors agglomération de Longueau. Le temps est ensoleillé. La chaussée est
sèche. Mr X., âgé de 51 ans est électrotechnicien. Au volant de sa Peugeot 206, il
vient de quitter son travail, à la zone industrielle et se dirige vers Paris (trajet
de 420 kms) afin de rentrer chez lui, comme chaque fin de semaine. Il est précédé
d'une file de voitures quand il aborde le rond point. Il ne connaît quasiment pas ce
chemin et cherche sa route en lisant les panneaux. Il roule à environ 20 km/h quand
il arrive au niveau du rond point. Il regarde sur sa gauche furtivement et ne voit
pas de véhicule. Il passe, alors qu'arrive au même moment un Renault Express qu'il
percute sur l'arrière droit. Le conducteur de la Peugeot 206 ne réagit qu'après-coup
en freinant, n'ayant pas vu L'Express. Il était en effet préoccupé par son
itinéraire. Les conducteurs sont ceinturés. Mr X. est indemne. Mr Y. est très
légèrement blessé. Les dépistages d'alcoolémie sont négatifs.
```

Figure 8-12- Description textuelle d'un cas réel d'accident

(pour la confidentialité des données, nous avons changé les marques des véhicules, les dates et le lieu de l'accident)

8.8.3 Métadonnées transformationnelles

La construction de ces métadonnées consiste à assigner chacun des 846 attributs à l'un des concepts du modèle séquentiel (i.e. *conduite normale long et court terme, rupture, urgence, choc*). Nous avons effectué une première version de classification des attributs avec chacune des deux équipes d'experts séparément. Nous avons noté une correspondance globale d'affectation entre les deux équipes de 85% (cf. Figure 8-14). Cette classification a été faite selon le patron de point de vue suivant :

```
<Vue transformationnelle, Concept transformationnel, Attributs>
avec
« Concept transformationnelle » ∈ {Situation avant la conduite, Conduite normale, Rupture, Urgence, Choc}.
```

Figure 8-13- Patron de l'élaboration des métadonnées transformationnelles

	Conduite Normale	Situation de rupture	Situation d'urgence	Situation de choc
Équipe A	330	245	56	205
Correspondances				
	87 %	81 %	79 %	96 %
Équipe B	291	288	67	198

Figure 8-14- Comparaison des deux versions de classifications (selon le modèle transformationnel) des attributs effectuées par les deux équipes d'experts

Les chiffres dans la Figure 8-14 correspondent aux nombre d'attributs affectés aux concepts par les deux équipes d'experts (A et B). Exemple : 330 attributs (resp. 291) ont été affectés au concept « *conduite normale* » par l'équipe A (resp. B). Cela correspond à 87% de correspondance d'affectation entre les deux équipes.

Deux mois plus tard, nous avons présenté les deux versions à l'une des deux équipes sans leur permettre d'identifier leur version antérieure. Nous leur avons demandé de comparer les deux versions et de refaire une troisième classification finale, celle que nous avons utilisée pour la construction des métadonnées transformationnelles.

La Figure 8-15 présente la distribution finale des attributs selon le point de vue transformationnel.

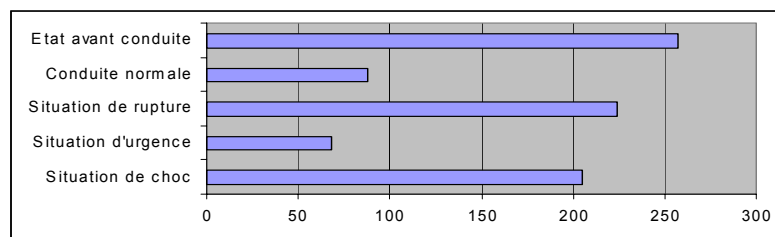


Figure 8-15- Distribution des attributs selon le point de vue transformationnel

La projection de chacun des accidents sur les méta-données est effectuée automatiquement selon le patron de point de vue suivant :

< N° Accident, Vue transformationnelle, Concept transformationnel, Attribut, Valeur(s)>.

La représentation de l'accident n°503 présenté dans la Figure 8-14 selon ce patron donne le tableau présenté dans la Figure 8-16. Une représentation en XML (cf. section 8.8.1) de la projection de l'accident n°503 sur les méta-données transformationnelles est donnée dans la Figure 8-17 :

PointDeVue	Concepts	Attributs	Valeurs
Vue Transformationnelle	Situation_avant_conduite	Age Modele_du_vehicule Situation_Professionnelle	51 Peugeot 206 Electricien
	Situation_Conduite_normale	Type_de_linfrastructure Conditions_de_surface Manoeuvre_avant_accident Vitesse_declaree	Intersection Sec aborde un rond point 20
	Situation_de_Rupture	Perception Difficulte_de_vision Type_du_masque	Defaillance Soleil Barriere anti-choc
	Situation_durgence	Action_avant_le_crash Action_apres_le_crash	Pas de reaction Freinage
	Situation_de_Choc	Premier_impact Deploiement_de_lairbag Cote_impacte	contre un vehicule Oui Frontal

Figure 8-16- Représentation de l'accident n°503 selon le point de vue transformationnel

```

<?xml version="1.0"?>
<Base_de_donnees_daccidents>
<Accident>
<N_Accident> N 503 </N_Accident>
<Point2vue>
  <Nom_Point2vue> Vue_Transformationnelle </Nom_Point2vue>

  <Concept>
    <NomDuConcept>Situation_avant_la_conduite</NomDuConcept>
    <Attributs>
      <Age>51</Age><Situation_Professionnelle>Electricien</Situation_Professionnelle>
      <Modele_du_vehicule> Peugeot 206 </Modele_du_vehicule>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Situation_de_Conduite_normale </NomDuConcept>
    <Attributs>
      <Type_de_linfrastructure>
        Intersection</Type_de_linfrastructure><Conditions_de_surface>Sec
      </Conditions_de_surface><Vitesse_declaree>20</Vitesse_declaree>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Situation_de_Rupture </NomDuConcept>
    <Attributs>
      <Perception> Defaillance </Perception> <Difficulte_de_vision> Soleil
      </Difficulte_de_vision> <Type_du_masque> Barriere anti-choc </Type_du_masque>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Situation_durgence </NomDuConcept>
    <Attributs>
      <Action_avant_le_crash>Pas de reaction</Action_avant_le_crash> <Action_apres_le_crash>
        Freinage </Action_apres_le_crash>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Situation_de_Choc </NomDuConcept>
    <Attributs>
      <Premier_impact> contre un vehicule </Premier_impact> <Deploiement_de_lairbag> Oui
      </Deploiement_de_lairbag><Cote_impacte> Frontal </Cote_impacte>
    </Attributs>
  </Concept>
</Point2vue>
</Accident>
</Base_de_donnees_daccidents> >
</Accident>
</Base_de_donnees_daccidents>
    
```

Figure 8-17- Projection de l'accident n°503 sur les métadonnées transformationnelles

8.8.4 Métadonnées ontologiques

Comme nous l'avons vu dans la section 8.7, le point de vue ontologique est instrumenté de trois modèles : le modèle « CVE », le modèle « systèmes de sécurité » et le modèle « véhicule orienté sécurité ». La construction des métadonnées ontologiques consiste à assigner chacun des 846 attributs aux concepts correspondants dans chacun des modèles. Prenons l'exemple du modèle CVE dont les concepts sont : {Conducteur, Véhicule, Environnement, interaction Cond_Veh, interaction Cond_Envir, interaction Veh_Envir, CVE, Piéton, Impliqué, interaction Cond_Piét, interaction Veh_Piét, interaction Envir_Piét}. La même procédure utilisée pour la construction des métadonnées transformationnelles a été utilisée pour la construction des métadonnées ontologiques. La classification des attributs selon ce point de vue a été faite selon le patron suivant :

```
<Vue ontologique, Concept ontologique, Attributs>
avec :
« Concept ontologique » ∈ {Conducteur, Véhicule, Environnement, interaction Cond_Veh,
interaction Cond_Envir, interaction Veh_Envir, CVE, Piéton, Impliqué, interaction
Cond_Piét, interaction Veh_Piét, interaction Envir_Piét}.
```

Figure 8-18- Patron de l'élaboration des métadonnées transformationnelles

La distribution des différents attributs selon ce point de vue est présentée dans la Figure 8-19 :

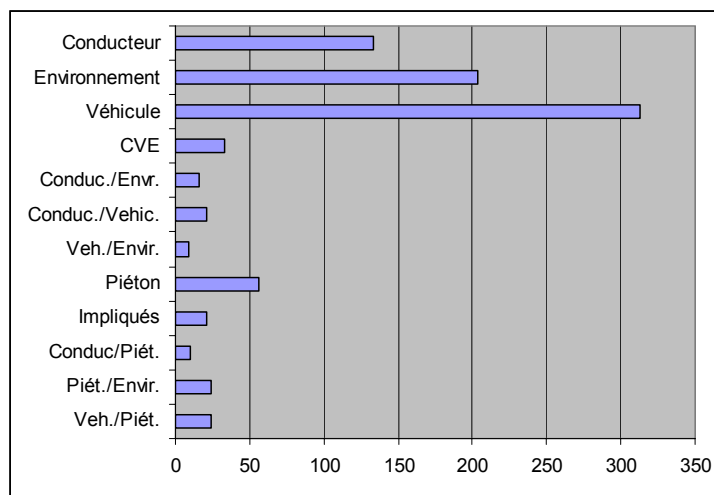


Figure 8-19- Distribution des attributs selon le point de vue ontologique

Une représentation automatique de chacun des accidents est effectuée selon le patron suivant :

:

```
<N° Accident, Vue Ontologique, Concept ontologique, Attribut, Valeur(s)>
```

Le résultat de la projection de l'accident n°503 sur les métadonnées ontologiques est donné dans le tableau suivant (cf. Annexe A-3.3 pour la représentation en XML) :

PointDeVue	Concepts	Attributs	Valeurs
Vue_ontologique	Conducteur	Age Situation_Professionnelle	51 Electricien
	Véhicule	Modele_du_vehicule Deploiement_de_lairbag Cote_impacte	Peugeot 206 Oui Frontal
	Environnement	Type_de_linfrastructure Conditions_de_surface	Intersection Sec
	Cond_Veh	Vitesse_declaree Action_avant_le_crash Action_apres_le_crash	20 Pas de reaction Freinage
	Cond_Envir	Perception Difficulte_de_vision Type_du_masque	Defaillance Soleil Barriere anti-choc
	CVE	Premier_impact	contre un vehicule

Figure 8-20- Représentation de l'accident n°503 selon le point de vue ontologique

8.8.5 Métadonnées fonctionnelles

La construction de ces métadonnées consiste à assigner chacun des 846 attributs à l'une des composantes du modèle fonctionnel (i.e. perception, diagnostic, pronostic, décision, action). La même procédure utilisée pour la construction des métadonnées transformationnelles a été utilisée pour la construction des métadonnées fonctionnelles (et des autres métadonnées). La classification des attributs selon ce point de vue a été faite selon le patron suivant :

<Vue Fonctionnelle, Concept Fonctionnel, Attributs>
 avec :
 « Concept fonctionnelle » ∈ {Perception, Diagnostic, Pronostique, Décision, Action}.

Figure 8-21- Patron de l'élaboration des métadonnées transformationnelles

Certains attributs peuvent concerner plusieurs étapes fonctionnelles. Par exemple, « l'état d'alcoolémie », peut avoir une influence sur toutes les étapes fonctionnelles. Pour cela une classe « globale » a été définie pour regrouper ces attributs. La distribution des différents attributs selon ce point de vue est présentée dans la Figure 8-22:

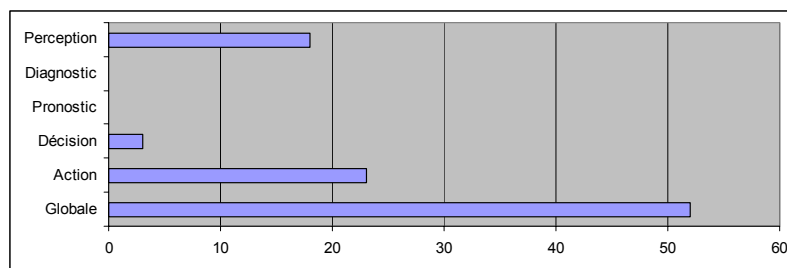


Figure 8-22- Distribution des attributs selon le point de vue fonctionnel

Une représentation automatique de chacun des accidents est effectuée selon le patron suivant :

<N° Accident, Vue Fonctionnelle, Concept Fonctionnel, Attribut, Valeur(s)>

La projection de l'accident n°503 sur les métadonnées donne la Figure 8-23 (cf. Annexe A-3.4 pour la représentation en XML) :

PointDeVue	Concepts	Attributs	Valeurs
Vue_Fonctionnelle	Etape_de_Perception	Action_perception Defaillance_perception Visibilite Difficulte_de_vision	regarde a gauche Contr.visuel_inadapte Reduite Soleil
	Etape_de_Diagnostic	Estimation_de_danger Vitesse_estimee	etat de securite 20
	Etape_de_Pronostic	Defaut_danticipation	Oui
	Etape_de_Decision	Temps_de_reaction_necessaire	2s
	Etape_dAction	Action_avant_crash Action_apres_crash	Pas de reaction Freinage

Figure 8-23- Représentation de l'accident n°503 selon le point de vue fonctionnel

En regardant la distribution des attributs selon le point de vue fonctionnel dans la Figure 8-22, nous constatons qu'il y a très peu d'informations dans les EDA concernant les étapes de « diagnostic », « pronostic » et « décision ». Pour pallier ce problème, nous avons défini dans le cadre d'un projet de fin d'études d'ingénieur une étude dont l'objectif est d'enrichir ce modèle [Ben Hassine, 2004]. De nouveaux attributs ont été donc ajoutés pour mieux caractériser ces trois étapes (cf. Annexe A-2.). Les études menées dans [Van Elslande et Alberton, 1997] sur la production de l'erreur humaine nous ont aidé à définir ces nouveaux attributs et les intégrer dans les aussi La construction des méta-données avec notre approche ASMEC a permis non seulement de structurer les données selon les différents points de vue du méta-modèle, mais aussi d'avoir une vision synthétique des la base EDA et avoir une idée sur la complétude des attributs caractérisant un accident dans cette base. Il s'agit donc d'une approche qui permet d'enrichir la base de données.

8.8.6 Métadonnées téléologiques

La construction de ces métadonnées consiste à assigner chacun des 846 attributs à l'une des composantes du modèle téléologique (i.e. navigation, contrôle latéral, contrôle longitudinal, guidage latéral, guidage longitudinal). La classification des attributs selon ce point de vue a été faite selon le patron suivant :

```
<Vue Téléologique, Tâche, Attributs>
avec :
« Tâche » ∈ {navigation, contrôle latéral, contrôle longitudinal, guidage latéral,
guidage longitudinal}.
```

Figure 8-24- Patron de l'élaboration des métadonnées transformationnelles

La distribution des différents attributs selon le point de vue téléologique est présentée dans la Figure 8-25 :

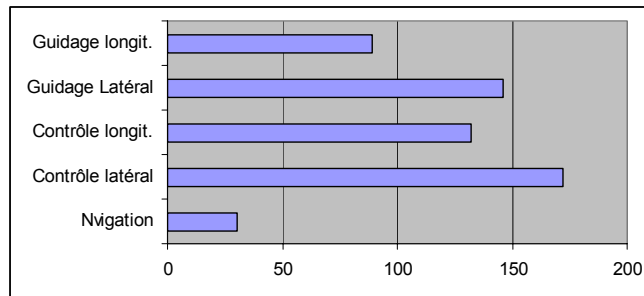


Figure 8-25- Distribution des attributs selon le point de vue téléologique

Une représentation automatique de chacun des accidents selon les méta-données téléologiques est effectuée selon le patron suivant :

```
<N° Accident, Vue Téléologique, Concept téléologique (tâche), Attribut,
Valeur(s)>
```

La projection de l'accident n°503 sur les métadonnées donne la Figure 8-26 (cf. Annexe A-3.5 pour la représentation en XML) :

PointDeVue	Concepts	Attributs	Valeurs
Vue_Téléologique	Navigation	Action_perception	regarde a gauche
		Defaillance_perception	Contr.visuel_inadapte
		Visibilite	Reduite
		Difficulte_de_vision	Soleil
	Contrôle_latéral	Action_avant_le_crash	Pas de reaction
	Contrôle_longitudinal	Action_apres_le_crash	Freinage
Action_avant_le_crash		Pas de reaction	
Guidage_latéral	Manceuvre_avant_accident	aborde un rond point	
Guidage_longitudinal	Manceuvre_avant_accident	suivre une fille de voitures	

Figure 8-26- Projection d'un accident sur les métadonnées téléologique.

8.9. Les liens entre les différents points de vue

Les liens entre les différents points de vue du méta-modèle (niveaux 1 et 2) dans ASMEC sont d'abord sémantiques et ils sont assurés par le fondement systémique des ces points de vue. Ces liens sont formalisés grâce aux métadonnées et plus précisément grâce à l'affectation multi-vues des attributs. Par exemple, l'attribut « *coup de volant* » caractérise en même temps la « *phase d'urgence* » du modèle séquentiel de l'axe transformationnel du point de vue objet observé et la tâche de « *perception* » du modèle de traitement d'information de l'axe fonctionnel du point de vue objet observé. Dans l'annexe A-5, nous donnons une formalisation de ces relations entre les différents points de vue. Ces relations permettent une lecture (interprétation) multi-vues de l'accident. Nous avons illustré cela à travers un cas d'accident réel que nous avons projeté sur les différentes méta-données (cf. Figure 8-17, Figure 8-20, Figure 8-23 et Figure 8-26).

Dans la Figure 8-27, nous donnons, à travers un diagramme d'objets, des exemples d'instanciation du point de vue objet observé du modèle ASMEC. On peut lire sur cette figure que ce point de vue est décrit par les quatre axes systémiques. Les modèles *CVE* et *séquentiel* instrumentent respectivement les axes ontologique et transformationnel. Ils sont à leur tour instrumentés par des concepts. Nous avons donné comme exemple d'attributs « *l'age* » qui caractérise les deux concepts « *conducteur* » et « *conduite normale* » qui appartiennent à deux modèles différents.

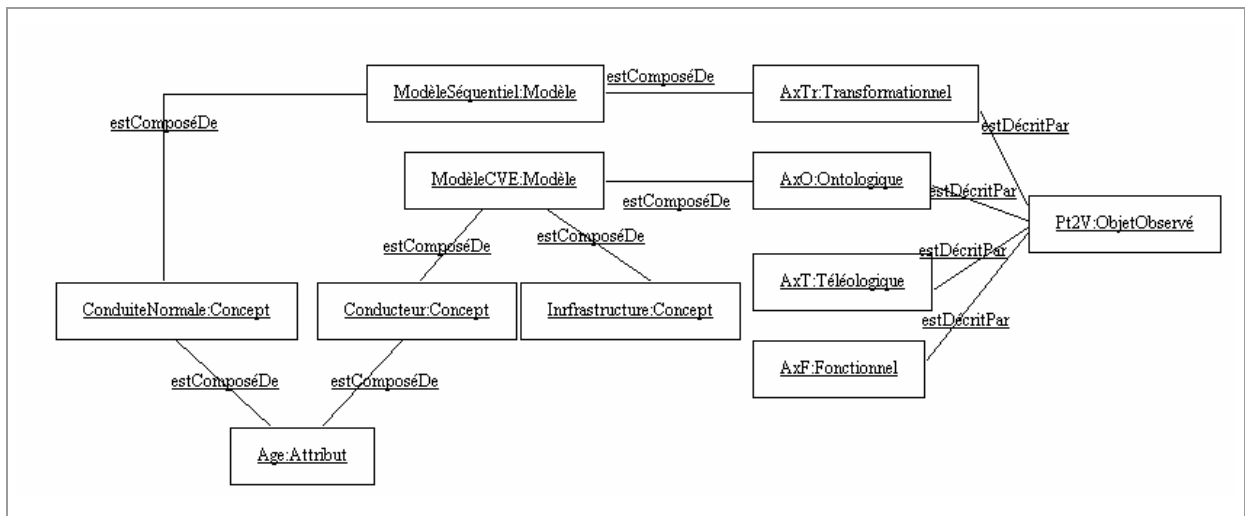


Figure 8-27- Diagramme d'objets :exemples d'instanciation du diagramme de classes d'ASMEC présenté dans la Figure 8-3.

8.10. Synthèse

Dans ce chapitre, nous avons commencé par présenter les objectifs d'ASMEC (cf. section 8.1) qui consistent à pallier les limites de la méthode CommonKADS constatés dans le chapitre précédent (cf. section 7.4). Ces limites sont relatives à la difficulté de spécifier certaines connaissances statiques indépendantes de la tâche d'élaboration de STA et à la difficulté de modéliser des connaissances selon plusieurs points de vue. Nous avons ensuite choisi une stratégie de modélisation (cf. section 8.2) dirigée par un méta-modèle après avoir montré ses avantages par

rapport à une approche ascendante dirigée par les données. La méthodologie de modélisation que nous avons choisie a consisté en trois phases :

- Le *cadrage du modèle* : nous avons construit dans un premier temps un *méta-modèle général* que nous avons instrumenté par trois points de vue (cf. section 8.5) : *observateur* (l’accidentologue qui résout la tâche d’élaboration des STA), *objet observé* (l’accident) et *contexte d’observation* (développement des systèmes de sécurité). Chacun de ces points de vue est décrit à son tour par les quatre sous-points de vue (ou axes) systémiques (cf. section 8.5) : ontologique, fonctionnel, transformationnel et téléologique. Dans une deuxième étape, nous avons défini des modèles du domaine par *isomorphie* avec le méta-modèle. Concrètement, nous avons instrumenté chacun des axes dans chacun des points de vue par des modèles du domaine (cf. section 8.6). Nous avons montré que la plupart des modèles fournis par CommonKADS (e.g. organisation, communication, agent, etc.) permettent d’instrumenter les axes des points de vue *observateur* et *contexte d’observation*. Mais, ce n’est pas le cas du point de vue objet observé. Nous nous sommes donc focalisés sur ce point de vue et nous avons identifié les modèles qui ont permis son instrumentation (cf. section 8.6.2) ;
- Le *développement du modèle* : il s’agit d’instrumenter les modèles choisis par une correspondance *homomorphique* avec les traits perçus du phénomène modélisé. Concrètement, nous avons instrumenté chacun des modèles par des concepts du domaine (cf. section 8.7) ;
- L’*opérationnalisation du modèle* : il s’agit de rendre les modèles opérationnels pour pouvoir les utiliser (en ECD, par exemple) et les interpréter. La technique que nous avons développée pour opérationnaliser les modèles consiste à développer des *métadonnées*. Il s’agit de caractériser les concepts par les attributs (cf. section 8.8) en prenant, un par un, les attributs (947) caractérisant un accident dans les BD et les assigner aux différents concepts dans les différents modèles. De nouveaux attributs ont été également définis (cf. Annexe A-2). Ce travail a été effectué avec les six experts participant à l’étude (2 équipes de 3 personnes). Les métadonnées ont été représentées à l’aide du langage XML.

Il est important de noter que pour créer des méta-données relatives aux nouveaux modèles pour une étude donnée, nous n’avons pas besoins de classier tous les attributs existants dans la BD. Il suffit de créer des méta-données en se basant uniquement sur les attributs qui seront utilisés dans l’étude. Il faut noter aussi que les méta-données non seulement peuvent être réutilisées, mais aussi personnalisées par chacun des utilisateurs. Autrement dit, un utilisateur peut élaborer ses propres méta-données en effectuant lui-même une nouvelle classification des attributs en fonction des concepts. Un utilisateur peut aussi avoir accès aux méta-données élaborées par les autres.

Chapitre 9.

AICEF : une Approche d'Incorporation des Connaissances Expertes dans la Fouille de données

AICEF est la deuxième approche que nous développons dans SAFE-Next. Elle permet l'utilisation des méta-données développées par ASMEC dans le chapitre précédent (Chapitre 8) pour l'incorporation des connaissances expertes dans le processus d'ECD et plus particulièrement dans les phases de préparation des données et d'interprétation des résultats (cf. Figure 9-1).

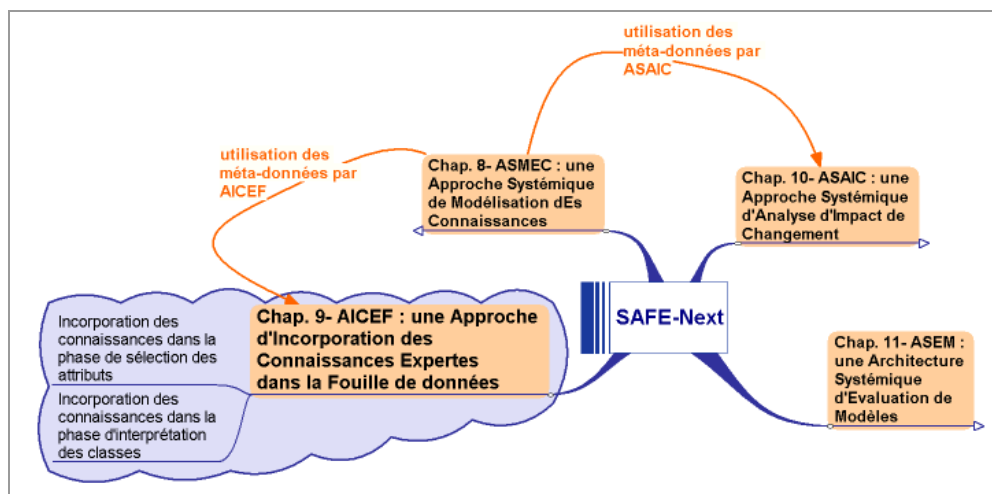


Figure 9-1- AICEF : approche d'incorporation des connaissances dans le processus d'ECD (deuxième approche dans SAFE-Next)

9.1. Utilisation des métadonnées pour la sélection des attributs

Pour effectuer une classification d’accidents, l’expert commence par sélectionner les attributs qui sont pertinents vis-à-vis de l’objectif de l’étude. Dans cette section, nous montrons comment les méta-données d’ASMEC permettent d’intégrer les connaissances expertes dans la première phase du processus ECD à travers une sélection multi-vues des attributs. Les liens entre le modèle de connaissances d’ASMEC et la base de données EDA sont assurés par le fait que *les attributs dans ASMEC correspondent aux colonnes (i.e. attributs collectés sur l’accident) dans la BD*. Ce lien est représenté dans la Figure 9-2 :

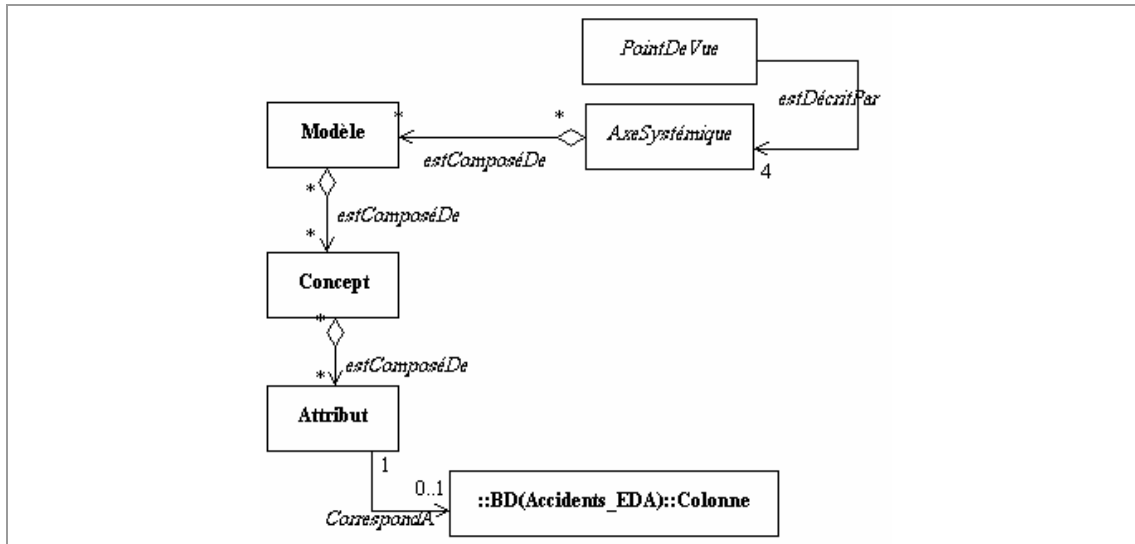


Figure 9-2- Lien entre le modèle de connaissances dans ASMEC et la base de données : les attributs dans le modèle ASMEC correspondent aux colonnes dans la BD.

Pour montrer comment les méta-données permettent l’incorporation des connaissances dans la phase de sélection des attributs, prenons l’exemple suivant : il s’agit d’une étude qui a pour objectif de « *comprendre les différentes configurations du problème de perte de contrôle* ». L’un des experts a formulé la requête suivante qui concerne les différents concepts dont il faut tenir compte :

```

(Ontologique v Fonctionnel v Transformationnel v Téléologique) ^ (Conducteur/Véhicule v
Conducteur/Environnement) ^ (Situation de rupture v Situation d’urgence) ^ (Perception v
Diagnostic v Pronostic)
    
```

Figure 9-3- Exemple de requête de sélection d’attributs pour l’élaboration de STA pour l’étude de perte de contrôle

Cette requête exprime la réponse d’un expert à la question suivante : « *Quels sont les axes systémiques, les modèles et les concepts du domaine qui pourraient être pertinents pour l’étude ?* ». La Figure 9-4 présente l’interface que nous avons créée et sur laquelle l’expert exprime sa requête en cochant les cases qui correspondent à sa réponse :

Figure 9-4- Interface d'ACEF aidant l'utilisateur à traduire son objectif en fonction des concepts du domaine et selon différents points de vue : sélection multi-vues des attributs.

Dans cette interface (cf. Figure 9-4), nous retrouvons les différents *modèles* instrumentant les *axes systémiques* au niveau du *point de vue objet observé*. Nous retrouvons également les *concepts* instrumentant ces modèles. Par exemple, le *modèle CVE* instrumente l'*axe ontologique* et il est instrumenté par les concepts « *conducteur* », « *véhicule* », etc. Cette interface traduit automatiquement la réponse de l'expert sous forme d'une requête SQL que nous présentons dans l'Annexe A-4.

Puisque les méta-données font le lien entre les concepts et les attributs, cette requête permet de sélectionner les attributs qui répondent au besoin de l'expert. Ainsi, les attributs sont sélectionnés selon le choix de l'utilisateur en combinant plusieurs points de vue existant dans le domaine et qui ont été identifiés par le méta-modèle (cf. §8.6.2.3 section 8.7) et formalisés par les experts (cf. section 8.8). Il s'agit donc d'un outil pour traduire, par l'expert, l'objectif de l'étude en fonction des concepts du domaine.

Pour évaluer l'apport de l'utilisation des métadonnées dans la phase de sélection des attributs, nous avons comparé des études qui ont été menées sans et avec leur utilisation. Parmi ces apports, rappelons les points suivants (voir [Ben Ahmed *et al.*, 2003] pour plus de détails) :

- (i) L'utilisation des métadonnées nous a permis de détecter des points de vue qui ont été implicites dans la sélection purement experte sans utilisation des modèles du domaine. Cela était le cas du modèle de traitement d'information. Nous nous sommes rendu compte que très peu d'attributs sélectionnés caractérisaient ce point de vue. En systématisant la sélection selon les différents points de vue, le risque d'oublier un point de vue pertinent est réduit ;
- (ii) Le nombre d'attributs est réduit automatiquement, mais en utilisant des connaissances expertes déjà implémentées par les métadonnées ;

- (iii) L’expert peut toujours effectuer une sélection manuelle des attributs et dans ce cas là les métadonnées peuvent lui servir d’outil pour vérifier la pertinence de ces attributs et la « complétude » de la sélection (i.e. si elle couvre les différents aspects de l’accidents qui sont nécessaires pour l’étude) ;
- (iv) Non seulement la qualité de la sélection est améliorée (i.e. sélection multi-vues), mais le temps est nettement réduit grâce à l’interface de sélection (cf. Figure 9-4). Effectuer manuellement la même tâche de sélection multi-vues des attributs est difficile, voire impossible.

Les attributs sélectionnés grâce à AICEF font ensuite l’objet de tâches classiques de préparation de données : fusionner certains attributs pour éviter la redondance, regroupement de certaines modalités, etc. (cf. section 6.1.1).

9.2. Utilisation des métadonnées pour l’interprétation des résultats de classification

A travers l’état de l’art que nous avons effectué sur la problématique d’interprétation de résultats issus de data mining (cf. section 6.1.3), nous avons constaté un réel manque de méthodes génériques traitant cette problématique. A cela s’ajoute le fait que les méthodes existantes se restreignent à des interprétations statistiques ou à des techniques de visualisation (qui sont, elles aussi, basées sur des critères statistiques) pour rendre la tâche d’interprétation plus interactive. Mais, nous n’avons pas trouvé de travaux qui essaient de modéliser les connaissances du domaine, autres que ceux basés sur des critères statistiques, et de les incorporer dans la tâche d’interprétation de classes.

Nous avons donc décidé de mener une réflexion plus profonde sur cette problématique. Cela nous a permis de catégoriser les tâches d’interprétation en deux types : *interprétation intra-classes* et *interprétation inter-classes*. Ensuite, nous avons défini une approche permettant l’intégration des connaissances du domaine à travers les métadonnées développées dans le Chapitre 8.

9.2.1 Interprétation intra-classes

Nous parlons d’*interprétation intra-classe* car il s’agit d’une interprétation de la classe en se basant uniquement sur des données relatives à la classe elle-même sans tenir compte des autres classes. Par exemple, on ne fait pas de comparaison des classes.

La sortie de l’application d’un algorithme de classification automatique se présente généralement sous forme d’un tableau. La Figure 9-5 en présente un exemple. Il s’agit des premières lignes d’un tableau contenant le résultat d’une classification que nous avons effectuée sur un échantillon de 717 accidents en utilisant 25 attributs pour obtenir 18 classes. Ce tableau contient 1771 lignes et 6 colonnes à analyser par l’expert (donc près de 10500 informations).

Libellés des variables	Modalités caractéristiques	% de la modalité dans l'échantillon	% de la modalité dans la classe	% de la classe dans la modalité	Valeur-Test
LocChoc	Hors chaussée	26.64	96.72	30.89	12.26
typeChoc	Tonneau renversement	21.76	78.69	30.77	9.92
Obp	Obp=sol	18.97	68.85	30.88	8.91
vehiprior2	Vehicule seul	29.15	72.13	21.05	7.16
sitacc	prob contrôle ve 21#	32.50	73.77	19.31	6.79
critini	Guidance infrastr 5#	15.62	44.26	24.11	5.51
evini	gene extérieur/d 14#	5.72	22.95	34.15	4.68
manident	Section courante 17#	24.83	49.18	16.85	4.19
typacc	typacc=pilotabilité	55.51	80.33	12.31	4.09
atm	atm=Clair/normal	55.79	80.33	12.25	4.04
surf	surf=Sec	62.62	85.25	11.58	3.89
typlieu	typlieu=H-Agg sur RD	47.98	70.49	12.50	3.58
fondef	fondef=Action	9.07	22.95	21.54	3.30
manident	Changement file 16#	6.14	18.03	25.00	3.26
typdef	Realisation inco 20#	33.61	52.46	13.28	3.04
mask	mask=Pas de masque	65.13	81.97	10.71	2.86
critini	Perte contrôle tr 8#	17.85	32.79	15.63	2.83
mecdef	mecdef=Panique	5.72	14.75	21.95	2.57
mecdef	Activite annexe 27#	7.67	16.39	18.18	2.23
evini	Drogue medicamen 11#	2.51	8.20	27.78	2.21

Figure 9-5- Description d'une classe à l'aide des modalités (Sortie du logiciel SPAD d'une classification des accidents)

Pour interpréter une classe telle que celle présentée dans la Figure 9-5, nous avons identifié deux types de connaissances : des *connaissances propres à la tâche de classification* et des *connaissances indépendantes de cette tâche*. Nous nous sommes rendu compte que le premier type sert à manipuler essentiellement des critères statistiques. Nous avons pu formaliser ce type de connaissances à l'aide de règles dans la base de connaissances (cf. section 7.1.2). Le deuxième est celui que nous avons formalisé à travers les métadonnées. Nous expliquons ci-dessous comment ces deux types de connaissances sont utilisés pour l'interprétation intra-classe.

9.2.1.1 Interprétation statistique : utilisation de connaissances propres à la tâche de classification

Les classes sont décrites à l'aide de modalités (Figure 9-5) et à chacune de ces modalités est attaché un critère appelé *Valeur-Test*, le *pourcentage des individus décrits par une modalité donnée dans une classe (MOD/CLASS)* ainsi que le *pourcentage d'une classe dans une modalité (CLASS/MOD)* (cf. §7.1.2.2 pour la définition des ces critères). En se basant sur ces critères statistiques, nous avons défini deux notions de pertinence pour les modalités qui caractérisent une classe : le *pouvoir discriminant* et le *pouvoir descriptif* d'une modalité. En utilisant ces deux notions, nous avons pu distinguer quatre types de modalités que nous avons présentés dans le paragraphe 7.1.2.2 (i.e. les *modalités discriminantes et caractérisantes*, les *modalités discriminantes et non-caractérisantes*, les *modalités non-discriminantes mais caractérisantes* et les *modalités non-discriminantes et non caractérisantes*).

En plus des notions du *pouvoir discriminant* et *descriptif* des modalités, d'autres critères ont été définis : des critères relatifs au *nombre de classes* (qui, généralement, ne doit pas dépasser 40 classes dans la pratique), à la *représentativité d'une classe* par rapport à l'échantillon de l'étude, etc.

Ces connaissances ont été formalisées à l'aide de règles dans la base de connaissances (cf. section 7.1.2). La Figure 9-6 en présente un exemple.

```

attribut_d-étude.modalité.Valeur_test > 2                OR
attribut_d-étude.modalité.Valeur_test <-2              AND
attribut_d-étude.modalité.poids_mod_dans_class > 30
a-pour-abstraction
attribut_d-étude.modalité.pertinence = 'discriminante et caractérisante';
    
```

Figure 9-6- Exemple de règle pour la sélection de modalités discriminantes caractérisantes pour la caractérisation des classes.

```

classe_accidents.poids > 50
IMPLIQUE
classification.qualité = 'refaire la classification en modifiant les
attributs ou les paramètres de classification';
    
```

Figure 9-7- Exemple de règle pour assurer l’homogénéité des classes en ce qui concerne leur représentativité dans l’échantillon d’étude.

Ces règles sont opérationnalisées sous forme de requêtes à l’aide du langage XSL (cf. section 8.8.1) permettant l’automatisation de la sélection des modalités pertinentes pour la caractérisation des classes. Il s’agit d’un processus itératif : on commence généralement par sélectionner les *modalités discriminantes et caractérisantes*. Si la description de la classe est satisfaisante vis-à-vis de l’objectif de l’étude on arrête le processus, sinon on passe aux *modalités discriminantes et non-caractérisantes*, etc.

9.2.1.2 Interprétation systémique : utilisation des métadonnées

Les *variables de classification* qui sont dans la première colonne du tableau de description des classes (cf. Premier tableau, Figure 9-9) correspondent aux *attributs* dans le modèle de connaissances d’ASMEC (cf. Figure 9-8).

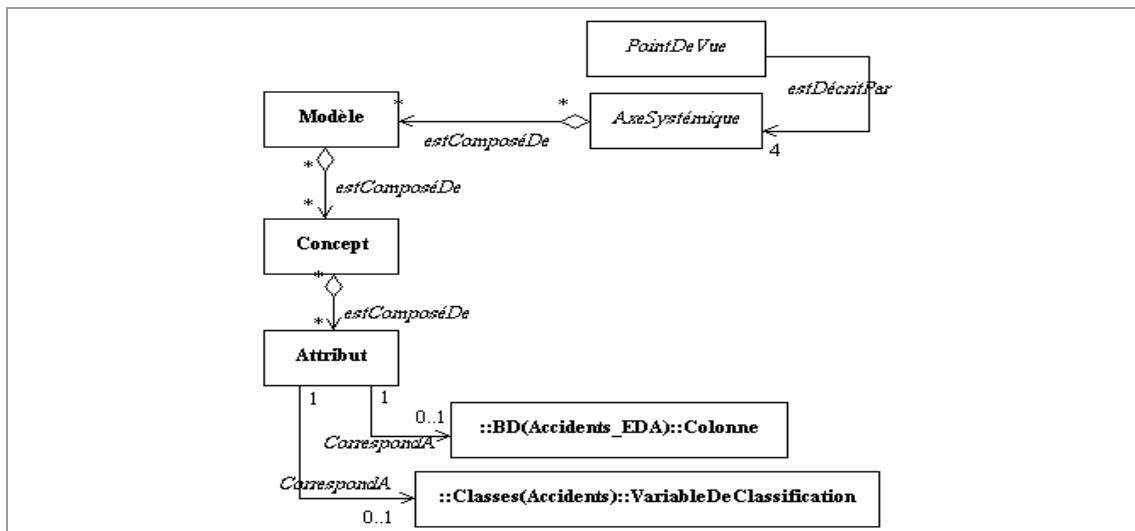


Figure 9-8- Lien entre le modèle de connaissances dans ASMEC et la description des classes : les attributs dans le modèle ASMEC correspondent aussi aux variables de classification.

Comme on peut le remarquer, ces variables, telles qu’elles sont présentées dans le tableau des descriptions des classes, sont relatives à des informations sur différents aspects et concepts de l’accident et ne sont pas structurées selon un point de vue donné. La correspondance entre ces *variables* et les *attributs* du modèle ASMEC permet de projeter automatiquement chaque classe sur

les différents modèles formalisés dans les méta-données. En d'autres termes, les méta-données permettent une lecture multi-vues et multi-niveaux d'abstraction des classes

Pour effectuer cette projection multi-vues des classes sur les méta-données, nous avons utilisé le langage XSL (cf. la fin de la section 8.8.1). La Figure 9-9 représente le résultat de cette projection : il s'agit de la représentation de la même classe présentée dans le premier tableau de la figure, mais selon les modèles formalisés dans les méta-données (CVE, traitement d'information, séquentiel et tâche de conduite).

En plus de la lecture multi-vues des classes, nous permettons à l'expert d'accéder automatiquement à certaines informations dont il a besoin pour l'interprétation des classes. Nous permettons d'afficher pour chaque classe les descriptions textuelles des trois accidents les plus proches de son centre Figure 9-9. L'expert a accès aussi aux reconstructions dynamiques de ces trois accidents. L'apport de l'interface de la Figure 9-9 consiste en [Ben Ahmed *et al.*, 2003]: (i) *L'amélioration de la qualité de l'interprétation* : lecture multi-vues des classes. (ii) *La réduction de son coût* : l'accès manuel à ces informations est une tâche très coûteuse. Nous avons fait le test suivant : nous avons pris la même classe et l'avons projetée manuellement selon les différents modèles. Nous avons recherché ensuite les autres informations dans la BD (descriptions textuelles, reconstructions, etc.). Ce travail manuel nous a pris *2,5 heures pour 1 seule classe*, alors qu'il nous a pris *1 seconde pour 18 classes* en utilisant AICEF.

Libellés des variables	Modalités caractéristiques	% de la modalité dans l'échantillon	% de la modalité dans la classe	% de la classe dans la modalité	Valeur-Test
LocChoc	Hors chaussee	26.64	96.72	30.89	12.26
typeChoc	Tonneau renversement	21.76	78.69	30.77	9.92
Obp	Obp=sol	18.97	68.85	30.88	8.91
vehiprior2	Vehicule seul	29.15	72.13	21.05	7.16
sitacc	prob contrôle ve 21#	32.50	73.77	19.31	6.79
critini	Guidance infrastr 5#	15.62	44.26	24.11	5.51
evini	gene exterieur/d 14#	5.72	22.95	34.15	4.68
manident	Section courante 17#	24.83	49.18	16.85	4.19
typacc	typacc=pilotabilite	55.51	80.33	12.31	4.09
atm	atm=Clair/normal	55.79	80.33	12.25	4.04
surf	surf=Sec	62.62	85.25	11.58	3.89
typlieu	typlieu=H-Agg sur RD	47.98	70.49	12.50	3.58
fondef	fondef=Action	9.07	22.95	21.54	3.30
manident	Changement file 16#	6.14	18.03	25.00	3.26
typdef	Realisation inco 20#	33.61	52.46	13.28	3.04
mask	mask=Pas de masque	65.13	81.97	10.71	2.86
critini	Perte contrôle tr 8#	17.85	32.79	15.63	2.83
mecdef	mecdef=Panique	5.72	14.75	21.95	2.57
mecdef	Activite annexe 27#	7.67	16.39	18.18	2.23
evini	Drogue medicamen 11#	2.51	8.20	27.78	2.21

Modèle Ontol.	Conducteur	Véhicule	Environnement	Cond/Veh	Cond/Env	Env/Vé
	Realisation_incor Changement_file gene_exterieur typacc=pilotabilite Activite_Annexe Drogue_medicament mecdef=Panique fondef=Action	Tonneau_renversement Obp=sol	atm=Clair/normal surf=Sec typlieu=H-Agg_ou_RD mask=Pas_de_masque Activite_Annexe Section_courante	prob_contrôle_veh surf=Sec Perte_contrôle_trans	Guidance_infrastr surf=Sec Changement_file	Obp=sol surf=Sec Vehicule_

Modèle Transf.	Etat long terme LT	Cond. normale CT	Rupture	Urgence	Choc
		atm=Clair/normal surf=Sec typlieu=H-Agg_ou_RD Section_courante	prob_contrôle_veh Guidance_infrastr mask=Pas_de_masque Realisation_incor Vehicule_seul Changement_file Activite_annexe	mecdef=Panique	Obp=sol surf=Sec Tonneau renversement Obp=sol Hors_chaussee

Modèle Fonct.	Perception	Diagnostic	Pronostique	Décision	Action	Globale
	mask=Pas_de_masque			mecdef=Panique	fondef=Action Realisation_incor Perte_contrôle_trans	Activite_annexe Drogue_medicament/f

Modèle Téléo.	Navigation	Guidage latéral	Guidage longit.	Controle latéral	Controle lo
	mask=Pas_de_masque	Guidance_infrastr Section_courante	Guidance_infrastr	Perte_contrôle_trans Tonneau_renversement	

CIRCONSTANCES EDA 73013967
Centre de rang 2
[Afficher la reconstruction](#)
Peugeot 306, seule en cause
L'accident a lieu le mercredi 27 septembre 2000, vers 14h30 sur la départementale 934, hors agglomération d'Arvillers, dans le sens Amiens-Roye. A cet endroit, la route est composée de deux fois deux voies séparées par un terre-plein herbeux d'environ 10 mètres de largeur. Il y a du soleil, la chaussée est sèche et le vent souffle à environ 35-40 km/h.
Mme X., âgée de 36 ans, déléguée médicale, vient de quitter Amiens au volant de sa Peugeot 306, et se dirige vers Noyon. Elle est très fatiguée, préoccupée et n'a pas mangé depuis la veille au soir. Elle perçoit un véhicule devant elle qu'elle décide de doubler. Elle se met sur la file de gauche et déclare rouler à environ 110 km/h. Elle dépasse l'autre véhicule, qu'elle perçoit dans son rétroviseur quand elle est pris d'un étourdissement. En ouvrant les yeux, elle se rend compte que sa voiture mord l'accotement à gauche. Elle freine doucement en débrayant et donne un coup de volant sur la droite afin de stabiliser le véhicule mais, elle se met en travers de la route, traverse les deux voies et se retrouve sur l'accotement herbeux à droite. Elle percute un arbre avec son flanc gauche. Le test d'alcoolémie est négatif
Mme X, ceinturée est gravement blessée.

CIRCONSTANCES EDA 71913951
Centre de rang 3
[Afficher la reconstruction](#)
Renault 9
L'accident a lieu le lundi 22 mai 2000, vers 15h55, sur la départementale 934, en direction de Roye, à la hauteur de Domart-sur-la-Luce. L'accident se produit en courbe à gauche et en descente. A cet endroit, la départementale 934 comporte deux fois deux voies à sens unique et chaussées séparées. Le temps est couvert. La chaussée est sèche.
Melle X., âgée de 19 ans, lycéenne, effectue le trajet de retour entre son lycée amiénois et son domicile, situé à proximité de Bouchoir. Elle doit déposer sa passagère à Roye. Elles sont en retard d'une demi-heure. Melle X. discute avec son amie et détourne le regard de la route quelques instants pour la regarder. Elle circule alors en descente et aborde une courbe à gauche, qu'elle ne perçoit pas et donc ne peut négocier. Elle entre tout droit dans l'accotement droit herbeux. Melle X. se rend immédiatement compte de sa sortie de route et donne un petit coup de volant " sec " vers la gauche pour revenir sur la chaussée. Elle n'a pas le temps de tenter quoi que ce soit d'autre : sa Renault 9 est en travers, elle traverse les deux voies, entre dans l'accotement gauche, heurte talus, effectue des tonneaux et s'immobilise sur le pavillon sur ce même accotement.
Melle X., et sa passagère sont ceinturées, elles sont très légèrement blessées. Aucun dépistage d'alcoolémie n'est pratiqué. Cependant, Melle X. déclare n'avoir consommé aucune boisson alcoolisée et ne présente aucun signe d'ébriété.

Figure 9-9- Grille systémique : projection automatique des résultats de classification sur les métadonnées (lecture au niveau des modalités)
(Pour la confidentialité des données, nous avons changé les marques des véhicules, les dates et les lieux des accidents).

Les projections automatiques des résultats sur la *grille systémique* (Figure 9-9) fournissent un langage entre l'accidentologue et le concepteur. En effet, parmi les modèles de projection, on trouve les deux modèles de représentation de l'accident selon le point de vue du concepteur qui est plus orienté véhicule et systèmes de sécurité. En effet, les modèles d'analyse de l'accident ne sont pas forcément les mêmes que ceux utilisés par l'expert infrastructure ou par le psychologue. Cette grille projette la classe d'accidents dans l'espace de représentation du concepteur ce qui facilite son utilisation par ce dernier pour inférer de nouvelles prestations de sécurité.

Les interfaces d'affichage des résultats dans AICEF ont été conçues en collaboration avec les experts. Cela nous a permis d'identifier les différents types d'information dont l'expert a besoin dans la phase d'interprétation. En plus des projections selon les différents points de vue, les interfaces d'AICEF affichent les différents critères de pertinence des modalités les plus utilisés : Valeur-Test, %MOD/CLASS, poids de la classe dans l'échantillon, description textuelle des trois accidents les plus centrés dans la classe et lien hypertexte vers la restitution cinématique des trois accidents les plus centrés (Figure 9-9).

9.2.2 Interprétation inter-classes

Pendant la phase d'interprétation des classes, nous avons remarqué que les experts essaient, notamment, de comprendre la classe en la comparant aux autres classes. Cela leur permet de trouver les classes les plus proches, de regrouper certaines classes et de voir ce qui discrimine vraiment la classe en question des autres classes. Ces informations sont bien sûr combinées avec les informations issues de l'interprétation intra-classe (cf. section 9.2.1).

Cette tâche est coûteuse car le nombre de classes peut être élevé (il peut atteindre 40) et le nombre d'attributs et de modalités caractérisant les classes peut être élevé aussi. Nous avons proposé alors l'utilisation d'une technique de visualisation permettant de synthétiser les informations dont l'expert a besoin pour mener une comparaison efficace en terme de temps et de qualité. La technique proposée utilise les *treillis de Galois* [Barbut et Monjardet, 1970] que nous introduisons dans le paragraphe suivant.

9.2.2.1 Treillis de Galois : présentation générale

Le *treillis de Galois* [Barbut et Monjardet, 1970] ou *treillis de concepts* [Wille, 1982] est une structure mathématique permettant de représenter les classes non disjointes sous-jacentes (hyper-rectangles) à un ensemble d'objets. Nous l'introduisons en définissant ses notions de base.

Définition 9.1- Contexte

C'est un triplet $C = (O, A, I)$, où O et A sont des ensembles et $I \subseteq O \times A$ est une relation.

La Figure 9-10 montre un exemple de contexte représenté par $C = (O, A, I)$ avec $O = \{1, 2, 3, 4, 5\}$: ensemble d'objets et $A = \{a, b, c, d, e, f, g, h, i\}$: ensemble d'attributs caractérisant les objets.

Définition 9.2- Contexte binaire

Un contexte (O, A, I) est dit binaire si les valeurs de A ne peuvent prendre que les valeurs 0 ou 1. Ces derniers indiquent respectivement l’absence ou la présence de l’attribut concerné dans la description de l’objet concerné.

Définition 9.3- Correspondance de Galois

Soit f l’application qui à tout élément o de O associe $f(o) = \{a \in A, (o, a) \in I\}$, qui représente le sous-ensemble de A en relation avec o selon la relation I . Soit g l’application qui à tout élément a de A associe $g(a) = \{o \in O, (o, a) \in I\}$, qui représente le sous-ensemble de O en relation avec a selon la relation I . Le couple (f, g) est appelé correspondance de Galois.

Définition 9.4- Concept

Le couple $\{A_1, O_1\}$ est un concept si $f(O_1) = A_1$ et $g(A_1) = f(O_1)$. O_1 est appelé l’**extension** du concept $\{A_1, O_1\}$ et A_1 est appelé l’**intention** du concept $\{A_1, O_1\}$. $\{A_1, O_1\}$ est appelé aussi concept formel avec A_1 comme **concept-objet** et O_1 comme **concept-attribut**.

Définition 9.5- Sous-concept/Sur-concept

La relation de sous-concept/sur-concept dénotée ‘ \leq ’ est la relation d’ordre définie entre les concepts de la façon suivante : $\{A_1, O_1\} \leq \{A_2, O_2\} \Leftrightarrow A_1 \subseteq A_2$ et $O_1 \subseteq O_2$.

Définition 9.6- Treillis de Galois

L’ensemble L de tous les concepts, munis de la relation d’ordre \leq , possède la structure mathématique de treillis et est appelé treillis de Galois $L(C)$ du contexte C .

Pour plus de clarté, nous présentons respectivement dans la Figure 9-10 et la Figure 9-11 un exemple de contexte ainsi que le treillis de Galois qui lui correspondant.

O\A	a	b	c	d	e	f	g	h	i
1	1		1			1		1	
2	1		1				1		1
3	1			1			1		1
4		1	1			1		1	
5		1			1		1		

Figure 9-10- La matrice binaire décrivant la relation I du contexte $C = (O, A, I)$

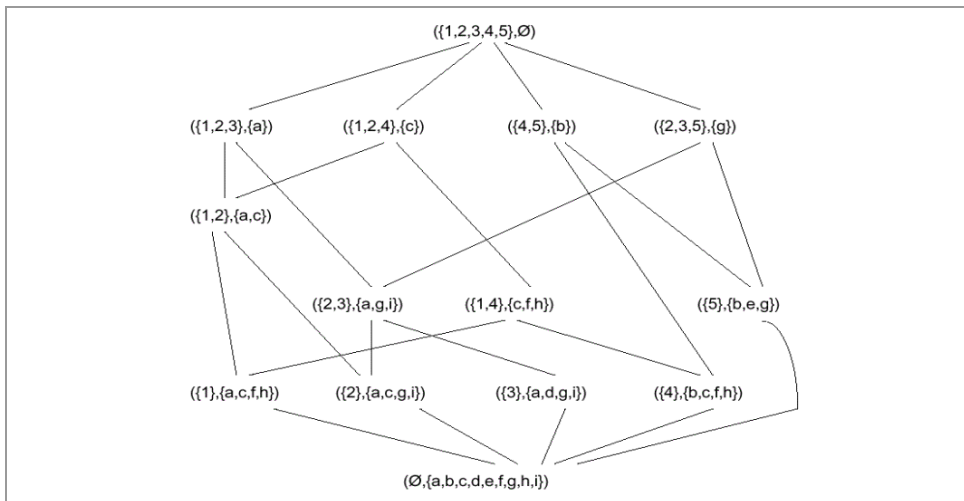


Figure 9-11- Le treillis de Galois $L(C)$ du contexte précédent

Selon la Figure 9-11, le nœud $\{1,4\}; \{c,f,h\}$ signifie que les objets 1 et 4 partagent les attributs c, f et h .

Pour éviter la confusion entre la notion de *concept de treillis* (Définition 9.4) et la notion de *concept du domaine* qui instrumente les différents modèles dans ASMEC (cf. section 8.4), nous remplaçons cette dernière par le terme de *concept systémique*. Rappelons que ces concepts systémiques sont caractérisées par des attributs.

Nous définissons aussi la notion de *concept systémique actif* comme suit :

Définition 9.7- Concept systémique actif dans une classe d'accidents

C'est un concept du domaine dont le nombre d'attributs (ou de modalités) pertinents le caractérisant dépasse un seuil défini par l'expert. Cela se traduit au niveau de la grille systémique par la présence d'un nombre d'attributs (ou de modalités) dans la case correspondante au concept supérieur au seuil fixé.

Prenons un exemple dans la Figure 9-9 : « *Conducteur* » et « *Véhicule* » sont deux concepts systémiques du modèle CVE. Si on fixe le nombre 3 comme seuil, le concept « *Conducteur* » est actif, alors que le concept « *Véhicule* » ne l'est pas.

9.2.2.2 Utilisation des treillis de Galois pour l'interprétation

Nous partons de la définition de la notion de contexte (Définition 9.1) et nous ajoutons une autre composante P au contexte contenant des conditions qui doivent être remplies par la relation I . En d'autres termes $\boxed{(o, a) \in I \text{ ssi } P(o, a) = \text{vrai}}$.

Soit alors le contexte binaire $C = (O, P, A, I)$ qui correspond à une classification que nous avons effectuée où :

- O : représente l'ensemble des classes obtenues ;
- P : contient des critères spécifiant les seuils des pouvoirs discriminant et caractérisant des modalités (cf. §9.2.1.1 pour les définitions). Les seuils utilisés par les experts dans la pratique sont : $2 < \text{valeur-test} < -2$ et $\text{MOD/CLASS} > 50\%$. Donc $P(o, a) = \text{vrai}$ ssi $(2 < \text{valeur-test} < -2 \text{ et } \text{MOD/CLASS} > 50\%)$;
- A : l'ensemble de toutes les modalités vérifiant les critères dans P . Il s'agit dans notre cas d'étude des modalités discriminantes et caractérisantes (cf. §9.2.1.1 pour la définition) pour au moins une classe ;
- $I \subseteq O \times A$ une relation binaire vérifiant : $\boxed{(o, a) \in I \text{ ssi } P(o, a) = \text{vrai}}$;
- f : l'application qui à toute classe o de O associe $f(o) = \{a \in A, (o, a) \in I\}$, qui affecte à une classe o l'ensemble des modalités discriminantes et caractérisantes.

Dans le treillis de Galois¹ présenté dans la Figure 9-12, les nœuds sont appelés les *concepts du treillis* (cf. Définition 9.4). L'*intention* (i.e. la caractérisation) d'un concept correspond aux *attributs* qui le caractérisent. Graphiquement, elle correspond aux attributs qui se trouvent au dessus du concept (i.e. les sur-concepts). L'*extension* d'un concept correspond aux classes. Graphiquement, elle est constituée de toutes les classes qui se trouvent au dessous du concept (i.e. les sous-concepts).

¹ L'outil que nous avons utilisé pour tracer les treillis s'appelle « ConExp ». Il est en libre accès sur le web.

Prenons un exemple : le nœud qui se trouve à droite, en haut, est un concept dont $l'intention = \{probl_contrôle_véh, Véhicule_seul\}$ et $l'extension = \{classe11, classe3, classe9\}$. Grâce à la dualité de la représentation classe/attribut fournie par les treillis, nous avons réussi à afficher, en un seul graphique, plusieurs informations qui proviennent de différentes sources et dont la combinaison est fondamentale pour l'interprétation. Ces informations portent sur l'expression des *dépendances inter-classes*, *inter-attributs* et *entre les classes et les attributs*. En ce qui concerne les *dépendances inter-classes*, par exemple, on peut lire facilement sur la Figure 9-12 le fait que les classes $\{3, 9, 11\}$ sont proches l'une de l'autre alors que les classes $\{5, 10, 17\}$ et $\{18\}$ sont complètement isolées car elles ne partagent aucune modalité (i.e. nœud dans le graphique) entre elles. En ce qui concerne les *dépendances inter-attributs*, on peut lire le fait que les modalités « *probl_contrôle_véh* » et « *Véhicule_seul* » sont deux modalités co-occurentes dans les classes $\{3, 9, 11\}$.

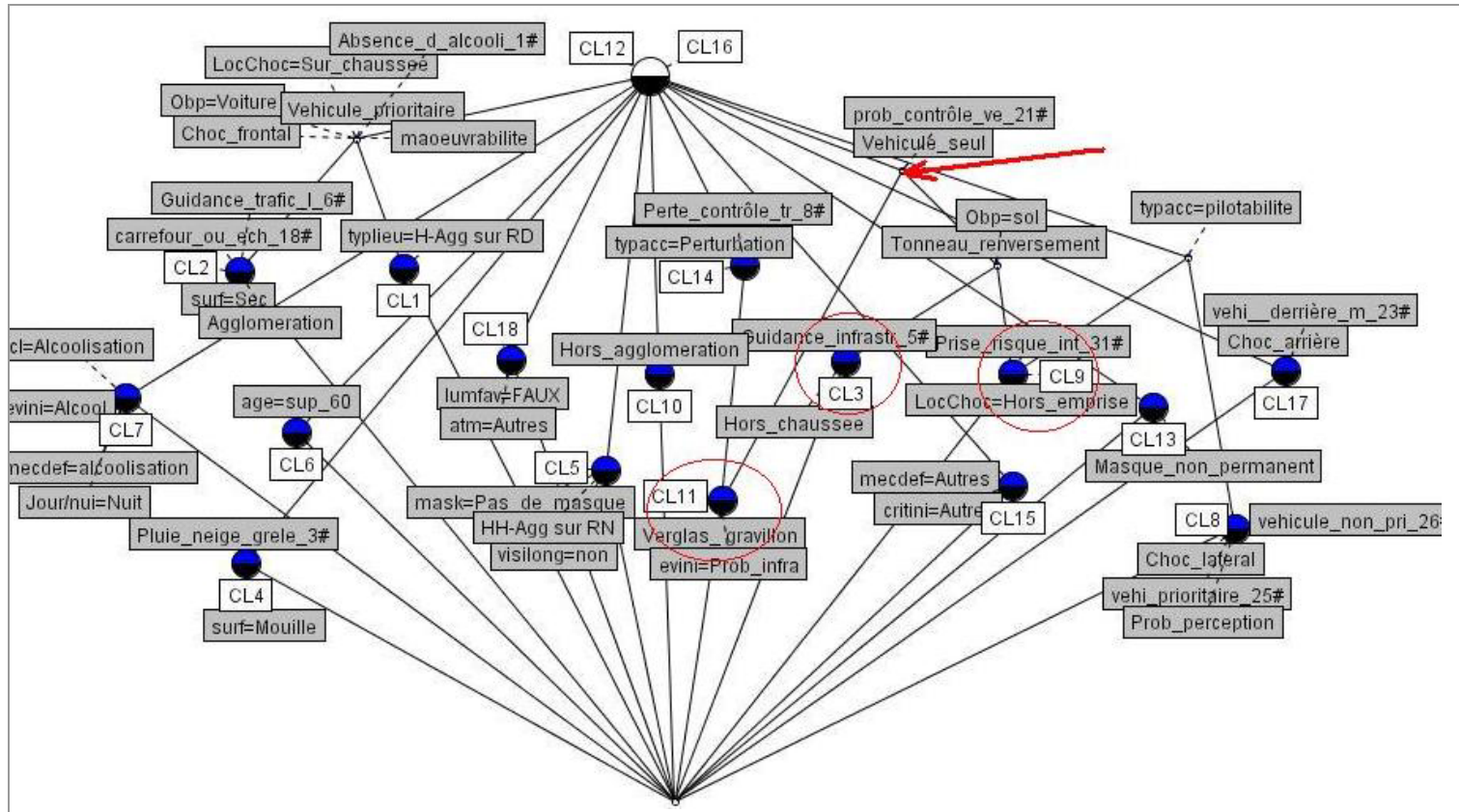


Figure 9-12- Treillis de Galois $L(C)$ correspondant à la classification C .

A partir du treillis de Galois présenté dans la Figure 9-12, certaines informations nécessaires à l'interprétation des classes, et qui auraient nécessité beaucoup du temps et d'effort pour être extraites, sont plus facilement accessibles. La Figure 9-13 illustre un exemple des avantages de l'utilisation des treillis pour comparer les classes. Il s'agit du même treillis de la Figure 9-12, mais en sélectionnant le nœud qui correspond à la co-occurrence des deux modalités {« prob_contrôle_véh » et « véh_seul »}. On s'aperçoit rapidement que les classes {3, 9 et 11} partagent ces deux modalités. Donc, ces trois classes peuvent former une *super-classe* caractérisée par les deux modalités partagées. A un niveau plus fin, on peut regrouper uniquement les classes {3 et 9} qui partagent les quatre modalités {« prob_contrôle_véh », « véh_seul », « obstacle=sol », « tonneau_reversement »}. Ce graphique fournit en même temps ce qui réunit les classes et ce qui les différencie.

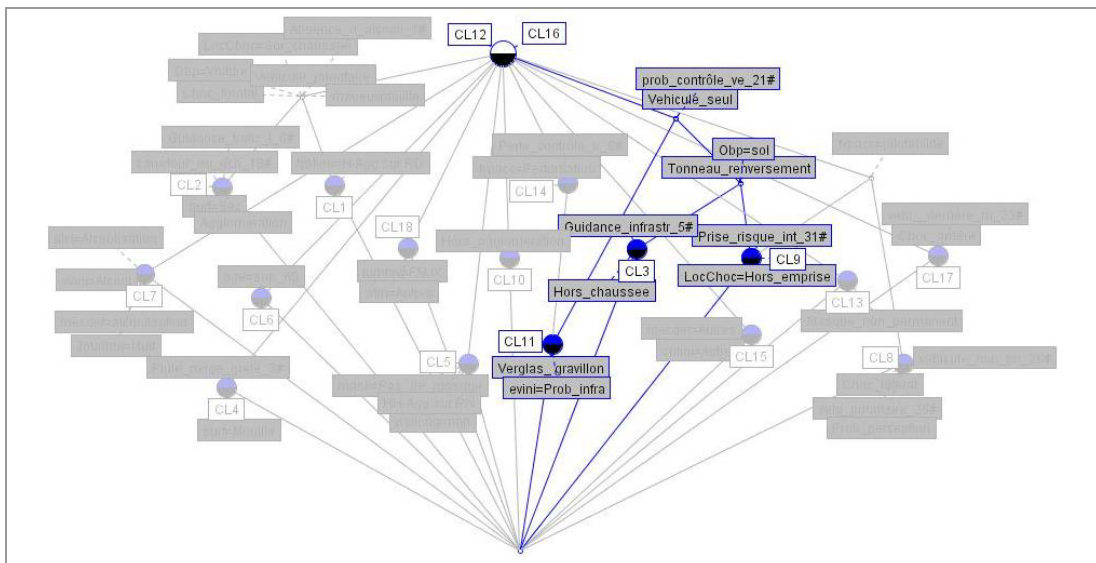


Figure 9-13- Exemple d'utilisation du treillis pour l'interprétation des classes au niveau des modalités.

Avec l'introduction des techniques des treillis dans le processus d'interprétation des classes, de nouvelles connaissances ont commencé à se mettre en place. En effet, on a remarqué que lors de la lecture des treillis de nouvelles règles ont émergé. Nous avons donc procédé à la formalisation de ces nouvelles règles pour les ajouter dans la base de connaissances. A chaque utilisation de cette technique de lecteur des classes, de nouvelles règles pourraient surgir pour enrichir la base de connaissances.

Parmi les règles que nous avons pu identifier, citons :

- Plus la classe correspond à un concept qui se trouve à un niveau élevé dans le treillis, plus elle est générale. Autrement dit, elle correspond à une généralisation de toutes les classes qui se trouvent dans la même branche à un niveau plus bas du treillis (i.e. ses sous-concepts). Par exemple : la classe 11 (Figure 9-12) correspond à un sous-concept de la classe 14. cette dernière est donc une généralisation de la première ;
- Plus le nombre de sous-concepts et de sur-concept d'une classe est faible, plus la classe est séparée des autres. C'est le cas, par exemple, des classes 4, 7 et 6 qui ne partagent aucun nœud avec les autres classes ;

- Plus la modalité correspond à un concept qui se trouve à un niveau élevé du treillis, plus elle est transversale. C’est-à-dire qu’elle correspond à une problématique transversale, qui concerne beaucoup de classes. C’est le cas de la modalité « *prob_contrôle_véhicule* » qui concerne les *classes* {3, 9 et 11};
- Chacun des nœuds du treillis correspond à une classe dont l’extension contient tous les accidents des extensions des sous-concepts. La caractérisation de cette classe est donnée par l’intention du concept auquel elle correspond. Ainsi le treillis fournit une classification à plusieurs niveaux d’abstraction ;
- Si plusieurs modalités correspondent au même concept qui se trouve à un niveau haut du treillis, ces modalités ont plus de chance d’être des facteurs d’accidents corrélés entre eux.

9.2.3 Interprétation multi-niveaux d’abstraction

Nous avons défini quatre niveaux d’abstraction pour l’interprétation d’une classe qui sont issus de l’architecture d’ASMEC (cf. section 8.4). Cela offre à l’expert une interprétation sur les niveaux : *modèles*, *concepts systémiques* du domaine (i.e. composantes des modèles), *attributs* et *modalités* (valeurs des attributs). Dans le paragraphe précédent (§ 9.2.2.2), nous avons présenté l’utilisation des treillis pour l’interprétation au niveau des *modalités*. Nous présentons ci-dessous leur utilisation pour l’interprétation sur les autres niveaux de granularité.

9.2.3.1 Interprétation au niveau « modèles »

La projection d’une classe sur la grille systémique présentée dans la Figure 9-9 permet à l’expert d’identifier rapidement la vue systémique la plus pertinente pour l’interprétation de la classe. Quand sur la grille le nombre de *concepts systémiques actifs* (cf. Définition 9.7) au niveau de l’un des modèles d’interprétation dépasse un seuil fixé par l’expert, cela signifie que le point de vue correspondant est pertinent pour l’interprétation. L’inverse, c’est-à-dire « *nombre (concepts systémiques actifs) < seuil* » signifie : ou bien que ce point de vue n’est pas pertinent pour l’interprétation ou bien que les attributs qui ont été choisis à la base pour effectuer la classification ne recouvrent pas cette vue. C’est le cas, par exemple, du point de vue fonctionnel fourni par le modèle de traitement d’information (Figure 9-9).

Ce niveau d’interprétation permet donc de déterminer quels sont parmi les modèles d’interprétation utilisés ceux qui sont pertinents pour interpréter une classe donnée.

9.2.3.2 Interprétation au niveau « concepts systémiques »

En se basant toujours sur la grille systémique et en regardant plus précisément les *concepts systémiques du domaine* qui constituent chacune des vues (e.g. « conducteur », « véhicule » pour le point de vue ontologique), plusieurs informations peuvent être extraites. En effet, si des concepts systémiques sont *co-occurents*, c’est-à-dire actifs (Définition 9.7) plusieurs fois et en même temps, il y a des chances pour qu’il y ait des interactions entre ces concepts. Cela peut être complété par une visualisation à l’aide des treillis. Dans la Figure 9-14, par exemple, on remarque que dans les classes {12, 18, 6, 1, 2, 11, 3 et 9} les deux concepts systémiques « *perception* » et « *environnement* » sont actifs.

Ceci paraît logique, puisque les problèmes de perception sont dans plusieurs cas reliés à des problèmes d'environnement (e.g. soleil, masque de visibilité, etc.).

La visualisation d'une classification à l'aide d'un treillis se fait en définissant le contexte binaire suivant : $C = (O, P, A, I)$ avec :

- O : représente l'ensemble des 18 classes obtenues ;
- A : l'ensemble des *concepts systémiques* du domaine instrumentant les différents modèles d'interprétation (cf. section 8.6.2) ;
- P : définie par $P(o, a) = \text{vrai ssi } a \text{ est un concept actif}$, le seuil est défini par l'expert ;
- $I \subseteq O \times A$ une relation binaire vérifiant : $(o, a) \in I \text{ ssi } P(o, a) = \text{vrai}$.

La Figure 9-14 donne le treillis de Galois correspondant à ce contexte. Ce niveau d'interprétation permet de détecter des concepts systémiques co-occurents au niveau d'un seul point de vue ou plusieurs points de vue.

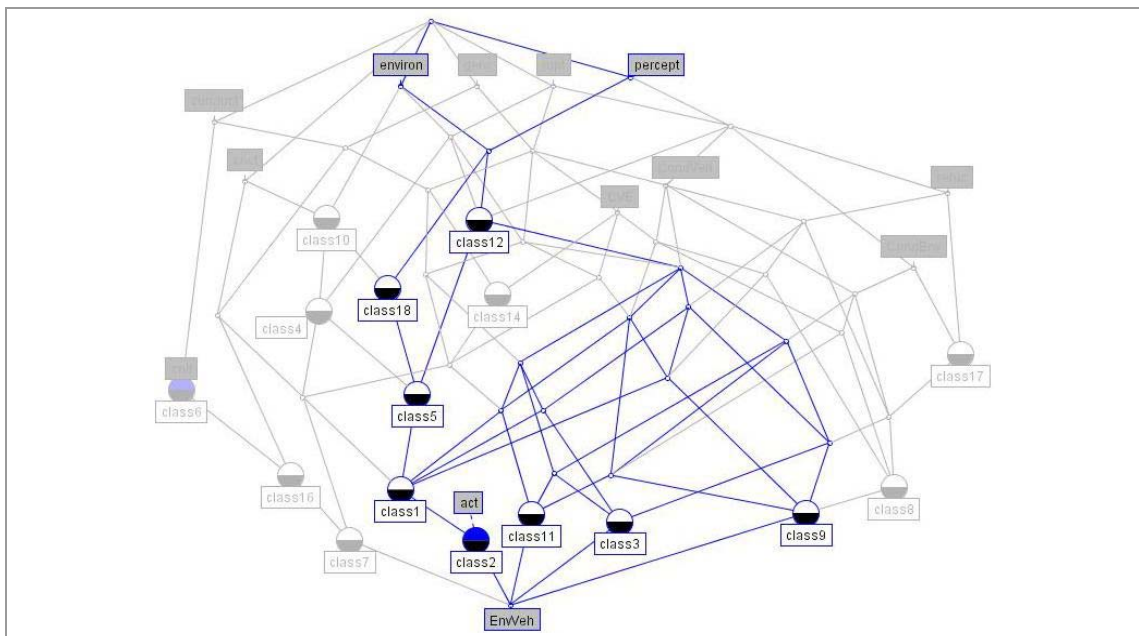


Figure 9-14- Interprétation des classes au niveau des concepts systémiques (i.e. concepts du domaine) : les concepts « environnement » et « perception » sont partagés par les classes 12, 18, 6, 1, 2, 11, 3 et 9.

Pour connaître, par exemple, les classes dans lesquelles il y a co-occurrence des concepts systémiques « environnement » et « perception », il suffit de sélectionner le premier nœud commun entre ces deux concepts. Toutes les classes recherchées sont alors visualisées. L'avantage de cette visualisation à l'aide du treillis est de faire ressortir des pistes qui pourraient être ensuite explorées par l'expert notamment en affinant l'analyse : passer à l'interprétation au niveau des *attributs* et des *modalités* (voir ci-dessous).

9.2.3.3 Interprétation au niveau « attributs »

La grille systémique peut être représentée au niveau des attributs. C'est-à-dire que chacun des concepts systémiques est caractérisé par les attributs pertinents à la place des modalités. Un attribut est jugé pertinent si l'une de ses modalités est jugée pertinente selon les critères statistiques relatifs

aux notions du *pouvoir discriminant* et *caractérisant* (cf. § 9.2.1.1). Ces attributs se trouvent dans la première colonne du tableau de description des classes (Figure 9-5).

Modèle Ontol.	Conducteur	Véhicule	Environnement	Cond/Veh	Cond/Env	Env/Veh
	TypDéfail Manoeuvre évèn_intiateur typAccident Mécanism_Défail évèn_intiateur Mécanism_Défail FonctDéfail	TypChoc TypObstacle	CondAmbiant EtatSurf TypLieu PrésenceMask Mécanism_Défail Manoeuvre	SituatAccid EtatSurf TacheCritiq	TacheCritiq EtatSurf Manoeuvre	TypObstacle EtatSurf Vehicule_seul
Modèle Transf.	Etat long terme LT	Cond. normale CT	Rupture	Urgence	Choc	
		CondAmbiant EtatSurf TypLieu Manoeuvre	SituatAccid TacheCritiq PrésenceMask TypDéfail Vehicule_seul Manoeuvre MévanismDéfail	MévanismDéfail	TypObstacle EtatSurf TypChoc TypObstacle LocChoc	
Modèle Fonct.	Perception	Diagnostic	Pronostic	Décision	Action	Globale
	PrésenceMask			MévanismDéfail	FonctDéfail TypDéfail TacheCritiq	MévanismDéfail évèn_intiateur/font>
Modèle Téléo.	Navigation	Guidage latéral	Guidage longit.	Controle latéral	Controle longit.	
	PrésenceMask	TacheCritiq Manoeuvre	TacheCritiq	TacheCritiq TypChoc		

Figure 9-15- Grille systémique : projection automatique des résultats de classification sur les métadonnées au niveau des attributs.

La Figure 9-15 présente la même grille que celle de la Figure 9-9, mais au niveau attributs au lieu des modalités. Cette grille sert à l’interprétation intra-classe (cf. section 9.2.1.2).

Pour pouvoir comparer les caractérisations des classes par les attributs pertinents, nous utilisons les treillis correspondants au contexte binaire suivant : $C = (O, P, A, I)$ avec :

- O : représente l’ensemble des 18 classes obtenues ;
- A : l’ensemble des *attributs pertinents* ;
- P : définie par $P(o, a) = \text{vrai ssi } a \text{ est un attribut pertinent}$;
- $I \subseteq O \times A$ une relation binaire vérifiant : $(o, a) \in I \text{ ssi } P(o, a) = \text{vrai}$.

Le treillis correspondant à ce contexte est donné dans la Figure 9-16.

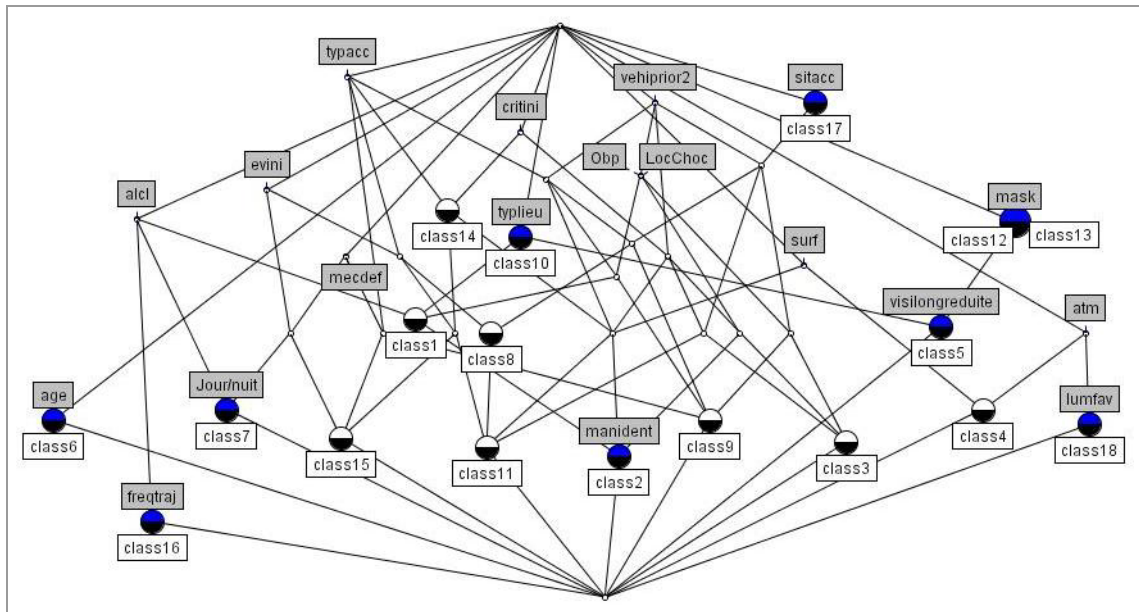


Figure 9-16- Interprétation des classes au niveau des attributs.

Ce treillis permet de visualiser tous les liens entre les classes au niveau des attributs. Cette visualisation permet d'identifier des *problématiques transversales* qui peuvent servir pour définir de nouvelles études en accidentologie.

Définition 9.8- Problématique transversale en accidentologie

Elle correspond à un ou plusieurs facteurs d'accidents (i.e. attributs) qui concernent plusieurs classes dans le treillis.

Considérons comme exemple la Figure 9-17 : elle représente le même treillis que celui dans la Figure 9-16, mais en sélectionnant l'attribut « *critini* »¹. Le treillis nous montre que cet attribut concerne les classes {2, 3, 11, 14, 15}. « *La tâche critique initiale* » peut donc correspondre à une problématique transversale qui pourrait faire l'objet d'une nouvelle étude. Le rôle du treillis est de faire ressortir ce genre de problématiques. Ce treillis permet aussi de comparer les classes et de voir ce qui les réunit ainsi que ce qui les différencie au niveau des attributs. Ces informations sont ensuite approfondies avec le dernier niveau de lecture, celui des modalités (cf. § 9.2.2.2).

¹ L'attribut « *critini* » concerne la *tâche critique initiale* que le conducteur était en train d'exécuter au moment de l'accident, juste avant la rupture.

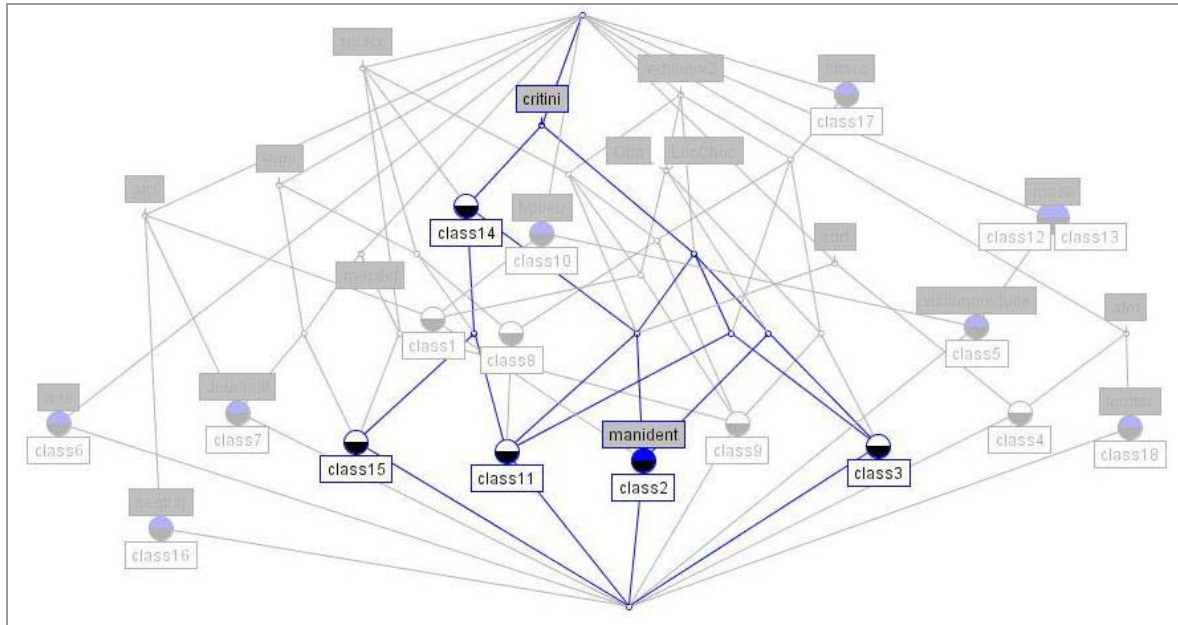


Figure 9-17- Exemple d'utilisation du treillis pour l'interprétation des classes au niveau des attributs.

9.3. Synthèse

Dans ce chapitre, nous avons montré comment AICEF utilise les métadonnées développées dans la section 8.8 pour incorporer les connaissances expertes du domaine dans la première et dernière étape du processus ECD (i.e. préparation des données et interprétation des résultats).

Grâce à AICEF, nous permettons à l'expert d'effectuer une sélection multi-vues des attributs. Le rôle de l'expert est de traduire l'objectif de son étude en fonction des concepts systémiques du domaine qui ont été identifiés pour instrumenter les différents modèles. Par exemple, l'étude dont l'objectif est de comprendre les différentes configurations du problème de perte de contrôle se traduit en fonction des concepts du domaine de la manière suivante: $\{(Fonctionnel \vee Transformationnel \vee Téléologique) \wedge (Conducteur_Véhicule \vee Conducteur_Environnement) \wedge (Situation\ de\ rupture \vee Situation\ d'urgence) \wedge (Perception \vee Diagnostic \vee Pronostiques)\}$. Ensuite, puisque chacun de ces concepts est déjà caractérisé par des attributs de la base de données grâce aux métadonnées, une requête est créée puis exécutée automatiquement pour proposer à l'expert l'ensemble d'attributs qui sont a priori pertinents. Le reste du processus de sélection est formalisé dans les structures d'inférence déjà présentées dans la Figure 7-10, section **Erreur ! Source du renvoi introuvable.**

En ce qui concerne la deuxième partie de ce chapitre, elle concerne l'intégration, via les métadonnées, des connaissances du domaine dans l'étape d'interprétation des classes. Pour répondre à cet objectif, nous avons proposé deux outils. Le premier est appelé « grille systémique ». Il fournit une projection automatique des classes selon les différents modèles d'interprétation qui instrumentent le méta-modèle (cf. section 8.6.2). Le deuxième outil est basé sur les treillis de Galois qui, grâce à la dualité objet/attribut de leur représentation, offrent un moyen efficace pour exprimer

les dépendances *inter-classes*, *inter-attributs* et *entre les classes et les attributs*. Nous avons développé deux interfaces graphiques facilitant l'utilisation de ces deux outils.

Nous avons défini trois types d'interprétation :

- (i) Une *interprétation intra-classe* qui consiste à comprendre et à trouver une « bonne » caractérisation d'une classe donnée en utilisant la grille systémique ;
- (ii) Une *interprétation inter-classes* qui consiste à comprendre une classe en la comparant avec les autres et ce en utilisant les treillis de Galois ;
- (iii) Une *interprétation statistique* qui se base sur des critères statistiques pour juger la pertinence des attributs.

Nous avons développé aussi une architecture multi-niveaux d'abstraction pour la lecture des classes. Cette architecture offre une lecture aux niveaux des « modèles », des « concepts systémiques » du domaine, des « attributs » caractérisant les concepts et des « modalités » des attributs. L'interprétation au « niveau modèles » (cf. §9.2.3.1) consiste à trouver le(s) modèle(s) pertinent(s) pour interpréter une classe donnée en utilisant la grille systémique. L'interprétation au « niveau concepts » (cf. §9.2.3.2) consiste à : d'une part, trouver les concepts pertinents en utilisant la grille systémique et d'autre part, trouver des interactions entre les concepts en utilisant les treillis de Galois. De même, l'interprétation au « niveau attributs » (cf. §9.2.3.3) consiste à trouver les attributs pertinents, les problématiques transversales d'étude et les interactions entre les facteurs accidentogènes. Quant à l'interprétation au « niveau modalités » (cf. § 9.2.2.2), elle consiste à trouver une meilleure caractérisation des classes, à comparer les classes et à créer de nouvelles classes en regroupant celles qui sont les plus proches. La Figure 9-18 illustre ces quatre niveaux d'interprétation des classes.

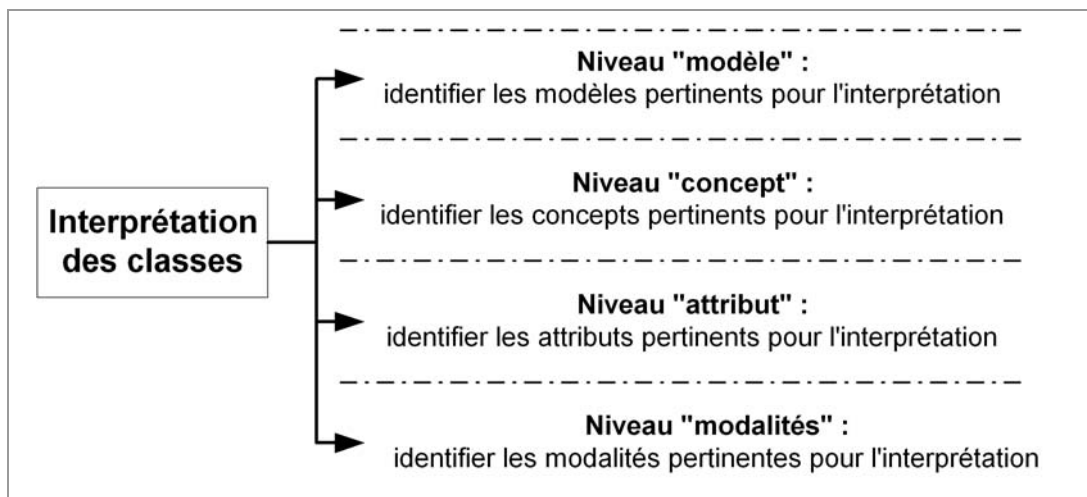


Figure 9-18- Lecture multi-niveaux d'abstraction des classes.

Chapitre 10.

ASAIC : une Approche Systémique d’Analyse d’Impact de Changement

ASAIC est la troisième approche que nous développons dans SAFE-Next. Elle permet l’utilisation des méta-données développées par ASMEC (cf. Chapitre 8) pour l’incorporation des connaissances du domaine dans l’analyse multi-vues et multi-niveaux d’abstraction de l’impact de changement dans le domaine de sécurité routière. (cf. Figure 10-1).

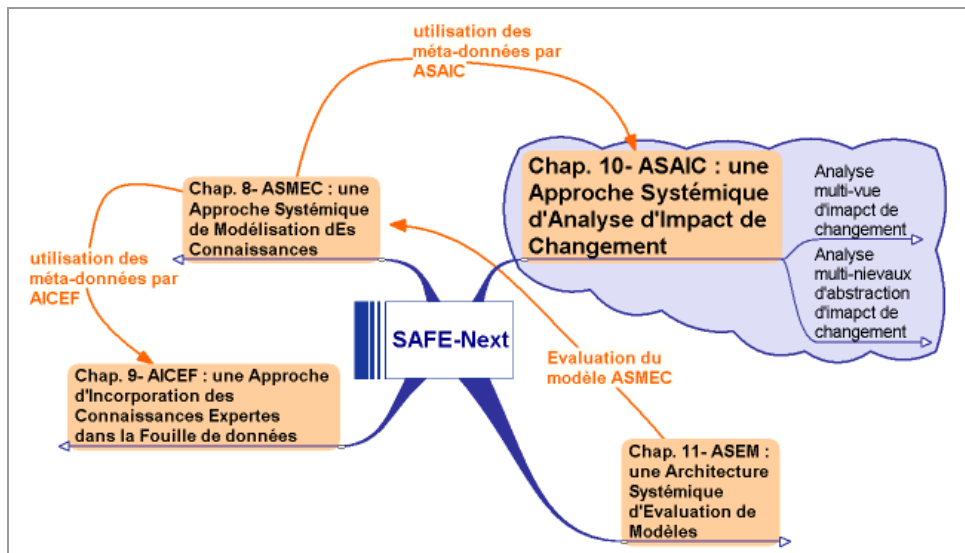


Figure 10-1- ASAIC : utilisation des méta-données pour l’analyse d’impact de changement (troisième approche dans SAFE-Next)

10.1. Besoin d'analyse d'impact d'un changement sur le système CVE

Comme nous l'avons déjà noté dans la section 3.5, l'accident de la route ne peut pas se réduire au crash, mais il est dû à plusieurs facteurs qui portent sur le conducteur et/ou le véhicule et/ou l'environnement. Dans l'environnement nous intégrons l'infrastructure, les conditions ambiantes, le trafic et le contexte socio-politique, économique et technologique. Un changement au niveau de l'un de l'un des trois sous-système (i.e. conducteur, véhicule et environnement) peut affecter la sécurité routière en général. En effet, un conducteur dans un véhicule équipé d'un ABS ou d'un ESP peut se sentir plus en sécurité et donc prendre plus de risques dans sa conduite. Cela peut affecter dramatiquement le comportement du système CVE. Les configurations des accidents et le besoin en sécurité peuvent donc changer. Cela génère, par retour d'expérience, de nouveaux besoins à l'entrée du processus de conception des systèmes de sécurité.

Ainsi, développer des SSE sans analyser à l'avance leurs impacts sur le comportement du conducteur, sur le comportement du véhicule et sur les interactions entre les différentes composantes du système CVE, peut engendrer des effets pervers.

Des techniques d'Ingénierie d'Analyse d'Impact de Changement (IAIC) (Engineering Change Impact Analysis) existent dans la littérature. Nous en faisons un état de l'art et nous montrons ensuite comment SAFE-Next fournit une approche d'IAIC que nous appelons ASAIC (Approche Systémique d'Analyse d'Impact de Changement).

10.2. Processus d'Ingénierie d'Analyse d'Impact de Changement (IAIC)

De nombreuses techniques d'IAIC ont été développées pour répondre au besoin des entreprises qui veulent faire face à l'évolutivité de l'environnement dans lequel elles existent. L'anticipation des changements et de leur impact est fondamentale pour qu'une organisation survive.

Dans [Pikosz et Malmqvist, 1998], on distingue deux catégories de changements auxquels une organisation peut faire face. La première concerne des changements affectant la structure d'un système tels que le changement d'une composante, le changement de la fonction d'une composante, l'introduction d'une nouvelle composante, etc. La deuxième catégorie concerne des changements affectant le processus du développement du produit tels que le changement des spécifications client, difficulté de fabrication d'une partie du produit, problème de qualité, etc.

L'un des problèmes clés pour faire face à un changement est de savoir tous les éléments (structurels ou fonctionnels) qui sont impactés ainsi que les conséquences de ces changements. Une description générique d'un processus d'IAIC dans le domaine de la conception peut se faire à travers les phases suivantes [Ma et al., 2003; Madhavji Nazim, 1992] :

- Identifier le type de changement : développer un nouveau système, intégrer une nouvelle technologie, un nouveau besoin client, etc. ;

- Evaluer l’impact du changement sur les différents éléments concernés : produit, ressources, processus, etc. ;
- Définir un scénario de changement en tenant compte des objectifs du changement ainsi que des éléments impactés ;
- Evaluer les plans alternatifs de changement et choisir le plus adapté ;
- Effectuer le changement sur tous les éléments concernés ;
- Capitaliser les détails de changement pour une réutilisation future et informer toutes les parties concernées (personnes, services, etc.).

Une illustration du processus d’IAIC pour la conception est donnée dans la Figure 10-2 [Ma *et al.*, 2003]. La différence entre les différentes approches d’IAIC se situent essentiellement au niveau des techniques utilisées dans la case « *analyseur d’impact* » de cette figure (cf. Figure 10-2). Dans la section suivante (section 10.3), nous présentons brièvement quelques approches utilisées pour l’AIC. Ensuite, nous développons dans la section 10.4 notre approche utilisant le modèle des connaissances ASMEC pour l’AIC.

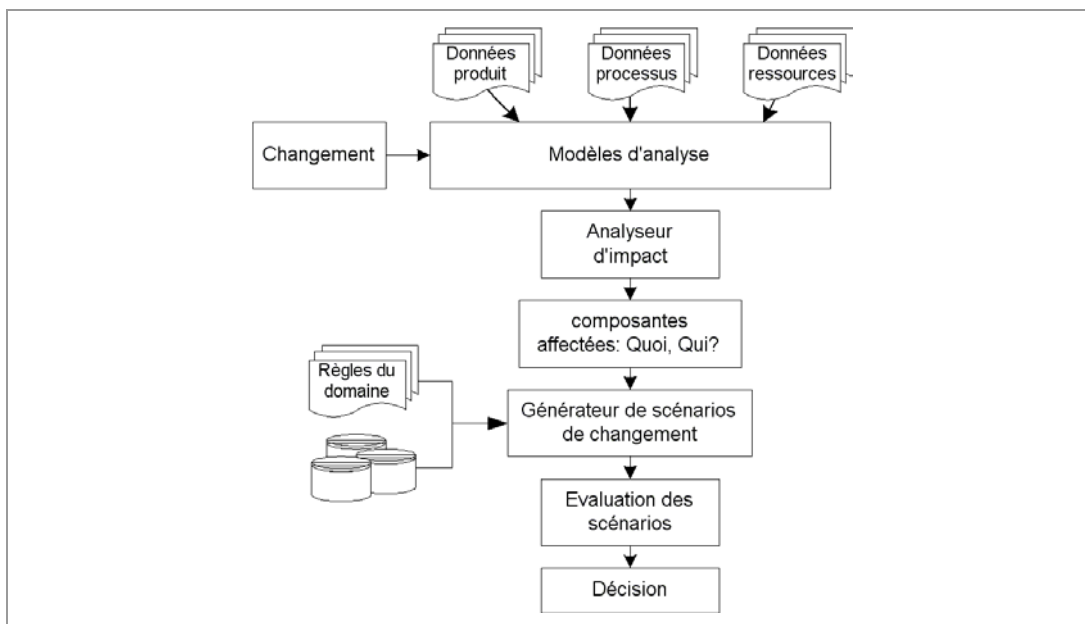


Figure 10-2- Processus d’analyse d’impact de changement (adaptée de [Ma *et al.*, 2003])

10.3. Différentes approches d’IAIC

L’IAIC est largement appliquée dans de nombreux domaines : maintenance en génie logiciel, gestion de projets, analyse des risques, processus de conception, etc. Une comparaison entre les différentes approches existantes avec des cas d’étude peut être trouvée dans [Pikosz et Malmqvist, 1998].

Parmi ces approches, nous citons le *Design Rational* (DR). Il est défini comme l’explication des raisons pour lesquelles un artefact est conçu tel qu’il est [Regli *et al.*, 2000]. Cela concerne en particulier les connaissances des décisions prises lors de la conception et les raisons et les

compromis qui les justifient [Longueville, 2003]. Parmi les méthodes utilisées en DR, nous citons la QOC [Buckingham Shum, 1997; Buckingham Shum et Sylvain, 2000] qui utilise les trois objets Questions, Options et Critères pour capitaliser les choix effectués en conception afin de permettre leur réutilisation dans l'analyse d'un changement. Nous citons aussi la méthode DRCS [Klein, 1993] qui spécifie un langage de description de la logique de conception (cf. [Klein, 1993] pour une description détaillée de cette méthode). Dans [Bose, 1998], l'approche DR a été utilisée pour : identifier les dépendances entre les décisions architecturales, valider et justifier des choix de connexions (de cause, d'ordre, de conflit, etc.) entre les différentes composantes et expliquer les propriétés des composantes. Cela permet ensuite l'analyse d'un changement au niveau de l'une des composantes. Dans [Gorti *et al.*, 1998], une représentation orientée objet du produit a été développée pour une identification automatique de l'interdépendance spatiale entre les différentes composantes d'un produit. Elle permet aussi d'identifier automatiquement les personnes ainsi que les unités organisationnelles concernées par le changement pour qu'elles soient informées.

Nous avons distingué plusieurs points de vue d'AIC. Des *approches orientées produit* [Murdock *et al.*, 1997] : elles se basent sur une représentation du produit identifiant les dépendances structurelles et/ou fonctionnelles entre ces différentes composantes pour analyser l'impact d'un changement. Des *approches orientées processus* [Blessing, 1996] : elles se basent sur une représentation du processus pour analyser l'impact d'un changement. D'autres approches intègrent plusieurs points de vue tels que l'approche intégrée *processus/produit/ressources* présentée dans [Ma *et al.*, 2003].

10.4. Elaboration d'un modèle d'analyse d'impact d'un changement sur le système CVE

Le point commun entre les différentes approches présentées dans la section précédente (section 10.3) réside dans le fait qu'elles cherchent toutes à construire des connaissances sur l'artefact étudié (produit, processus, organisation, ressource). Ces connaissances peuvent porter sur les relations de dépendance entre les composantes de l'artefact, les attributs de l'artefact, les attributs des tâches de conception, les décisions prises lors de la conception, etc.

En sécurité routière, les connaissances en accidentologies sur le système CVE et son comportement sont fondamentales pour analyser l'impact de l'introduction d'un système de sécurité. Étant donné la complexité du système, plusieurs points de vue sont nécessaires pour anticiper tous les types de changement possibles. Nous utilisons alors le modèle multi-vues et multi-niveaux d'abstraction de connaissances que nous avons développé dans ASMEC pour analyser l'impact d'un changement sur le système CVE (cf. section 8.4). La Figure 10-3 montre comment les différents points de vue de la systémique sont utilisés dans l'AIC. Les méta-données permettent de relier les différents points de vue et donc d'analyser l'impact selon ces différentes vues.

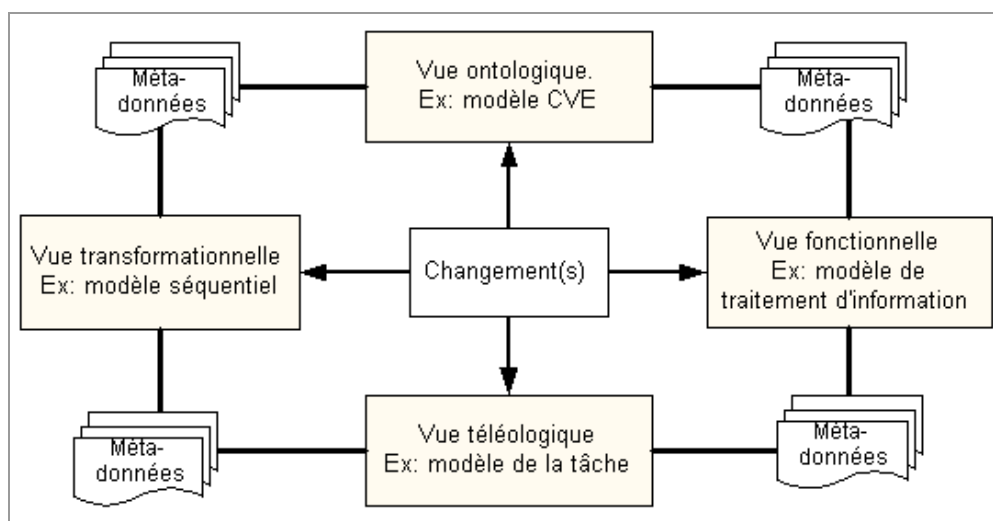


Figure 10-3- Un modèle d'AIC sur le système CVE

Par rapport au processus classique d'AIC présenté dans la Figure 10-2, le modèle de la Figure 10-3 se situe dans la case « *modèles d'analyse* ». Quant à l'utilisation de ce modèle (i.e. la case « *analyseur d'impact* » dans la Figure 10-2), nous proposons les étapes suivantes :

- Faire une première analyse selon les quatre points de vue systémiques du méta-modèle : cela permet de savoir si le changement affecte des aspects ontologiques et/ou fonctionnels et/ou transformationnels et/ou téléologique ;
- Identifier le(s) modèle(s) adapté(s) pour l'analyse (e.g. modèle CVE, modèle séquentiel, etc.) ;
- Identifier les concepts pertinents pour l'analyse (e.g. conducteur, phase de rupture, etc.) ;
- Identifier les attributs des concepts pertinents pour l'analyse ;

Dans notre cas, les attributs correspondent aux « *composantes affectées* » dans le processus générique d'AIC présenté dans la Figure 10-2. Nous avons choisi de décrire le fonctionnement de notre « *analyseur d'impact* » à travers la structure d'inférence que nous présentons dans la Figure 10-4.

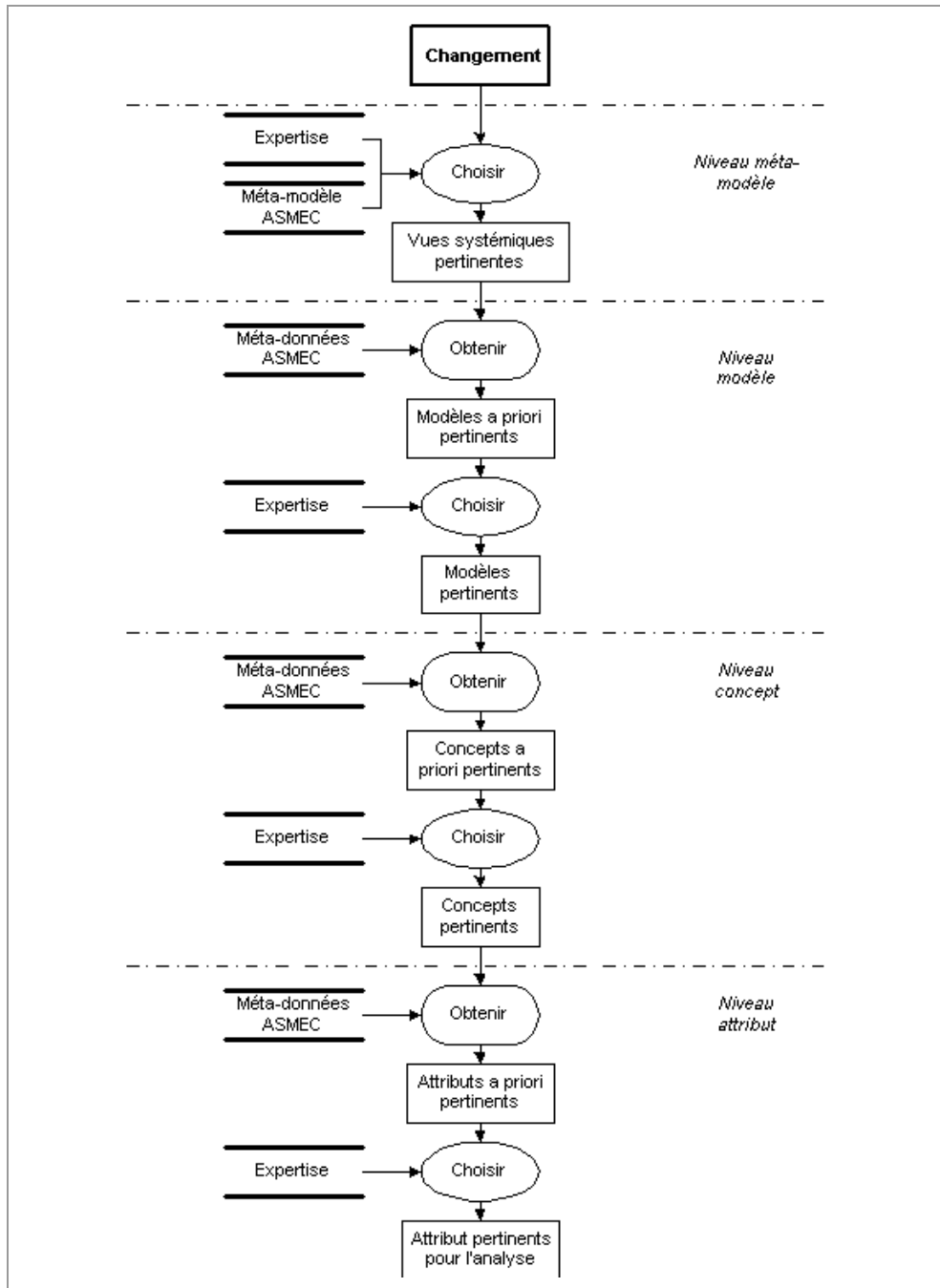


Figure 10-4- Structure d'inférence dans l'analyseur d'impact

En plus de ses aspects multi-vues et multi-niveaux d'abstraction, notre analyseur d'impact permet une *analyse interactive* « expert/méta-données ». En effet, à un niveau donné de l'architecture (méta-modèle, modèles, concepts, attributs), l'expert choisit les éléments (i.e. modèles, concepts, attributs) qui peuvent être affectés par un changement. Grâce aux méta-données d'ASMEC (cf. les sections 8.6, 8.7 et 8.8), nous permettons à l'expert l'accès automatique à tous les éléments qui sont relatifs à l'élément choisi et qui appartiennent au niveau d'abstraction suivant. Par exemple, si l'expert décide que *l'aspect ontologique* risque d'être affecté par le changement, tous les *modèles* représentant ce point de vue (e.g. modèle CVE, modèle des systèmes de sécurité, etc.) sont alors sélectionnés et proposés à l'expert. Ce dernier choisit alors un ou plusieurs modèles d'analyse (e.g. CVE). Tous les *concepts* concernant les modèles choisis (e.g. conducteur, véhicule, environnement)

sont alors sélectionnés et affichés. Au dernier niveau, nous permettons à l’expert d’obtenir tous les *attributs* qui caractérisent un concept donné (e.g. conducteur) selon un point de vue donné (e.g. ontologique) et qui risquent d’être affectés par le changement. Des décisions sont alors prises (par l’expert) pour anticiper l’impact du changement.

Dans la section suivante, nous présentons un extrait du cas d’étude d’utilisation de notre modèle d’AIC que nous avons effectué dans [Ben Ahmed et al., 2004b].

10.5. Cas d’étude : analyse de l’impact de l’ESP sur le comportement du système CVE

Ce travail a été effectué avec une équipe de deux experts. Il avait pour objectif l’analyse de l’impact de l’ESP (Electronic Stability Programm) sur le comportement du système CVE de point de vue de l’accidentologie. L’ESP est un système qui permet de contrôler la stabilité d’un véhicule. Sa fonction principale est de détecter les survirages ou sous-virages et de corriger la trajectoire du véhicule pour éviter les pertes de contrôle. Nous allons présenter le déroulement du cas d’étude ainsi que les résultats en fonction des différents niveaux d’abstraction de l’architecture ASMEC (cf. section 8.4). Les résultats sont présentés dans la Figure 10-5.

- *Analyse au niveau du méta-modèle* : il s’agit d’identifier quels sont les aspects systémiques qui peuvent être affectés par le changement. Selon les experts, les quatre aspects *ontologique*, *fonctionnel*, *transformationnel* et *téléologique* sont concernés par ce changement ;
- *Analyse au niveau des modèles* : il s’agit de choisir les modèles pertinents au niveau de chacun des points de vue pour l’analyse de l’impact du changement. Les modèles qui ont été choisis sont : le *modèle CVE* (pour le point de vue ontologique), le *modèle séquentiel* (pour les points de vue transformationnel), le *modèle de traitement de l’information* (pour le point de vue fonctionnel) et le *modèle de la tâche de conduite* (pour le point de vue téléologique) ;
- *Analyse au niveau des concepts*: il s’agit de choisir parmi les concepts des modèles choisis ceux qui sont pertinents pour l’étude. Les concepts « *véhicule* », « *interaction conducteur/véhicule* » et « *interaction véhicule/environnement* » ont été sélectionnés dans le modèle CVE. Les concepts « *diagnostic* » et « *action* » ont été sélectionnés dans le modèle de traitement d’information. Les concepts « *urgence* » et « *choc* » ont été sélectionnés dans le modèle séquentiel et les concepts « *navigation* » et « *contrôle* » ont été sélectionnés dans le modèle de la tâche ;
- *Analyse au niveau des attributs* : il s’agit de sélectionner parmi les attributs qui caractérisent les concepts sélectionnés ceux qui sont pertinents. Nous en présentons quelques exemples dans la Figure 10-5.

Aspects systémiques affectés									
Niveau "méta-modèle"	Aspect ontologique			Aspect fonctionnel		Aspect transformationnel		Aspect téléologique	
Modèles pertinents pour l'analyse									
Niveau "modèle"	Modèle "CVE"			Modèle de traitement d'information		Modèle séquentiel		Modèle de la tâche	
Concepts pertinents pour l'analyse									
Niveau "concept"	Conduc./Véh.	Véhicule	Véh./Envir.	Etape de diagnostic	étape d'action	Situation d'urgence	situation de choc	Contrôl latéral	Guidage latéral
Attributs pertinents pour l'analyse									
Niveau "attributs"	*controlabilité véhicule	* interaction ESP/freinage? * interaction ESP/pression des pneu	* Adhérence * Dynamique véhicule en virage	* Perception du danger dans un virage	* Angle volant * Vitesse volant	* Angle volant * Vitesse volant	* nouvelle config. Accid? nouvelle config. De choc?	* Angle volant * Vitesse volant	* Dynamique véhicule en virage * ESP/sortie de voie?

Figure 10-5- Extrait des résultats de l'application d'ASAIC pour l'analyse de l'impact de l'ESP en utilisant

Ce travail nous a permis de déterminer les attributs de l'accident qui sont affectés par l'introduction de l'ESP. Ça nous a permis aussi de dégager des problématiques sous forme de questions que les accidentologues se posent. Nous avons identifié deux types de questions : le premier est relatif à des questions dont la réponse peut être donnée par un tiers, notamment les concepteurs. Exemple : « *interaction ESP/freinage ?* ». Le deuxième type est relatif à des questions qui suggèrent de mener des expérimentations pour trouver des réponses. Exemple : « *influence de l'existence de l'ESP sur la perception du danger par le conducteur dans un virage ?* ».

10.6. ASAIC : vers une approche globale d'analyse de changement

10.6.1 Motivations de l'élaboration d'ASAIC

A travers le modèle que nous avons développé dans la section 10.4, nous fournissons un outil qui permet l'analyse de l'impact d'un changement sur le comportement du système CVE. Cependant, dans la réalité, l'impact n'affecte pas uniquement le système CVE, mais également le processus de conception des véhicules chez les concepteurs ainsi le métier de l'accidentologie au LAB.

Prenons le même exemple de l'introduction d'un nouveau système de sécurité dans le véhicule : ce changement, en plus de son influence sur le comportement du système CVE, a des impacts sur le processus de développement du véhicule ainsi que les ressources mises en œuvre chez les constructeurs automobile. De même, il peut affecter le métier de l'accidentologie lui-même. En effet, le LAB doit développer de nouvelles méthodes et techniques pour : l'évaluation du système, l'identification de son impact sur le système CVE et identifier les nouveaux besoins en terme de sécurité. Par exemple, de nouvelles études sont à mener, de nouvelles variables sont à définir et collecter, etc. Cela peut même affecter l'organisation du LAB : nouvelles ressources mobilisées, nouvelles antennes de collecte, nouvelle organisation de travail, etc.

Dans [Ben Ahmed et al., 2004a], nous avons montré que le modèle intégré produit/processus/ressources qu'on trouve dans la majorité des recherches en IAIC [Gorti et al., 1998; Ma et al., 2003; Murdock et al., 1997; Szykman et al., 1999] se focalise sur le « processus de conception » du produit et néglige celui relatif à son « utilisation » ainsi que son « évaluation ». Nous n'avons pas

trouvé de méthodes permettant l'analyse systématique de ces impacts sur le « processus d'utilisation du produit développé ». Ce dernier correspond dans notre cas au *processus de conduite*. Le modèle que nous avons développé dans la section 10.4 a pour objectif de remédier à cette limite. Nous n'avons pas trouvé non plus de méthodes permettant l'analyse de l'impact d'un changement sur le « processus d'évaluation du produit ». Ce dernier correspond dans notre cas au *métier de l'accidentologie*.

La première motivation de l'élaboration d'un modèle intégré est donc d'identifier les différents points de vue nécessaires à une analyse « complète » d'un changement en sécurité routière. Notre approche ASAIC consiste donc à intégrer :

- Un modèle de l'impact d'un changement sur le système CVE (i.e. *système d'utilisation*). Nous avons développé ce modèle dans la section 10.4 ;
- Un modèle d'analyse de l'impact d'un changement sur le *système de conception* des véhicules ;
- Un modèle d'analyse de l'impact d'un changement sur le métier de l'accidentologie (i.e. *système d'évaluation*) ;

Dans les sections suivantes, nous présentons une brève description des modèles d'AIC sur le système de conception des véhicules ainsi que sur le métier de l'accidentologie. (cf. [Ben Ahmed *et al.*, 2004a] pour une description plus détaillée).

10.6.2 Un modèle d'analyse d'impact de changement sur le système de conception

De nombreux modèles d'AIC en conception existent dans la littérature [Gorti *et al.*, 1998; Ma *et al.*, 2003; Murdock *et al.*, 1997; Szykman *et al.*, 1999]. Nous ne proposons pas de nouveaux modèles, mais nous avons juste choisi un modèle qui nous a paru pertinent pour une activité telle que la conception des véhicules. Il s'agit d'un modèle qui intègre les trois points de vue *produit/processus/ressource* pour effectuer une AIC. C'est le modèle proposé dans [Ma *et al.*, 2003]. Dans le paragraphe suivant, nous montrons comment la vision système reste générique par rapport au modèle choisi.

Une *représentation ontologique* du processus de conception consiste en une représentation des ressources et des structures mises en place. Elle concerne la description des informations sur les personnes, les équipes, les unités organisationnelles, les outils (e.g. logiciels, machines). Une *représentation fonctionnelle* porte sur une description des tâches et des activités de conception en utilisant des techniques « classiques » comme la WBS (Work Breakdown Structure). Une *représentation transformationnelle* du processus de conception consiste à représenter les différents états par lesquels passe l'artefact durant le processus. Des représentations hiérarchiques de la taxonomie du produit comme celles fournies par le *modèle d'information produit* développé dans [Murdock *et al.*, 1997] et [Szykman *et al.*, 1999] peuvent être utilisées. Une *représentation téléologique* d'un processus de conception consiste en une *analyse dirigée par les objectifs* : chacune des composantes du produit, chacune des activités et chacune des unités de ressource est affectée à un objectif donné. Les *matrices de connexion* et les *matrices d'affectation* permettent de représenter ce point de vue. Les *matrices de connexion* représentent les liaisons entre les « composantes du produit » et

« les tâches de conception » : chaque composante est reliée aux tâches qui conduisent à son élaboration. Les *matrices d'affectation* représentent les liens entre les « ressources » et les « tâches de conception » : chaque ressource est affectée aux tâches qui les utilisent. Nous présentons ce modèle dans la Figure 10-6.

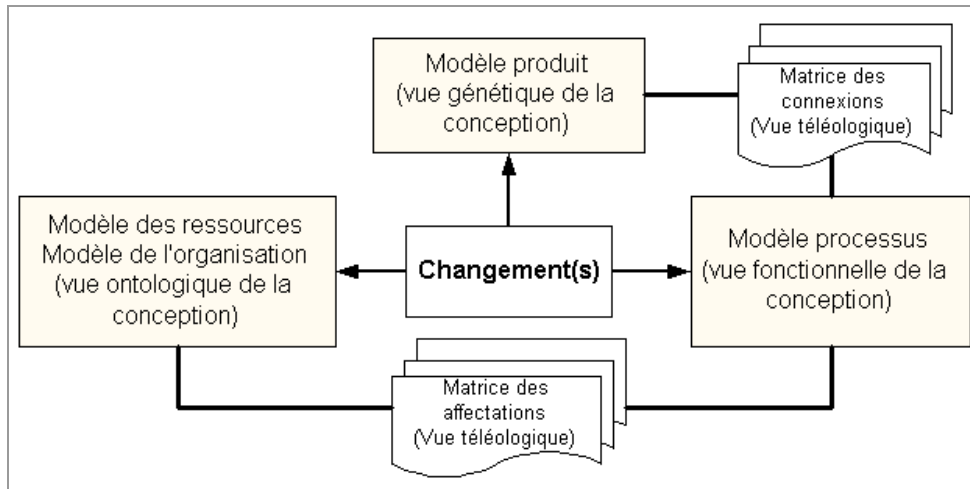


Figure 10-6- Un modèle d'analyse d'impact de changement sur le système de conception

L'utilisation de ce modèle se fait selon le processus décrit dans la section 10.2 et présenté dans la Figure 10-2 : identification du type de changement, évaluation de l'impact du changement sur les différents éléments concernés (produit, ressources, processus). Les matrices de connexion et les matrices d'affectation permettent de propager l'impact, définir des scénarios de changement, évaluer les scénarios, effectuer le changement sur tous les éléments concernés et capitaliser les détails de changement. Une description plus détaillée des différents modèles (produit, processus et ressource) ainsi que l'utilisation dans l'AIC est donnée dans [Ma *et al.*, 2003].

10.6.3 Un modèle d'analyse d'impact de changement sur le métier de l'accidentologie

Une représentation systémique du métier de l'accidentologie au LAB peut être utilisée pour l'analyse de l'impact d'un changement sur ce métier. Une *représentation ontologique* consiste en une représentation de l'organisation du LAB (structures, équipes, personnes, etc.) ainsi que celle des outils utilisés (matériels et immatériels). Une *représentation transformationnelle* consiste en une représentation des différentes possibilités d'évolution du LAB. Une *représentation fonctionnelle* consiste en une représentation du processus de fonctionnement au LAB (collecte de données, expérimentation, études cliniques, etc.). Une *représentation téléologique* consiste en une représentation des objectifs du LAB (stratégiques, tactiques et opérationnels) et en leurs relations avec le processus de fonctionnement ainsi que l'organisation. Ceci peut être effectué grâce à des *matrices d'affectation* (cf. Figure 10-7) qui permettent de : (i) affecter à chaque objectif du LAB les ressources (personne, outils, etc.) et les activités (e.g. collecte de données, analyse de données, etc.) qui permettent d'atteindre cet objectif. (ii) affecter à chaque activité les ressources nécessaires pour son exécution.

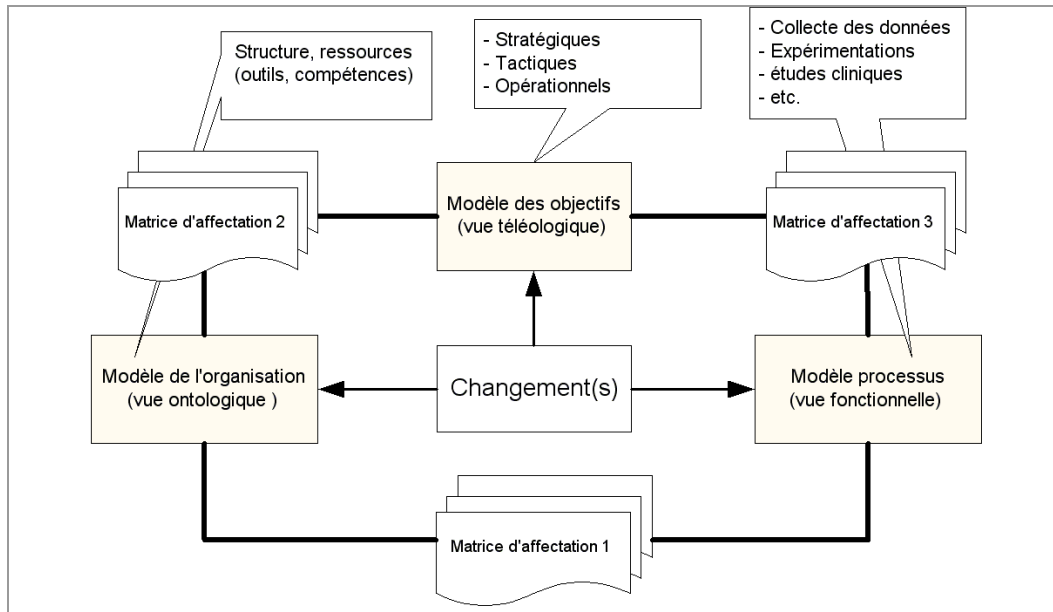


Figure 10-7- Un modèle d'analyse d'impact de changement sur le métier de l'accidentologie

10.6.4 ASAIC : un modèle intégré d'analyse d'impact de changement en sécurité routière

Nous proposons à travers ASAIC une approche d'analyse systémique de l'impact d'un changement en sécurité routière qui se base sur un modèle multi-vues. Ce dernier intègre les modèles présentés dans les sections 10.4, 10.6.2 et 10.6.3. En d'autres termes, il s'agit d'une AIC selon les trois points de vue :

- *Point de vue de l'utilisation* : il permet une AIC sur le système CVE ;
- *Point de vue de la conception* : ce point de vue intègre à son tour trois sous-points de vue : produit/processus/ressource ;
- *Point de vue de l'évaluation* : il permet une AIC sur le métier de l'accidentologie. Ce point de vue intègre, à son tour, trois points de vue : objectif/organisation/processus.

La Figure 10-8 illustre l'architecture d'ASAIC.

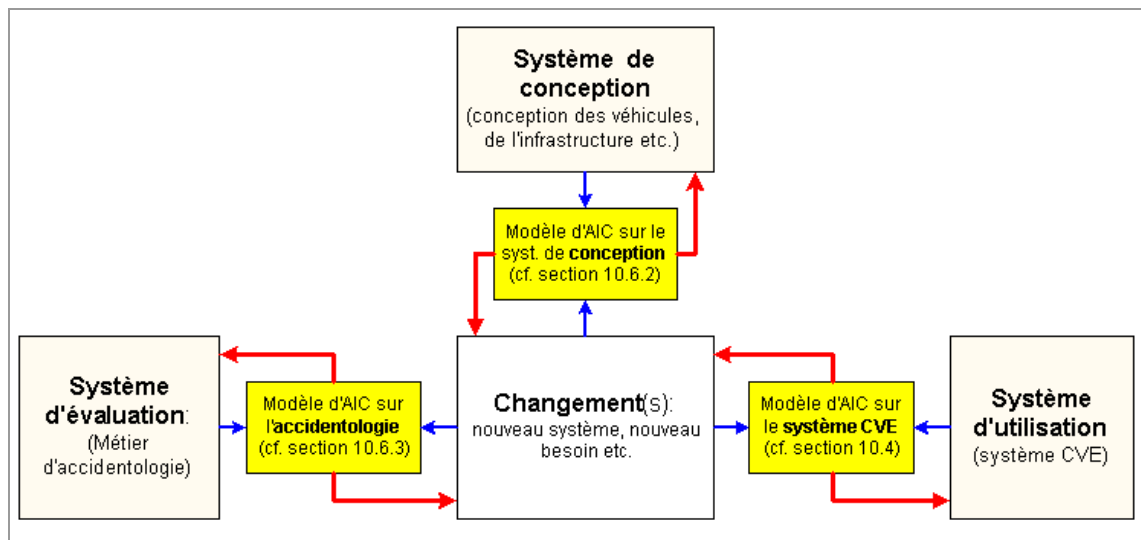


Figure 10-8- Un modèle intégré pour l'AIC en sécurité routière.

Selon la Figure 10-8, un changement au niveau de l'un des trois systèmes (i.e. conception, utilisation et évaluation) peut engendrer des effets sur les deux autres. Un nouveau système de sécurité (i.e. changement chez les constructeurs), par exemple, entraîne des nouveaux besoins en terme de sécurité routière (i.e. changement de comportement du système CVE). Cela entraîne de nouveaux besoins dans le processus de conception et en accidentologie en terme de méthodes d'étude. La décision prise à la suite d'un AIC peut consister à effectuer un nouveau changement qu'il faut analyser par le même processus. L'intégration des trois points de vue que nous proposons permet d'anticiper et contourner les effets pervers. Seul le modèle d'AIC sur le système CVE a été validé à travers un cas d'étude (cf. section 10.5). L'application des deux autres modèles fait partie des perspectives de nos travaux.

10.7. Synthèse

Dans ce chapitre, nous avons présenté la troisième méthode développée dans l'approche SAFE-Next. Il s'agit d'une Approche Systémique d'Analyse de l'Impact d'un Changement (ASAIC). Nous avons commencé par un succinct état de l'art sur les méthodes d'AIC. Nous avons vu que la majorité des méthodes existantes se focalisent sur des objets appartenant au processus de développement d'un produit (e.g. produit, processus, ressource, organisation). Pour remédier à cette limite, nous proposons dans ASAIC d'intégrer le *système de conception*, le *système d'utilisation* et le *système d'évaluation*. Dans notre cas :

- Le *système de conception* correspond au système de conception des véhicules. Ce point de vue intègre trois sous-point de vue : produit/processus/ressource (cf. section 10.4) ;
- Le *système d'utilisation* correspond au système CVE (cf. section 10.6.2) ;
- Le *système d'évaluation* correspond au métier de l'accidentologie au LAB. Ce point de vue intègre trois sous-point de vue : objectifs/organisation/processus (cf. section 10.6.3) ;

Chapitre 11.

ASEM : une Architecture Systémique d’Evaluation de Modèles

ASEM est la quatrième approche que nous développons dans SAFE-Next. Elle permet une *évaluation a priori* (i.e. au cours de la modélisation) et *a posteriori* (après la modélisation) de modèles (cf. Figure 11-1). Elle permet aussi *d’analyser la relation intra-critères d’évaluation de modèles* et *la relation entre les critères d’évaluation de connaissances et les critères d’évaluation de modèles*. ASEM est utilisée pour l’évaluation du modèle de connaissances dans ASMEC. Nous revenons dans la section suivante aux motivations de l’élaboration d’ASEM.

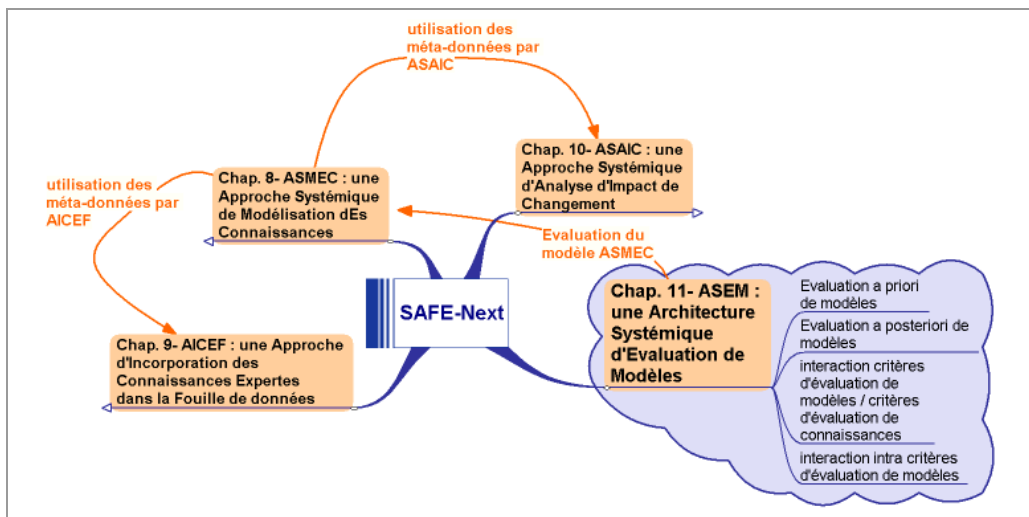


Figure 11-1- ASEM : approche d’évaluation de modèles (quatrième approche dans SAFE-Next)

Comme nous l’avons déjà présenté dans la Définition 5.4, *la modélisation est une action d’élaboration et de construction intentionnelle de modèles susceptibles de rendre intelligible un phénomène perçu complexe et elle est intrinsèque à toute tâche humaine* [Le Moigne, 1999]. Cette définition est en

accord avec la conception constructiviste (cf. section 2.1 et 3.3) de la modélisation que nous avons choisi d'appliquer. Nous comprenons la modélisation comme un processus de construction basé sur une interaction entre le sujet et l'objet à modéliser ainsi que le contexte de modélisation. Cette interaction est contrôlée par des contraintes de modélisation qui sont relatives au système modélisé lui-même, à l'objectif de modélisation et au contexte de modélisation (économique, social, technologique, etc.). Une représentation de la tâche de modélisation selon un formalisme SADT (Structured Analysis and Design Technique) est donnée dans la Figure 11-2.

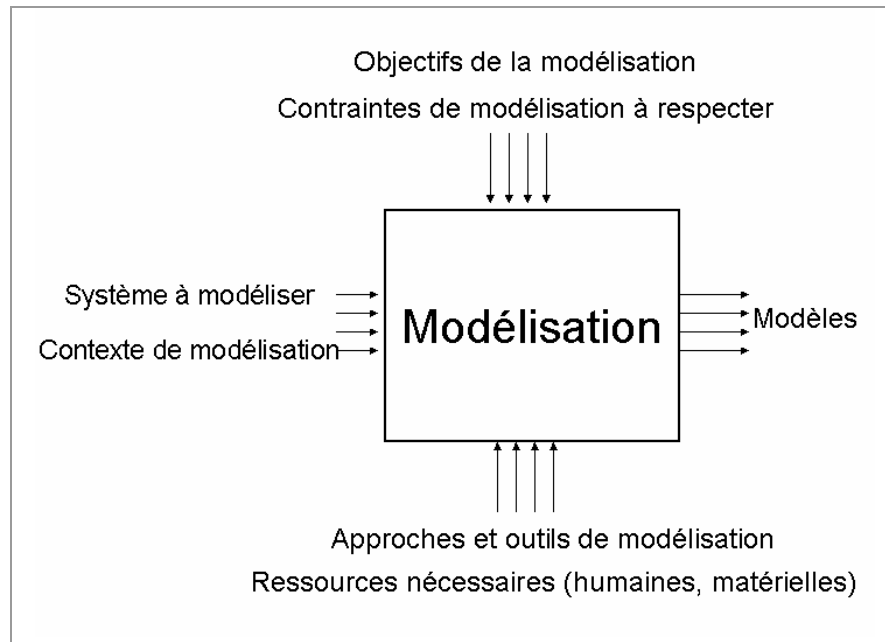


Figure 11-2- La tâche de modélisation

11.1. Motivations de l'élaboration d'ASEM

Conformément au constructivisme, nous pensons que durant notre travail de modélisation notre perception du phénomène modélisé (e.g. l'accident, la connaissance, etc.) peut changer. Donc, les modèles qui en résultent changent et évoluent avant d'atteindre un « état final » qui nous « satisfait ». Mais, puisque la modélisation dépend de son contexte et que ce dernier est en évolution continue, ces notions d'état final et de satisfaction deviennent relatives. Ceci sous-entend l'existence d'un « état cible » et d'une évaluation de l'écart entre l'état du modèle à un moment donnée de sa construction et cet état cible.

Bien que nous présentions ce chapitre après avoir présenté les chapitres concernant la modélisation, nous avons construit ce modèle d'évaluation de modèles parallèlement au processus de modélisation. En effet, nous avons choisi dès le début de développer cette notion d'état cible et de l'exprimer sous forme de critères d'évaluation d'un modèle. Ces critères ont pour objectif de :

- Contrôler, durant le processus de modélisation, le respect de contraintes relatives à l'intégration de l'objectif de l'utilisateur du modèle, l'évolutivité, l'utilisabilité du modèle, etc.

- Sélectionner un modèle existant pour les tâches de *modélisation dirigée par les modèles* (cf. section 8.3) ;
- Valider et évaluer a posteriori des modèles que nous avons construits (cf. Chapitre 10).

Ce que nous proposons donc concrètement, c'est de mener parallèlement au processus de modélisation un processus d'évaluation. Cette évaluation peut être explicite comme dans le cas de choix entre deux alternatives (e.g. choisir entre deux formalismes de représentation : on choisit celui qui répond le mieux aux critères déjà fixés). Mais, elle peut être aussi implicite car nous pensons que rien qu'être conscient des critères qu'il faut remplir, change notre perception et manière d'aborder la modélisation. Cela fait tendre le résultat de notre modélisation vers *l'état cible*. Cette hypothèse est animée par le principe même du constructivisme et particulièrement par ce qui a été montré dans [Henry, 2003; Henry et Mark, 2003; Kirkhart, 2000] : l'influence de l'évaluation sur la perception des évaluateurs ainsi que sur leurs actions.

La deuxième motivation de l'élaboration d'un *modèle d'évaluation de modèle*, vient du constat que nous avons fait : l'évaluation des modèles de connaissances en IC est basée essentiellement sur la simulation de ces modèles en utilisant des langages d'opérationnalisation (cf. section 5.4). La simulation est une sorte d'*évaluation a posteriori*, c'est-à-dire qu'elle n'est applicable que lorsqu'une partie au moins du modèle est achevée. Nous pensons que cette approche peut être complétée par une *évaluation a priori*. Cette dernière doit être menée dans les phases en amont, d'autant plus que dans la réalité le modélisateur a des critères implicites qu'il essaie de respecter lors de la modélisation. Par exemple, le choix de l'approche de modélisation (e.g. ascendante ou descendante) (cf. section 5.3) et le choix du langage de représentation (cf. section 5.4) sont effectués le plus souvent en se basant sur des critères, même si ces critères ne sont pas toujours explicités. Cependant, en génie logiciel, un domaine pourtant très proche de l'IC, on se rend compte rapidement de l'abondance de travaux sur l'élaboration de critères d'évaluation d'un programme ou d'un outil informatique. Ces critères (maintenabilité, robustesse, etc.) ne remplacent en aucun cas l'évaluation par simulation, mais ils permettent de contrôler le processus du développement (pour choisir, par exemple, entre deux solutions alternatives). Ces critères peuvent aussi être utilisés dans une évaluation a posteriori quand il s'agit, par exemple, de comparer deux logiciels. Le besoin de disposer de tels critères en IC est donc certain.

Pour définir ces critères d'évaluation d'un modèle, revenons à sa définition :

Définition 11.1- Un modèle

Un modèle est défini dans une perspective constructiviste (cf. section 2.1 et 3.3) comme une représentation intelligible artificielle et symbolique d'un système ou d'un comportement. Il sert d'interface entre un sujet et un système réel pour la construction de la connaissance.

Selon la Définition 11.1, un modèle n'est pas un objectif en soit, mais un moyen pour construire une connaissance. La question sur l'évaluation d'un modèle peut se traduire en une question sur *l'évaluation de la connaissance*. La Figure 11-3 représente trois niveaux dans une tâche de modélisation : *niveau système*, *niveau modèle* et *niveau connaissance*. Un modèle représente l'interface qu'utilise un sujet pour construire des connaissances sur un système. Un modèle « *adéquat* » est un modèle qui permet de construire une *connaissance « adéquate »*.

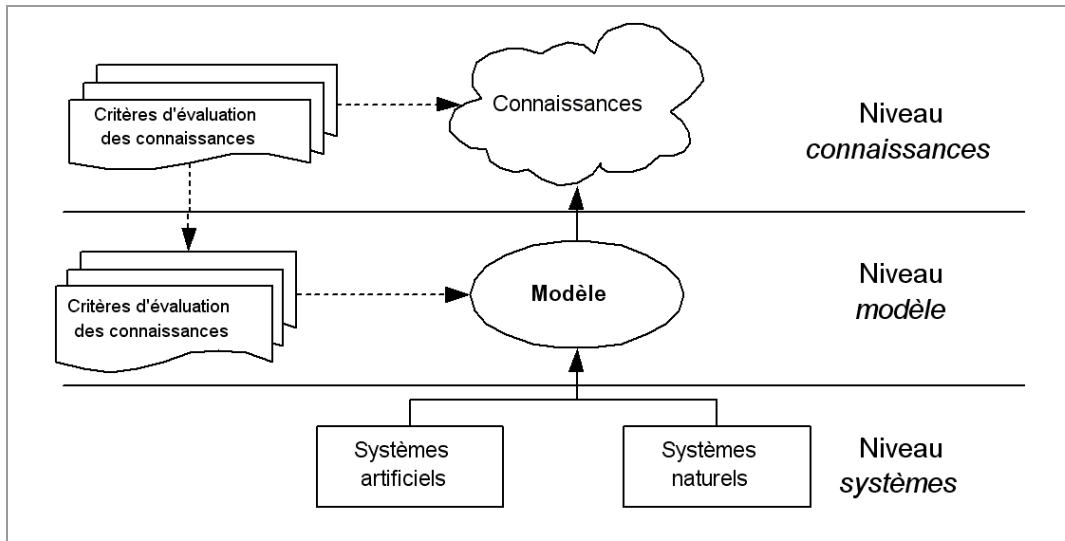


Figure 11-3- Une vue générale sur l'évaluation d'un modèle.

Les critères à vérifier par un modèle sont déclinés des critères qui doivent être vérifiés par la connaissance qu'on veut construire à travers le modèle.

Mais, comme on l'a vu dans la section 3.3, le constructivisme affirme le caractère relatif de la connaissance. Il convient donc de se demander si la notion de *connaissance adéquate* a un sens.

11.2. Critères d'évaluation de connaissances

Selon [Campbell, 1974; Heylighen, 1993,1997], l'une des questions fondamentales auxquelles l'épistémologie essaie de répondre est la suivante : *qu'est ce qui distingue une « vraie » connaissance (ou connaissance adéquate) d'une « fausse » connaissance (ou connaissance inadéquate)*.

Là encore, on retrouve le débat philosophique entre absolutisme et relativisme (cf. section 3.1). Puisque le positivisme insiste sur le caractère absolu, passif et permanent de la connaissance, il affirme la possibilité d'évaluer la connaissance à l'aide de critères objectifs et non ambigus. Cependant, le constructivisme, qui insiste sur le caractère relatif et évolutif de la connaissance développe, en plus de critères objectifs, des critères subjectifs plus contextuels et évolutifs [Campbell, 1974; Heylighen, 1993].

Selon [Dudezert, 2003], la notion de la « valeur des connaissances » reste ambiguë dans la littérature. Les travaux sont orientés vers l'évaluation de l'apport de la gestion des connaissances, l'apport des systèmes d'information, etc. et non le développement de critères d'évaluation des connaissances elles-mêmes. Nous nous intéressons dans ce chapitre aux travaux développés en cybernétique (cf. section 3.4) [Campbell, 1974; Heylighen, 1993,1997]. Pour définir des critères d'évaluation de connaissances, Campbell [Campbell, 1974] utilise dans ce qu'il appelle *épistémologie évolutionnaire (evolutionary epistemology)* le *principe sélectionniste* qui consiste à dire que : *une connaissance « adéquate » est une connaissance qui réussit à survivre, donc à être sélectionnée (utilisée)*.

Utilisant le même principe, Heylighen [Heylighen, 1997] distingue trois classes de *critères de sélection* : (i) des critères *objectifs*, (ii) des critères *subjectifs* et (iii) des critères *inter-subjectifs*.

11.2.1 Les critères objectifs

Ce sont des critères pour juger l'objectivité ou « la réalité » de la connaissance. Ces critères peuvent être utilisés, par exemple, pour tester des hypothèses, vérifier des prédictions, etc. Parmi ces critères, citons [Heylighen, 1997] :

Invariance

Dans la connaissance sur un phénomène donné, il y a une partie « solide/objective » qui persiste même si la perception (i.e. contexte, temps, etc.) évolue. Trois types d'invariance peuvent être distingués. (1) *Invariance par rapport aux modalités* : la perception est la même quelque soit le moyen de perception (e.g. vue, toucher, etc.), le point de vue de la perception, etc. ; (2) *Invariance par rapport au temps* : la perception est la même si elle est effectuée à des moments différents ; (3) *Invariance par rapport aux personnes* : la perception est la même si elle est effectuée par des personnes différentes.

Distinctivité

Une perception « réelle » est celle qui peut être caractérisée en détails, structurée selon un schéma cohérent et distinct. Le rêve, par exemple, est non « réel » car il s'agit d'un ensemble de perceptions floues. Les phénomènes qui ne sont pas distinguables sont non informatifs et donc ne peuvent pas être considérés comme réels.

Contrôlabilité

Une connaissance qui réagit différemment vis-à-vis d'actions différentes qui sont effectuées sur elle a plus de chance d'être réelle comparée à une connaissance qui réagit aléatoirement ou qui ne réagit pas.

11.2.2 Les critères subjectifs

Ils sont relatifs à la capacité de la connaissance à être assimilée par un individu. La théorie quantique, par exemple, malgré son caractère objectif, est assimilée par peu de gens, d'où l'intérêt des critères subjectifs. La capacité cognitive d'un sujet est limitée et l'apprentissage est basé sur l'intensité des liens entre la connaissance en question et des connaissances déjà acquises. La complexité de la connaissance, son conflit avec les connaissances existantes, etc. risquent de surcharger le sujet et réduire la chance de cette connaissance d'être sélectionnée et donc de survivre. Les critères subjectifs développés dans [Heylighen, 1997] sont :

Utilité individuelle

Elle postule que le sujet va uniquement faire l'effort d'apprendre et de mémoriser une idée qui peut l'aider à atteindre son objectif.

Simplicité

Plus une idée est complexe, plus la charge cognitive pour son assimilation est importante et donc plus sa chance d'être sélectionnée est faible.

Cohérence

La facilité avec laquelle un système cognitif arrive à assimiler une nouvelle idée dépend du support qu'elle reçoit des idées qui sont déjà assimilées. En d'autres termes, les idées qui ne sont pas connectées à des idées déjà existantes sont difficilement assimilables et donc difficilement sélectionnées.

Nouveauté (et curiosité)

Les idées nouvelles et inattendues tendent à attirer les attentions et donc stimule l'énergie cognitive ce qui facilite leur assimilation.

11.2.3 Les critères inter-subjectifs

Ils sont basés sur le fait que les idées sont facilement assimilables quand elles font l'objet d'une interaction entre plusieurs individus (inter-subjectivité). Parmi ces critères, on cite [Heylighen, 1997] :

Publicité (ou propagande)

Elle est relative à l'intensité des efforts que les sujets ont investis pour diffuser l'idée et la rendre connue.

Expressivité

Elle dépend de la capacité de la connaissance à être exprimée dans un langage clair et facile à transmettre et à assimiler.

Formalité

C'est la capacité d'une idée à être formulée indépendamment du contexte ce qui réduit son ambiguïté. Plusieurs individus peuvent donc la comprendre de la même manière. Cette idée est facile à transmettre et donc à survivre.

Utilité collective

Certaines connaissances peuvent être utiles pour la collectivité même si elles ne le sont pas pour un individu isolé. Les standards techniques, code moral, etc. sont des exemples qui ont de la valeur pour la collectivité et elles sont sélectionnées au niveau des groupes.

Autorité

Les connaissances qui sont issues de personnes qui sont reconnues comme étant des experts ou qui sont des leaders sont plus facilement acceptées par la communauté.

11.3. Critères d'évaluation d'un modèle

Comme nous l'avons déjà noté, les critères d'évaluation d'un modèle sont les critères qui permettent au modèle de générer des connaissances « adéquates » i.e. qui satisfont les critères que nous avons présentés dans la section 11.2.

Mais, comment peut-on s'assurer de la complétude des critères que nous allons développer ? Pour cela, nous utilisons l'approche systémique en assimilant le modèle à un système qui a une *ontologie* et une *téléologie*, qui *fonctionne* et *évolue*. Ainsi, l'évaluation d'un modèle est déclinée sur les quatre axes systémiques permettant de satisfaire certains critères au niveau de la connaissance qui en résulte. Nous nous sommes inspirés de l'évaluation en génie logiciel pour définir certains critères.

Comme illustration, nous prenons le modèle de l'architecture SAFE-Next présentée dans la section 8.4. Nous rappelons que les critères que nous développons ci-dessous ont été développés parallèlement à SAFE-Next et qu'ils ont influencé (directement ou indirectement) la phase de modélisation.

11.3.1 Evaluation de l’ontologie d’un modèle

Définition 11.2- Ontologie d’un modèle

Elle est constituée de ses différentes entités ainsi que du langage (i.e. le formalisme (e.g. UML, graphes conceptuels, XML, etc.), vocabulaire, grammaire, etc.) utilisé pour décrire ces entités.

Le Tableau 11.1 présente les critères que nous avons définis.

Critères d’évaluation de l’ontologie : <i>Critères relatifs aux composantes du modèle</i>	Cas de SAFE-Next
Cohérence : degré d’uniformité, de standardisation du formalisme de représentation du modèle et l’absence de contradictions entre ses différentes parties.	Choix de XML, UML, CML2
Conformité : degré de respect à des standards, conventions, réglementations, théories, etc.	Choix de XML, UML, CML2
Auto-descriptivité : capacité de l’ontologie du modèle de contenir des informations qui expliquent le fonctionnement ainsi que les objectifs du modèle.	Choix de XML, UML, CML2
Indépendance du domaine : degré d’indépendance du modèle du domaine et du contexte de son développement (discipline, modélisateur, objectif, etc.).	Généricité de l’architecture ASMEC
Complétude : degré de complétude des différentes parties du modèle	Choix de la systémique
Interdépendances des composantes : lien entre les différentes composantes du modèle	Généricité de l’architecture ASMEC

Tableau 11.1- Critère d’évaluation de l’ontologie d’un modèle
(Critères relatifs aux composantes du modèle)

Le choix des langages XML (cf. section 8.8), UML (cf. section 8.10) et CML2 (cf. Chapitre 7 pour représenter respectivement les métadonnées), l’architecture de SAFE-Next et les différents types de connaissances était animé par un souci de *consistance*, de *conformité* et d’*auto-descriptivité*. En effet, en plus du degré d’uniformité, de standardisation de ces langages, ils sont sémantiquement riches au point de permettre la compréhension des modèles uniquement à travers leur représentation (exemples : notions de hiérarchie dans XML, classe dans UML, etc.).

D’un autre côté, l’architecture d’ASMEC (cf. section 8.4) assure une certaine indépendance de toute l’approche par rapport au domaine de l’accidentologie. Car, il s’agit d’une architecture de modélisation dirigée par un méta-modèle général à plusieurs niveaux d’abstraction (méta-modèle, modèle, concept, etc.) qui sont tous indépendants du domaine de développement (i.e. l’accidentologie). L’approche systémique confère à SAFE-Next une certaine complétude au niveau de la représentation d’un phénomène complexe comme l’accident.

Dans l’ontologie du modèle nous intégrons aussi son *interface avec l’utilisateur*. Les critères d’évaluation de cette interface sont présentés dans le Tableau 11.2.

Critères d'évaluation de l'ontologie <i>Critères relatifs à l'interface modèle-utilisateur</i>	Cas de SAFE-Next
Utilisabilité : facilité avec laquelle un utilisateur peut utiliser un modèle (préparer les entrées, simuler, interpréter les résultats, etc.)	Implication des experts dans la conception des interfaces
Attractivité : capacité des attributs du modèle (relatifs à la forme et aux fonctions) à le rendre plus attractif vis-à-vis de l'utilisateur.	Implication des experts dans la conception des interfaces
Compréhensibilité : capacité du modèle à faciliter pour l'utilisateur de savoir son adéquation à un problème donné. La documentation d'un modèle, par exemple, améliore ce critère.	Indépendance du domaine & modélisation au niveau cognitif
Apprentissage : capacité du modèle de permettre à l'utilisateur d'apprendre rapidement son utilisation.	Implication des experts dans la conception des interfaces
Testabilité : facilité de validation du modèle.	Implication des experts dans la conception des interfaces
Opérabilité : capacité de permettre à l'utilisateur sa manipulation et son contrôle.	Implication des experts dans la conception des interfaces

Tableau 11.2- *Critères d'évaluation de l'ontologie d'un modèle*
 (Critères relatifs à l'interface modèle-utilisateur)

Les experts ont été impliqués tout au long de la phase de conception des interfaces d'utilisation de SAFE-Next et surtout celles relatives à l'interprétation des résultats (cf. Figure 9-9 dans la section 9.2.1). Cela nous a permis d'identifier et de faire figurer sur ces interfaces les informations principales dont l'expert a besoin : différents points de vue, critères de pertinences des modalités, poids de la classe dans l'échantillon, description textuelle des trois accidents les plus centrés dans la classe, lien hypertexte vers la restitution cinématique des trois accidents les plus centrés, etc. (Figure 9-9).

11.3.2 Evaluation du fonctionnement d’un modèle

Définition 11.3- Fonctionnement d’un modèle

Il décrit son interaction avec son environnement pour satisfaire les objectifs de l'utilisateur. Il est aussi relatif au comportement du modèle sous des contraintes sévères (données manquantes, fausse manipulation, etc.).

Nous présentons des critères d’évaluation du fonctionnement d’un modèle dans le Tableau 11.3.

Critères d’évaluation du fonctionnement	Cas de SAFE-Next
Robustesse : capacité du modèle à fonctionner dans des conditions contraignantes (entrées invalides, perturbation, etc.). Ce critère se décline en plusieurs sous critères : tolérance d’erreur, tolérance de défaillance, fiabilité, prédisposition à l’erreur	Possibilité de modification des métadonnées & Contrôle total par l’expert
Tolérance d’erreur : capacité du modèle à travailler normalement malgré la présence d’entrée erronée.	L’intégration de plusieurs points de vue lors de l’élaboration des métadonnées réduit le risque d’erreur.
Tolérance de défaillance (ou stabilité ou survivabilité) : capacité du modèle à travailler normalement malgré la défaillance de l’un de ses composants.	Les approches intégrées dans SAFE-Next (i.e. ASMEC, AICEF, ASAIC et ASEM) sont indépendantes
Prédisposition à l’erreur : capacité du modèle à permettre à l’utilisateur d’introduire intentionnellement ou par mauvais usage des erreurs.	Possibilité pour un utilisateur d’élaborer, en utilisant SAFE-Next, ses propres modèles.
Contrôlabilité : capacité du modèle à réagir différemment dans des situations différentes.	
Généralité : capacité de réaliser plusieurs fonctions.	Généralité de SAFE-Next (cf. section 12.2)

Tableau 11.3- Critères d’évaluation du fonctionnement d’un modèle

Les métadonnées représentent dans SAFE-Next l’objet central. En effet, les tâches de sélection multi-vues des attributs (cf. section 9.1) ainsi que d’interprétation multi-vues des résultats (cf. section 9.2) sont essentiellement basées sur les métadonnées. Des erreurs au niveau des métadonnées induisent donc des erreurs dans les deux tâches. Pour contourner ce problème, SAFE-Next n’est pas un outil de résolution automatique de problème, mais une méthode interactive où l’expert garde le rôle central. Nous avons fait de sorte qu’à tout moment l’expert puisse personnaliser ces métadonnées en fonction de sa perception. Il suffit de changer l’affectation des attributs en fonction des concepts. En ce qui concerne la généralité de SAFE-Next, elle est due au fait qu’elle puisse être utilisée pour la modélisation des connaissances, la sélection d’attributs et l’interprétation des classes. D’autres fonctionnalités de SAFE-Next (capitalisation des connaissances, formation de nouveaux opérateurs, analyse d’impact de changement, etc.) seront développées dans la section 12.2.1.

11.3.3 Evaluation de l'évolutivité d'un modèle

Définition 11.4- L'évolution d'un modèle

Elle est caractérisée par les transformations (structurelles ou fonctionnelles) d'un modèle dues à des changements internes (e.g. nouvelle composante ajoutée au modèle) ou externes (e.g. nouvelle application). L'évolutivité quant à elle, caractérise la capacité d'un modèle à s'adapter à ces transformations internes et externes.

Critères d'évaluation du l'évolutivité d'un modèle	Cas de SAFE-Next
Nouveauté : la nouveauté d'un modèle tend à attirer les attentions et donc stimule l'énergie cognitive ce qui facilite son assimilation	Apports de SAFE-Next en ECD (cf. section 12.2.3)
Modifiabilité : capacité d'un modèle à être modifié. Ce critère dépend des critères suivants : flexibilité, extensibilité, maintenabilité, testabilité.	Modifiabilité des métadonnées
Flexibilité : facilité de sa modification pour l'utiliser dans d'autres contextes (i.e. environnements, objectifs, etc.).	Apports de SAFE-Next en KM (cf. section 12.2.1)
Extensibilité : facilité de modifier le modèle pour améliorer son fonctionnement.	Modifiabilité des métadonnées
Maintenabilité : facilité de modification du modèle pour le corriger.	Modifiabilité des métadonnées
Remplaçabilité : capacité à être remplacé par un autre modèle ou remplacer l'une de ces composantes.	Interaction de SAFE-Next avec CommonKADS, KOD, etc.
Adaptabilité : capacité à s'adapter à de nouveaux besoins pour de nouveaux utilisateurs, donc à un nouveau contexte, sans avoir besoin de le modifier.	Apports de SAFE-Next en KM et en conception (cf. sections 12.2.1 et 12.2.4)
Réutilisabilité : la capacité d'un modèle à permettre la réutilisation de l'un de ses composants pour d'autres objectifs.	Apports de SAFE-Next en KM et en conception (cf. sections 12.2.1 et 12.2.4)

Tableau 11.4- Critères d'évaluation de l'évolutivité d'un modèle

Comme nous l'avons expliqué dans la section précédente, la possibilité de modifier, voire de personnaliser les métadonnées confère à SAFE-Next la modifiabilité, l'extensibilité et la maintenabilité. En ce qui concerne les autres critères, nous montrons dans les sections 12.2.1 et 12.2.4 les différentes applications possibles de SAFE-Next (exemple : en gestion des connaissances et en analyse d'impact de changement) montrant son adaptabilité et réutilisabilité pour des objectifs autres que ceux qui étaient à l'origine de sa création.

11.3.4 Evaluation de la téléologie d'un modèle

Définition 11.5- La téléologie d'un modèle

C'est l'objectif de la construction du modèle. L'évaluation de la téléologie consiste à évaluer le décalage entre les objectifs fixés et ceux atteints.

Critères d'évaluation de la téléologie d'un modèle	Cas de SAFE-Next
Précision : capacité du modèle à fournir des réponses à un degré de précision fixé.	
Efficacité : capacité de modèle à fournir, sous certaines conditions, une réponse performante relativement aux ressources utilisées	
Effectivité : à quel degré le modèle répond aux besoins de l'utilisateur ?	Implication de l'expert

Tableau 11.5- Critères d'évaluation de la téléologie d'un modèle

L'implication de l'expert, que ce soit dans l'élaboration des métadonnées (cf. section 8.8) ou dans la phase de conception des interfaces d'utilisation (cf. section 9.2.1.2 et Figure 9-9) fait que SAFE-Next est *effective* : les solutions qu'elle apporte sont proches du besoin des experts. Elle permet aussi d'améliorer la *précision* et l'*efficacité*. En effet, face à l'abondance des informations dans les EDA, elle permet d'effectuer un filtrage et éviter ainsi le problème fréquent qu'est la surcharge d'informations. L'utilisateur peut concentrer ses efforts sur les informations pertinentes. L'automatisation de l'interprétation multi-vues des classes réduit considérablement le temps consacré d'habitude à cette tâche et améliore sa qualité.

11.4. Relation critères d'évaluation d'un modèle /critères d'évaluation de connaissances

Puisqu'un modèle n'est pas un objectif en soit, mais un moyen pour construire des connaissances, un « *bon* » modèle est celui qui permet de construire des connaissances « *adéquates* ». En d'autres termes, les critères d'évaluation d'un modèle (cf. section 11.3) permettent de remplir les critères d'évaluation des connaissances (cf. section 11.2).

Il convient donc de trouver des liens entre les différents types de critères concernant un modèle et les connaissances recherchées. Nous n'avons pas trouvé dans la littérature des travaux traitant cette problématique. En se basant uniquement sur notre modeste expérience, nous avons essayé de trouver ces liens que nous présentons dans le *Tableau 11.6*.

Nous proposons dans ce tableau l'approche plus que le contenu car ce dernier est basé uniquement sur notre observation et n'a pas été l'objet d'une validation élargie. Nous envisageons de proposer ce tableau ainsi que les différentes définitions des critères (cf. sections 11.2 et 11.3) à plusieurs personnes (chercheurs et industriels) pour voir si nous pouvons généraliser les résultats. Nous voyons ce tableau comme un point de départ pour créer une architecture évolutive et adaptable selon les domaines d'application.

11.5. Relation intra-critères d'évaluation d'un modèle

Nous sommes partis de l'hypothèse qui affirme l'interdépendance entre les critères d'évaluation d'un modèle : la modification d'un critère peut entraîner la modification d'autres. Là encore, nous n'avons pas trouvé de travaux étudiant cette problématique. Comme dans la section précédente, nous proposons un tableau dont les résultats sont à prendre avec précaution car ils n'ont pas encore fait l'objet d'une validation élargie.

Le Tableau 11.6 peut être lu dans les deux sens : horizontal et vertical. Dans le sens vertical, prenons l’exemple du critère « *invariance* » des connaissances (première colonne) : **pour améliorer ce critère (i.e. signe +)**, on peut **améliorer** les critères « *consistance*, *auto-descriptivité*, *indépendance du domaine*, etc. » et/ou **dégrader (signe -)** le critère de « *complétude* » du modèle. Dans le sens horizontal, prenons l’exemple du critère « *indépendance du modèle de son domaine* » (troisième ligne) : **l’amélioration de ce critère peut entraîner au niveau de la connaissance l’amélioration** des critères « *invariance*, *simplicité*, *cohérence*, etc. » et/ou la **dégradation** des critères « *distinctivité*, *contrôlabilité*, *formalité*, etc. »

				Critères d'évaluation des connaissances													
				Critères objectif			Critères subjectifs				Critères inter-subjectifs						
				Invariance	Distinctivité	Contrôlabilité	Utilité individuelle	Simplicité	Cohérence	Nouveauté	Publicité	Expressivité	Formalité	Utilité collective	Autorité		
Critères d'évaluation de modèles	évaluation de l'ontologie	Formalisme de représentation	Consistance	+				+	+			+	+	+			
			Auto-descriptivité	+	+			+				+	+				
			Indépend_domaine	+	-	-	-	+	+			+	+	-	+	-	
			Attractivité					+				+			+		
			Complétude	-	+	+	+	-	-			+		-			
			Conformité	+	+			+	+			+	+	+	+		
			Interdép_composants	+	+	-	+	+	-								
	Interface Modèle/Utilisateur	Utilisabilité	+			+					+		+	+			
		Réutilisabilité	+			+					+		+	+			
		Compréhensibilité	+			+	+				+		+	+			
		Apprentissage	+			+	+				+		+	+			
		Testabilité	+			+	+				+		+	+			
		Opérabilité	+			+	+				+		+	+			
		Evaluation du fonctionnement	Robustesse	Tolérance d'erreur	+	-	-	+						+	+		
	Tolérance défaillance			+	+	+	+						+	+			
	Stabilité			+	+	+	+							+	+		
	Prédisposition à l'erreur			+	-	-	+							+	+		
	Contrôlabilité			+	+	+	+										
	Evaluation de l'évolutivité	Modifiabilité	Généralité		-	-	-	+				+		+	+		
			Nouveauté								+	+		+	+	+	
Flexibilité				-	-	-					+		+	+			
Maintenabilité													+	+			
Extensibilité				-	-	-					+		+	+			
Adaptabilité				-	-	-	+				+		+	+			
Evaluation de la téléologie		Remplaçabilité		-	-	-	+							+			
		Précision	-	+	+	+				+		-					
		Efficacité				+				+			+				
		Effectivité				+				+			+				

Tableau 11.6 - Relations entre les critères d'évaluation d'un modèle (liste à gauche) et les critères d'évaluation des connaissances (liste du haut) (+ : amélioration d'un critère, - : dégradation d'un critère).

Comme dans le Tableau 11.6, le Tableau 11.7 peut être lu dans les deux sens horizontal et vertical. Dans le sens vertical, prenons l'exemple du critère « *flexibilité* » d'un modèle (19^{ème} colonne) : **pour améliorer ce critère (signe +)**, on peut **améliorer** les critères « *consistance, indépendance du domaine, conformité, etc.* » et/ou **dégrader (signe -)** les critères « *complétude, interdépendance des composantes, etc.* » du modèle (voir les définitions dans le **Tableau 11.1**). Dans le sens horizontal, prenons l'exemple du même critère « *flexibilité* » (19^{ème} ligne) : l'amélioration de ce critère peut entraîner l'amélioration des critères « *attractivité, utilisabilité, réutilisabilité, etc.* » et/ou la dégradation des critères « *contrôlabilité, précision, etc.* »

		Evaluation de l'ontologie d'un modèle														Evaluation du fonctionnement					Evaluation de l'évolutivité					Evaluation de la téléologie			
		Formalisme de représentation							Interface Modèle/Utilisateur							Robustesse					Modifiabilité								
		Consistance	Auto-descriptivité	Indépend_domaine	Attractivité	Complétude	Conformité	Interdép_compos	Utilisabilité	Réutilisabilité	Compréhensibilité	Apprentissage	Testabilité	Opérabilité	Tolérance d'erreur	Tolérance de défaillance	Stabilité	Prédisposition à l'erreur	Contrôlabilité	Généralité	Flexibilité	Maintenabilité	Extensibilité	Adaptabilité	Remplaçabilité	Précision	Efficacité	Effectivité	
évaluation de l'ontologie	Formalisme de représentation	Consistance	+	+	+		+		+	+	+		+							+	+	+	+						
		Auto-descriptivité		+	+				+	+	+	+		+						+	+	+	+	+					
		Indépend_domaine		+			-		+	+	+	+								+	+	+	+	+			-	-	-
		Attractivité									+	+							+										
		Complétude			-	+			+	+	+	+							+								+	-	+
		Conformité		+	+	-			+	+	+	+		+								+	+	+	+				
évaluation de l'ontologie	Interface Modèle/Utilisateur	Utilisabilité								+	+			+	+														
		Réutilisabilité												+	+														
		Compréhensibilité												+	+					+	+	+	+	+					
		Apprentissage												+	+														
		Testabilité												+	+														+
		Opérabilité												+	+							+	+	+	+				
Evaluation du fonctionnement	Robustesse	Tolérance d'erreur																											
		Tolérance de défaillance																											
		Stabilité																											
		Prédisposition à l'erreur																											
		Contrôlabilité																											
		Généralité																											
Evaluation de l'évolutivité	Modifiabilité	Flexibilité																											
		Maintenabilité																											
		Extensibilité																											
		Adaptabilité																											
		Remplaçabilité																											
Evaluation de la téléologie		Précision																											
		Efficacité																											
		Effectivité																											

Tableau 11.7 - Relations intra-critères d'évaluation d'un modèle. L'amélioration (signe +) d'un critère, peut engendrer l'amélioration ou la dégradation (signe -) d'un autre

11.6. Synthèse

Ce chapitre à un double objectif :

- Développer un *modèle d'évaluation de modèles* : cela a été motivé par notre constat de la rareté (voire l'absence) de travaux sur l'évaluation de modèles en IC, autres que ceux se basant sur la simulation. Cela nous a paru contradictoire avec le fait que l'IC soit fondée essentiellement sur la modélisation (cf. Définition 2.5). La deuxième motivation est relative aux travaux sur la théorie de l'évaluation qui ont montré que la perception de l'objet évalué, le modèle en l'occurrence, peut changer au cours de l'évaluation. Pour élaborer ce modèle, nous sommes partis de la perception constructiviste d'un modèle qui le conçoit comme une interface entre un sujet et le monde réel pour construire des connaissances (cf. Définition 11.1 et Figure 11-3). Nous avons commencé alors par faire une revue bibliographique des critères que doit remplir une connaissance « adéquate ». Selon la vision *sélectionniste* de l'*épistémologie évolutionnaire* de Campbell [Campbell, 1974], *une connaissance « adéquate » est une connaissance qui réussit à survivre, donc à être sélectionnée*. Ces critères ont été ensuite déclinés en critères d'évaluation de modèle. Nous avons proposé ensuite (sections 11.4 et 11.5) des relations entre les critères d'évaluation d'un modèle et ceux d'évaluation de connaissances ainsi que des relations intra-critères d'évaluation d'un modèle ;
- Fournir *un moyen de validation a priori de SAFE-Next* : nous avons choisi dès le départ de mener parallèlement au processus de modélisation et d'élaboration de SAFE-Next un processus d'évaluation. Ce choix est justifié par ce qui a été montré dans [Henry, 2003; Henry et Mark, 2003; Kirkhart, 2000] sur l'influence de l'évaluation sur la perception des évaluateurs ainsi que sur leurs actions. Les critères que nous avons développés tout au long des travaux présentés dans ce mémoire nous ont servi en quelque sorte de contraintes de modélisation et critères de choix (entre des techniques, des méthodes, des langages, etc.) quand il y a eu plusieurs alternatives.

Chapitre 12.

Validation et apports de SAFE-Next

12.1. Démarche de validation

Bien que la *tâche de validation* soit souvent présentée à la fin du processus de conception de l'objet à valider (i.e. méthode, système, produit, etc.), nous ne la concevons pas comme une étape qu'on effectue à la fin. Nous la considérons comme un processus qui se construit et s'effectue parallèlement au processus de construction de l'objet à évaluer lui-même.

12.1.1 Points à valider

Les différents points à valider dans nos travaux sont :

- Les apports de SAFE-Next en Gestion des Connaissances en tant qu'approche de capitalisation des connaissances, création de nouvelles connaissances, création de culture partagée et formation de nouveaux opérateurs ;
- Les apports de SAFE-Next en IC, grâce à ASMEC, en tant qu'approche de modélisation multi-vues des connaissances en utilisant une *approche mixte* : *descendante* (dirigée-par-un-méta-modèle) et *ascendante* (dirigée-par-les-données).
- Les apports de SAFE-Next, grâce à ASEM, en tant qu'approche d'évaluation de modèles de connaissances ;
- Les apports de SAFE-Next en ECD, grâce à AICEF, en tant qu'approche proposant l'utilisation des techniques de méta-données pour l'incorporation des connaissances du domaine dans le processus d'ECD ;
- Les apports de SAFE-Next en conception de nouveaux produits, grâce à ASAIC, en tant qu'approche de : (i) construction d'un espace de conception et (ii) d'analyse d'impact des

changements pour anticiper les conséquences de l'utilisation d'un nouveau produit (e.g. système de sécurité embarqué) sur l'environnement de son utilisation (e.g. système CVE).

12.1.2 Méthode de validation

Nous avons effectué deux types de validation de SAFE-Next : *validation a priori* et *validation a posteriori*.

12.1.2.1 Validation a priori

Les *critères d'évaluation de modèles* développés dans ASEM (cf. Chapitre 11) tels que la *consistance*, *l'évolutivité*, *l'utilisabilité*, la *robustesse*, etc. nous ont servi de contraintes à respecter durant le processus d'élaboration et d'implémentation de SAFE-Next. Ces critères ont influencé les choix méthodologiques et techniques que nous avons effectués. La validation de SAFE-Next selon le modèle ASEM a été effectuée en même temps que l'élaboration de ce modèle dans le Chapitre 11 ;

12.1.2.2 Validation a posteriori

Deux méthodes de validation ont été effectuées dans ce cadre :

- La première a consisté en la *validation du modèle multi-niveaux d'abstraction et multi-vues de connaissances*, ASMEC, en l'appliquant au domaine de l'accidentologie. Cela a été réalisé à travers des réunions de travail avec les deux équipes d'experts à Amiens et à Evreux¹. Nous avons déjà décrit le déroulement de ces travaux de validation dans les sections 8.6, 8.7 et 8.8 : l'instanciation des différents points de vue du modèle de connaissances d'ASMEC (cf. section 8.4) par des *modèles*, des *concepts* et des *attributs* du domaine de l'accidentologie. Cela nous a permis d'éprouver d'un côté *l'acceptabilité* de SAFE-Next vis-à-vis des experts et d'un autre côté son *applicabilité* sur des *données réelles*. Nous avons noté un ensemble d'apports de notre approche que nous développons dans la section 12.2.1 ;
- La deuxième validation a consisté en trois cas d'études :
 - L'objectif du premier était d'évaluer *l'apport de SAFE-Next dans la phase de sélection des attributs*. Nous sommes partis d'une étude existante [Riviere, 2003] que nous avons refaite en utilisant AICEF (cf. section 9.1). Dans [Riviere, 2003], 25 attributs ont été sélectionnés en utilisant l'expertise. En utilisant AICEF, nous avons effectué une projection de ces attributs sur nos métadonnées multi-vues. Cela nous a permis de montrer que non seulement la qualité de la sélection est améliorée grâce à la sélection multi-vues des attributs, mais également que le coût (en terme de temps) était nettement réduit grâce à l'interface de sélection (cf. Figure 9-4). Effectuer manuellement la même sélection multi-vues est une tâche difficile, voire impossible ; les résultats de ce cas d'étude sont présentés à la fin de la section 9.1. Nous en faisons une synthèse dans la section 12.2.3 ;

¹ voir la section 4.2

- Le deuxième cas d'étude avait pour objectif l'évaluation de l'apport de SAFE-Next dans la *phase d'interprétation des classes*. Nous sommes partis aussi des résultats d'une étude faite au LAB dans le cadre d'un projet de fin d'études d'ingénieurs [Nasri, 2003] que nous avons encadré. Nous avons pris les classes obtenues et les avons projetées sur les méta-données. Nous avons montré dans la section 9.2 l'apport des différentes interprétations (intra et inter classe) multi-vues et se situant sur plusieurs niveaux d'abstraction (accéder plus facilement à des informations pertinentes, réduire le coût de l'interprétation, etc.). Nous en faisons une synthèse dans la section 12.2.3.
- le dernier cas d'étude nous a permis d'évaluer l'utilisabilité de SAFE-Next pour d'autres besoins qui n'ont pas été prévu. Il s'agit de son utilisation pour évaluer l'impact d'un nouveau système de sécurité sur le comportement du système CVE. Nous développons ce point dans la section 12.2.4.2.

D'autres apports de SAFE-Next ont été constatés sans faire l'objet d'une étude de validation particulière. En effet, nous avons remarqué que SAFE-Next a eu des effets annexes qui n'ont pas été prévus tels que : la *mutualisation des connaissances* entre les experts, la *capitalisation de connaissances*, *formation de nouveaux opérateurs*, etc. Nous détaillons ces effets dans la section 12.2.1.

12.2. Les apports de SAFE-Next

Les apports de SAFE-Next se situent à plusieurs niveaux et peuvent être présentés selon différents points de vue. Nous avons choisi de les présenter selon les différentes disciplines par rapport auxquelles nous avons positionné nos travaux au début de ce mémoire (cf. Chapitre 2), c'est-à-dire : le KM, l'IC, l'ECD et la conception de nouveaux produits. Ces apports ont été validés dans le domaine de l'accidentologie, mais comme nous allons le montrer ci-dessous, ils sont généralisables à d'autres domaines.

12.2.1 Apports de SAFE-Next en KM

12.2.1.1 SAFE-Next, une approche de capitalisation des connaissances

En terme de capitalisation des connaissances en accidentologie, SAFE-Next assure les points suivants :

- Elle formalise et structure les modèles les plus connus dans le domaine de l'accidentologie (CVE, séquentiel, traitement d'information, etc.) (cf. §8.6.2.3). Elle facilite leur utilisation dans des tâches d'ECD grâce aux métadonnées (cf. les sections 9.1 et 9.2) ou dans des tâche d'analyse d'impact de changement (cf. les sections 10.4 et 10.5) ;
- Elle a enrichi les points de vue existants par de nouveaux modèles tels que « *système de sécurité* » et « *véhicule orienté sécurité* » (cf. section 8.6.2) ;
- Elle capte, à travers les métadonnées, des connaissances sur chacun des attributs utilisés en accidentologie. Ces connaissances portent sur les différentes classifications de chacun des attributs selon les différents points de vue (cf. section 8.8). l'utilisation des métadonnées a

donc permis d'accéder à des connaissances implicites et a rendu leur formalisation plus facile.

Au démarrage de nos travaux, il y avait 6 experts (2 équipes composées de 3 experts chacune). Aujourd'hui, il en reste uniquement 2 et 3 nouveaux techniciens ont été recrutés. Les métadonnées ont donc permis de préserver une partie des connaissances des personnes qui sont parties. Les structures d'inférence et les modèles des tâches (cf. les section 7.2 et 7.3) ont joué aussi un rôle de capitalisation des connaissances : elles ont permis de formaliser le processus d'élaboration des STA. Ainsi, les nouveaux utilisateurs peuvent facilement accéder à des connaissances cruciales pour accélérer leur formation (cf. §12.2.1.4).

12.2.1.2 SAFE-Next, une approche de mutualisation et de partage des connaissances

Nous avons remarqué avec les experts du LAB et du CEESAR que SAFE-Next affecte positivement la perception des personnes qui ont participé à son élaboration et qui l'utilisent. Elle a joué un rôle de mutualisation des connaissances. En effet, elle a permis à chacun des experts de connaître les représentations de l'accident chez les autres. Cela a permis aussi « d'homogénéiser » les différentes représentations au niveau conceptuel : percevoir l'accident à travers les vues de la systémique (cf. §12.2.2.4).

Lors de l'affectation d'un attribut aux différents concepts concernés, il peut y avoir un *conflit de points de vue* entre les experts. Cependant, cela nous a permis aux experts de voir s'ils partageaient les mêmes définitions. Pour le modèle séquentiel (Figure 8-14), par exemple, nous avons noté une correspondance globale d'affectation entre les deux équipes de 85%, donc une discordance de 15% sur les 945 attributs. Cela nous a amené à faire une réunion commune entre les deux équipes pour aboutir à un consensus. Nous avons cherché ce consensus car nous avons voulu construire des STA qui tiennent compte des différents points de vue. Cependant, le consensus n'est pas un objectif dans l'absolu et c'est pour cela que SAFE-Next permet à chacun des experts d'avoir sa propre version de méta-données (i.e. sa propre classification d'attributs). Cela permettrait à chacun des experts, si besoin il y a, de créer ses propres modèles selon son domaine, ses objectifs, etc. Elle lui offre aussi la possibilité de partager ses modèles avec les autres. Dans les modèles que nous avons formalisés (cf. section 8.6.2.3), certains sont initialement propres aux accidentologues (CVE, séquentiel, etc.), d'autres sont a priori propres aux concepteurs (« *véhicule orienté sécurité* » et « *systèmes de sécurité* »). L'intégration de ces différents modèles par ASMEC a créé une interface et un langage entre les deux métiers (i.e. accidentologie et conception) (cf. Figure 9-9 et §8.6.2.3).

12.2.1.3 SAFE-Next, une approche de création de nouvelles connaissances

SAFE-Next offre une vision synthétique des différents types de connaissance du domaine et selon différents points de vue, ce qui permet de détecter des « trous » où il faut créer de nouvelles connaissances. Cela s'est manifesté à deux niveaux, celui des *points de vue* et celui des *attributs* :

Enrichir les points de vue existants. De nouveaux modèles ont été développés : « *pertinence générale des attributs* », « *fiabilité des attributs* », « *systèmes de sécurité* » et « *véhicule orienté sécurité* » (cf. § 8.6.2.3). Les deux premiers modèles sont d'une valeur certaine pour un utilisateur novice pour la

sélection des attributs adéquats à son étude. Les deux derniers sont fondamentaux pour projeter des connaissances en accidentologie dans l'espace de représentation du concepteur.

Enrichir la BD par de nouveaux attributs. La distribution des attributs selon le point de vue fonctionnel, par exemple, (cf. Figure 8-22) a permis aux experts de constater qu'il y a très peu d'informations dans le EDA concernant les étapes de « *diagnostic* », « *pronostic* » et « *décision* ». La construction des méta-données dans SAFE-Next permet donc d'avoir une idée sur la complétude des attributs caractérisant un accident dans la base EDA. Pour pallier ce problème, nous avons défini dans le cadre d'un projet de fin d'études d'ingénieur une étude dont l'objectif est d'enrichir ce modèle [Ben Hassine, 2004]. De nouveaux attributs ont été donc ajoutés (cf. Annexe A-2.). De même, grâce aux deux nouveaux modèles « *pertinence générale des attributs* » et « *fiabilité des attributs* », nous avons constaté l'existence d'attributs qui ne sont pas renseignés (i.e. on ne collecte pas leurs valeurs) ou qui sont renseignés, mais qui ne sont pas fiables et d'autres qui sont renseignés, mais qui ne sont pas utilisés dans les études. Nous avons choisi de ne pas supprimer ces attributs de la BD EDA. A la place, nous avons décidé avec les experts d'accorder à ces attributs des *niveaux de pertinence* (i.e. *très important, important, plutôt important, peu important et pas important*) (cf. § 8.6.2.3) afin d'en tenir compte lors de la sélection des attributs dans une étude donnée.

12.2.1.4 SAFE-Next, un outil de formation

Nous sommes persuadés que SAFE-Next pourrait faire un excellent moyen de formation des nouveaux embauchés au LAB et au CEESAR. Ceci est motivé par deux points. Premièrement, les connaissances capitalisées par SAFE-Next (modèles multi-vues, base de connaissances, connaissances d'inférences, connaissances de tâches, etc.) bénéficient, grâce à la systémique, d'une certaine complétude par rapport aux différentes représentations de l'accident et différentes tâche d'élaboration de STA. Ce contenu fait donc un excellent point de départ à la fois riche et synthétique pour un nouvel opérateur.

Deuxièmement, SAFE-Next est plus qu'une résolution d'un problème donnée à un moment donné (e.g. celui des STA), c'est aussi une façon de voir l'accident. Pour les opérateurs novices, SAFE-Next peut donc agir en tant que système de formation afin d'accélérer leur apprentissage et en tant que vision synthétique pour les personnes qui l'utilisent. Toutes ces remarques ont été observées et confirmées par les experts eux-mêmes. Reste maintenant la réalisation d'un outil intégré orienté formation (interfaces adéquates, etc.).

12.2.2 Apports de SAFE-Next en IC

12.2.2.1 SAFE-Next : une approche mixte

Par rapport aux trois grandes familles d'approches de modélisation des connaissances en IC (*ascendantes, descendantes et mixtes*) (cf. section 5.3), SAFE-Next se situe comme une *approche mixte*. En effet, la modélisation est guidée par un méta-modèle multi-vues, donc descendante, dont l'instrumentation fait appel à des méthodes descendantes ou ascendantes car pour instrumenter le méta-modèle par des modèles, on peut sélectionner un modèle (ou une partie d'un modèle) existant dans une bibliothèque ou créer un nouveau modèle en partant des données. L'instrumentation des structures d'inférences de préparation de données (cf. section 7.2.2) est un exemple

d'instrumentation par une approche descendante. En effet, nous avons sélectionné des composantes du *modèle d'évaluation* et du *modèle de conception de configuration* fournis par la bibliothèque CommonKADS [Breuker et Van de Velde, 1994; Schreiber *et al.*, 1994a]. Cependant, l'instrumentation de la structure d'inférence de la *tâche de classification* des accidents (cf. §7.2.2.2) a été effectué selon une approche ascendante, car nous avons mené une étape d'élicitation pour identifier les tâches, les inférences, etc. Dans la bibliothèque de CommonKADS, il n'existe pas de modèles de classification (il en existe un pour le classement). L'instrumentation du méta-modèle de *l'objet observé* (i.e. l'accident) a été fait aussi selon une approche ascendante car nous avons effectué une étape d'élicitation pour identifier les modèles du domaine utilisés par les experts (CVE, séquentiel, etc.) (cf. section 8.6.2). L'instrumentation des modèles par des concepts et l'instrumentation des concepts par des attributs (cf. sections 0 et 8.8) se base essentiellement sur l'élicitation.

Ainsi, SAFE-Next bénéficie des avantages des deux approches (ascendante et descendante) (cf. section 5.3.1 et 5.3.2) en générant des modèles qui correspondent aux points de vue existants grâce à l'élicitation, mais en même temps des points de vue génériques qui guident la phase d'élicitation grâce au méta-modèle. Cela permet de réduire le coût de la phase d'élicitation et d'assurer une certaine complétude au niveau des points de vue (cf. §12.2.2.5 pour plus de détails sur la notion de complétude).

12.2.2.2 Résoudre l'ambiguïté de la modélisation des connaissances

En ce qui concerne la modélisation des connaissances, presque toutes les méthodes en IC utilisent le même paradigme de modélisation qui consiste à distinguer *connaissances statiques* et *connaissances dynamiques*. Cependant, il y a une certaine ambiguïté en ce qui concerne la définition de ces deux notions. Dans certaines méthodes, MKSM [Ermine *et al.*, 1996] par exemple, on assimile les connaissances statiques au *modèle des concepts*, c'est-à-dire aux connaissances du domaine indépendantes de la tâche de résolution du problème (cf. §5.3.2.2). Cependant, on trouve dans la pratique des connaissances statiques telles que les *règles de la base des connaissances* (cf. section 7.1.2), mais qui ne correspondent pas à des concepts. C'est le cas, par exemple, des « *règles_décision_pertinence_attribut* » qui sont des connaissances statiques du domaine, mais qui ne sont pas des concepts. L'inverse est faux aussi, car tous les concepts ne représentent pas les connaissances statiques. « *Accident* » et « *Classe* » sont tous les deux des concepts, mais le premier, contrairement au deuxième, est statique et indépendant de la tâche de résolution du problème de construction des STA. Dans CommonKADS [Schreiber *et al.*, 1994a], cette ambiguïté n'existe pas car on ne fait pas d'équivalence entre connaissances statiques et concepts du domaine et ce grâce aux rôles de connaissances (cf. section 7.2.1). Les rôles statiques peuvent concerner des concepts, des règles, des modèles du domaine, etc. Cependant, au niveau des rôles statiques, on ne distingue pas ceux qui sont propres à la tâche de résolution de problème et ceux qui en sont indépendants et propres au phénomène étudié. Par exemple, « *règles_décision_pertinence_attribut* » et « *modèle séquentiel* » sont deux rôles statiques, mais le premier est propre à la *tâche d'élaboration des STA* alors que le deuxième en est indépendant.

Dans SAFE-Next, nous avons essayé de résoudre cette ambiguïté grâce à la distinction entre le point de vue *objet observé* qui concerne les *connaissances propres au phénomène étudié et indépendantes de l'application* et le point de vue *observateur* qui concerne les *connaissances propres à l'application*. Cette distinction facilite la tâche d'élaboration des SBC car les méthodes d'élicitation et d'implémentation des deux types de connaissances sont différentes. Ensuite, *la distinction entre connaissances statiques et connaissances dynamiques se fait au niveau de chacun des points de vue objet observé et observateur*. Ainsi, « *règles_décision_pertinence_attribut* » est une *connaissance statique* qui appartient au *point de vue observateur* et « *modèle séquentiel* » est une *connaissance statique*, mais qui appartient au *point de vue objet observé* (car elle est indépendante de la tâche de résolution de problème). La même chose est assurée pour les concepts : « *Accident* » est un concept propre à *l'objet observé*, alors que « *Classe* » est un concept propre au *point de vue observateur*.

12.2.2.3 *Elaboration d'un schéma multi-vues du domaine : application de la systémique à la modélisation des connaissances*

Une autre différence majeure entre SAFE-Next et les méthodes existantes se situe au niveau de la définition des concepts. A notre connaissance, aucune des méthodes d'IC identifiées dans la littérature (cf. section 5.3) n'a identifié des catégories de concepts sachant que dans la pratique on peut distinguer des concepts qui portent sur des aspects physiques, d'autres sur des fonctions, etc. Grâce au deuxième niveau du méta-modèle (cf. section 8.4), i.e. les quatre axes systémiques appliqués au *point de vue objet observé*, nous permettons d'identifier des concepts *ontologiques* (exemple : « *pneu* »), *fonctionnels* (exemple : « *perception* »), *transformationnel* (exemple : « *rupture* ») et *téléologiques* (exemple : *navigation*). Sans la vue systémique, tous ces concepts seraient mélangés. Cette vision permet d'élaborer un *schéma multi-vues du domaine* ce qui permet de résoudre l'un des points faibles dans CommonKADS (cf. section 8.1). De même, l'application des axes systémiques au *point de vue observateur* permet de distinguer des connaissances sur *les ontologies de la méthode de résolution de problème* (exemple : rôle de connaissances), *ses fonctions* (exemple : structures de contrôle), *ses transformations* (exemple : structure d'inférences) et *ses téléologies* (tâches).

12.2.2.4 *Les métadonnées : un outil efficace pour l'acquisition des connaissances*

La construction des métadonnées a consisté en la caractérisation des concepts dans les différents modèles par des attributs en utilisant des attributs existant et/ou en créant de nouveaux attributs (cf. section 8.8). Ces métadonnées sont ainsi un moyen efficace pour capter et formaliser les connaissances des experts concernant les attributs, puisqu'il s'agit d'une description experte de ces attributs. Le fait de dire que l'attribut « *âge* », par exemple, caractérise le concept « *conducteur* » dans le modèle « *CVE* » et le concept « *conduite normale long terme* » dans le « *séquentiel* » est un moyen de formalisation d'un certain type de connaissances. Comme nous l'avons déjà noté respectivement dans les paragraphes §12.2.1.1 et §12.2.1.2, ces métadonnées ont permis d'accéder à des connaissances implicites et ont permis de partager les représentations de l'accident chez les différents experts et ce à différents niveaux d'abstraction (i.e. modèles, concepts et attributs).

12.2.2.5 Assurer la complétude des points de vue lors de l'acquisition des connaissances

Grâce au méta-modèle dans ASMEC (cf. section 8.4) et à l'acquisition des connaissances dirigée par un méta-modèle (cf. section 8.3), SAFE-Next permet d'assurer une certaine complétude des connaissances élicitées. En effet, grâce au point de vue *objet observé*, SAFE-Next se base sur la modélisation du phénomène lui-même au lieu de modéliser la perception du phénomène par les experts. Les points de vue systémiques indiquent à l'ingénieur de connaissances ce qu'il faut éliciter pour appréhender la complexité du phénomène. Dans le cas contraire, le risque d'oublier ou de ne pas pouvoir identifier un point de vue est important (cf. §8.6.2.3). En effet, dans le projet ACACIA [Dieng *et al.*, 1996] (cf. §5.3.3.1), le modèle de « *la tâche de conduite* » (cf. section 4.3.4) qui correspond au point de vue téléologique, par exemple, n'a pas été identifié sachant que ce point de vue est crucial pour la tâche de diagnostic car il permet de comprendre l'erreur comme étant un écart entre ce que le conducteur a voulu faire (sa téléologie) et ce qu'il a fait. L'absence de ce modèle est due au fait qu'ACACIA s'est basé essentiellement sur une étape d'élicitation pour la construction des modèles du domaine. Il suffit alors qu'un expert ayant ce point de vue ne participe pas à l'étude ou que ce point de vue ne soit pas utilisé dans l'entreprise pour que ce point de vue ne soit pas identifié. Ce risque est réduit dans SAFE-Next, grâce à ASMEC, puisque la question d'instrumenter le point de vue téléologique est systématique.

12.2.2.6 ASEM : un modèle général d'évaluation de modèles en IC

Bien que la notion de *modèle* en IC soit centrale, nous avons noté la rareté des travaux sur l'évaluation des modèles. C'est dans ce contexte qu'ASEM a été élaboré. Nous avons proposé des critères d'évaluation d'un modèle selon quatre points de vue, ceux de la systémique. Nous avons proposé ensuite (sections 11.4 et 11.5) des relations entre les critères d'évaluation d'un modèle et ceux d'évaluation de connaissances. Ces relations ont pour objectif de guider la tâche de modélisation car ils permettent de déterminer les critères à respecter au niveau du modèle pour respecter un ensemble de critères au niveau des connaissances recherchées. Nous avons défini aussi des relations intra-critères d'évaluation d'un modèle. Ils ont pour objectif de connaître l'effet de l'amélioration (resp. la dégradation) d'un critère donné sur les autres critères.

Dans ASEM nous proposons plus qu'un modèle d'évaluation : nous proposons une « *conception* » de l'évaluation. En effet, en plus du fait qu'ASEM peut être utilisé pour l'évaluation *a posteriori* d'un modèle déjà élaboré, nous proposons de l'utiliser parallèlement au processus de modélisation. Cela est motivé par le fondement constructiviste et cybernétique de nos travaux qui affirme que le résultat d'une opération (l'évaluation dans ce cas) influe sur son opérateur (i.e. l'évaluateur) (cf. section 11.1). Ce postulat a été vérifié et validé dans plusieurs études [Henry, 2003; Henry et Mark, 2003; Kirkhart, 2000].

12.2.3 Apports de SAFE-Next en ECD

Les apports essentiels de SAFE-Next en ECD se situent au niveau de la première et dernière étape du processus, i.e. *sélection des attributs* (cf. section 6.3.1) et *interprétation des résultats* (cf. section 6.3.2). En effet, la majorité des techniques pour effectuer ces deux tâches se base sur des critères statistiques, souvent indépendants du domaine. SAFE-Next offre alors un outil efficace d'incorporation des connaissances du domaine dans le processus d'ECD. Grâce aux méta-données (cf. section 8.8), à l'interprétation multi-niveaux d'abstraction (cf. section 9.2.3) et à l'utilisation des treillis de Galois (cf. §9.2.2.2), l'utilisateur des techniques d'ECD peut *sélectionner des attributs* et/ou *interpréter* les résultats en fonction de n'importe quel modèle existant ou construit par l'utilisateur lui-même. Face à l'abondance des informations dans les EDA, SAFE-Next permet d'effectuer un filtrage et éviter ainsi le problème fréquent qu'est la surcharge d'information. Ainsi, l'usager peut concentrer ses efforts sur les informations pertinentes. Cela permet d'améliorer *l'efficacité* (i.e. réduction du temps et du coût) ainsi que *l'effectivité* (meilleure qualité grâce à l'aspect multi-vue) du processus d'élaboration des STA.

12.2.4 Apports de SAFE-Next en conception de nouveaux produits

12.2.4.1 Utilisation de SAFE-Next pour spécifier l'espace de conception

Comme nous l'avons déjà noté tout au long de ce mémoire et particulièrement dans les sections 2.4 et 4.4, les STA développés par SAFE-Next fournissent un outil performant pour *comprendre l'étiologie de l'accident et des lésions, afin de proposer des contre-mesures qui seront utilisées comme des entrées dans le processus de conception*. Il s'agit d'un langage entre les accidentologues et les concepteurs des systèmes de sécurité embarqués. En effet, le développement de nouveaux systèmes de sécurité embarqués dans les véhicules requiert des connaissances sur le processus de conduite en général et en accidentologie en particulier. Comme nous l'avons montré dans le paragraphe §8.6.2.3, certains modèles de représentation de l'accident utilisé dans le processus d'ECD (préparation des données et interprétation des résultats) sont propres aux accidentologues (e.g. modèle CVE, modèle séquentiel, etc.), alors que d'autres sont propres aux concepteurs (e.g. modèle « véhicule orienté sécurité » et « systèmes de sécurité »). Ainsi, les points de vue des accidentologues, aussi bien que les points de vue des concepteurs ont été intégrés dans le processus de construction des STA.

Par rapport à un cycle classique de développement d'un produit, SAFE-Next se situe dans les techniques de *préconception*. Elle permet de développer des connaissances sur le phénomène (ou système) étudié pour le comprendre et pouvoir proposer des solutions. Sans ces connaissances, des méthodes classiques d'aide à la conception comme l'analyse fonctionnelle (externe et interne), la méthode APTE et la QFD (Query Function Deployment) ou des méthodes d'aide à l'innovation comme TRIZ ne peuvent pas être appliquées. Ces méthodes permettent la représentation et la structuration des fonctionnalités du nouveau système, donc elles supposent une connaissance préalable de ces fonctionnalités. Ainsi, elles sont des méthodes de *structuration et représentation de l'espace de conception*. Mais, ce dont nous avons besoin dans notre cas, ce sont des méthodes de *construction de cet espace de conception*.

12.2.4.2 Utilisation de SAFE-Next pour l'analyse d'impact de changement

Comme nous l'avons montré dans le Chapitre 10 et plus particulièrement dans la section 10.6.4, SAFE-Next offre, à travers ASAIC, une approche pour l'analyse multi-vues de l'impact d'un changement. Elle a posé le cadre pour anticiper ou évaluer a posteriori les impacts d'un changement selon les trois points de vue « *processus de conception* », « *processus d'utilisation* » et « *processus d'évaluation* ». La majorité des approches que nous avons pu identifier dans la littérature se focalisent sur le processus de conception lors de l'analyse de changement. Nous avons montré que les trois points de vue que nous proposons sont inter-reliés : un changement au niveau de l'un des trois processus engendra des changements aux niveaux des deux autres (cf. section 10.6.1). Ils sont donc cruciaux pour appréhender la complexité de l'analyse d'un impact, particulièrement dans des domaines complexes comme celui de l'accidentologie.

Synthèse de la partie III

Dans le premier chapitre de cette partie (Chapitre 7), nous avons appliqué la méthode CommonKADS [Schreiber *et al.*, 1994a] pour élaborer un modèle de connaissances pour l'élaboration de STA. Trois types de connaissances ont alors été identifiés et formalisés : des *connaissances du domaine*, des *connaissances d'inférences* et des *connaissances de la tâche*. Nous avons montré à la fin de ce premier chapitre les limites de CommonKADS pour formaliser certains types de connaissances, plus précisément des connaissances du domaine qui sont indépendantes de la tâche d'élaboration des STA. Ces connaissances, qui correspondent à des *rôles de connaissances statiques* selon le formalisme de CommonKADS, restent sous forme de « *boite noire* ». De plus, nous montrons les limites de CommonKADS quand il s'agit d'élaborer un *schéma multi-vues du domaine* (cf. section 7.1.1).

La suite de la partie III (i.e. du Chapitre 7 au Chapitre 12) a été consacrée à l'élaboration de notre approche SAFE-Next (Systemic Approach For Enhanced kNowledge EXTraction). Nous avons doté SAFE-Next de quatre approches principales : ASMEC, AICEF, ASAIC et ASEM pour fournir un cadre global, respectivement, de modélisation des connaissances en IC, d'incorporation des connaissances du domaine en ECD, d'analyse d'impact de changement en conception et d'évaluation de modèles en général. Une synthèse sur chacune de ces approches ainsi que leurs apports est fournie dans la conclusion générale.

Conclusion générale

Synthèse générale de nos travaux

Aujourd'hui, l'informatisation des saisies de données et la puissance des systèmes de collecte conduisent à la construction de grandes Bases de Données (BD). L'exploitation de ces millions de données en management, en marketing, en biologie et dans beaucoup d'autres domaines fait appel à des techniques d'Extraction de Connaissances de Données (ECD). La complexité des données, du domaine d'application et des connaissances recherchées rendent fondamentale la combinaison des approches expertes et automatisées pour l'extraction de connaissances valables, nouvelles et potentiellement utiles.

C'est typiquement le cas dans le domaine de l'accidentologie quand il s'agit de construire des Scénarios-Types d'Accidents (STA). En partant de l'hypothèse qu'à un groupe d'accidents similaires correspondent des contre-mesures similaires, les STA offrent un support de connaissances et un outil efficace de retour d'expérience pour définir et évaluer des Systèmes de Sécurité. En effet, ils permettent de passer de la compréhension de chaque accident, pris individuellement, à une vue plus générale des problèmes de sécurité sur l'échantillon d'accidents étudiés. Ces STA peuvent être élaborés manuellement au cas par cas par une approche purement experte ou en utilisant des techniques d'ECD comme la classification automatique (i.e. trouver des groupes d'accidents similaires). Cependant, ces deux possibilités présentent de nombreuses limites. D'un côté, la quantité importante des données, leur nature multi-sources (entretiens, photos et reconstruction dynamique, etc.) et le besoin de construire des STA à la demande (selon un objectif donné) rendent une *construction experte* de ces STA très coûteuse. D'un autre côté, la complexité de l'accident, la pluridisciplinarité de l'accidentologie (médecine, psychologie, mécanique, ingénierie de la route, épidémiologie, etc.) et le besoin de combiner plusieurs points de vue d'analyse rendent une *construction automatique* de ces STA inadaptée en terme de qualité des résultats.

Dans ce travail de thèse, nous proposons la formalisation et l'incorporation des connaissances expertes dans le processus d'ECD comme un moyen de combiner les approches expertes et automatisées. Nous développons alors une nouvelle approche, appelée SAFE-Next (Systemic Approach For Enhanced kNowledge EXtraction, approche systémique pour l'extraction des connaissances) que nous généralisons pour fournir un cadre global de modélisation multi-vues des

connaissances en IC, un moyen d'incorporation des connaissances expertes dans le processus d'ECD, un cadre pour l'analyse d'impact de changement lors de la conception d'un nouveau produit et un moyen d'évaluation de modèles en général. Ainsi, nous avons doté notre approche SAFE-Next de quatre approches :

ASMEC (Approche Systémique de Modélisation des Connaissances)

ASMEC est la première approche que nous avons développée dans SAFE-Next (cf. Chapitre 8). Elle permet la modélisation des connaissances selon une *architecture multi-vues et multi-niveaux d'abstraction* (cf. Figure IV.1 et Figure IV.2) : *méta-modèle, modèle, concept, attribut* et *instance* (cf. figure ci-dessous). Le *méta-modèle* (1^{er} niveau) est constitué de deux couches. La première est instrumentée par trois points de vue : *objet observé* (i.e. accident), *observateur* (i.e. expert chargé de l'élaboration des STA) et *contexte de l'observation* (i.e. utilisation des STA pour développer des systèmes de sécurité). Chacun de ces points de vue est, à son tour, instrumenté dans la deuxième couche par les quatre points de vue systémiques : *ontologique, fonctionnel, transformationnel* et *téléologique*.

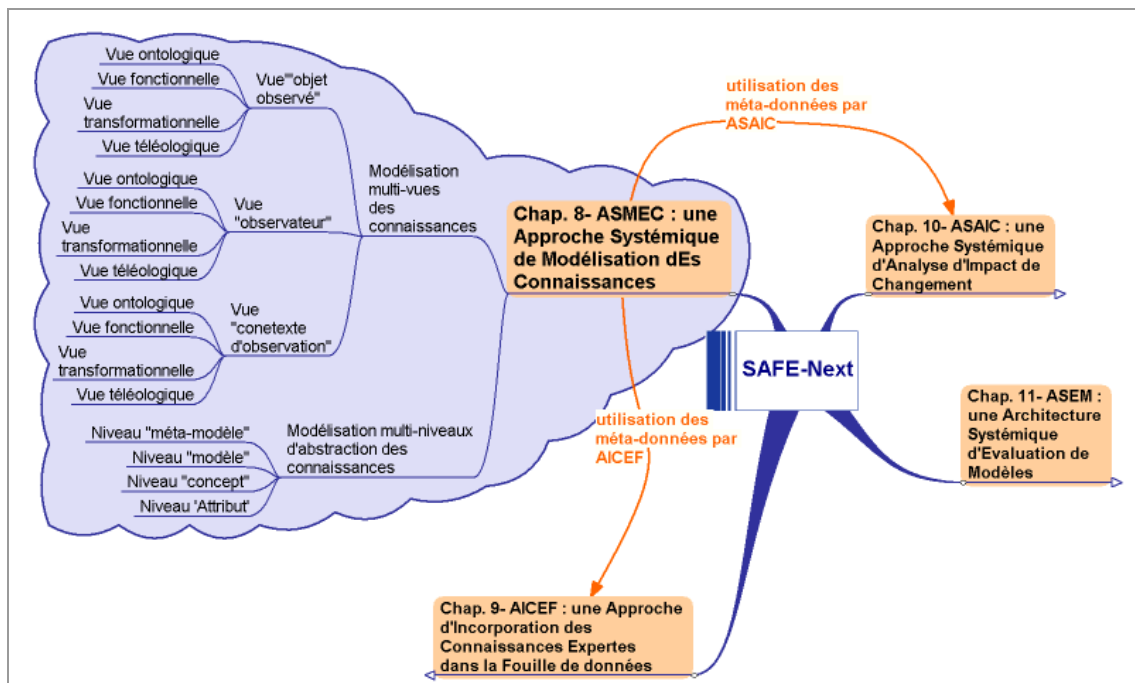


Figure IV.1- ASMEC : approche systémique de modélisation multi-vues de connaissances (première approche dans SAFE-Next)

Dans le deuxième niveau d'abstraction de notre architecture de modélisation, chacun des points de vue de la deuxième couche du méta-modèle est instrumenté par un ou plusieurs *modèles* du domaine (cf. modèle UML, Figure IV.2). Chaque modèle est instrumenté par des *concepts* (3^{ème} niveau). Chaque concept est, à son tour, caractérisé par des *attributs* (4^{ème} niveau) dont l'instanciation donne le dernier niveau d'abstraction (les *instances*). Cette architecture est à la fois générale et fine pour pouvoir modéliser non seulement des *connaissances du domaine* (grâce au point de vue *objet observé*), mais aussi des connaissances relatives à la *méthode de résolution de problème* (grâce au point de vue *observateur*) et au *contexte de l'étude* (grâce au point de vue *contexte de l'observation*).

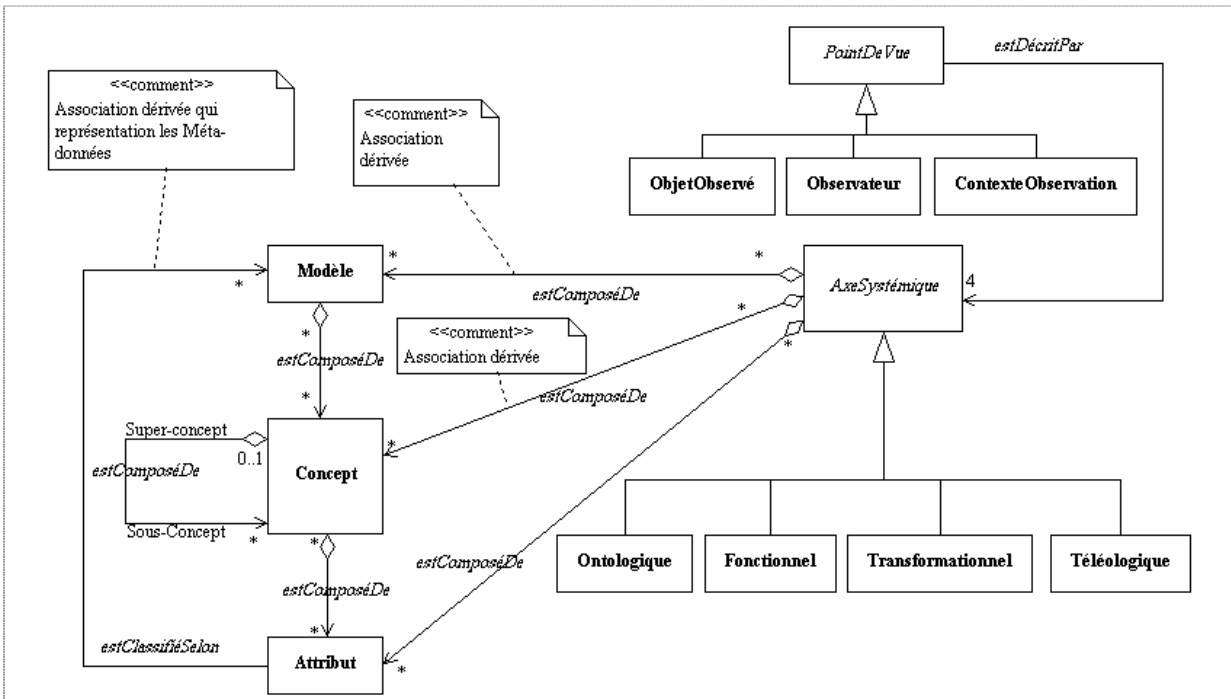


Figure IV.2- Modèle UML de l'architecture d'ASMEC : modélisation dirigée par un méta-modèle multi-vues et multi-niveaux d'abstraction.

AICEF (Approche d'Incorporation des Connaissances Expertes dans la Fouille de données)

AICEF est la deuxième approche que nous avons développée dans SAFE-Next (cf. Chapitre 9). Nous proposons à travers AICEF un moyen d'incorporation des connaissances expertes dans les deux tâches de *préparation des données* et d'*interprétation des classes*. Pour cela, nous utilisons les méta-données que nous avons élaborées dans ASMEC (cf. Figure IV.3).

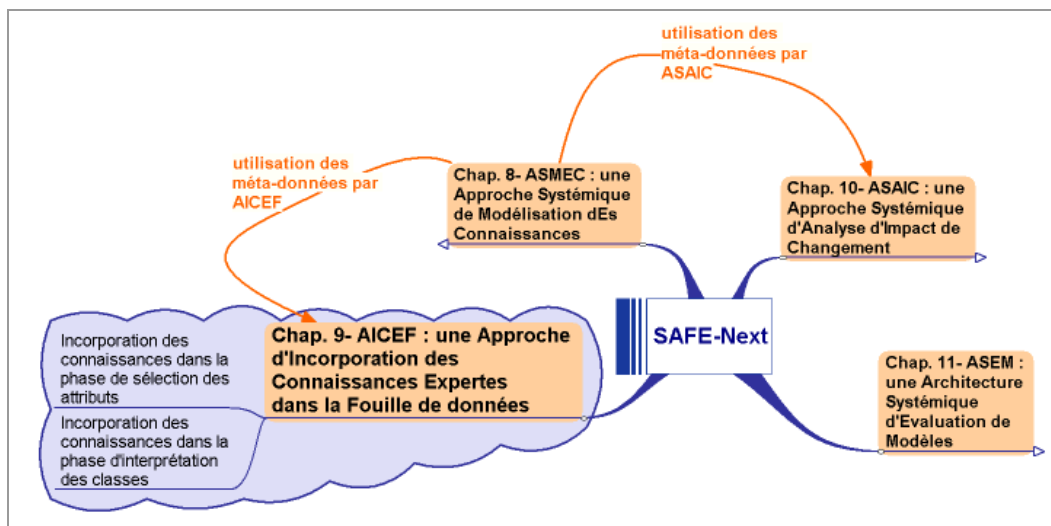


Figure IV.3- AICEF : approche d'incorporation des connaissances dans le processus d'ECD (deuxième approche dans SAFE-Next)

En effet, l'architecture d'ASMEC, vue du *niveau des attributs* (4^{ème} niveau de l'architecture), peut être perçue comme une classification multi-vues de ces attributs. Le fait d'affecter un attribut à

plusieurs concepts, un concept à plusieurs modèles et un modèle à plusieurs points de vue, est une sorte d'explicitation de connaissances expertes sur ces attributs. En d'autres termes, il s'agit d'une création de « données sur les données », d'où la notion de *méta-données*. Puisque la sélection des attributs dans la première phase d'ECD (i.e. préparation des données) nécessite des connaissances sur ces attributs, les méta-données offrent un outil efficace afin de répondre à cet objectif. De même, puisque à la sortie d'un programme de classification les classes sont caractérisées par ces mêmes attributs, les méta-données offrent donc un outil efficace pour l'incorporation des connaissances expertes dans la phase d'interprétation des classes. Nous avons implémenté cette méthode dans un outil utilisant le langage XML pour la représentation des méta-données. Cet outil permet une sélection multi-vues des attributs et une interprétation multi-vues des classes d'accidents pour élaborer des STA.

ASAIC (Approche Systémique d'Analyse d'Impact de Changement)

ASAIC est la troisième approche que nous avons développée dans SAFE-Next (cf. Chapitre 10). Elle propose l'utilisation des méta-données développées par ASMEC pour une *analyse multi-vues et multi-niveaux d'abstraction de l'impact de changement* (e.g. introduction d'un nouveau système de sécurité) selon les trois points de vue génériques suivants : « processus d'utilisation », « processus de conception » et « processus d'évaluation ». Dans notre cas, ces derniers correspondent, respectivement, au « comportement du système CVE », « processus de conception des véhicules » et « métier de l'accidentologie » (cf. Figure IV.4). L'élaboration d'ASAIC a été motivée par le fait qu'un changement au niveau de l'un des composantes du système CVE peut affecter le comportement du système et donc le besoin en sécurité routière. Ces changements doivent être anticipés par les concepteurs des systèmes de sécurité et par le métier de l'accidentologie pour évaluer ses impacts.

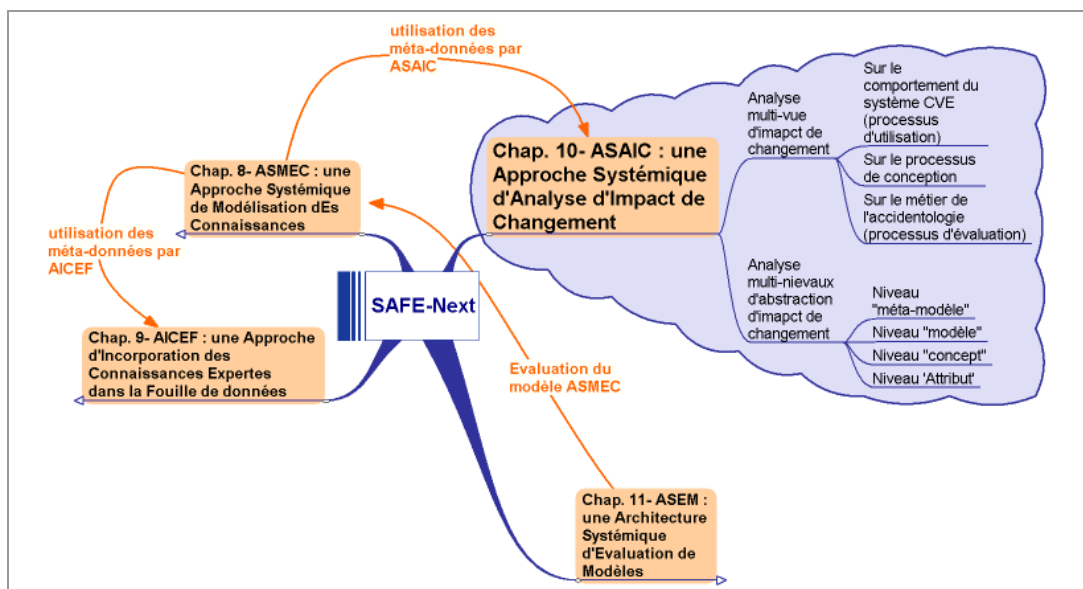


Figure IV.4- ASAIC : utilisation des méta-données pour l'analyse d'impact de changement (troisième approche dans SAFE-Next)

ASEM (Approche Systémique d'Evaluation de Modèle)

ASEM est la quatrième approche que nous avons développée dans SAFE-Next (cf. Chapitre 11). Elle fournit un *modèle général d'évaluation a priori* (i.e. au cours de la modélisation) et *a posteriori* (après la modélisation) de modèles de connaissances (cf. Figure IV.5). Elle permet aussi d'*analyser les relations intra-critères d'évaluation de modèles* et les *relations entre les critères d'évaluation de connaissances et les critères d'évaluation de modèles*.

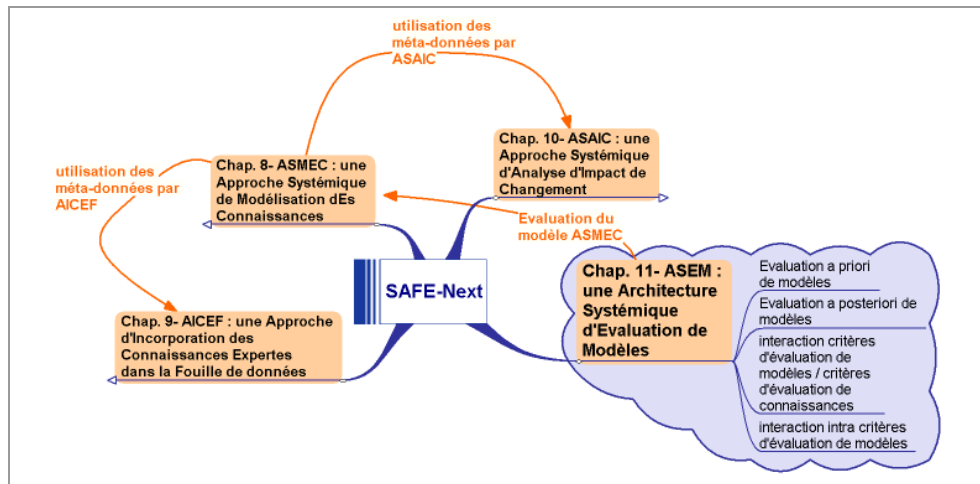


Figure IV.5- ASEM : approche d'évaluation de modèles (quatrième approche dans SAFE-Next)

La première motivation de l'élaboration d'ASMEC est notre constat de la rareté de travaux sur l'évaluation de modèles en IC, autres que ceux se basant sur l'opérationnalisation et la simulation de ces modèles. Cependant, en génie logiciel, un domaine pourtant très proche de l'IC, la simulation n'est pas le seul moyen de validation. Nous considérons que l'opérationnalisation et la simulation des modèles permettent une *évaluation a posteriori* et nous montrons l'intérêt de développer des moyens *d'évaluation a priori*. La deuxième motivation de l'élaboration d'ASEM est issue de notre vision systémique de la modélisation elle-même : nous pensons que durant notre travail de modélisation, notre perception du phénomène modélisé (e.g. l'accident, la connaissance, etc.) peut changer, et donc, les modèles qui en résultent changent. D'où l'intérêt de mener un processus d'évaluation parallèlement au processus de modélisation.

La Figure IV.6 donne une vue synthétique de l'ensemble de nos travaux :

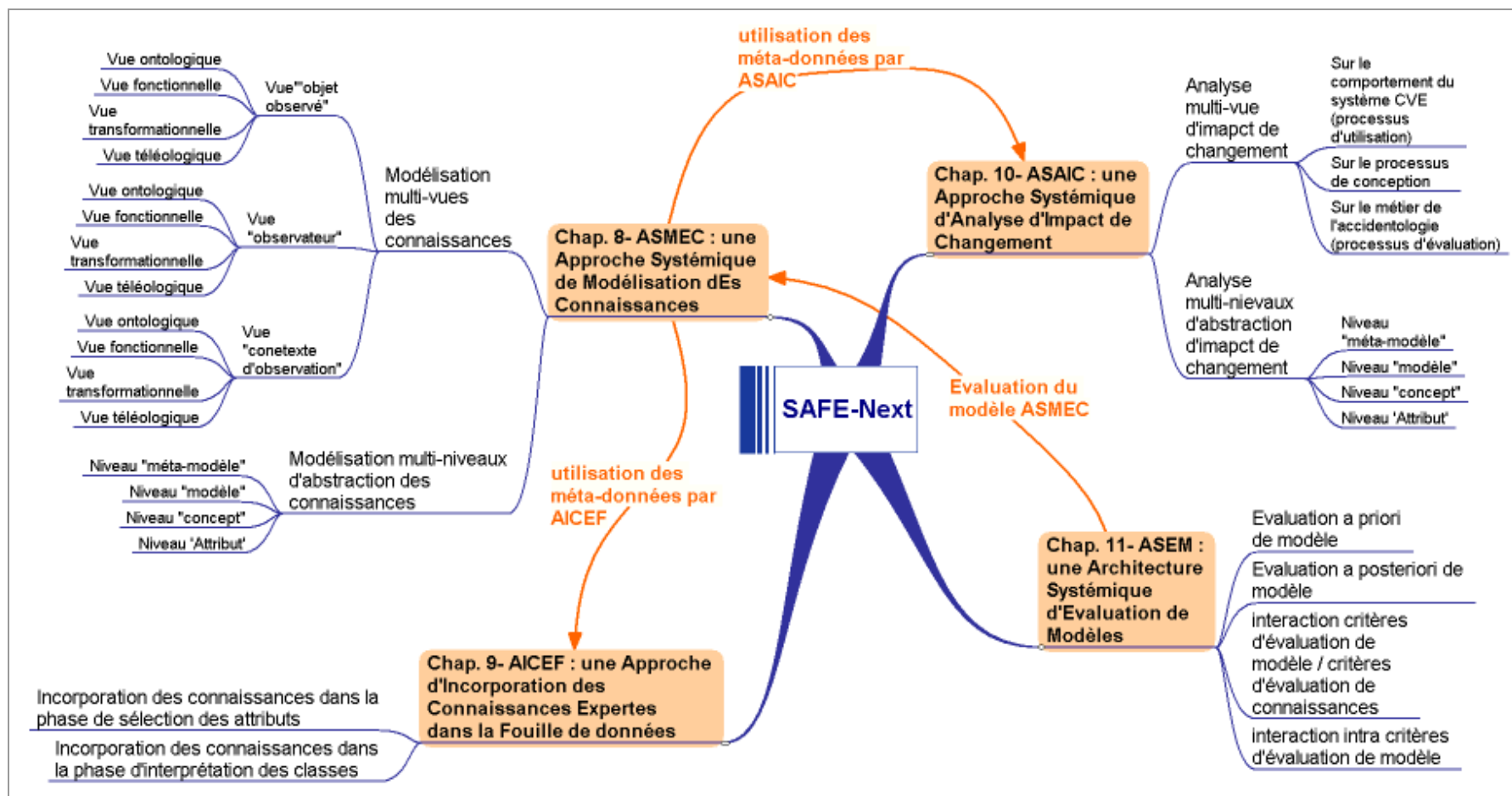


Figure IV.6- Synthèse de l'ensemble des approches dans SAFE-Next.

Synthèse des contributions

Les fondements épistémologique et méthodologique de nos travaux sont respectivement le *constructivisme* et la *systémique* (ou *cybernétique*). En se basant sur ces fondements, nos travaux de recherches ont conduit à des contributions réparties en quatre domaines :

En gestion des connaissances en accidentologie, SAFE-Next fournit, grâce à ASMEC et AICEF, un outil efficace pour l'élaboration des STA permettant une meilleure analyse et compréhension de l'accident. Elle fournit à travers les STA un moyen de communication entre les métiers de l'accidentologie et de conception afin de proposer des contre-mesures efficaces pour lutter contre l'insécurité routière. SAFE-Next fournit aussi un moyen de capitalisation des connaissances (cf. §12.2.1.1). En effet, grâce à ASMEC, elle regroupe et structure les modèles existants dans un domaine donné (e.g. l'accidentologie) et offre une vision synthétique et accessible des différents types de connaissances. Grâce à l'utilisation des métadonnées, elle permet d'accéder à des connaissances implicites et de les formaliser. Elle permet aussi de préserver ces connaissances en cas de départ des experts (cf. §12.2.1.1). Ainsi, les nouveaux utilisateurs peuvent facilement accéder à des connaissances cruciales pour accélérer leur formation (cf. §12.2.1.4). Elle a joué aussi un rôle de mutualisation et de partage des connaissances (cf. §12.2.1.2). De plus, SAFE-Next, grâce à la vision synthétique qu'elle fournit des différents types de connaissance selon différents points de vue, permet de détecter des « trous » où il faut créer de nouvelles connaissances (cf. §12.2.1.3). Elle permet alors *d'enrichir les points de vue existants* (de nouveaux modèles en accidentologie ont été développés) et *d'enrichir la BD EDA par de nouveaux attributs*.

En ingénierie des connaissances, SAFE-Next propose, grâce à ASMEC, un modèle général multi-vues et multi-niveaux d'abstraction de modélisation des connaissances pour le développement des Systèmes à Base de Connaissances (SBC). Elle permet d'élaborer des schéma multi-vues du domaine (cf. §12.2.2.3) en assurant une certaine complétude au niveau de points de vue (cf. §12.2.2.5) et fournit, grâce aux méta-données un moyen efficace d'acquisition de connaissances (cf. §12.2.2.4). Par rapport aux trois grandes familles d'approches de modélisation des connaissances en IC (*ascendantes*, *descendantes* et *mixtes*), SAFE-Next se situe comme une *approche mixte* (cf. §12.2.2.1). Ainsi, elle bénéficie des avantages des deux approches (ascendante et descendante) en générant des modèles qui correspondent aux points de vue existants grâce à l'élicitation, mais en même temps des points de vue génériques qui guident la phase d'élicitation et ce grâce au méta-modèle d'ASMEC. SAFE-Next fournit aussi, grâce à ASEM, un moyen

d'évaluation des modèles de connaissances (cf. §12.2.2.6). En plus du fait qu'ASEM peut être utilisé pour l'évaluation *a posteriori* d'un modèle déjà élaboré, nous proposons de l'utiliser parallèlement au processus de modélisation ce qui permet d'améliorer la qualité de ces modèles.

En extraction des connaissances des données, SAFE-Next propose à travers AICEF l'utilisation des métadonnées multi-vues, réalisées selon le modèle d'ASMEC, pour l'incorporation des connaissances expertes du domaine dans la première et la dernière phase du processus ECD (i.e. préparation des données et interprétation des résultats) (cf. §12.2.3). La majorité des techniques existantes pour effectuer ces deux tâches se base sur des critères statistiques, souvent indépendants du domaine. Grâce aux méta-données, AICEF permet l'incorporation des connaissances du domaine dans ces deux tâches. Face à l'abondance des informations, AICEF permet d'effectuer un filtrage multi-vues et multi-niveaux d'abstraction des données lors du processus d'ECD et éviter ainsi le problème fréquent qu'est la surcharge d'information. Cela permet d'améliorer l'efficacité (i.e. réduction du temps et du coût) ainsi que l'effectivité (meilleure qualité grâce à l'aspect multi-vues) du processus d'ECD.

En conception de nouveaux systèmes, SAFE-Next fournit à travers les STA un moyen de communication entre les accidentologues et les concepteurs des systèmes de sécurité pour définir et évaluer les prestations de sécurité. Par rapport à un cycle de développement d'un produit, SAFE-Next se situe dans les techniques de préconception. Elle permet de développer des connaissances sur le phénomène (ou système) étudié pour le comprendre et pouvoir proposer des solutions et des contre-mesures. Il s'agit donc d'une approche permettant la construction de l'espace de conception (cf. §12.2.4.1). SAFE-Next fournit aussi, grâce à ASAIC, une approche pour l'analyse multi-vues de l'impact d'un changement (cf. §12.2.4.2). Elle permet d'anticiper ou d'évaluer *a posteriori* les impacts d'un changement selon les trois points de vue « *processus de conception* », « *processus d'utilisation* » et « *processus d'évaluation* ». La majorité des approches que nous avons pu identifier dans la littérature se focalise sur le processus de conception. L'ensemble des contributions de SAFE-Next est résumé dans la Figure IV.7 :

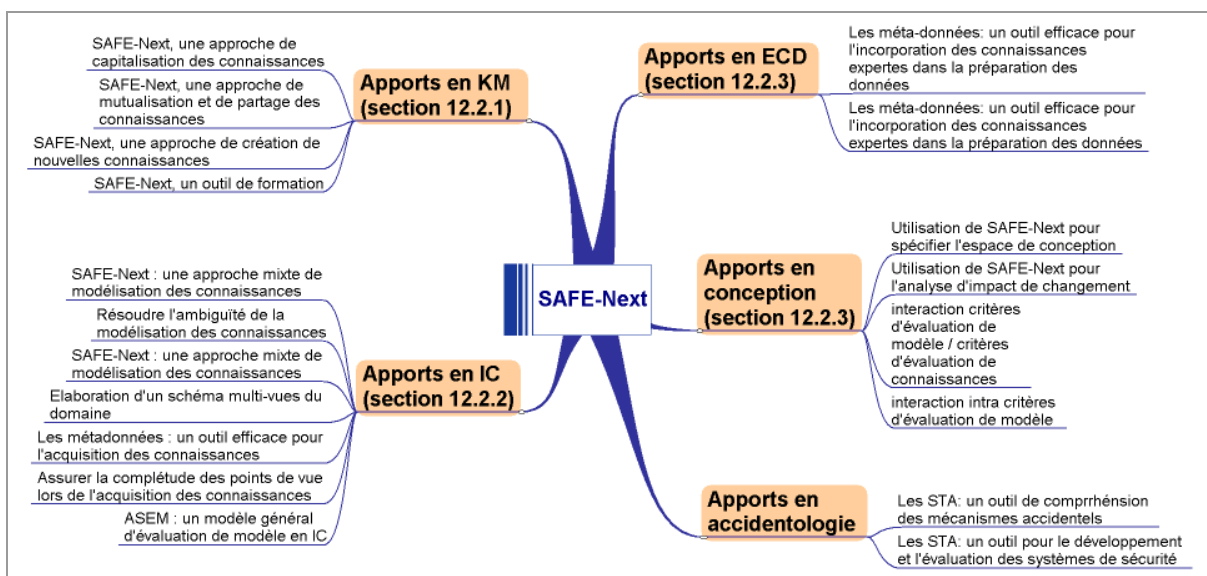


Figure IV.7- Une synthèse des contributions de SAFE-Next en KM, IC, ECD, conception et accidentologie.

Discussion & perspectives

Utilisation de SAFE-Next pour le développement des points de vue *observateur* et *contexte de l'observation*

Comme nous l'avons présenté dans la section 8.4, le premier niveau d'abstraction du méta-modèle d'ASMEC est composé de trois points de vue : *objet observé* (e.g. l'accident), *observateur* (e.g. expert chargé de la tâche de l'élaboration des STA) et *contexte d'observation* (e.g. élaboration des STA pour le développement des systèmes de sécurité). Dans ce mémoire, nous avons utilisé SAFE-Next uniquement pour développer le *point de vue objet observé* en tenant compte des deux autres points de vue. Le développement du point de vue *objet observé* correspond au développement du *schéma du domaine* (cf. §8.6.2.3). Mais, nous avons montré à travers l'architecture globale d'ASMEC (cf. section 8.4), la spécification du méta-modèle (cf. section 8.5) et l'instrumentation du méta-modèle (cf. section 8.6), qu'ASMEC ne sert pas uniquement au développement du schéma du domaine. Les deux autres points de vue peuvent être développés grâce à l'architecture de notre approche. La Figure 8-10 (page 112) illustre bien cela. Nous pensons alors qu'il pourrait être intéressant d'appliquer le modèle multi-vues d'ASMEC pour développer les points de vue *observateur* et *contexte de l'observation*. Cela pourrait permettre, comme dans le cas du point de vue *objet observé*, d'organiser les modèles existants dans la littérature selon les différents points de vue de la systémique et selon différents niveaux de granularité, de créer de nouveaux modèles et/ou d'enrichir des modèles existants.

Application de SAFE-Next dans un domaine où on ne dispose ni de BD, ni d'attributs

Dans ce mémoire, nous avons appliqué et validé SAFE-Next en accidentologie où on dispose de données (les bases de données EDA). Il convient donc de poser la question suivante : *qu'est-ce qu'on fait dans le cas où les attributs ne seraient pas « complets » (c'est-à-dire que certains aspects du phénomène ne sont pas décrits par des attributs) ou lorsqu'on ne dispose plus d'attributs ?*

Nous pensons que SAFE-Next reste applicable dans ces deux cas. En effet, dans le premier (i.e. incomplétude d'attributs), les points de vue systémiques d'ASMEC peuvent guider la définition ou la recherche de nouveaux attributs. C'est ce que nous avons effectué pour enrichir le modèle

fonctionnel dont trois concepts (« *diagnostic* », « *pronostic* » et « *décision* ») étaient sous-caractérisés (cf. section 8.8.5). Dans le deuxième cas (i.e. absence d'attributs), il faut commencer par construire une ontologie du domaine (concepts, relations, etc.). Des méthodes comme KOD [Vogel, 1989], MIKE [Angele *et al.*, 1996] ou PROTEGEE [Puerta *et al.*, 1992] peuvent être utilisées pour la construction de l'ontologie. SAFE-Next pourrait aider dans ce premier temps à la structuration multi-vues des ontologies en utilisant ASMEC. Dans un deuxième temps, elle pourrait être appliquée pour extraire des connaissances des données construites exactement comme nous l'avons fait pour élaborer les STA en utilisant les données d'EDA. Un fort intérêt a été manifesté pour SAFE-Next par le service de maintenance chez PSA Peugeot Citroën. L'idée est d'utiliser le méta-modèle SAFE-Next, de l'instrumenter par des modèles du domaine de la maintenance, trouver les concepts instrumentant les modèles et définir les attributs relatifs à ces concepts. L'objectif à terme est de construire une BD d'incidents qui sera exploitée selon le même principe de construction des STA pour élaborer des scénarios de défaillance. Cela va nous permettre d'évaluer l'utilisabilité de SAFE-Next pour construire d'abord une ontologie multi-vues et aller jusqu'à construire des bases de données multi-vues.

Application de SAFE-Next dans d'autres domaines

SAFE-Next a prouvé son efficacité et ses apports en accidentologie (cf. section 12.2). Comme on peut le vérifier à travers l'architecture d'ASMEC (cf. section 8.4) et notamment la représentation UML de cette architecture (cf. Figure 8-3, page 102), SAFE-Next permet de modéliser et d'opérationnaliser des connaissances indépendamment du domaine dans lequel elle a été développée (i.e. l'accidentologie). Nous pensons alors qu'elle pourrait avoir des contributions significatives dans d'autres domaines comme le marketing, la maintenance, l'affectation de prêts par les banques, etc. En marketing, par exemple, le traitement des bases de données clients par SAFE-Next permettra de créer des profils de clients (l'équivalent au STA). Cela peut aider les entreprises à mieux comprendre les besoins de leurs clients, adapter leurs produits, spécifier leur communication envers les clients, etc. En maintenance, SAFE-Next pourrait être appliquée pour créer des catégories de défaillances à partir d'une base d'incidents. Cela permettrait aux entreprises de mieux définir les contres mesures adéquates.

Application d'ASEM pour l'évaluation d'autres types de modèles

Nous avons appliqué notre modèle d'évaluation de modèle, ASEM, uniquement sur les modèles que nous avons réalisés en accidentologie. Dans une perspective de généralisation, il est indispensable de l'appliquer pour évaluer d'autres types de modèles. Cette perspective est en cours de réalisation : en collaboration avec d'autres chercheurs du Laboratoire Génie Industriel (LGI), nous allons considérer d'autres types de modèles (en mathématique, en physique, etc.) et nous allons appliquer ASME.

Intégrer les différents outils développés dans AICEF dans un seul SBC

Pour l'incorporation des connaissances du domaine dans le processus d'ECD, AICEF utilise plusieurs techniques et outils (treillis de Galois, grille systémique, interprétation multi-niveaux d'abstraction, interface de sélection multi-vues des attributs, etc.). Ces outils doivent être intégrés dans un seul système pour faciliter leur utilisation. Cette perspective est en cours de réalisation.

Développer un outil de formation en accidentologie basé sur SAFE-Next

Nous avons montré dans les contributions (cf. section 12.2.1) comment SAFE-Next a fourni un moyen de capitalisation des connaissances en regroupant et organisant les modèles existants en accidentologie, en formalisant et préservant des connaissances implicites, en mutualisant et homogénéisant les perceptions des experts et en offrant une vision synthétique des différents types de connaissances et selon différents points de vue. Tous ces atouts offrent la possibilité de développer un outil de formation en accidentologie basé sur SAFE-Next. Cette perspective est en cours de discussion avec le CEESAR et le LAB. Cet outil aura pour vocation de faciliter aux nouveaux utilisateurs l'accès à des connaissances cruciales pour accélérer leur formation.

Elargir ASAIC pour analyser l'impact de changement sur le processus de conception ainsi que sur le processus d'évaluation

Comme nous l'avons montré dans la section 10.6.1, un changement au niveau du système CVE (e.g. introduction d'un nouveau système de sécurité) n'affecte pas uniquement le comportement de ce système, mais également le *processus de conception des véhicules* chez les concepteurs ainsi le *métier de l'accidentologie* au LAB. Dans la section 10.6.4, nous avons montré comment SAFE-Next permet, grâce à ASAIC, d'anticiper ou d'évaluer a posteriori les impacts d'un changement selon les trois points de vue « *processus de conception* », « *processus d'utilisation* » et « *processus d'évaluation* ». Seule l'analyse d'impact selon le point de vue « *processus d'utilisation* », qui correspond au comportement du système CVE, a été appliquée et validée dans nos travaux à travers un cas d'étude (évaluation de l'impact de l'ESP sur le système CVE). Il est donc souhaitable d'appliquer ASAIC sur les deux autres points de vue qui correspondent au *processus de conception des véhicules* et au *métier de l'accidentologie*. Dans la section 10.6.4, nous avons posé uniquement le cadre d'un modèle intégré en sécurité routière pour l'analyse de changement. La réalisation complète de ce modèle est au stade de définition au sein du LAB.

Références bibliographiques

- [Agrawal *et al.*, 1993] R. Agrawal, T. Imielinski et A. Swami. Mining Associations between Sets of Items in Massive Databases. *ACM-SIGMOD, Conference on Management of Data*, Washington, 1993.
- [Ahlberg et Wistrand, 1995] C. Ahlberg et E. Wistrand. IVEE : An information Visualization & Exploration Environment. *Proceedings on Information Visualization*, p 66-73, 1995.
- [Anderberg, 1973] M. R. Anderberg. Cluster Analysis for application, Academic Press, New York, 1973.
- [Angele *et al.*, 1996] J. Angele, D. Fensel et R. Studer. Domain and task modeling in MIKE, Chapman & Hall, 1996.
- [Anjewierden et Schreiber, 2004] A. Anjewierden et A. T. Schreiber. *CML2 Syntax*. <http://www.swi.psy.uva.nl/projects/kads22/cml2doc.html>, 2004.
- [Ashby, 1965] W. R. Ashby. An introduction to cybernetics, ed. Hall, C., London, 1965.
- [Aussenac-Gilles *et al.*, 1992] N. Aussenac-Gilles, J.-P. Krivine et J. Sallantin. L'acquisition des connaissances pour les systèmes à base de connaissances. *Editorial de la Revue Intelligence Artificielle*, vol. 6(1-2), p. 7-18, 1992.
- [Aussenac-Gilles et Matta, 1994] N. Aussenac-Gilles et N. Matta. Making a method of problem solving explicit with MACAO. *International Journal of Human-Computer Studies*, vol. 40, p. 193-219, 1994.
- [Baker, 1960] J. S. Baker. Experimental case studies of Traffic Accidents., Traffic Institute, Northwestern University, Evanston, Illinois, 1960.
- [Ball et Hall, 1967] G. H. Ball et D. J. Hall. A clustering techniques for summarizing multivariate data. *Behavioral Sciences*, vol. 12, p. 153-155, 1967.
- [Ballay, 1999] J. F. Ballay. Le rôle de l'information et des hommes dans la gestion des connaissances. Gestion des documents et gestion des connaissances. *Hermès science-publications*, vol. 3(3-4), p. 301-320, 1999.
- [Ballay, 1997] J.-F. Ballay. Capitalisation collective du savoir-faire. l'expérience du projet diademe à EDF. *Actes du Deuxième Congrès Industriel*, Albi, EDF-GDF., 1997.
- [Bar et Page, 2002] F. Bar et Y. Page. An empirical classification of lane departure crashes for the identification of relevant counter-measures. *46th AAAM Annual Conference*, Tempe, Arizona, 2002.
- [Barbut et Monjardet, 1970] M. Barbut et B. Monjardet. Ordre et Classification, Hachette Université, Paris, 1970.
- [Ben Ahmed *et al.*, 2002] W. Ben Ahmed, M. Mekhilef, M. Bigand et Y. Page. *Revue des méthodes et techniques de classification automatique en data mining*, Cahier de Recherche, LGI, Ecole Centrale Paris, 2002.
- [Ben Ahmed *et al.*, 2003] W. Ben Ahmed, M. Mekhilef, M. Bigand et Y. Page. Intégration des connaissances du domaine pour la fouille de données complexes. *Extraction et Gestion de Connaissances*, Université Blaise Pascal, Clermont Ferrand, France, 2003.
- [Ben Ahmed *et al.*, 2004a] W. Ben Ahmed, M. Mekhilef, M. Bigand et Y. Page. An Integrated Framework For Engineering Change Impact Analysis In Safety System Design Process. *5th International Conference On Integrated Design And Manufacturing In Mechanical Engineering*, University of Bath, Bath, UK, 2004a.
- [Ben Ahmed *et al.*, 2004b] W. Ben Ahmed, M. Mekhilef, M. Bigand et Y. Page. A Knowledge-Based System For Change Impact Analysis. *ASME, Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Salt Lake City, Utah, USA, 2004b.
- [Ben Ahmed et Yannou, 2003] W. Ben Ahmed et B. Yannou. Polysemy of Values or Conflict of Interests: A Multi-Disciplinary Analysis. *International Journal of Value-Based Management*, Kluwer Academic Publishers, vol. 16, p. 153-179, 2003.

- [Ben Hassine, 2004] M. Ben Hassine. *Contribution à la définition d'un système à base de connaissances pour l'élaboration des scénarios d'accidents*, Rapport interne, LAB PSA Peugeot-Citroën Renault, 2004.
- [Benzécri, 1973] J.-P. Benzécri. *L'analyse des données, tome 1: La taxonomie. Tome 2: L'analyse des correspondances* (2de. Éd. 1976), Dunod, Paris, 1973.
- [Bertalanffy, 1969] L. v. Bertalanffy. *General system theory: foundations, development, applications*, George Braziller, New York, 1969.
- [Bigand *et al.*, 1999] M. Bigand, D. Ndiaye, D. Corbeel et J. P. Bourey. Un référentiel pour les systèmes flexibles de production manufacturières. *Revue Internationale d'Ingénierie des systèmes de Production Mécanique*, vol. 3, p. 3-10, 1999.
- [Blessing, 1996] L. T. M. Blessing. Design Process Capture and Support. *Proceedings of the 2nd Workshop on Product Structuring*, Delft, 1996.
- [Booch *et al.*, 1999] G. Booch, R. J. et J. I. Unified Modelling Language User Guide, Addison Wesley Professional, 1999.
- [Bose, 1998] P. Bose. Change Analysis in an Architectural Model: A Design Rationale Based Approach. *Proceeding of the third international workshop on software architecture*, Orlando, Florida, USA, 1998.
- [Brachman et Anand, 1996] R. Brachman et T. Anand. The Process of Knowledge Discovery in Databases: A Human-Centered Approach. *In Advances in Knowledge Discovery and Data Mining, 37-58*, eds. U. Fayyad, G. Piatet-sky- Shapiro, P. Smyth, and R. Uthurusamy, Menlo Park, Calif, 1996.
- [Brenac, 1997] T. Brenac. *L'analyse séquentielle de l'accident de la route: comment la mettre en pratique dans les diagnostics de sécurité routière, Outil et méthode*, Rapport de recherche n°3, INRETS, 1997.
- [Brenac et Fleury, 1999] T. Brenac et D. Fleury. Le concept de scénario type d'accident de la circulation et ses applications. *Recherche Transport Sécurité*, vol. 63, p. 63-77, 1999.
- [Brenac et Megherbi, 1996] T. Brenac et B. Megherbi. Diagnostic de sécurité routière sur une ville : intérêt de l'analyse fine de procédures d'accidents tirées aléatoirement. *Recherche Transport Sécurité*, vol. 52, p. 59-71, 1996.
- [Breuker et Van de Velde, 1994] J. Breuker et W. Van de Velde. CommonKADS Library For Expertise Modeling, Reusable Problem Solving Components, IOS Press, Amsterdam, 1994.
- [Bruaux, 2001] S. Bruaux. *Modélisation des connaissances appliquée au calage d'un code de simulation des trajectoires de navires*. Rapport de DEA d'informatique, Laboratoire de Recherche en Informatique Appliquée, Université de Picardie, 2001.
- [Bruaux *et al.*, 2003] S. Bruaux, G. Kassel et G. Morel Étude critique de la méthode CommonKADS, application au calage de codes de calcul. *Ingénierie des Connaissances, IC'03*, Laval, France., 2003.
- [Buckingham Shum, 1997] S. Buckingham Shum. Balancing Formality With Informality: User Centred Requirement For Knowledge Management Technologies. *Artificial Intelligence In Knowledge Management, AAAI Spring Symposium*, Stand Ford University, Knowledge Media Institute, 1997.
- [Buckingham Shum et Sylvain, 2000] S. Buckingham Shum et M. Sylvain. Rapid Knowledge Construction: a Case Study in a Corporate Contingency Planning Using Collaborative Hypermedia. *Knowledge Management Beyond the Hype*, Birmingham, UK, 2000.
- [Calvanese *et al.*, 2001] D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi et R. Rosati. Data integration in data warehousing. *Int. Journal. of Cooperative Information Systems*, vol. 10, p. 237-271, 2001.
- [Cameron, 1997] M. Cameron. *Accident Data Analysis to Develop Target Groups for Countermeasures*, Rapport N°46 & 47, Monash University Accident Research Centre, 1997.

- [Campbell, 1974] D. T. Campbell. Evolutionary Epistemology. in *Schilpp P.A. (ed.), The Philosophy of Karl Popper*, p. 413-463, Open Court Publish., La Salle, Ill, 1974.
- [Cantzler, 1996] O. Cantzler. *Une architecture conceptuelle pour la pérennisation d'histoire globaux de conception de produits industriels complexes*. PhD. Thesis, Laboratoire Génie Industriel, Ecole Centrale Paris, 1996.
- [Caroll, 1995] J. M. Caroll. Scenario-Based Design: Envisioning Work and Technology in System Development, John Wiley and Sons, New York, 1995.
- [Caroll, 1998] J. M. Caroll. Scenario-Based Design. in *Helander M., Landauer T.K.,Prabhu P., Handbook of Human-Computer Interaction. 2nd edition, Ch. 17, p.*, North-Holland, Amsterdam, 1998.
- [Carré et Dekker, 1990] B. Carré et L. Dekker. The point of view notion for multiple inheritance. *ACM ECOOP/OOPSLA'90 Conference, Ottawa, Canada, 1990*.
- [CEESAR, 2000] CEESAR. *Etudes Détaillées d'Accident (EDA). Livre de codage de l'information dans les bases de données*, Rapport interne, CEESAR Centre Européen d'Etude de Sécurité et Analyse des Risques, 2000.
- [Cellier, 1990] J.-M. Cellier. L'erreur humaine dans le travail, In Leplat J., De terssac G. Les facteurs humains de la fiabilité dans les système complexes, Octares, 1990.
- [Charlet *et al.*, 2000] J. Charlet, M. Zacklad, G. Kassel et D. Bourigault. Ingénierie des connaissances, évolution récentes et nouveaux défis, Eyrolles, Paris, 2000.
- [Chovan *et al.*, 1994] J. D. Chovan, L. Tijerina, J. H. Everson, J. A. Pierowicz et D. L. Hendricks. *Examination of Intersection, Left Turn Across Path Crashes and Potential IVHS Countermeasures.*, Rapport N° : DOT HS 808 154, National Highway Traffic Safety Administration, 1994.
- [Compte-Sponville, 1998] A. Compte-Sponville. Philosophie de la valeur. *Valeur, Marché et Organisation*, Nantes, p 15-26, 1998.
- [Condillac, 1776] E. B. Condillac. Le commerce et le Gouvernement considérés relativement l'un à l'autre, Paris, 1776.
- [David *et al.*, 1993] J.-M. David, J.-P. Krivine et R. Simmons. Second Generation Expert Systems, Springer Verlag, Bonn, Allemagne, 1993.
- [De Keyser, 1985] V. De Keyser. L'erreur humaine. *La recherche*, vol. 20(216), p. 1444-1455, 1985.
- [Dhar et Tuzhilin, 1993] V. Dhar et A. Tuzhilin. Abstract-driven pattern discovery in databases. *IEEE Transactions on Knowledge and Data Engineering*, vol. 6, p. 926-938, 1993.
- [Diday, 1971] E. Diday. La méthode des nuées dynamiques. *Revue statistique Appliquée*, vol. 19(2), p. 19-34, 1971.
- [Dieng, 1997] R. Dieng. *Comparison of Conceptual Graphs for Modelling Knowledge of Multiple Experts : Application to Traffic Accident Analysis*, Projet Acacia, Rapport de recherche, n°3161, INRIA, 1997.
- [Dieng *et al.*, 1996] R. Dieng, A. Giboin, C. Amerge, O. Corby, S. Despres, L. Alpay, S. Labidi et S. Lapalut Building of a Corporate Memory for Traffic Accident Analysis. *Proceedings of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW'96)*, Banff, Canada, p 35.31-35.20, 1996.
- [Dudezert, 2003] A. Dudezert. *La valeur des connaissances en entreprise : recherche sur la conception de méthodes opératoires d'évaluation des connaissances en organisation*. Thèse de Doctorat, LGI, Ecole Centrale Paris, 2003.
- [Dunham, 1996] R. Dunham. Knowledge: the new business battleground. *KM Metazine*, vol. 1, 1996.
- [Duribreux-Cocquebert et Houriez, 2000] M. Duribreux-Cocquebert et B. Houriez. Application industrielle d'une approche mixte de modélisation des connaissances. in *Charlet et al. (Eds.), Ingénierie des connaissances, évolution récentes et nouveaux défis*, p. 25-42, Eyrolles, Paris, 2000.

- [Ermine *et al.*, 1996] J.-L. Ermine, M. Chaillot, P. Bigeon, B. Charreton et D. Malavieille. MKSM : Méthode pour la gestion des connaissances. *Ingénierie des systèmes d'information, AFCET Hermès*, vol. 4(4), p. 541-575, 1996.
- [Everitt, 1993] B. Everitt. *Cluster Analysis* (3rd ed.), Edward Arnold, London, 1993.
- [Famili *et al.*, 1997] A. Famili, W. M. Shen, R. Weber et E. Simoudis. Data Preprocessing and Intelligent Data Analysis. *Intelligent Data Analysis*, vol. 1, p. 3-23, 1997.
- [Fayyad *et al.*, 1996] U. Fayyad, G. Piatetsky-Shapiro et P. Smyth. The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications Of The ACM*, vol. 39(11), p. 27-34, 1996.
- [Fensel *et al.*, 1991] D. Fensel, J. Angele et D. Landes. Knowledge Representation and Acquisition Language (KARL). *Proceedings of the 11th International Workshop on Expert Systems and Their Applications, Tools and Techniques*, Avignon, France, p 821-833, 1991.
- [Fensel et Van Harmelen, 1994] D. Fensel et F. Van Harmelen. A comparison of languages which operationalize and formalize KADS models of Expertise. *The Knowledge Engineering Review*, vol. 9(2), p. 105-146, 1994.
- [Ferrandez, 1995] F. Ferrandez. L'étude détaillée d'accident orientée vers la sécurité primaire, Presses des Pont et Chaussées, 1995.
- [Ficher, 1936] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Of Eugenics*, vol. 7, p. 179-188, 1936.
- [Filip *et al.*, 2004] C. Filip, G. Wets et I. Thomas. Road Traffic Accident Clustering With Categorical Attributes. *83rd Annual Meeting*, Washington D.C., 2004.
- [Fleury, 1998] D. Fleury. Sécurité et urbanisme, la prise en compte de la sécurité routière dans l'aménagement urbain, Presse de l'école nationale des ponts et chaussées, 1998.
- [Fleury *et al.*, 1991] D. Fleury, C. Fline et J. F. Peytavin. Diagnostic local de sécurité, outils et méthodes, Editions SETRA, Collection Etudes de sécurité, Bagnaux, 1991.
- [Foret-Bruno *et al.*, 2001] J.-Y. Foret-Bruno, J.-F. Huère et Y. Page. *Projet de recherche E.LY.S.A. : Epidémiologie Routière et Analyse des Situations Accidentelles*, Rapport interne, LAB, 2001.
- [Forgy, 1965] E. W. Forgy. Cluster analysis of multivariate data : efficiency versus interpretability of classification. *Biometrics*, vol. 21(3), p. 768, 1965.
- [Gandon et Dieng, 2001] F. Gandon et R. Dieng. Ontologie pour un système multi-agents dédié à une mémoire d'entreprise. *Ingénierie des Connaissances, IC'2001*, Grenoble, France, p 1-21, 2001.
- [Garrido De Ceita, 2002] A. Garrido De Ceita. Extension de la méthodologie CommonKADS pour la spécification du comportement des systèmes réactifs. *Journées des Doctorants du LSIS (UMR CNRS 6168)*, Marseille, France, 2002.
- [Geurts *et al.*, 2003] K. Geurts, G. Wets, T. Brijs et K. Vanhoof. *Clustering and profiling traffic roads by means of accident data*, Rapport N° RA-2003-27, Limburg University, 2003.
- [Gorti *et al.*, 1998] S. R. Gorti, A. Gupta, G. J. Kim, R. D. Sriram et A. Wong. An Object-Oriented Representation for Product and Design Processes. *Computer-Aided Design*, vol. 30(7), p. 489-501, 1998.
- [Grinstein *et al.*, 2001] G. Grinstein, M. Trutschl et U. Cvek High-Dimensional Visualizations. *Proceedings of the Visual Data Mining Workshop, KDD'2001*, 2001.
- [Halkidi *et al.*, 2001] M. Halkidi, Y. Batistakis et M. Vazirgiannis. Clustering algorithms and validity measures. *Proceedings of the SSDBM Conference*, Virginia, USA, 2001.
- [Halkidi *et al.*, 2002] M. Halkidi, Y. Batistakis et M. Vazirgiannis. Cluster Validity Methods: Part I. *SIGMOD Record*, vol. 31(2), p. 40-45, 2002.

- [Han et Fu, 1996] J. Han et Y. Fu. Attribute-oriented induction in data mining. *Advances in Knowledge Discovery*, Cambridge, MA, p 399-421, 1996.
- [Han et Kamber, 2000] J. Han et M. Kamber. *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2000.
- [Haumer et al., 1998] P. Haumer, P. Heymans et K. Pohl. *An Integration of Scenario-Based Requirements Elicitation and Validation Techniques*, Technical Report, CREWS Report Series No. 98-28, RWTH Aachen, LS Informatik V, 1998.
- [Hellerstein, 1997] J. M. Hellerstein. Special Issue on Data Reduction Techniques. *Bulletin of the Technical Committee on Data Engineering*, vol. 20(4), 1997.
- [Henry, 2003] G. T. Henry. Influential Evaluation. *American Journal of Evaluation*, vol. 24(4), p. 515-524, 2003.
- [Henry et Mark, 2003] G. T. Henry et M. M. Mark. Beyond use: Understanding assessment's influence on attitudes and actions. *American Journal of Evaluation*, vol. 24(3), p. 293-314, 2003.
- [Heylighen, 1993] F. Heylighen. Selection Criteria for the Evolution of Knowledge. in: *Proc. 13th Int. Congress on Cybernetics, Association Internat. de Cybernétique*, Namur, 1993.
- [Heylighen, 1997] F. Heylighen. Objective, subjective and intersubjective selectors of knowledge. *Evolution and Cognition*, vol. 3(1), p. 63-67, 1997.
- [Heylighen, 2003] F. Heylighen. *Web Dictionary of Cybernetics and Systems*. 2003.
- [Heylighen et Joslyn, 2001] F. Heylighen et C. Joslyn. Cybernetics and Second Order Cybernetics. in *R.A. Meyers (ed.), Encyclopedia of Physical Science & Technology (3rd ed.)*, p. 155-177, Academic Press, New York, 2001.
- [Huang, 1997a] Z. Huang. Clustering Large Data Sets with Mixed Numeric and Categorical Values. *Proceedings of The First Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Singapore, 1997a.
- [Huang, 1997b] Z. Huang. A fast clustering algorithm to cluster very large categorical data sets in data mining. *SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, Tucson, Arizona, 1997b.
- [ISDF, 1994] ISDF. Institut de sûreté de fonctionnement, L'état de l'art dans le domaine de la fiabilité humaine, Octares, 1994.
- [Istennes, 1997] Z. Istennes. *ZOLA: un langage réflexif pour représenter et opérationnaliser des modèles conceptuels*. Thèse de Doctorat, Département Informatique, Université de Nantes, 1997.
- [Jacob-delouis et Krivine, 1995] I. Jacob-delouis et J.-P. Krivine. LISA: un langage rééflexif pour opérationnaliser les modèles d'expertise. *Revue d'Intelligence Artificielle*, vol. 9, p. 53-88, 1995.
- [Jain et Dubes, 1988] A. K. Jain et R. C. Dubes. *Algorithms for Clustering Data*, Prentice Hall, 1988.
- [Jain et al., 1999] A. K. Jain, M. Narasimha Murty et P. J. Flynn. Data clustering: A review. *ACM Computing Surveys*, vol. 31(3), p. 264-323, 1999.
- [Jarke et al., 1998] M. Jarke, T. Bui et J. M. Carroll. Scenario management: an interdisciplinary approach. *Requirements Engineering*, vol. 3(4), p. 155-173, 1998.
- [Jouini et al., 2003] W. Jouini, W. Ben Ahmed, M. Mekhilef, M. Bigand et Y. Page. *Les méthodes et techniques d'Extraction de Connaissances de Bases de données*, Cahier De Recherche, LGI, Ecole Centrale Paris, 2003.
- [Karbach et VoB, 1993] W. Karbach et A. VoB. MODEL-K for prototyping and strategic reasoning at the knowledge level. in *David, J.M., Krivine, J.P., and Simmons, R., (Eds.), Second Generation Expert Systems*, p. 721-745, Springer Verlag, Berlin, 1993.
- [Kassagi, 2002] M. Kassagi. *Caractérisation expérimentale du comportement des conducteurs en situation d'urgence pour la spécification de systèmes de sécurité active*. Thèse de doctorat, Laboratoire Génie Industriel, Ecole Centrale Paris, 2002.

- [Kassel *et al.*, 2000] G. Kassel, M.-H. Abel et C. Gréboval-Barry. Programmer au niveau connaissance en DEF-*. in *Charlet et al. (Eds), Ingénierie des connaissances, évolution récentes et nouveaux défis*, p. 145-160, Eyrolles, Paris, 2000.
- [Kaufman et Rousseeuw, 1990] L. Kaufman et P. Rousseeuw. *Finding Groups in Data : An Introduction to Cluster Analysis*, John Wiley and Sons, 1990.
- [Kazman *et al.*, 1996] R. Kazman, G. Abowd, L. Bass et P. Clements. Scenario-Based Analysis of Software Architecture. *IEEE Software*, vol. 13(16), p. 47-55, 1996.
- [Keim et Kriegel, 1996] D. A. Keim et A. P. Kriegel. Visualization techniques for mining large databases: a comparison. *IEEE transaction on knowledge and data Engineering*, vol. 8(6), p. 923-938, 1996.
- [Kirkhart, 2000] K. Kirkhart. Reconceptualizing evaluation use: An integrated theory of influence. in *V. Caracelli & H. Preskill (Eds.) The expanding scope of evaluation use. New directions for evaluation*, p. 5-24, San Francisco: Jossey-Bass, 2000.
- [Klein, 1993] M. Klein. Capturing Design Rationale in Concurrent Engineering Teams. *IEEE Computer*, vol. 26(1), p. 39-47, 1993.
- [Kurucz *et al.*, 1977] Kurucz, Morrow, Fogarty, Janicek et Klapper. The Effectiveness of ABS in Real Life Accidents. *14th international technical conference on Enhancing Safety Vehicles*, Munich, Allemagne, 1977.
- [Kurucz, 1977] M. F. Kurucz, J. The Effectiveness of ABS in Real Life Accidents. *14th international technical conference on Enhancing Safety Vehicles*, Munich, Allemagne, 1977.
- [Labidi, 1995] S. Labidi. *Ingénierie de la connaissance dans le cadre de projets multi-experts: méthodes, techniques et outils*. Thèse de Doctorat, Université de Nice - Sophia Antipolis, 1995.
- [Lamiri *et al.*, 2004] M. Lamiri, W. Ben Ahmed, M. Mekhilef, M. Bigand et Y. Page. *Interprétation des résultats issus classification automatique*, Cahier de Recherche, LGI, Ecole Centrale Paris, 2004.
- [Le Coz et Page, 2003] J.-Y. Le Coz et Y. Page. La démarche accidentologique au service de l'évolution de la sécurité des véhicules. *Revue de la gendarmerie Nationale*, vol. 207, 2003.
- [Le Moigne, 1974] J. L. Le Moigne. *La théorie du système général*, P. U. F., Paris, 1974.
- [Le Moigne, 1999] J.-L. Le Moigne. *La modélisation des systèmes complexes*, Dunod, 1999.
- [Le Roux, 1994] B. Le Roux. *Eléments d'une approche constructive de la modélisation et de la réutilisation en acquisition de connaissances*. Université Paris VI, 1994.
- [Lebart *et al.*, 1997] L. Lebart, A. Morineau et M. Piron. *Statistique exploratoire multidimensionnelle*, 2ème éd., Dunod, Paris, 1997.
- [Leite *et al.*, 2000] J. Leite, J. Doorn et K. GN. A Scenario Construction Process. *Requirements Engineering*, vol. 5, p. 38-61, 2000.
- [Leplat, 1985] J. Leplat. *Erreur humaine, fiabilité humaine*, Armand Colin, 1985.
- [Leplat et Hoc, 1992] J. Leplat et J. M. Hoc. Tâche et activité dans l'analyse psychologique des situations. in *Leplat J., L'analyse du travail en psychologie ergonomique- Tome 1*, p., Octares, 1992.
- [Lerman, 1970] I. C. Lerman. *Les bases de la classification automatique*, Gauthier-Villars, Paris, 1970.
- [Linster, 1992] M. Linster. *Knowledge Acquisition Based on Explicit Methods of Problem Solving*. Thèse de Doctorat, University of Kaiserslautern, 1992.
- [Longueville, 2003] B. Longueville. *Capitalisation des processus de décision dans les projets d'innovation : Application à l'automobile*. Thèse de Doctorat, Laboratoire Génie Industriel, Ecole Centrale Paris, 2003.
- [Ma *et al.*, 2003] S. Ma, B. Song, W. Feng Lu et C. F. Zhu. A Knowledge-Supported System For Engineering Change Impact Analysis. *Proceedings of DETC'03 ASME 2003, Design Engineering Technical Conferences*, Chicago, Illinois, USA, 2003.

- [MacQueen, 1967] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. *Proc. Symp. Math. Stat. And probability (5th)*, Univ. Of California, Berkeley, p 281-297, 1967.
- [Madhavji Nazim, 1992] H. Madhavji Nazim. Environment Evolution: The Prism Model of Changes. *IEEE Transaction on Software Engineering*, vol. 18(5), p. 380-392, 1992.
- [Malaterre et Fontaine, 1992] G. Malaterre et H. Fontaine. Les aides à la conduite : quels enjeux pour la sécurité ? *Recherche Transport Sécurité*, vol. 35, p. 43-53, 1992.
- [Manco *et al.*, 2004] G. Manco, C. Pizzuti et D. Talia. Eureka!: an interactive and visual knowledge discovery tool. *Journal of Visual Languages & Computing*, vol. 15(1), p. 1-35, 2004.
- [Marca et McGowan, 1987] D. A. Marca et C. L. McGowan. SADT, Structured Analysis and Design Technique, McGraw-Hill, New York, 1987.
- [Marcaillou, 1995] S. Marcaillou. *Intégration de la notion de points de vue dans la modélisation par objets - Le langage VBOOL*. Thèse de Doctorat, Université Paul Sabatier de Toulouse, 1995.
- [Marx, 1969] K. Marx. Value, Price and Profit. Edited by Eleanor Marx Aveling, International Publishers, New York, 1969.
- [Maturana et Varela, 2002] H. Maturana et F. Varela. The Tree of Knowledge: The Biological Roots of Human Understanding, Shambhala Publications, Boston, 2002.
- [McClean *et al.*, 2000] S. McClean, B. Scotney et M. Shapcott. Using background knowledge in the aggregation of imprecise evidence in databases. *Data & Knowledge Engineering*, vol. 32, p. 131-143, 2000.
- [Megherbi, 1999] B. Megherbi. *Scénarios types d'accidents de la circulation sur autoroute : élaboration, méthodes de reconnaissance et application pour le diagnostic et la prévention*. Thèse de Doctorat, Ecole nationale des ponts et chaussées, 1999.
- [Miller, 1995] J. G. Miller. Living Systems, University Press of Colorado, 1995.
- [Morin et Le Moigne, 1999] E. Morin et J. L. Le Moigne. L'intelligence de la complexité, L'Harmattan, Paris, 1999.
- [Motta, 1998] E. Motta. An Overview of the OCML Modelling Language. *8 th Workshop on Knowledge Engineering: Methods & Languages KEML 98*, Karlsruhe, Germany, 1998.
- [Motta *et al.*, 1990] E. Motta, T. Rajan et M. Eisenstadt. Knowledge Acquisition as a Process of Model Refinement. *Knowledge acquisition*, vol. 2(1), p. 21-49, 1990.
- [Murdock *et al.*, 1997] J. W. Murdock, S. Szykman et R. D. Sriram. An Information Modeling Framework to Support Design Databases and Repositories. *Proceedings of the ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Sacramento, California, USA, 1997.
- [Najm *et al.*, 2001] W. G. Najm, J. D. Smith et D. L. Smith. *Analysis of Crossing Path Crashes*, Rapport N° DOT HS 809 423, National Highway Traffic Safety Administration, 2001.
- [Nakache et Confais, 2000] J.-P. Nakache et J. Confais. Méthode de classification avec illustration SPAD et SAS, Centre International de Statistique et d'Informatique Appliquées (CISIA.CERESTA), Montreuil, 2000.
- [Nasri, 2003] S. Nasri. *Application du data mining en accidentologie*, Rapport interne, LAB PSA Peugeot-Citroën Renault, 2003.
- [Neboit, 1995] M. Neboit. Aides avancées à la conduite automobile. *Le Travail Humain*, vol. 58(2), 1995.
- [Newell, 1982] A. Newell. The Knowledge Level. *Artificial Intelligence*, vol. 18, p. 87-127, 1982.
- [Ng et Han, 1994] R. T. Ng et J. Han. Efficiency and effective clustering method for spatial data mining. *Int. Conf. Very Large Data Bases*, Santiago, Chile, p 144-155, 1994.

- [Nicolet *et al.*, 1989] J.-L. Nicolet, A. Carino et W. J.-C. Catastrophes ? Non Merci !, La prévention des risques technologiques et humains, Masson, 1989.
- [OMS, 2004] OMS. *Journée mondiale de la Santé : l'accident de la route n'est pas une fatalité !* <http://www.who.int/mediacentre/news/releases/2004/pr24/fr/>, 2004.
- [Page, 2002] Y. Page. *Elaboration de scénarios types d'accident pour le développement des systèmes de sécurité active embarqués dans les véhicules*, Rapport interne, LAB PSA Peugeot-Citroën Renault, 2002.
- [Page *et al.*, 2004] Y. Page, R. Driscoll, J.-Y. Le Coz et T. Hermitte. Combination of statistical and case-by-case approach for accident situations classification. *FISITA*, Spain, 2004.
- [Palmeri et Blalock, 2000] T. J. Palmeri et C. Blalock. The role of background knowledge in speeded perceptual categorization. *Cognition*, vol. 77, p. 45-57, 2000.
- [Perron, 1997] T. Perron. *Méthode d'analyse de sécurité primaire automobile pour la spécification fonctionnelle et l'évaluation prévisionnelle d'efficacité de systèmes d'évitement d'accidents*. Thèse de Doctorat, Laboratoire Génie Industriel, Ecole Centrale Paris, 1997.
- [Piaget, 1970] J. Piaget. *L'épistémologie génétique*, Presses Universitaires de France, Que sais-je?, Paris, 1970.
- [Pikosz et Malmqvist, 1998] P. Pikosz et J. Malmqvist. A Comparative Study of Engineering Change Management in Three Swedish Engineering Companies. *Proceedings of the 1998 ASME Design Engineering Technical Conferences*, Atlanta, USA, 1998.
- [Potts, 1995] C. Potts. Using schematic scenarios to understand user needs. *Proceedings of DIS'95 - Symposium on designing interactive systems: processes, practices and techniques*. ACM Press, University of Michigan, p 247-256, 1995.
- [Puerta *et al.*, 1992] A. R. Puerta, J. W. Egar, S. W. Tu et M. A. Musen. A multiple-method Knowledge acquisition shell for the automatic generation of Knowledge acquisition tools. *Knowledge acquisition*, vol. 4, p. 171-196, 1992.
- [Pyle, 1999] D. Pyle. *Data Preparation For Data Mining*, Morgan Kaufmann Publishers, 1999.
- [Ragland et Zabyshny, 2003] D. R. Ragland et A. A. Zabyshny. *Intersection Decision Support Project: Taxonomy of Crossing-Path Crashes at Intersections Using GES 2000 Data*, Rapport N°, Institute of Transportation Studies, U.C. Berkeley Traffic Safety Center, University of California, 2003.
- [Rahm et Do, 2000] E. Rahm et H. H. Do. Data Cleaning: Problems and Current Approaches. *IEEE Data Engineering Bulletin*, vol. 23(4), 2000.
- [Rasmussen, 1983] J. Rasmussen. Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, And Cybernetics*, vol. 13(3), p. 257-266, 1983.
- [Reason, 1993] J. Reason. *L'erreur humaine*, Presses universitaires de France, 1993.
- [Regli *et al.*, 2000] W. C. Regli, X. Hu et W. Sun. A Survey on Design Rationale Systems: Approaches, Representation, Capture and Retrieval. *Engineering With Computers*, vol. 16, p. 209-235, 2000.
- [Ribière, 1999] M. Ribière. *Représentation et gestion de multiples points de vue dans le formalisme des graphes conceptuels*. Thèse de doctorat, Ecole Doctorale-Science pour l'ingénieur, Université de Nice-Sophia Antipolis, 1999.
- [Richard, 2003] H. Richard. *Approaches to Inquiry - Key Concepts*. http://www.icra-edu.org/objects/anglolearn/Inquiry-Key_Concepts1.pdf, 2003.
- [Riviere, 2003] C. Riviere. *Configurations accidentelles, identification de prestations sécuritaires et évaluation a priori de l'efficacité de ces prestations*, Rapport d'avancement, LAB, PSA Peugeot Citroën/ Renault, 2003.
- [Sachinopoulou, 2001] A. Sachinopoulou. *Multidimensional Visualization*, Espoo, VTT Electronics, 2001.

- [Saporta, 1990] G. Saporta. *Probabilités, Analyse des données et Statistiques*, Technip, Paris, 1990.
- [Scheringer *et al.*, 2001] M. Scheringer, T. Vögl, J. Von Grote, B. Capaul, R. Schubert et K. Hungerbühler. Scenario-Based Risk Assessment of Multi-Use Chemicals: Application to Solvents. *Risk Analysis*, vol. 21(3), p. 481-497, 2001.
- [Schreiber *et al.*, 1994a] G. Schreiber, H. Akkermans, R. De Hoog, N. Shadbolt, W. Van de Velde et B. Wielinga. *Knowledge Engineering and Management: The CommonKADS methodology*, MIT Press, Cambridge, Massachusetts, London, 1994a.
- [Schreiber *et al.*, 1994b] G. Schreiber, B. Wielinga, H. Akkermans, W. Van de Velde et A. Anjewierden. CML: The CommonKADS Conceptual Modeling Language. *8th European Knowledge Acquisition Workshop EKAW'94*, Allemagne, p 283-300, 1994b.
- [Schreiber *et al.*, 1993] G. Schreiber, B. Wielinga et J. Breuker. KADS. A Principled Approach to Knowledge-Based System Development, Vol. 11 of Knowledge-Based Systems Book Series, Academic Press, London, 1993.
- [Schreiber *et al.*, 1994c] G. Schreiber, B. Wielinga, D. H. R., H. Akkermans et W. Van de Velde. CommonKADS : A comprehensive methodology for KBS development. *IEEE Expert*, vol. 9(6), p. 28-37, 1994c.
- [Smith, 1776] A. Smith. *An Inquiry into the Nature and Causes of the Wealth of Nations*, Whitestone, Dublin, 1776.
- [Sneath et Sokal, 1973] P. H. A. Sneath et R. P. Sokal. *Numerical Taxonomy*, Freeman, London, 1973.
- [Sohn et Lee, 2003] S. Y. Sohn et S. H. Lee. Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accident in Korea. *Safety Science*, vol. 41(1), p. 1-14, 2003.
- [Sohn et Shin, 2001] S. Y. Sohn et H. W. Shin. Pattern Recognition for Road Traffic Accident Severity in Korea. *Ergonomics*, vol. 44(1), p. 107-117, 2001.
- [Sowa, 1984] J. F. Sowa. *Conceptual Structures, information processing in mind and machine*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1984.
- [Studer *et al.*, 1998] R. Studer, V. R. Benjamins et D. Fensel. Knowledge Engineering: Principals and methods. *Data & Knowledge Engineering*, vol. 25, p. 161-197, 1998.
- [Sutcliffe, 1998] A. Sutcliffe. Scenario-based requirements analysis. *Requirements Engineering*, vol. 3(1), p. 48-65, 1998.
- [Szykman *et al.*, 1999] S. Szykman, J. W. Racz et R. D. Sriram. The Representation of Function in Computer-Based Design. *Proceedings of the 1999 ASME Design Engineering Technical Conferences*, Las Vegas, Nevada, 1999.
- [Tchounikine *et al.*, 2000] P. Tchounikine, Z. Istennes et F. Trichet. ZOLA: un langage permettant une approche flexible de l'opérationnalisation du modèle conceptuel d'un système à base de connaissances. in Charlet *et al.* (Eds), *Ingénierie des connaissances, évolution récentes et nouveaux défis*, p. 129-144, Eyrolles, Paris, 2000.
- [Trichet, 1998] F. Trichet. *DSTM : un environnement de modélisation et d'opérationnalisation de la démarche de résolution de problèmes d'un Système à Base de Connaissances*. Thèse de Doctorat en Informatique, Laboratoire IRIN, Université de Nantes, Faculté des Sciences et des Techniques, 1998.
- [Van Elslande et Alberton, 1997] P. Van Elslande et L. Alberton. *Scénarios-types de production de l'erreur humaine dans l'accident de la route, problématique et analyse qualitative*, Rapport de recherche N°218, INRETS, 1997.
- [Van Harmelen et Balder, 1992] F. Van Harmelen et J. Balder. (ML)²: A Formal Language for KADS Models of Expertise. *Knowledge Acquisition*, vol. 4, p. 127-161, 1992.
- [Van Helvert et Fowler, 2003] J. Van Helvert et C. Fowler. *Scenario based User Needs Analysis*, Chimera Working Paper N°2003-02, Colchester: University of Essex, 2003.

- [Van Lancker, 2004] L. Van Lancker. XML + XSL. <http://www.ccim.be/ccim328/xml/index.htm>, 2004.
- [Verzat, 2000] C. Verzat. *L'apprentissage collectif*, Support de cours DEA-GSI, LGI, Ecole Centrale Paris, 2000.
- [Vogel, 1989] C. Vogel. Knowledge Oriented Design (KOD): la mise en oeuvre, Editions Masson, 1989.
- [Von Foerster, 1995] H. Von Foerster. The Cybernetics of Cybernetics (2nd edition), Future Systems Inc., Minneapolis, 1995.
- [Von Foerster et Poerksen, 2002] H. Von Foerster et B. Poerksen. Understanding Systems: Conversation on Epistemology and Ethics, Kluwer/Plenum, New York, 2002.
- [Walras, 1874] L. Walras. *Éléments d'économie politique pure*, Paris, 1874.
- [Warner, 1981] J.-C. Warner. Le facteur humain dans la conduite de grands systèmes. *Le progrès Technique*, vol. 21, 1981.
- [Wetter, 1990] T. Wetter. First Order Logic Foundation of the KADS Conceptual Model. in *Wielinga, B. et al. Current Trends in Knowledge Acquisition*, p. 356-375, IOS Press, Amsterdam, 1990.
- [Wielinga et al., 1994] B. J. Wielinga, Y. Hans Akkermans, H. Hassan, O. Olsson, K. Orsvärn, G. Schreiber, P. Terpstra, W. Van de Velde et S. Wells. *Expertise Model Definition Document*, ESPRIT Project P5248 KADSII, Document Id. KADSII/M2/UvA/026/5.0, University of Amsterdam, 1994.
- [Wiener, 1948] N. Wiener. *Cybernetics or control and communication in the animal and the machine*, 2nd Edition, MIT Press, Cambridge, MA, 1948.
- [Wille, 1982] R. Wille. Restructuring Lattice Theory. in *Ivan Rival, editor: Symposium on Ordered Sets*, p. 445-470, University of Calgary, Boston, 1982.
- [Wong, 1982] M. A. Wong. A hybrid clustering method for identifying high density clusters. *Journal of American Statistic Association*, vol. 77, p. 841-847, 1982.
- [Zaït et Messatfa, 1997] M. Zaït et H. Messatfa. A comparative study of clustering methods. *Future Generation Computer System*, vol. 13, p. 149-159, 1997.
- [Zhang et al., 2002] L. Zhang, C. Tang, Y. Song, A. Zhang et M. Ramanathan. *VizCluster and Its Application on Clustering Gene Expression Data*, Department of Computer Science and Engineering, State University of New York at Buffalo, 2002.

Annexe

A-1. Modèle de connaissances de l'élaboration des STA en accidentologie

A-1.1 Connaissances du domaine

A-1.1.1. Schéma du domaine

```

KNOWLEDGE-MODEL Scénarios_accidents;

/* Modèle de connaissances en accidentologie pour l'élaboration de
   scénarios-Types d'accident de la route

   AUTHOR:                Walid Ben Ahmed, LAB PSA Peugeot Citrën Renault & LGI, Ecole Centrale
   Paris
   LAST MODIFIED: 25 May 2004
*/

/*****
/* I. Connaissances Domaine*/
*****/

DOMAIN-KNOWLEDGE accidentology-domain;

/*.....*/
/*      I.1- Schéma du domaine      */
/*.....*/

DOMAIN-SCHEMA élaboration_scénario-schema;

/*-----*/
/*      I.1.a- Types d'informations sur l'accident */
/*-----*/

CONCEPT accident;
  DESCRIPTION:
    "l'accident peut être décrit à travers interaction des troisi composantes: impliqués, véhicules et
    environnement";
  ATTRIBUTES:
    num accident : NATURAL;
END CONCEPT accident;

CONCEPT BD_EDA_accidents;
  DESCRIPTION:
    "Base de données d'accidents";
  ATTRIBUTES:
    num accident : NATURAL;
END CONCEPT BD_EDA_accidents;

BINARY-RELATION appartient;
  DESCRIPTION:
    "un accident appartient à la BD EDA";
  ARGUMENT-1: accident;
  CARDINALITY: 1+;
  ARGUMENT-2: BD_EDA_accidents;
  CARDINALITY: 1;
END BINARY-RELATION appartient;

CONCEPT attribut;

```

```
DESCRIPTION:
"l'accident peut être décrit à travers interaction des troisi composantes: impliqués, véhicules et
environnement";
ATTRIBUTES:
nom_attribut : STRING;
modalité : UNIVERSAL;
nombre_modalités : UNIVERSAL;
importance_générale : UNIVERSAL;
liste_actions: {supprimer, regrouper_modalités, supprimer_modalités};
END CONCEPT attribut;

BINARY-RELATION description_accident;
DESCRIPTION:
"une classe est regroupement d'accidents présentant une similitude dans leurs déroulements";
ARGUMENT-1: accident;
CARDINALITY: 1+;
ARGUMENT-2: attribut;
CARDINALITY: 1+;
ATTRIBUTES:
attribut : UNIVERSAL;
modalité : UNIVERSAL;
END BINARY-RELATION description_accident;

BINARY-RELATION est_caractérisé_par;
DESCRIPTION:
"une classe est regroupement d'accidents présentant une similitude dans leurs déroulements";
ARGUMENT-1: impliqué;
CARDINALITY: 1+;
ARGUMENT-2: attribut;
CARDINALITY: 1+;
END BINARY-RELATION est_caractérisé_par;

BINARY-RELATION est_caractérisé_par;
DESCRIPTION:
"une classe est regroupement d'accidents présentant une similitude dans leurs déroulements";
ARGUMENT-1: véhicule;
CARDINALITY: 1+;
ARGUMENT-2: attribut;
CARDINALITY: 1+;
END BINARY-RELATION est_caractérisé_par;

BINARY-RELATION est_caractérisé_par;
DESCRIPTION:
"une classe est regroupement d'accidents présentant une similitude dans leurs déroulements";
ARGUMENT-1: environnement;
CARDINALITY: 1+;
ARGUMENT-2: attribut;
CARDINALITY: 1+;
END BINARY-RELATION est_caractérisé_par;

BINARY-RELATION reconstruction_accident;
DESCRIPTION:
"une classe est regroupement d'accidents présentant une similitude dans leurs déroulements";
ARGUMENT-1: accidentologue;
CARDINALITY: 1+;
ARGUMENT-2: accident;
CARDINALITY: 1+;
ATTRIBUTES:
num_accident : STRING;
qualité_reconstr : STRING;
END BINARY-RELATION reconstruction_accident;

CONCEPT impliqué;
DESCRIPTION:
"les personnes impliquées dans l_accident";
ATTRIBUTES:
Nombre_tués : NATURAL;
Nombre_blessés_graves : NATURAL;
Nombre_blessés_légers : NATURAL;
Nombre_indemnes : NATURAL;
SUPER-TYPE-OF:
conducteur, piéton, occupant, blessé;
DISJOINT: YES;
COMPLETE: YES;
END CONCEPT impliqué;

CONCEPT piéton;
DESCRIPTION:
"les personnes impliquées dans l_accident";
END CONCEPT piéton;

CONCEPT occupant;
DESCRIPTION:
"les personnes impliquées dans l_accident";
END CONCEPT occupant;

CONCEPT blessé;
DESCRIPTION:
"les personnes impliquées dans l_accident";
```

```

END CONCEPT blessé;

CONCEPT conducteur;
  DESCRIPTION:
    "Il s'agit d'une table dans la BD EDA contenant la description du conducteur avant, durant et
    après l'accident";
  ATTRIBUTES:
    Conducteur_professionnel : oui_non-value;
    Catégorie_usager: {véhicule_léger, poids_lourds, véhicule_utilitaire, ambulance, taxi,
    Motocycliste, Cycliste, Piéton};
  HAS-PARTS:
    déplacement_en_cours; situation_urgence; expérience_conduite;
END CONCEPT conducteur;

CONCEPT état_conducteur;
  DESCRIPTION:
    "Il s'agit d'une description de l'état du conducteur avant, durant l'accident";
  SUPER-TYPE-OF:
    état_stable, facteurs_médicaux, état_transitoire, état_court_terme;
  DISJOINT: YES;
  COMPLETE: YES;
END CONCEPT état_conducteur;

VALUE-TYPE oui_non-value;
  TYPE: NOMINAL;
  VALUE-LIST: {oui, non};
END VALUE-TYPE oui_non-value;

CONCEPT déplacement_en_cours;
  DESCRIPTION:
    "Il s'agit d'une description de l'état du conducteur avant la conduite";
  ATTRIBUTES:
    Nationalité : STRING;
    Situation_famille : {Célibataire, Vie_commune, Marié, Séparé, Divorcé, Veuf};
    Date_naissance : STRING;
    Usager_local : oui_non-value;
END CONCEPT déplacement_en_cours;

CONCEPT état_stable;
  DESCRIPTION:
    "Il s'agit d'une description de l'état du conducteur avant la conduite";
  ATTRIBUTES:
    Nationalité : STRING;
    Situation_famille : {Célibataire, Vie_commune, Marié, Séparé, Divorcé, Veuf};
    Date_naissance : STRING;
    Usager_local : oui_non-value;
END CONCEPT état_stable;

CONCEPT facteurs_médicaux;
  DESCRIPTION:
    "Il s'agit d'une description des facteurs médicaux relatif aux conducteur au moment de
    l'accident";
  ATTRIBUTES:
    problème1_santé : problème-value;
    problème2_santé : problème-value;
END CONCEPT facteurs_médicaux;

VALUE-TYPE problème-value;
  TYPE: NOMINAL;
  VALUE-LIST: {Pas_problème, Problème_cardiaque, Problème_rénal, Problème_tension, Diabète,
    Epilepsie, tétanie, Handicap_physique, Troubles_mentaux, Insomnie_chronique, Maux, Vertiges,
    étourdissement, Crampes, Problèmes_auditifs, Problèmes_visuels};
END VALUE-TYPE problème-value;

CONCEPT état_transitoire;
  DESCRIPTION:
    "Il s'agit d'une description de l'état du conducteur juste avant la conduite";
  ATTRIBUTES:
    Alcoolisation : {absence_alcoolisation, Doute_alcoolisation, alcoolisation};
    Résultat_dépistage : REAL;
    Circonstances_alcoolisation : {A_la_maison, chez_quelqu_un, restaurant_ou_bar, Dans_véhicule,
    travail, plusieurs_endroits};
    Rôle_drogue : oui-non-vlaue;
    prise_médicaments : oui-non-vlaue;
END CONCEPT état_transitoire;

CONCEPT état_court_terme;
  DESCRIPTION:
    "Il s'agit d'une description de l'état du conducteur juste avant l'accident";
  ATTRIBUTES:
    Temps_depuis_le_réveil : STRING;
    Temps_depuis_dernier_repas : STRING;
    Consistance_dernier_repas : {Repas_onsistant, Collation};
    Rôle_suspecté_état_corporel : oui-non-value;
END CONCEPT état_court_terme;

CONCEPT déplacement_en_cours;
  DESCRIPTION:
    "Il s'agit d'une description des objectifs et des circonstances du déplacement en cours";

```

```
    ATTRIBUTES:
Motif_déplacement : STRING;
Sens_trajet1 : {Aller, Retour};
Distance_prévue : NATURAL;
Utilisation_balladeur : oui-non-value;
END CONCEPT état_court_terme;

CONCEPT situation_urgence;
DESCRIPTION:
  "Il s'agit d'une description de la situation quand le conducteur aurait dû entreprendre une
manœuvre de récupération";
ATTRIBUTES:
  Vitesse_déclarée : STRING;
  Perception_danger : oui-non-value;
  Action_réactive_déclarée : STRING;
END CONCEPT situation_urgence;

CONCEPT expérience_conduite;
DESCRIPTION:
  "Il s'agit d'une description de la situation quand le conducteur aurait dû entreprendre une
manœuvre de récupération";
SUPER-TYPE-OF:
  permis_conduire, nombre_infraction, accidents_précédents;
DISJOINT: YES;
COMPLETE: YES;
END CONCEPT expérience_conduite;

CONCEPT permis_conduire;
DESCRIPTION:
  "Il s'agit d'une description de la situation quand le conducteur aurait dû entreprendre une
manœuvre de récupération";
ATTRIBUTES:
  Validité_permis : STRING;
  Ancienneté_permis : STRING;
  Autres_permis : STRING;
END CONCEPT permis_conduire;

CONCEPT nombre_infraction;
DESCRIPTION:
  "Il s'agit d'une description de la situation quand le conducteur aurait dû entreprendre une
manœuvre de récupération";
ATTRIBUTES:
  Nombre_infractions_vitesse : NATURAL;
  Nombre_infractions_alcoolémie : NATURAL;
  Autres_infractions : STRING;
  Crédit_points_permis : NATURAL;
END CONCEPT nombre_infraction;

CONCEPT accidents_précédents;
DESCRIPTION:
  "Il s'agit d'une description de la situation quand le conducteur aurait dû entreprendre une
manœuvre de récupération";
ATTRIBUTES:
  Nombre_accidents_corporel_en_tant_que_non_responsable : NATURAL;
  Nombre_accidents_corporel_en_tant_que_responsable : NATURAL;
  Temps_écoulé_depuis_dernier_accident : NATURAL;
END CONCEPT accidents_précédents;

VALUE-TYPE ralisation-value;
TYPE: NOMINAL;
VALUE-LIST: {Non-réalisé, Réalisé};
END VALUE-TYPE ralisation-value;

CONCEPT accidentologue;
DESCRIPTION:
  "Identification de l'accidentologue qui a effectué le codage";
SUPER-TYPE-OF:
  expert_infrastructure, expert_véhicule, psychologue;
DISJOINT: YES;
COMPLETE: YES;
END CONCEPT accidentologue;

CONCEPT expert_infrastructure;
DESCRIPTION:
  "Identification de l'accidentologue qui a effectué le codage";
ATTRIBUTES:
  code_accidentologue : STRING;
  nom_accidentologue: STRING;
END CONCEPT expert_infrastructure;

CONCEPT expert_véhicule;
DESCRIPTION:
  "Identification de l'accidentologue qui a effectué le codage";
ATTRIBUTES:
  code_accidentologue : STRING;
  nom_accidentologue: STRING;
END CONCEPT expert_véhicule;

CONCEPT psychologue;
```

```

    DESCRIPTION:
    "Identification de l_accidentologue qui a effectué le codage";
    ATTRIBUTES:
    code_accidentologue : STRING;
    nom_accidentologue: STRING;
END CONCEPT psychologue;

BINARY-RELATION entretient;
    DESCRIPTION:
    "entretien effectué par l_accidentologue (psychologue) avec les conducteurs
    impliqués dans l_accident";
    ARGUMENT-1: psychologue;
    CARDINALITY: 1;
    ARGUMENT-2: conducteur;
    CARDINALITY: 1+;
    ATTRIBUTES:
    Premier_entretien: ralisation-value;
    Date_premier_entretien : STRING;
    Heure_premier_entretien : STRING;
END BINARY-RELATION entretiens;

BINARY-RELATION collectées_par;
    DESCRIPTION:
    "entretien effectué par l_accidentologue (psychologue) avec les conducteurs
    impliqués dans l_accident";
    ARGUMENT-1: attribut;
    CARDINALITY: 1;
    ARGUMENT-2: accidentologue;
    CARDINALITY: 1+;
END BINARY-RELATION collectées_par1;

CONCEPT véhicule;
    DESCRIPTION:
    "Il s_agit du véhicule avant et après le choc selon deux points de vue: sécurité primaire et
    sécurité secondaire";
    ATTRIBUTES:
    Type_véhicule : STRING;
    HAS-PARTS:
    véh_sécur_primaires; véh_sécur_secondaires;
END CONCEPT véhicule;

CONCEPT véh_sécur_primaires;
    DESCRIPTION:
    "Il s_agit d'une description orientée sécurité primaire du véhicule avant et après le choc";
    ATTRIBUTES:
    type_véhicule : STRING;
    HAS-PARTS:
    véhicule_immatricule; caractéristiques_séc_primaires; contrôle_technique; défauts;
END CONCEPT véh_sécur_primaires;

BINARY-RELATION est_propriétaire;
    DESCRIPTION:
    "Description du rapport du conducteur au véhicule";
    ARGUMENT-1: conducteur;
    CARDINALITY: 1+;
    ARGUMENT-2: véhicule;
    CARDINALITY: 1+;
    ATTRIBUTES:
    Ancienneté_achat : STRING;
    Type_achat : {Achat_neuf, Achat_occasion};
    Kilométrage_depuis_achat : STRING;
END BINARY-RELATION est_propriétaire;

CONCEPT véhicule_immatricule;
    DESCRIPTION:
    "Il s_agit d'une description multi-point de vue systémique de l_accident";
    ATTRIBUTES:
    Marque_véhicule : STRING;
    Puissance_fiscale : STRING;
    Couleur_véhicule : STRING;
END CONCEPT véhicule_immatricule;

CONCEPT caractéristiques_séc_primaires;
    DESCRIPTION:
    "Il s_agit d'une description du véhicule";
    ATTRIBUTES:
    Type_carburant : {Essence_carburateur, Gazole, GPL, Electrique, Essence_injection, Méthane};
    Véhicule_spécifique : oui-non-value;
    HAS-PARTS:
    Commandes; Spécificités_conception; Aides_conduite; Performances; Charge_au_moment_accident;
END CONCEPT caractéristiques_séc_primaires;

CONCEPT Commandes;
    DESCRIPTION:
    "Il s_agit d'une description des commandes disponibles dans le véhicule";
    ATTRIBUTES:
    Position_levier_vitesse : STRING;
    Boîte_automatique : STRING;

```

```
Présence_auto-radio : oui-non-value;
END CONCEPT Commandes;

CONCEPT Spécificités_conception;
DESCRIPTION:
  "Il s'agit d'une description des commandes disponibles dans le véhicule";
ATTRIBUTES:
  Position_poste_conduite : {Gauche, Droite};
  Quatre_roues_directrices : oui-non-value;
  Direction_assistée : oui-non-value;
  Turbo : oui-non-value;
END CONCEPT Spécificités_conception;

CONCEPT Aides_conduite;
DESCRIPTION:
  "Il s'agit d'une description des aides à la conduites présentes dans le véhicule";
ATTRIBUTES:
  Equipement_conducteur_handicapé : oui-non-value;
  Aide_conduite : {Dispositif_anti-blocage_roues, Dispositif_anti-patinage,
  Régulateur_vitesse, Ordinateur_bord};
END CONCEPT Aides_conduite;

CONCEPT Performances;
DESCRIPTION:
  "Il s'agit d'une description des performances mécaniques du véhicule";
ATTRIBUTES:
  Cylindrée : NATURAL;
  Puissance : NATURAL;
  Rapport_Puissance_Poids : NATURAL;
  Vitesse_pointe : NATURAL;
END CONCEPT Performances;

CONCEPT Charge_au_moment_accident;
DESCRIPTION:
  "Il s'agit d'une description du chargement du véhicule au moment de l'accident";
ATTRIBUTES:
  Nombre_occupants : NATURAL;
  Poids_total_occupants : NATURAL ;
  Charge_coffre : NATURAL;
END CONCEPT Charge_au_moment_accident;

CONCEPT contrôle_technique;
DESCRIPTION:
  "Il s'agit d'une description des contrôles techniques effectués sur le véhicule";
ATTRIBUTES:
  Contrôle_technique_effectué : oui-non-value;
  Date_dernier_contrôle_technique : STRING;
  Nature_contrôle : STRING;
END CONCEPT contrôle_technique;

CONCEPT défauts;
DESCRIPTION:
  "Il s'agit d'une description des défauts constatés par l'accidentologue sur le véhicule";
ATTRIBUTES:
  Etat_garniture_pédales : {Bon_état, Défaut};
  Etat_essuie-glaces : {Bon_état, Défaut};
  SUPER-TYPE-OF:
    défaut_Direction, défaut_Eclairage_Signalisation, défaut_Pneus,
    défaut_Suspension_Amortissement;
  DISJOINT: YES;
  COMPLETE: YES;
END CONCEPT défauts;

CONCEPT défaut_Direction;
DESCRIPTION:
  "Il s'agit d'une description des défauts de direction constatés par l'accidentologue sur le véhicule";
END CONCEPT défaut_Direction;

CONCEPT défaut_Eclairage_Signalisation;
DESCRIPTION:
  "Il s'agit d'une description des défauts d'éclairage et de signalisation constatés par l'accidentologue sur le véhicule";
END CONCEPT défaut_Eclairage_Signalisation;

CONCEPT défaut_Pneus;
DESCRIPTION:
  "Il s'agit d'une description des défauts des pneus constatés par l'accidentologue sur le véhicule";
END CONCEPT défaut_Pneus;

CONCEPT défaut_Suspension_Amortissement;
DESCRIPTION:
  "Il s'agit d'une description des défauts de Suspension et d'Amortissement constatés par l'accidentologue sur le véhicule";
END CONCEPT défaut_Suspension_Amortissement;
```



```

CONCEPT défaut_Suspension_Amortissement;
  DESCRIPTION:
  "Il s'agit d'une description des défauts de Suspension et d'Amortissement constatés par
  l'accidentologue sur le véhicule";
END CONCEPT défaut_Suspension_Amortissement;

CONCEPT véh_sécur_secondaire;
  DESCRIPTION:
  "Il s'agit d'une table dans la BD EDA contenant la description du conducteur avant, durant et
  après l'accident";
  ATTRIBUTES:
  type_véhicule : STRING;
  Catégorie_usager: {véhicule_léger, poids_lourds, véhicule_utilitaire, ambulance, taxi,
  Motocycliste, Cycliste, Piéton};
  HAS-PARTS:
  caractéristiques_séc_second; ouvrant; vitrages; sellerie;
END CONCEPT véh_sécur_secondaire;

CONCEPT caractéristiques_séc_second;
  DESCRIPTION:
  "Il s'agit d'une description du véhicule";
END CONCEPT caractéristiques_séc_second;

CONCEPT ouvrant;
  DESCRIPTION:
  "Il s'agit d'une description du véhicule";
END CONCEPT ouvrant;

CONCEPT vitrages;
  DESCRIPTION:
  "Il s'agit d'une description du véhicule";
END CONCEPT vitrages;

CONCEPT sellerie;
  DESCRIPTION:
  "Il s'agit d'une table dans la BD EDA contenant la description du conducteur avant, durant et
  après l'accident";
  SUPER-TYPE-OF:
  sellerie_avant, sellerie_secondaire, sellerie_arrière;
  DISJOINT: YES;
  COMPLETE: YES;
END CONCEPT sellerie;

CONCEPT sellerie_avant;
  DESCRIPTION:
  "Il s'agit d'une description du véhicule";
END CONCEPT sellerie_avant;

CONCEPT sellerie_secondaire;
  DESCRIPTION:
  "Il s'agit d'une description du véhicule";
END CONCEPT sellerie_secondaire;

CONCEPT sellerie_arrière;
  DESCRIPTION:
  "Il s'agit d'une description du véhicule";
END CONCEPT sellerie_arrière;

CONCEPT choc;
  DESCRIPTION:
  "Il s'agit d'une description des chocs";
  ATTRIBUTES:
  Nombre_chocs : NATURAL;
  Position_finale_véhicule : {Sur_toit, Sur_côté_droit, Sur_côté_gauche, Sur_roues};
  SUPER-TYPE-OF: choc_principal, choc_secondaire;
  DISJOINT: YES;
  COMPLETE: YES;
END CONCEPT choc;

BINARY-RELATION est_impliqué;
  DESCRIPTION:
  "un véhicule peut subir plusieurs choc consécutifs et plusieurs véhicules peuvent être impliqués
  dans un choc";
  ARGUMENT-1: véhicule;
  CARDINALITY: 1+;
  ARGUMENT-2: choc;
  CARDINALITY: 1+;
END BINARY-RELATION est_impliqué;

CONCEPT choc_principal;
  DESCRIPTION:

```

```
"Il s'agit d'une description du choc principal subit par le véhicule concerné";
  ATTRIBUTES:
    Type_obstacle_principal : STRING;
    Type_choc_principal: type_choc_value;
    Vitesse_rapprochement : NATURAL;
  HAS-PARTS: Cotes_plan_accident;
END CONCEPT choc;

VALUE-TYPE type_choc_value;
  TYPE: NOMINAL;
  VALUE-LIST: {Choc_frontal, Choc_latéral, Choc_arrière, Tonneau, Renversement, Side_swipe};
END VALUE-TYPE type_choc_value;

CONCEPT Cotes_plan_accident;
  DESCRIPTION:
    "Il s'agit d'une description des cotes du plan où l'accident à eu lieu";
  ATTRIBUTES:
    Distance_freinage : REAL;
    Distance_arrêt : REAL;
END CONCEPT Cotes_plan_accident;

CONCEPT choc_secondaire;
  DESCRIPTION:
    "Il s'agit d'une description du choc secondaire subit par le véhicule concerné";
  ATTRIBUTES:
    type_obstacle_secondaire : STRING;
    Type_choc_secondaire: type_choc_value;
END CONCEPT choc_secondaire;

CONCEPT environnement;
  DESCRIPTION:
    "Il s'agit d'une description de l'environnement où l'accident a eu lieu";
  HAS-PARTS: trafic; conditions_ambiantes; équipement_route; infrastructure;
END CONCEPT environnement;

CONCEPT trafic;
  DESCRIPTION:
    "Il s'agit d'une description de l'environnement où l'accident a eu lieu";
END CONCEPT trafic;

CONCEPT conditions_ambiantes;
  DESCRIPTION:
    "Il s'agit d'une description de l'environnement où l'accident a eu lieu";
END CONCEPT conditions_ambiantes;

CONCEPT équipement_route;
  DESCRIPTION:
    "Il s'agit d'une description de l'environnement où l'accident a eu lieu";
END CONCEPT équipement_route;

CONCEPT infrastructure;
  DESCRIPTION:
    "Il s'agit d'une description de l'infrastructure où l'accident a eu lieu";
  HAS-PARTS:
    caractéristiques_lieu;          caractéristiques_surface;          caractéristiques_géométriques;
    tracés_lieu_accident;
END CONCEPT infrastructure;

CONCEPT caractéristiques_lieu;
  DESCRIPTION:
    "Il s'agit d'une description de l'infrastructure où l'accident a eu lieu";
END CONCEPT caractéristiques_lieu;

CONCEPT caractéristiques_surface;
  DESCRIPTION:
    "Il s'agit d'une description de l'infrastructure où l'accident a eu lieu";
  SUPER-TYPE-OF:
    aspects_visuels, aspects_quantitatifs;
  DISJOINT: YES;
  COMPLETE: YES;
END CONCEPT caractéristiques_surface;

CONCEPT aspects_visuels;
  DESCRIPTION:
    "Il s'agit d'une description de l'infrastructure où l'accident a eu lieu";
END CONCEPT aspects_visuels;

CONCEPT aspects_quantitatifs;
  DESCRIPTION:
    "Il s'agit d'une description de l'infrastructure où l'accident a eu lieu";
END CONCEPT aspects_quantitatifs;

CONCEPT tracés_lieu_accident;
  DESCRIPTION:
    "Il s'agit d'une description de l'infrastructure où l'accident a eu lieu";
```

```

HAS-PARTS:
  Section_rectiligne; Courbe_Virage; Profil_en_travers; Profil_en_long;
END CONCEPT tracés_lieu_accident;

CONCEPT Section_rectiligne;
  DESCRIPTION:
    "Il s'agit d'une description de l'environnement où l'accident a eu lieu";
END CONCEPT Section_rectiligne;

CONCEPT Courbe_Virage;
  DESCRIPTION:
    "Il s'agit d'une description de du virage où l'accident a eu lieu";
END CONCEPT Courbe_Virage;

CONCEPT Profil_en_travers;
  DESCRIPTION:
    "Il s'agit d'une description du Profil en travers de l'endroit où l'accident a eu lieu";
END CONCEPT Profil_en_travers;

CONCEPT Profil_en_long;
  DESCRIPTION:
    "Il s'agit d'une description du Profil en longueur de l'endroit où l'accident a eu lieu";
END CONCEPT Profil_en_long;

CONCEPT caractéristiques_géométriques;
  DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu lieu";
  SUPER-TYPE-OF:
    caract_Abord,    caract_Accotement,    caract_Chaussée,    caract_Intersection,    caract_Carrefour,
    visibilité_géométriques;
  DISJOINT: YES;
  COMPLETE: YES;
END CONCEPT caractéristiques_géométriques;

CONCEPT caract_Abord;
  DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu lieu";
  SUPER-TYPE-OF: Abord_gauche, Abord_droit;
  DISJOINT: YES;
  COMPLETE: YES;
END CONCEPT caract_Abord;

CONCEPT Abord_gauche;
  DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu lieu";
END CONCEPT Abord_gauche;

CONCEPT Abord_droit;
  DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu lieu";
END CONCEPT Abord_droit;

CONCEPT caract_Accotement;
  DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu lieu";
  SUPER-TYPE-OF: Accotement_gauche, Accotement_droit;
  DISJOINT: YES;
  COMPLETE: YES;
END CONCEPT caract_Accotement;

CONCEPT Accotement_gauche;
  DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu lieu";
END CONCEPT Accotement_gauche;

CONCEPT Accotement_droit;
  DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu lieu";
END CONCEPT Accotement_droit;

CONCEPT caract_Chaussée;
  DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu lieu";
END CONCEPT caract_Chaussée;

CONCEPT caract_Intersection;
  DESCRIPTION:

```

```

    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu
    lieu";
END CONCEPT caract_Intersection;

CONCEPT caract_Carrefour;
    DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu
    lieu";
END CONCEPT caract_Carrefour;

CONCEPT visibilité_géométriques;
    DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu
    lieu";
    SUPER-TYPE-OF: Visibilité_longitudinale, Visibilité_latérale;
    DISJOINT: YES;
    COMPLETE: YES;
END CONCEPT visibilité_géométriques;

CONCEPT Visibilité_longitudinale;
    DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu
    lieu";
END CONCEPT Visibilité_longitudinale;

CONCEPT Visibilité_latérale;
    DESCRIPTION:
    "Il s'agit d'une description des caractéristiques géométrique de l'endroit où l'accident a eu
    lieu";
END CONCEPT Visibilité_latérale;

/*-----*/
/*          I.1.b- Types de connaissances pour l'élaboration de scénarios d'accidents */
/*-----*/

CONCEPT objectif_de_l'étude;
    DESCRIPTION:
    "l'objectif d'élaboration des STA";
    ATTRIBUTES:
    description : UNIVERSAL;
END CONCEPT objectif_de_l'étude;

CONCEPT échantillon_d'étude;
    DESCRIPTION:
    "l'accident peut être décrit à travers interaction des troisi composantes: impliqués, véhicules et
    environnement";
    ATTRIBUTES:
    nombre_accidents : NATURAL;
END CONCEPT échantillon_d'étude;

RULE-TYPE critères_qualité_accident;
    ANTECEDENT:
    accident;
    CARDINALITY: 1;
    CONSEQUENT:
    échantillon_d'étude;
    CARDINALITY: 1;
    CONNECTION-SYMBOL:
    implique;
END RULE-TYPE critères_qualité_accident;

RULE-TYPE règles_abstraction_accident;
    ANTECEDENT:
    accident;
    CARDINALITY: 1;
    CONSEQUENT:
    description_abstraite_accident;
    CARDINALITY: 1;
    CONNECTION-SYMBOL:
    a-pour-abstraction;
END RULE-TYPE règles_abstraction_accident;

CONCEPT scénario_type_accidents;
    DESCRIPTION:
    "l'accident peut être décrit à travers interaction des troisi composantes: impliqués, véhicules et
    environnement";
    ATTRIBUTES:
    nombre_accidents : NATURAL;
    poids : REAL;
END CONCEPT scénario_type_accidents;

CONCEPT classe_accidents;
    DESCRIPTION:
    "l'accident peut être décrit à travers interaction des troisi composantes: impliqués, véhicules et
    environnement";
    ATTRIBUTES:
    nombre_accidents : NATURAL;
    poids : REAL;

```

```

END CONCEPT classe_accidents;

BINARY-RELATION classification;
  DESCRIPTION:
  "une classe est regroupement d'accidents présentant une similitude dans leurs déroulements";
  ARGUMENT-1: accident;
  CARDINALITY: 1+;
  ARGUMENT-2: classe_accidents;
  CARDINALITY: 1;
  ATTRIBUTES:
    intra-similarité : REAL;
    inter-similarité : REAL;
  qualité : STRING;
END BINARY-RELATION classification;

BINARY-RELATION classification;
  DESCRIPTION:
  "une classe est regroupement d'accidents présentant une similitude dans leurs déroulements";
  ARGUMENT-1: classe_accidents;
  CARDINALITY: 1;
  ARGUMENT-2: objectif_de_l'étude;
  CARDINALITY: 1;
  ATTRIBUTES:
    intra-similarité : REAL;
    inter-similarité : REAL;
  qualité : STRING;
END BINARY-RELATION classification;

BINARY-RELATION description_synthétique;
  DESCRIPTION:
  "une classe est regroupement d'accidents présentant une similitude dans leurs déroulements";
  ARGUMENT-1: scénario_type_accidents;
  CARDINALITY: 1;
  ARGUMENT-2: classe_accidents;
  CARDINALITY: 1;
END BINARY-RELATION description_synthétique;

RULE-TYPE critères_qualité_classification;
  ANTECEDENT:
  caractérisation_classe;
  CARDINALITY: 1;
  CONSEQUENT:
  scénario_type_accidents;
  CARDINALITY: 1;
  CONNECTION-SYMBOL:
  implique;
END RULE-TYPE critères_qualité_classification;

CONCEPT attribut_d'étude;
  DESCRIPTION:
  "l'accident peut être décrit à travers interaction des troi composantes: impliqués, véhicules et
  environnement";
  SUB-TYPE-OF: attribut;
  ATTRIBUTES:
  nom_attribut : STRING;
  modalité : UNIVERSAL;
  nombre_modalités : UNIVERSAL;
  valeur_propre : REAL;
  importance_générale : UNIVERSAL;
END CONCEPT attribut_d'étude;

BINARY-RELATION caractérisation_classe;
  DESCRIPTION:
  "une classe est regroupement d'accidents présentant une similitude dans leurs déroulements";
  ARGUMENT-1: classe_accidents;
  CARDINALITY: 1+;
  ARGUMENT-2: attribut_d'étude;
  CARDINALITY: 1+;
  ATTRIBUTES:
  modalité : UNIVERSAL;
  Valeur_test : REAL;
  poids_class_dans_mod : REAL;
  poids_mod_dans_class : REAL;
  pertinence : STRING;
END BINARY-RELATION caractérisation_classe;

RULE-TYPE règles_abstraction_classe;
  ANTECEDENT:
  caractérisation_classe;
  CARDINALITY: 1+;
  CONSEQUENT:
  caractérisation_classe;
  CARDINALITY: 1;
  CONNECTION-SYMBOL:
  a-pour-abstraction;
END RULE-TYPE règles_abstraction_classe;

RULE-TYPE règles_abstraction_attributs;
  ANTECEDENT:

```

```

attribut;
  CARDINALITY: 1;
  CONSEQUENT:
attribut;
  CARDINALITY: 1;
  CONNECTION-SYMBOL:
a-pour-abstraction;
END RULE-TYPE règles_abstraction_attributs;

RULE-TYPE règles_décision_pertinence_attribut;
  ANTECEDENT:
  attribut;
    CARDINALITY: 1+;
    CONSEQUENT:
  attribut_d-étude;
    CARDINALITY: 1;
    CONNECTION-SYMBOL:
  implique;
END RULE-TYPE règles_décision_pertinence_attribut;

RULE-TYPE règles_modification_attribut;
  ANTECEDENT:
  attribut;
    CARDINALITY: 1;
  CONSEQUENT:
  attribut;
    CARDINALITY: 1;
  CONNECTION-SYMBOL:
  implique;
END RULE-TYPE règles_modification_attribut;

END DOMAIN-SCHEMA élaboration_scénario-schema;

```

A-1.1.2. Base de connaissances

```

/*.....*/
/* I.2- Base de connaissances*/
/*.....*/

/*-----*/
/*          I.2.a-   Connaissances préparation des données: sélection d'attribut
                & sélection d'accident                               */
/*-----*/

KNOWLEDGE-BASE connaissances_préparation_données;

  USE:
  règles_abstraction_attributs FROM élaboration_scénario-schema,
  règles_décision_pertinence_attribut FROM élaboration_scénario-schema,
  règles_modification_attribut FROM élaboration_scénario-schema,
  critères_qualité_classification FROM élaboration_scénario-schema;

  EXPRESSIONS:
  /*Ces règles sont utilisées pour l'abstraction des attributs en ajoutant une information
  concernant leur pertinence pour l'étude.*/
  /*****/

  poids(accident.attribut.modalité = 'inconnu') > 60
    A-POUR-ABSTRACTION
  attribut.nbre_inconnue = 'très élevé';

  40 < poids(accident.attribut.modalité = 'inconnu') < 60
    A-POUR-ABSTRACTION
  attribut.nbre_inconnue = 'élevé';

  10 < poids(accident.attribut.modalité = 'inconnu') < 40
    A-POUR-ABSTRACTION
  attribut.nbre_inconnue = 'moyen';

  poids(accident.attribut.modalité = 'inconnu') < 10
    A-POUR-ABSTRACTION
  attribut.nbre_inconnue = 'acceptable';

  attribut.nombre_modalités > 30
    A-POUR-ABSTRACTION
  attribut.nombre_modalités = 'très élevé';

  20 < attribut.nombre_modalités < 30
    A-POUR-ABSTRACTION
  attribut.nombre_modalités = 'élevé';

  10 < attribut.nombre_modalités < 20
    A-POUR-ABSTRACTION
  attribut.nombre_modalités = 'moyen';

  attribut.nombre_modalités < 10
    A-POUR-ABSTRACTION
  attribut.nombre_modalités = 'acceptable';

```

```

/* règles de décision de la pertinence des attributs*/

attribut.nbre_inconnue = 'acceptable'          OR
attribut.nbre_inconnue = 'moyen'              AND
attribut.nombre_modalités = 'acceptable'      OR
attribut.nombre_modalités = 'acceptable'      AND
attribut.importance_générale = 'Très important' OR
ttribut.importance_générale = 'important'
    IMPLIQUE
attribut.pertinence_étude = 'attribut pertinent';

attribut.nbre_inconnue = 'très élevé'         AND
attribut.nombre_modalités = 'très élevé'     AND
attribut.importance_générale = 'très important'
    IMPLIQUE
attribut.pertinence_étude = 'attribut non pertinent';

attribut.nbre_inconnue != 'très élevé'       AND
attribut.nombre_modalités != 'très élevé'    AND
attribut.importance_générale = 'très important' OR
attribut.importance_générale = 'important'
    IMPLIQUE
attribut.pertinence_étude = 'attribut pertinent';

poids(accident.attribut.modalité = 'inconnu') > 60
    IMPLIQUE
attribut.pertinence = 'attribut non pertinent';

/*listes d'actions pour modifier un attribut lors de la sélection des attributs d'études*/
/*l'expert peut affecter un niveau de pertinence à l'attribut sans utiliser les règles ci-dessus,
il a donc la priorité sur les règles existantes*/

attribut.nombre_modalités = très_élevé      OR
attribut.nombre_modalités = élevé AND
attribut.pertinence = 'attribut non pertinent'
    IMPLIQUE
attribut.liste_actions = 'supprimer attribut';

attribut.nombre_modalités = très_élevé      OR
attribut.nombre_modalités = élevé AND
attribut.pertinence = 'attribut pertinent'
    IMPLIQUE
attribut.liste_actions = 'regrouper les modalités';

attribut.nbre_inconnue = très_élevé        OR
attribut.nbre_inconnue = élevé AND
attribut.pertinence = 'attribut non pertinent'
    IMPLIQUE
attribut.liste_actions = 'supprimer attribut';

attribut.nbre_inconnue = très_élevé        OR
attribut.nbre_inconnue = élevé AND
attribut.pertinence = 'attribut pertinent'
    IMPLIQUE
attribut.liste_actions = 'remplacer les valeurs inconnues';

40 < poids(accident.attribut.modalité = 'inconnu') < 60
    IMPLIQUE
attribut.pertinence = 'attribut non actif, mais descriptif';

/*règles d'abstraction de la description d'un accident*/
/*****/

poids(description_accident.attribut.modalité = 'inconnu') > 60
    IMPLIQUE
accident.nbre_inconnu = 'très_élevé';

40 < poids(description_accident.attribut.modalité = 'inconnu') < 60
    IMPLIQUE
accident.nbre_inconnu = 'élevé';

10 < poids(description_accident.attribut.modalité = 'inconnu') < 40
    IMPLIQUE
accident.nbre_inconnu = 'moyen';

poids(description_accident.attribut.modalité = 'inconnu') > 10
    IMPLIQUE
accident.nbre_inconnu = 'acceptable';

/*règles de sélection d'un accident pour la classification*/
/*****/

accident.nbre_inconnu = 'acceptable'      OR
accident.nbre_inconnu = 'moyen'
    IMPLIQUE
accident.qualité = 'accident pertinent';

accident.nbre_inconnu = 'très élevé'      OR

```

```
accident.nbre_inconnu = 'élevé'
      IMPLIQUE
accident.qualité = 'accident non pertinent';

END KNOWLEDGE-BASE connaissances_préparation_données;

/*-----*
-----*/
/*          I.2.b- Interprétation des classes: interprétation intra & statistique /
/*-----*
-----*/

KNOWLEDGE-BASE conaiss_interprétation_classe;

      USE:
critères_qualité_classification FROM élaboration_scénario-schema,
règles_abstraction_classe FROM élaboration_scénario-schema;
      EXPRESSIONS:

/* Règles pour l'interprétation intra-classe: abstraction des classe*/
/*-----*/

attribut_d-étude.modalité.Valeur_test > 2 OR
attribut_d-étude.modalité.Valeur_test <-2 AND
attribut_d-étude.modalité.poids_mod_dans_class > 30
      A-POUR-ABSTRACTION
attribut_d-étude.modalité.pertinence = 'discriminante et caractérisante';

attribut_d-étude.modalité.Valeur_test > 2 OR
attribut_d-étude.modalité.Valeur_test <-2 AND
attribut_d-étude.modalité.poids_mod_dans_class < 30
      A-POUR-ABSTRACTION
attribut_d-étude.modalité.pertinence = 'discriminante et non caractérisante';

attribut_d-étude.modalité.Valeur_test < 2 AND
attribut_d-étude.modalité.Valeur_test >-2 AND
attribut_d-étude.modalité.poids_mod_dans_class > 30
      A-POUR-ABSTRACTION
attribut_d-étude.modalité.pertinence = 'non discriminante et caractérisante';

attribut_d-étude.modalité.Valeur_test < 2 AND
attribut_d-étude.modalité.Valeur_test >-2 AND
attribut_d-étude.modalité.poids_mod_dans_class < -30
      A-POUR-ABSTRACTION
attribut_d-étude.modalité.pertinence = 'non discriminante et non caractérisante';

/* si pour toute les classes, on :*/
attribut_d-étude.modalité.poids_mod_dans_class > 30
      A-POUR-ABSTRACTION
attribut_d-étude.modalité.pertinence = 'problématique transversale';

/* Règles pour l'interprétation statistique des classes: règles d'évaluation d'une classe et
d'une classification*/
/*-----*/

classe_accidents.poids > 50
      IMPLIQUE
classification.qualité = 'refaire la classification en modifiant les attributs ou les paramètres
de classification';

classe_accidents.nombre > échantillon_d-étude.nombre
      IMPLIQUE
classification.qualité = 'refaire la classification en modifiant les attributs ou les paramètres
de classification';

classe_accidents.nombre < valeur_seuil
      IMPLIQUE
classification.qualité = 'refaire la classification ou regrouper cette classe avec d'autres
classes';

classification.intra-similarité = inf(classe_accidents.intra-similarité) AND
classification.inter-similarité = sup(classe_accidents.inter-similarité)
      IMPLIQUE
classification.qualité = 'choisir cette classification';

END KNOWLEDGE-BASE conaiss_interprétation_classe;

END DOMAIN-KNOWLEDGE
```


A-1.2 Connaissances des inférences

A-1.2.1 Préparation des données

A-1.2.1.1 Sélection des attributs d'étude

A-1.2.1.1.1 Les rôles de connaissances

```

/*****
/* II. Connaissances Inférences*/
*****/

INFERENCE-KNOWLEDGE inférences_élaboration_STA;

/*-----
-----*/
/*      Inférences sélection des attributs d'étude*/
/*-----
-----*/

/*-----Liste des Rôle de Connaissances dynamiques (intermédiaires) utilisés comme des
input/output des inférences-----*/

KNOWLEDGE-ROLE Besoin;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      objectif_de_l-étude;
END KNOWLEDGE-ROLE Besoin;

KNOWLEDGE-ROLE contraintes;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      objectif_de_l-étude;
END KNOWLEDGE-ROLE contraintes;

KNOWLEDGE-ROLE préférences;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      objectif_de_l-étude;
END KNOWLEDGE-ROLE préférences;

KNOWLEDGE-ROLE attributs;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      attribut;
END KNOWLEDGE-ROLE attributs;

KNOWLEDGE-ROLE attribut-d-étude;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      attribut_d-étude;
END KNOWLEDGE-ROLE attribut-d-étude;

KNOWLEDGE-ROLE description_attributs;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      attribut;
END KNOWLEDGE-ROLE description_attributs;

KNOWLEDGE-ROLE description_attribut;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      attribut;
END KNOWLEDGE-ROLE description_attribut;

KNOWLEDGE-ROLE description_abstraite_attribut;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      attribut;
END KNOWLEDGE-ROLE description_abstraite_attribut;

KNOWLEDGE-ROLE critères-1;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      critères_pertinence_attribut FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE critères-1;

KNOWLEDGE-ROLE critère-1;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      critère_sélectionné_pertinence_attribut
connaissances_préparation_données;
END KNOWLEDGE-ROLE critère-1;

KNOWLEDGE-ROLE valeur_critère-1;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      critères_pertinence_attribut FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE valeur_critère-1;

KNOWLEDGE-ROLE décision_pertinence-attribut;
  TYPE: DYNAMIC;

```

```

    DOMAIN-MAPPING:      oui_non-value;
END KNOWLEDGE-ROLE décision_pertinence-attribut;

KNOWLEDGE-ROLE décision_supprimer-1;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      oui_non-value;
END KNOWLEDGE-ROLE décision_supprimer-1;

KNOWLEDGE-ROLE violations-1;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      critère_sélectionné_pertinence_attribut
connaissances_préparation_données;
END KNOWLEDGE-ROLE violations-1;
FROM

KNOWLEDGE-ROLE liste_actions-1;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      listes_modifications_attributs_ou_modalités
connaissances_préparation_données;
END KNOWLEDGE-ROLE liste_actions-1;
FROM

KNOWLEDGE-ROLE action;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      modification_attributs_ou_modalités
connaissances_préparation_données;
END KNOWLEDGE-ROLE action;
FROM

KNOWLEDGE-ROLE attribut_modifié;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      attribut;
END KNOWLEDGE-ROLE attribut_modifié;

KNOWLEDGE-ROLE descr_attribut_modifié;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      attribut;
END KNOWLEDGE-ROLE descr_attribut_modifié;

KNOWLEDGE-ROLE échantillon-1;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      échantillon_d-étude;
END KNOWLEDGE-ROLE échantillon-2;

/*-----Liste des Rôle de Connaissances statiques utilisés comme des connaissances du
domaine dans les inputs des inférences-----*/

KNOWLEDGE-ROLE modèles_multivues_domaine;
  TYPE: STATIC;
  DOMAIN-MAPPING:      modèles_multivues_domaine;
END KNOWLEDGE-ROLE modèles_multivues_domaine;

KNOWLEDGE-ROLE liste-critères-1;
  TYPE: STATIC;
  DOMAIN-MAPPING:      règles_decision_pertinence_attribut
connaissances_préparation_données;
END KNOWLEDGE-ROLE liste-critères-1;
FROM

KNOWLEDGE-ROLE ensemble_attribut;
  TYPE: STATIC;
  DOMAIN-MAPPING:
  attribut;
END KNOWLEDGE-ROLE ensemble_attribut;

KNOWLEDGE-ROLE BD_EDA_accidents;
  TYPE: STATIC;
  DOMAIN-MAPPING:      BD_EDA_accidents;
END KNOWLEDGE-ROLE BD_EDA_accidents;

KNOWLEDGE-ROLE connaissances_d-abstraction-1;
  TYPE: STATIC;
  DOMAIN-MAPPING:
  règles_abstraction_attributs FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE connaissances_d-abstraction-1;

KNOWLEDGE-ROLE connaissances_sélection-1;
  TYPE: STATIC;
  DOMAIN-MAPPING:
  règles_decision_pertinence_attribut FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE connaissances_sélection-1;

KNOWLEDGE-ROLE connaissances_évaluation-1;
  TYPE: STATIC;
  DOMAIN-MAPPING:
  règles_abstraction_attributs FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE connaissances_évaluation-1;

KNOWLEDGE-ROLE connaissances_modification-1;
  TYPE: STATIC;
  DOMAIN-MAPPING:
  connaissances_modification-attribut FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE connaissances_modification-1;

```

A-1.2.1.1.2 Les inférences

```

/*-----Liste des inférences-----*/

INFERENCE opérationneliser;
  ROLES:
    INPUT:      Besoin;
    OUTPUT:     contraintes,      préférences;
    STATIC:     modèles_multivues_domaine;
  SPECIFICATION: "
  il s'agit de traduire/transformer le Besoin en spécification qui peuvent être utilisées dans
  l'étude pour la sélection des attributs ou l'évaluation du résultats finale, c-à-d les scénarios
  types d'accident. les contraintes définissent des restrictions dans le choix des attributs pour
  sélectionner les attributs pertinents. Comme il peut y avoir plusieurs attributs pertinents, les
  préférences permettent de choisir les attributs d'étude parmi les attributs pertinents";
END INFERENCE opérationneliser;

INFERENCE sélectionner-1;
  ROLES:
    INPUT:      contraintes,      ensemble_attribut;
    OUTPUT:     attributs;
  SPECIFICATION: "
  Il s'agit de sélectionner à partir de l'ensembles des attributs les attributs qui respectent les
  contraintes. Ces attributs sont pertinents a priori, car il reste d'autres évaluations";
END INFERENCE sélectionner-1;

TRANSFER-FUNCTION calculer-1;
  TYPE: OBTAIN;
  ROLES:
    INPUT:      BD_EDA_accidents, attributs;
    OUTPUT:     description_attributs;
END TRANSFER-FUNCTION calculer-1;

INFERENCE sélectionner-11;
  ROLES:
    INPUT:      attributs;
    OUTPUT:     description_attribut;
  SPECIFICATION: "
  Il s'agit de sélectionner aléatoirement un attribut";
END INFERENCE sélectionner-11;

INFERENCE abstraire-1;
  ROLES:
    INPUT:      description_attribut;
    OUTPUT:     description_abstraite_attribut;
    STATIC:     connaissances_d-abstraction-1;
  SPECIFICATION: "
  l'entrée est un attribut décrit à l'aide de critères statistiques. la sortie est composée du même
  attribut mais dont la description est enrichie avec de nouvelles variables permettant de juger la
  pertinence de l'attribut. les nouvelles variables sont déterminées en se basant sur les
  connaissances statiques d'abstraction.";
END INFERENCE abstraire-1;

INFERENCE spécifier-1;
  OPERATION-TYPE: rechercher;
  ROLES:
    INPUT:      description_abstraite_attribut;
    OUTPUT:     critères-1;
    STATIC:     liste-critères-1;
  SPECIFICATION: "
  Il s'agit de rechercher les critères qui peuvent être utilisés pour l'évaluation des attributs.
  Les même connaissances statiques d'abstraction sont utilisées";
END INFERENCE spécifier-1;

INFERENCE sélectionner-12;
  ROLES:
    INPUT:      critères-1;
    OUTPUT:     critère-1;
  SPECIFICATION: "
  Il s'agit de sélectionner l'un ou plusieurs de critères pour l'évaluation";
END INFERENCE sélectionner-12;

INFERENCE évaluer-1;
  ROLES:
    INPUT:      critère-1, description_abstraite_attribut;
    OUTPUT:     valeur_critère-1;
  SPECIFICATION: "
  déterminer la valeur de la critère (ou critère) choisi pour un cas donné. Les contraintes issus
  du Besoin interviennent dans l'évaluation.";
END INFERENCE évaluer-1;

INFERENCE combiner-1;
  ROLES:
    INPUT:      valeur_critère-1;
    OUTPUT:     décision_pertinence_attribut, violations-1;
    STATIC:     règles_décision_pertinence_attribut;
  SPECIFICATION:

```

```
"en combinant les résultats des évaluations selon les différents critères choisis, décider si
l'attribut est pertinent. Si oui, il est candidat pour être choisi comme attribut d'étude. Si
non, l'attribut est rejeté ";
END INFERENCE combiner-1;

INFERENCE critiquer-1;
ROLES:
  INPUT:      violations-1;
  OUTPUT:     liste_actions-1;
  STATIC:     connaissances_modification-1;
SPECIFICATION:
  "en regardant les critères de pertinences qui n'ont pas été respectés, déterminer les causes et
  la liste des actions à faire ";
END INFERENCE critiquer-1;

INFERENCE modifier-1;
ROLES:
  INPUT:      liste_actions-1, description_abstraite_attribut;
  OUTPUT:     attribut_modifié;
SPECIFICATION:
  "si l'attribut n'est pas supprimé, il est modifié en regroupant certaines modalités, en
  remplaçant les valeurs inconnues par la moyenne, etc. ";
END INFERENCE modifier-1;

TRANSFER-FUNCTION calculer-11;
TYPE: OBTAIN;
ROLES:
  INPUT:      attribut_modifié, BD_EDA_accidents;
  OUTPUT:     descr_attribut_modifié;
END TRANSFER-FUNCTION calculer-11;

INFERENCE choisir-1;
ROLES:
  INPUT:      attributs, préférences;
  OUTPUT:     attribut-d'étude;
SPECIFICATION:
  "il s'agit de sélectionner parmi les attributs qui vérifient les contraintes, donc qui sont
  pertinents les attributs d'étude en tenant compte des préférences";
END INFERENCE choisir-1;

INFERENCE instancier;
ROLES:
  INPUT:      BD_EDA_accidents, attribut-d'étude;
  OUTPUT:     échantillon-1;
SPECIFICATION:
  "construire l'échantillon-1 contenant des accidents décrits par les attributs d'étude";
END INFERENCE instancier;
```

A-1.2.1.1.3 La structure d'inférences

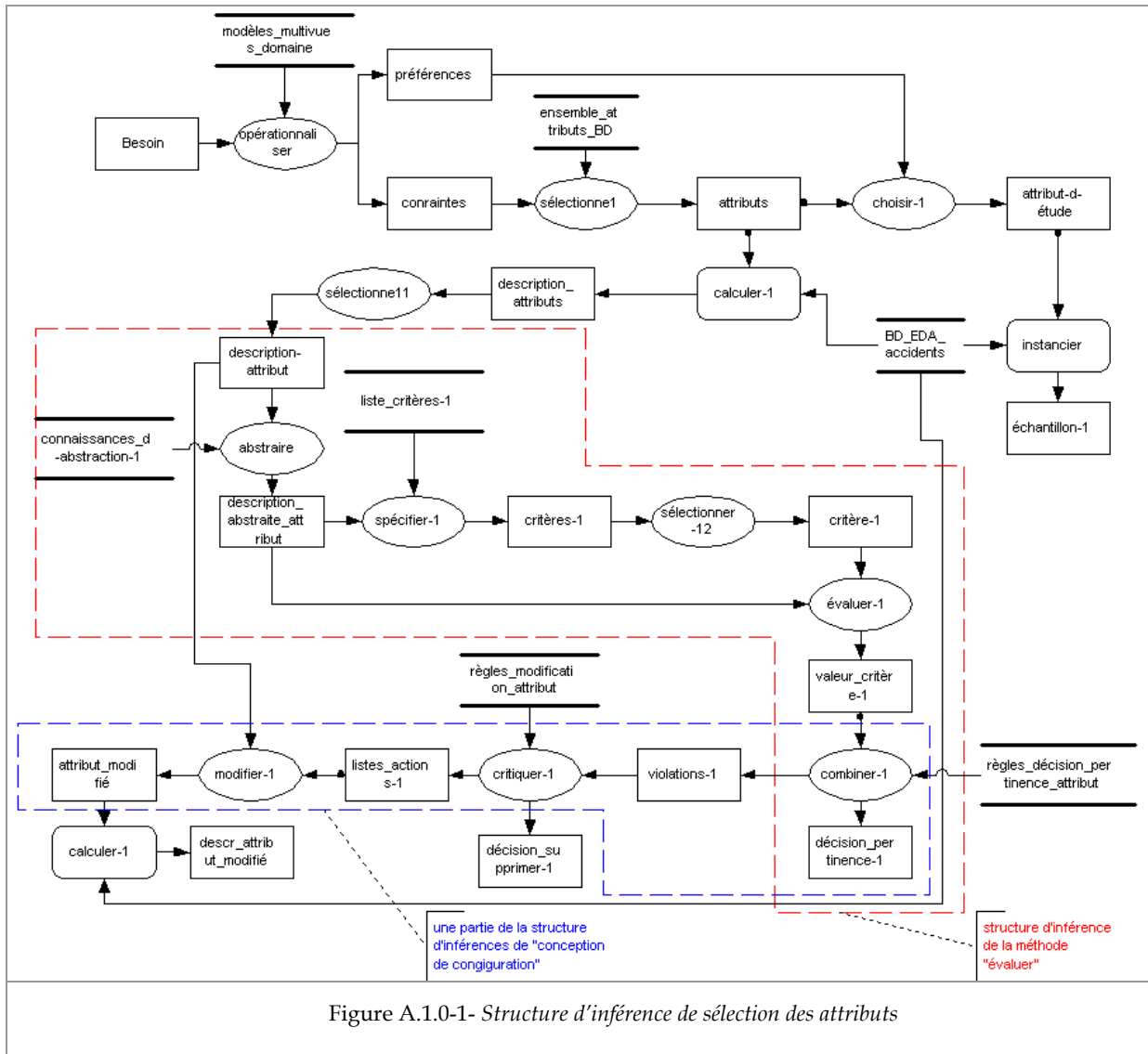


Figure A.1.0-1- Structure d'inférence de sélection des attributs

A-1.2.1.2 Sélection des accidents d'étude

A-1.2.1.2.1 Les rôles de connaissances

```

/-----*/
/*      Inférences de sélection des accidents pertinentes pour l'échantillon d'étude      */
/-----*/
/-----*/

KNOWLEDGE-ROLE description_accident;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:
    accident;
END KNOWLEDGE-ROLE description_accident;

KNOWLEDGE-ROLE description_abstraite_accident;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:
    accident;
END KNOWLEDGE-ROLE description_abstraite_accident;

KNOWLEDGE-ROLE connaissances_évaluation-2;
  TYPE: STATIC;
  DOMAIN-MAPPING:
    critères_qualité_accident FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE connaissances_évaluation-2;

KNOWLEDGE-ROLE liste-critères-2;
  TYPE: STATIC;
  DOMAIN-MAPPING:
    critères_qualité_accident FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE liste-critères-2;

```

```

KNOWLEDGE-ROLE critères-2;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:
  Liste_critère_évaluation_accident FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE critères-2;

KNOWLEDGE-ROLE critère-2;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:
  critère_sélectionné_qualité_accident FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE critère-2;

KNOWLEDGE-ROLE valeur_critère-2;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:
  valeur_critère_sélectionné_qualité_accident;
END KNOWLEDGE-ROLE valeur_critère-2;

KNOWLEDGE-ROLE décision_pertinence-accident;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      oui_non-value;
END KNOWLEDGE-ROLE décision_pertinence-accident;

KNOWLEDGE-ROLE décision_supprimer-2;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      oui_non-value;
END KNOWLEDGE-ROLE décision_supprimer-2;

KNOWLEDGE-ROLE violations-2;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      critères_qualité_accident FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE violations-2;

KNOWLEDGE-ROLE connaissances_modification-2;
  TYPE: STATIC;
  DOMAIN-MAPPING:
  connaissances_modification-accident FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE connaissances_modification-2;

KNOWLEDGE-ROLE liste_actions-2;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      listes_modifications_accidents FROM connaissances_préparation_données;
END KNOWLEDGE-ROLE liste_actions-2;

KNOWLEDGE-ROLE accident_modifié;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      accident;
END KNOWLEDGE-ROLE accident_modifié;

KNOWLEDGE-ROLE descr_accident_modifié;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      accident;
END KNOWLEDGE-ROLE descr_accident_modifié;

KNOWLEDGE-ROLE échantillon_d-étude;
  TYPE: DYNAMIC;
  DOMAIN-MAPPING:      échantillon_d-étude;
END KNOWLEDGE-ROLE échantillon_d-étude;

```

A-1.2.1.2.2 Les inférences

```

/*Inférences de sélection d'accident pertinent*/
/*-----*/

INFERENCE sélectionner-2;
  ROLES:
  INPUT:      échantillon-1;
  OUTPUT:     description_accident;
  SPECIFICATION:
  "Il s'agit de sélectionner accident par accident à partir de l'échantillon-1 contenant les
accidents de l'échantillon de base décrits à l'aide des attributs choisies pour l'étude";
END INFERENCE sélectionner-2;

INFERENCE abstraire_2;
  ROLES:
  INPUT:      description_accident;
  OUTPUT:     description_abstraite_accident;
  STATIC:     connaissances_d-abstraction-2;
  SPECIFICATION: "
  l'entrée est composée de l'ensemble des accidents décrits à l'aide de critères statistiques. la
sortie est composée des mêmes accidents mais dont la description est enrichie avec de nouvelles
variables permettant de juger la pertinence de l'attribut.";
END INFERENCE abstraire_2;

INFERENCE spécifier-2;
  OPERATION-TYPE: rechercher;
  ROLES:
  INPUT:      description_abstraite_accident;

```

```

        OUTPUT:      critères-2;
        STATIC:      liste-critères-2;
        SPECIFICATION:
        "Il s'agit de rechercher les critères qui peuvent être utilisés pour l'évaluation de la pertinence
d'un accident. Les mêmes connaissances statiques d'abstraction sont utilisées";
END INFERENCE spécifier-2;

INFERENCE sélectionner-21;
    ROLES:
    INPUT:          critères-2;
    OUTPUT:         critère-2;
    SPECIFICATION:
    "Il s'agit de sélectionner l'un ou plusieurs de critères pour l'évaluation";
END INFERENCE sélectionner-21;

INFERENCE évaluer-2;
    ROLES:
    INPUT:          critère-2, description_abstraite_accident;
    OUTPUT:         valeur_critère-2;
    SPECIFICATION: "
détérminer la valeur de la critère (ou critère) choisi pour un cas donné. Les contraintes issus
du Besoin interviennent dans l'évaluation.";
END INFERENCE évaluer-2;

INFERENCE combiner-2;
    ROLES:
    INPUT:          valeur_critère-2;
    OUTPUT:         décision_pertinence-accident, violations-2;
    STATIC:         critères_qualité_accident;
    SPECIFICATION:
    "en combinant les résultats des évaluations selon les différents critères choisi, décider si
l'accident est pertinent. Si oui, il est candidat pour être choisi comme accident d'étude. Si non,
l'accident est rejeté ";
END INFERENCE combiner-2;

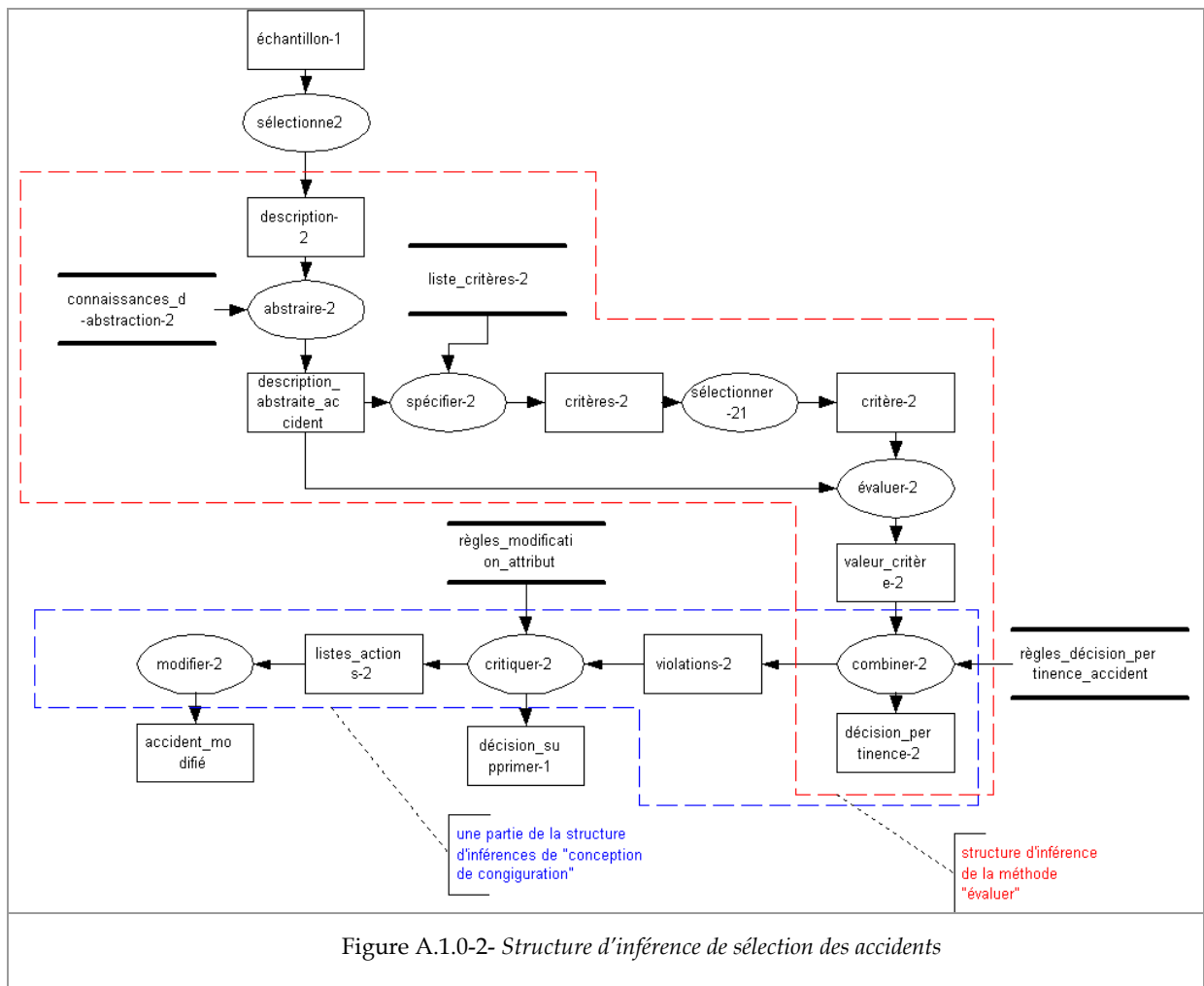
INFERENCE critiquer-2;
    ROLES:
    INPUT:          violations-2;
    OUTPUT:         liste_actions-2;
    STATIC:         connaissances_modification-2;
    SPECIFICATION:
    "en regardant les critères de pertinences qui n'ont pas été respectés, déterminer les causes et
la liste des actions à faire ";
END INFERENCE critiquer-2;

INFERENCE modifier-2;
    ROLES:
    INPUT:          liste_actions-2, description_abstraite_accident;
    OUTPUT:         accident_modifié;
    SPECIFICATION:
    "si l'accident n'est pas supprimé, il est modifié en remplaçant les valeurs inconnues par la
moyenne, etc. ";
END INFERENCE modifier-2;

TRANSFER-FUNCTION calculer-21;
    TYPE: OBTAIN;
    ROLES:
    INPUT:          accident_modifié, BD_EDA_accidents;
    OUTPUT:         descr_accident_modifié;
END TRANSFER-FUNCTION calculer-21;

```

A-1.2.1.2.3 La structure d'inférences



A-1.2.2. Classification

A-1.2.2.1 Les inférences

A-1.2.2.2 Les rôles de connaissances

A-1.2.2.3 La structure d'inférences

A-1.2.3. Interprétation des classes

A-1.2.3.1 Les rôles de connaissances

```

/*-----*/
----*/
/*      Inférences d'interprétation des classes pour construire les STA      */
/*-----*/
----*/

/*Rôles statiques et dynamiques*/
/*-----*/

KNOWLEDGE-ROLE description_classes;

```



```

        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        classe;
    END KNOWLEDGE-ROLE description_classes;

    KNOWLEDGE-ROLE description_classe;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        classe;
    END KNOWLEDGE-ROLE description_classe;

    KNOWLEDGE-ROLE connaissances_d-abstraction-3;
        TYPE: STATIC;
        DOMAIN-MAPPING:
        règles_abstraction_classe FROM connaiss_interprétation_classe;
    END KNOWLEDGE-ROLE connaissances_d-abstraction-3;

    KNOWLEDGE-ROLE description_abstraite_classe;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        description_abstraite_classe;
    END KNOWLEDGE-ROLE description_abstraite_classe;

    KNOWLEDGE-ROLE connaiss_interprétation_stat_intra;
        TYPE: STATIC;
        DOMAIN-MAPPING:
        critères_qualité_classification FROM connaiss_interprétation_classe;
    END KNOWLEDGE-ROLE connaiss_interprétation_stat;

    KNOWLEDGE-ROLE liste-critères-3;
        TYPE: STATIC;
        DOMAIN-MAPPING:
        critères_qualité_classification FROM connaiss_interprétation_classe;
    END KNOWLEDGE-ROLE liste-critères-3;

    KNOWLEDGE-ROLE critères-3;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        critères_qualité_classification FROM connaiss_interprétation_classe;
    END KNOWLEDGE-ROLE critères-3;

    KNOWLEDGE-ROLE critère-3;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        critères_qualité_classification FROM connaiss_interprétation_classe;
    END KNOWLEDGE-ROLE critère-3;

    KNOWLEDGE-ROLE valeur_critère-3;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        valeur_critère_sélectionné_qualité_classe;
    END KNOWLEDGE-ROLE valeur_critère-3;

    KNOWLEDGE-ROLE décision_pertinence-classe;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        oui_non-value;
    END KNOWLEDGE-ROLE décision_pertinence-classe;

    KNOWLEDGE-ROLE décision_supprimer-3;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        oui_non-value;
    END KNOWLEDGE-ROLE décision_supprimer-3;

    KNOWLEDGE-ROLE connaissances_modification-3;
        TYPE: STATIC;
        DOMAIN-MAPPING:
        connaissances_modification-classe FROM connaiss_interprétation_classe;
    END KNOWLEDGE-ROLE connaissances_modification-3;

    KNOWLEDGE-ROLE violations-3;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        critères_qualité_classification FROM connaissances_préparation_données;
    END KNOWLEDGE-ROLE violations-3;

    KNOWLEDGE-ROLE liste_actions-3;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        listes_modifications_classes FROM connaissances_préparation_données;
    END KNOWLEDGE-ROLE liste_actions-3;

    KNOWLEDGE-ROLE classe_modifiée;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        classe;
    END KNOWLEDGE-ROLE classe_modifiée;

    KNOWLEDGE-ROLE descr_classe_modifiée;
        TYPE: DYNAMIC;
        DOMAIN-MAPPING:
        classe;
    END KNOWLEDGE-ROLE descr_classe_modifiée;

    KNOWLEDGE-ROLE échantillon_d-étude;

```

```

TYPE: DYNAMIC;
DOMAIN-MAPPING:      échantillon_d-étude;
END KNOWLEDGE-ROLE échantillon_d-étude;

```

A-1.2.3.2 Les inférences

```

/*inférences de projection d'une classe sur un point de vue domaine choisi par l'utilisateur*/
/*-----*/

TRANSFER-FUNCTION obtenir-3;
TYPE: OBTAIN;
ROLES:
  INPUT:   classes;
  OUTPUT:  description_classes;
END TRANSFER-FUNCTION obtenir-3;

INFERENCE sélectionner-31;
ROLES:
  INPUT:   description_classes;
  OUTPUT:  description_classe;
  SPECIFICATION:
  "Il s'agit de sélectionner aléatoirement une classe";
END INFERENCE sélectionner-31;

INFERENCE abstraire_3;
ROLES:
  INPUT:   description_classe;
  OUTPUT:  description_abstraite_classe;
  STATIC:  connaissances_d-abstraction-3;
  SPECIFICATION: "
  l'entrée est composée de l'ensemble des classes décrites à l'aide de critères statistiques. la
  sortie est composée des mêmes classes mais dont la description est enrichie avec de nouvelles
  variables permettant de juger leurs pertinences.";
END INFERENCE abstraire_3;

INFERENCE spécifier-3;
OPERATION-TYPE: rechercher;
ROLES:
  INPUT:   description_abstraite_classe;
  OUTPUT:  critères-3;
  STATIC:  liste-critères-3;
  SPECIFICATION:
  "Il s'agit de rechercher les critères qui peuvent être utilisés pour l'évaluation de la pertinence
  d'un accident. Les mêmes connaissances statiques d'abstraction sont utilisées";
END INFERENCE spécifier-3;

INFERENCE sélectionner-3;
ROLES:
  INPUT:   critères-3;
  OUTPUT:  critère-3;
  SPECIFICATION:
  "Il s'agit de sélectionner l'un ou plusieurs de critères pour l'évaluation";
END INFERENCE sélectionner-3;

INFERENCE évaluer-3;
ROLES:
  INPUT:   critère-3, description_abstraite_classe;
  OUTPUT:  valeur_critère-3;
  SPECIFICATION: "
  déterminer la valeur de la critère (ou critère) choisi pour un cas donné. Les contraintes issus
  du Besoin interviennent dans l'évaluation.";
END INFERENCE évaluer-3;

INFERENCE combiner-3;
ROLES:
  INPUT:   valeur_critère-3;
  OUTPUT:  décision_pertinence-classe, violations-3;
  STATIC:  critères_qualité_classification, règles_abstraction_classe,
  critère_sélectionné_qualité_classe, modèles_multivues_domaine;
  SPECIFICATION:
  "en combinant les résultats des évaluations selon les différents critères choisi, décider si la
  classe est pertinente";
END INFERENCE combiner-3;

INFERENCE critiquer-3;
ROLES:
  INPUT:   violations-3;
  OUTPUT:  liste_actions-3;
  STATIC:  connaissances_modification-3;
  SPECIFICATION:
  "en regardant les critères de pertinences qui n'ont pas été respectés, déterminer les causes et
  la liste des actions à faire ";
END INFERENCE critiquer-3;

INFERENCE modifier-3;
ROLES:
  INPUT:   liste_actions-3, description_abstraite_classe;

```

```

OUTPUT: classe_modifiée;
SPECIFICATION:
"si la classe n'est pas supprimée, sa description est modifiée en choisissant d'autres
attributs/modalités descriptifs ";
END INFERENCE modifier-3;

TRANSFER-FUNCTION obtenir-31;
TYPE: OBTAIN;
ROLES:
INPUT: classe_modifiée;
OUTPUT: descr_classe_modifiée;
END TRANSFER-FUNCTION obtenir-31;

INFERENCE choisir-3;
ROLES:
INPUT: description_classes, préférences;
OUTPUT: scénario_type_accidents;
SPECIFICATION: "
Il s'agit de choisir les classes qui vont être utilisées dans l'étude parmi les classes obtenues.
Les préférences issus du Besoin de l'étude sont utilisé comme des inputs aussi ";
END INFERENCE choisir-3;

END INFERENCE-KNOWLEDGE
    
```

A-1.2.3.3 La structure d'inférences

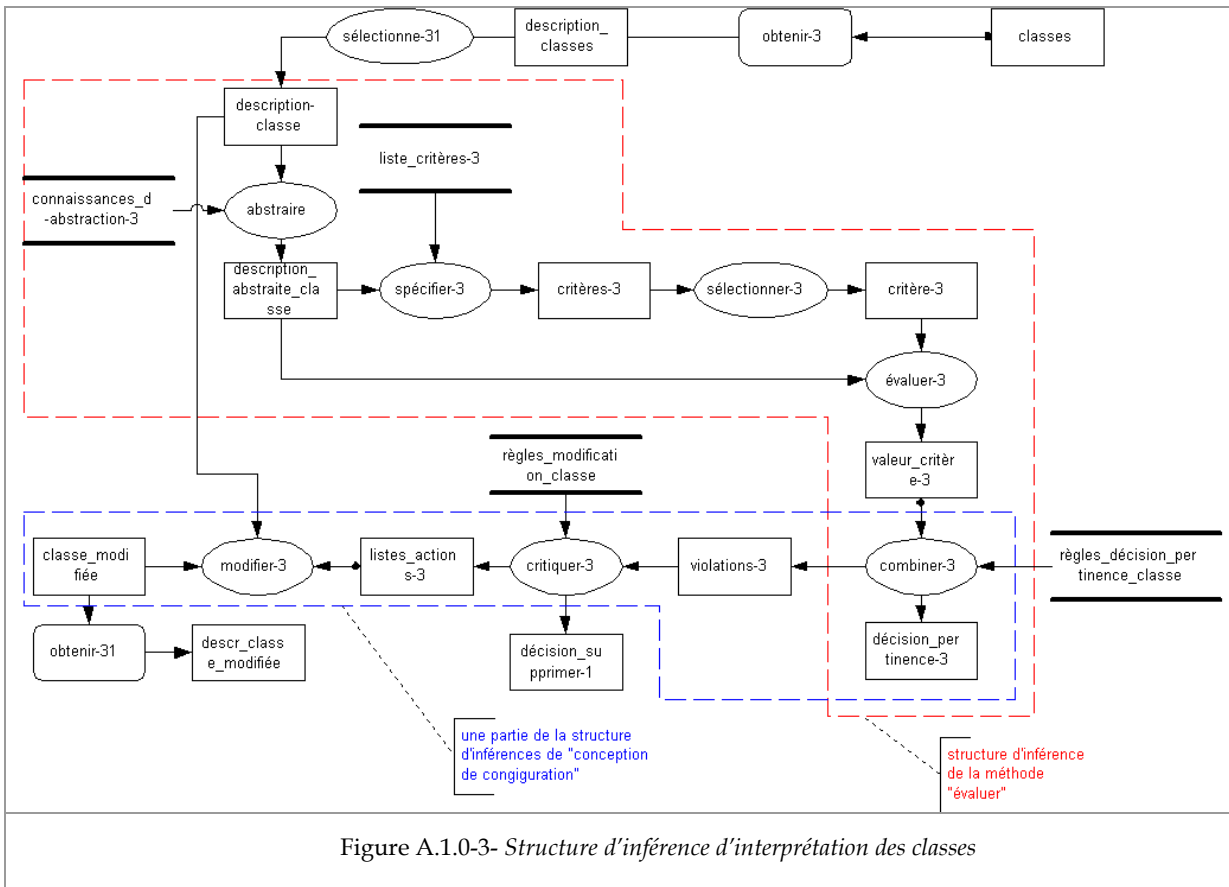


Figure A.1.0-3- Structure d'inférence d'interprétation des classes

A-1.3 Connaissance des la tâche

```

/*****
/* III. Connaissances tâches*/
*****/

TASK-KNOWLEDGE élaboration_scénarios;

TASK élaborer_des_STA;
DOMAIN-NAME: tâches_d-élaboration_de_scénarios_d-accidents;
GOAL: "
à partir d'un échantillon d'individus (e.g. accident de la route), élaborer des classes homogènes
et interprétables (e.g. des STA)";
    
```

```

    ROLES:
    INPUT:
    Besoin: "objectifs de l'étude pour laquelle on va élaborer des STA";
    BD_EDA_accidents: "il s'agit de la base de données d'accident (EDA) contenant des tables
accidents_x_attributs";
    modèles_multivues_domaine: "des modèles représentant différents points de vue du domaines";
    règles_abstraction_attributs: "règles utilisées pour l'abstraction des attributs";
    règles_décision_pertinence_attribut: "règles expertes pour juger la pertinence des attributs";
    règles_abstraction_classe: "règles d'abstraction des classes";
    critères_qualité_accident: "règles pour juger la qualité d'un accident lors de la sélection des
accidents";
    critères_qualité_classification : "critère pour juger la qualité d'un classe dans la phase
d'interprétation des classes";
    OUTPUT:
    description_synthétique: "descriptions synthétiques des classes jugées pertinentes vis-à-vis de
l'objectif de l'étude";
END TASK élaborer_des_STA;

TASK-METHOD préparer_classifier_interpréter;
REALIZES:
    élaborer_des_STA;
DECOMPOSITION:
    TASKS: préparer_données, classifier_données, interpréter_classes;
    ROLES:
    INTERMEDIATE:
    échantillon_d-étude: "échantillon accidents_x_attributs pertinents pour l'étude";
    classes_accidents: "Chaque classes regroupe des accidents qui sont similaire au sens de la
distance choisie pour la classification automatique";
    CONTROL-STRUCTURE:
    préparer_données(BD_EDA_accidents + Besoin + modèles_multivues_domaine +
règles_abstraction_attributs + règles_décision_pertinence_attribut + critères_qualité_accident ->
échantillon_d-étude);
    classifier_données(échantillon_d-étude -> classes_accidents);
    interpréter_classes (classes_accidents + règles_abstraction_classe +
critères_qualité_classification -> description_synthétique);
END TASK-METHOD assess-through-abstract-and-match;

/*-----*/
-----*/
/*      III.1- Tâche de préparation des données: sélection des attributs et des accidents      */
/*-----*/
-----*/

TASK préparer_données;
DOMAIN-NAME: construire_échantillon_d-étude;
GOAL: "
sélectionner les attributs pertinents pour l'étude. Ensuite sélectionner les accidents pertinents
et construire l'échantillons d'accidents pertinents pour l'étude";
ROLES:
INPUT:
Besoin: "objectifs de l'étude pour laquelle on va élaborer des STA";
BD_EDA_accidents: "il s'agit de la base de données d'accident (EDA) contenant des tables
accidents_x_attributs";
modèles_multivues_domaine: "des modèles représentant différents points de vue du domaines";
règles_abstraction_attributs: "règles utilisées pour l'abstraction des attributs";
règles_décision_pertinence_attribut: "règles expertes pour juger la pertinence des attributs";
règles_modification_attribut: "règles expertes permettant de déterminer les actions adéquates à
effectuer pour améliore la pertinence d'un attribut (e.g. regrouper mdalités, remplacer valeurs
inconnues, etc.) ou le supprimer";
critères_qualité_accident: "règles pour juger la qualité d'un accident lors de la sélection des
accidents";
OUTPUT:
échantillon_d-étude: "échantillon accidents_x_attributs pertinents pour l'étude";
END TASK préparer_données;

TASK-METHOD décrire_abstraire_évaluer_modifier;
REALIZES:
    préparer_données;
DECOMPOSITION:
    INFÉRENCES: opérationnaliser;
    TASKS: décrire_attribut, sélectionner_attribut_d-étude, sélectionner_accidents_d-étude;
    ROLES:
    INTERMEDIATE:
    contraintes: "elles définient des restrictions de choix pour définir des éléments valables (e.g.
attributs, accidents ou classes). ces contraintes peuvent être exprimés à travers des valeurs
seuils de certains critères ";
    préférences: "elles permettent de choisir entres des éléments qui sont tous valables";
    CONTROL-STRUCTURE:
    opérationnaliser (Besoin + modèles_multivues_domaine -> préférences + contraintes);
END TASK-METHOD décrire_abstraire_évaluer_modifier;

/******/

TASK décrire_attribut;
DOMAIN-NAME: description_des_attributs;
GOAL: "
obtenir une description des attributs à l'aide de paramètres statistiques permettant leur
évaluation";
ROLES:

```

```

INPUT:
BD_EDA_accidents: "il s'agit de la base de données d'accident (EDA) contenant des tables";
contraintes: "elles définissent des restrictions de choix pour définir des éléments valables (e.g.
attributs, accidents ou classes). ces contraintes peuvent être exprimés à travers des valeurs
seuils de certains critères ";
OUTPUT:
description_attribut: "une description des attributs à l'aide de paramètres statistiques
permettant son évaluation";
END TASK décrire_attribut;

TASK-METHOD sélectionner_calculer;
REALIZES:
décrire_attribut;
DECOMPOSITION:
INFERENCES: sélectionner-1, calculer-1;
ROLES:
INTERMEDIATE:
attributs: "l'ensemble des attributs existants dans la base de données EDA et desquels on choisi
les attributs pour une étude donnée";
CONTROL-STRUCTURE:
sélectionner-1 (BD_EDA_accidents + contraintes -> attributs);
calculer-1 (attributs + BD_EDA_accidents -> description_attribut);
END TASK-METHOD sélectionner_calculer;

/*****/

TASK sélectionner_attribut_d'étude;
DOMAIN-NAME: sélectionner_attribut_d'étude;
GOAL:
"sélectionner les attributs pertinents et construire l'échantillon duquel on va sélectionner les
accidents pertinents pour construire l'échantillon d'étude";
ROLES:
INPUT:
description_attributs: "l'ensemble des attributs décrits par des paramètres statistiques
permettant leur évaluation";
règles_abstraction_attributs: "règles utilisées pour l'abstraction des attributs";
règles_décision_pertinence_attribut: "règles expertes pour juger la pertinence des accidents";
règles_modification_attribut: "règles expertes permettant de déterminer les actions adéquates à
effectuer pour améliorer la pertinence d'un attribut (e.g. regrouper modalités, remplacer valeurs
inconnues, etc.) ou le supprimer";
BD_EDA_accidents: "il s'agit de la base de données d'accident (EDA) contenant des tables";
OUTPUT:
échantillon-1: "échantillon d'accidents décrit par les attributs pertinents pour l'étude";
END TASK sélectionner_attribut_d'étude;

TASK-METHOD abstraire_évaluer_modifier_attribut;
REALIZES:
sélectionner_attribut_d'étude;
DECOMPOSITION:
INFERENCES: sélectionner-11, abstraire-1, spécifier-1, sélectionner-12, évaluer-1, combiner-1,
critiquer-1, modifier-1, calculer-11, choisir-1, instancier;
ROLES:
INTERMEDIATE:
description_attribut: "une description d'un attribut à l'aide de paramètres statistiques
permettant son évaluation";
description_abstraite_attribut: "une description abstraite de l'attribut facilitant son
évaluation";
critères-1: "liste de critères d'évaluation";
critère-1: "critère choisi pour l'évaluation";
valeur_critère-1: "valeur du critère";
décision_pertinence_attribut: "décider si oui ou non l'attribut est pertinent";
décision_supprimer-1: "décider s'il faut écarter l'attribut de l'étude";
violations-1: "déterminer pourquoi l'attribut n'est pas pertinent";
liste_actions-1: "définir une liste d'action de modification à effectuer pour améliorer la
pertinence de l'attribut";
attribut_modifié: "description de l'attribut après sa modification";
descr_attribut_modifié: "description de l'attribut après sa modification";
CONTROL-STRUCTURE:
WHILE NOT EMPTY description_attributs DO
sélectionner-11 (description_attributs -> description_attribut);
description_attributs := description_attributs DELETE description_attribut;
WHILE NEW-SOLUTION abstraire(description_attribut -> description_abstraite_attribut) DO
/* réutiliser la description abstraite comme input de l'inférence abstraire, jusqu'à ce
qu'il n'y a plus la possibilité de faire des abstractions */
description_attribut := description_abstraite_attribut;
END WHILE
spécifier-1 (description_abstraite_attribut + liste-critères-1 -> critères-1);
REPEAT
sélectionner-12 (critères-1 -> critère-1);
évaluer-1 (description_abstraite_attribut + critère-1 -> valeur_critère-1);
résultats_évaluation-1 := valeur_critère-1 ADD résultats_évaluation-1;
UNTIL
HAS-SOLUTION combiner-1 (résultats_évaluation-1 + règles_décision_pertinence_attribut ->
décision_pertinence_attribut + violations-1);
END REPEAT
critiquer-1 (violations-1 + règles_modification_attribut -> décision_supprimer-1 +
liste_actions-1);
IF décision_supprimer-1 == 'oui'
THEN

```

```

/*les attributs supprimés ne seront pas utilisés dans l'étude*/
attributs := attributs DELETE attribut;
ELSE
  modifier-1 (description_attribut + liste_actions-1 -> attribut_modifié);
  calculer-11 (BD_EDA_accidents + attribut_modifié -> descr_attribut_modifié);
  /*après modification, les attributs seront réévalués*/
  description_attributs := descr_attribut_modifié ADD description_attributs;
END IF
END WHILE
/* la liste des attributs qui restent dans l'objet 'attributs' est la liste d'attributs d'étude.
elle est donc utilisée pour constituer l'échantillon-1 duquel on va sélectionner les accidents
pertinents pour construire l'échantillon final, i.e. échantillon_d-étude*/

choisir-1 (préférences + attributs -> attributs_d-étude);
instancier (attributs + BD_EDA_accidents -> échantillon-1);

END TASK-METHOD abstraire_évaluer_modifier_attribut;

/*****/

TASK sélectionner_accidents_d-étude;
  DOMAIN-NAME: sélectionner_accidents_d-étude;
  GOAL:
  "sélectionner les accidents pertinents et construire l'échantillon d'étude";
  ROLES:
  INPUT:
  BD_EDA_accidents: "il s'agit de la base de données d'accident (EDA) contenant des tables
accidents_x_attributs";
  échantillon-1: "échantillon d'accidents décrits par les attributs pertinents pour l'étude";
  critères_qualité_accident: "critères pour juger la pertinence d'un accident";
  règles_abstraction_accidents: "règles utilisées pour l'abstraction des accidents";
  règles_modification_accident: "règles expertes permettant de déterminer les actions adéquates à
effectuer pour améliorer la pertinence d'un accident (e.g. remplacer valeurs inconnues, etc.) ou
le supprimer";
  OUTPUT:
  échantillon_d-étude: "échantillon accidents_x_attributs qui va être utilisé dans l'étude";
END TASK sélectionner_accidents_d-étude;

TASK-METHOD sélectionner_abstraire_évaluer_modifier_accident;
  REALIZES:
  sélectionner_accidents_d-étude;
  DECOMPOSITION:
  INFERENCE: sélectionner-2, abstraire-2, spécifier-2, sélectionner-21, évaluer-2, combiner-2,
critiquer-2, modifier-2;
  ROLES:
  INTERMEDIATE:
  description_accident: "une description d'un accident à l'aide de paramètres statistiques
permettant son évaluation";
  description_abstraite_accident: "description abstraite d'un accident facilitant son évaluation";
  critères-2: "critères d'évaluation";
  critère-2: "le critère choisi pour l'évaluation";
  valeur_critère-2: "la valeur du critère";
  décision_pertinence_accident: "décider si oui ou non l'accident est pertinent pour l'étude";
  décision_supprimer-2: "décider s'il faut écarter l'accident de l'étude";
  violations-2: "déterminer pourquoi l'accident n'est pas pertinent";
  liste_actions-2: "définir une liste d'actions de modification à effectuer pour améliorer la
pertinence de l'accident";
  accident_modifié: "description de l'accident après sa modification";
  descr_accident_modifié: "description de l'accident après sa modification";
  CONTROL-STRUCTURE:
  calculer-2 (échantillon-1 -> description_accidents);
  WHILE NOT EMPTY description_accidents DO
    sélectionner-2 (description_accidents -> description_accident);
    description_accidents := description_accidents DELETE description_accident;
    WHILE NEW-SOLUTION abstraire(description_accident -> description_abstraite_accident) DO
      /* réutiliser la description abstraite comme input de l'inférence abstraite, jusqu'à ce
qu'il n'y a plus la possibilité de faire des abstractions */
      description_accident := description_abstraite_accident;
    END WHILE
    spécifier-2 (description_abstraite_accident + liste-critères-2 -> critères-2);
    REPEAT
      sélectionner-21 (critères-2 -> critère-2);
      évaluer-2 (description_abstraite_accident + critère-2 -> valeur_critère-2);
      résultats_évaluation-2 := valeur_critère-2 ADD résultats_évaluation-2;
    UNTIL
      HAS-SOLUTION combiner-2 (résultats_évaluation-2 + critères_qualité_accident ->
décision_pertinence_accident + violations-2);
    END REPEAT
    critiquer-2 (violations-2 + règles_modification_accident -> décision_supprimer-2 +
liste_actions-2);
    IF décision_supprimer-2 == 'oui'
    THEN
      /*les attributs supprimés ne seront pas utilisés dans l'étude*/
      échantillon-1 := échantillon-1 DELETE description_accident;
    ELSE
      modifier-2 (description_accident + liste_actions-2 -> accident_modifié);
      calculer-21 (BD_EDA_accidents + accident_modifié -> descr_accident_modifié);
      /*après modification, les attributs seront réévalués*/
      description_accidents := descr_accident_modifié ADD description_accidents;
    END IF
  END WHILE

```

```

        END IF
    END WHILE
    échantillon_d-étude := échantillon-1;
END TASK-METHOD sélectionner_abstraire_évaluer_modifier_accident;
/*-----*/
/*      III.2- Tâche de classification des accidents      */
/*-----*/

TASK classifieur_données;
    DOMAIN-NAME: construire_échantillon_d-étude;
    GOAL: "
    sélectionner les attributs pertinents pour l'étude. Ensuite sélectionner les accidents pertinents
    et construire l'échantillons d'accidents pertinents pour l'étude";
    ROLES:
        INPUT:
            échantillon_d-étude: "échantillon accidents_x_accidents pertinents pour l'étude";
        OUTPUT:
            classes_accidents: "Chaque classes regroupe des accidents qui sont similaire au sens de la
            distance choisie pour la classification automatique";
END TASK classifieur_données;

/*-----*/
/*      III.3- Tâche d'interprétation des classes      */
/*-----*/

TASK interpréter_classes;
    DOMAIN-NAME: élaboration_scénarios;
    GOAL:
    "évaluer la pertinence statistique et sémantique des classes et synthétiser leurs descriptions
    pour générer des scénarios-type d'accidents";
    ROLES:
        INPUT:
            description_classes: "chacune des classes est décrite des modalités affectées de critères
            statistiques";
            connaissances_d-abstraction-3: "règles utilisées pour l'abstraction des classes en trouvant des
            descriptions plus faciles à utiliser dans l'évaluation de ces classe";
            connaissances_interprétation_stat_intra: "règles aidant à l'interprétation statistique et sémantique
            des classes";
            connaissances_modification-classe: "règles expertes permettant de déterminer les actions adéquates à
            effectuer pour améliorer la pertinence d'une classe (e.g. trouvant d'autre modalités
            discriminantes, changer les attributs utilisés dans la classification, etc.) ou la supprimer";
        OUTPUT:
            scénario_type_accidents: "descriptions synthétiques des classes et pertinentes vis-à-vis du
            Besoin de l'étude";
END TASK interpréter_classes;

TASK-METHOD abstraire_évaluer_modifier_attribut;
    REALIZES:
        interpréter_classes;
    DECOMPOSITION:
    INFERENCE: obtenir-3, sélectionner-3, abstraire-3, spécifier-3, sélectionner-31, évaluer-3,
    combiner-3, critiquer-3, modifier-3, obtenir-31, choisir-3;
    ROLES:
        INTERMEDIATE:
            description_classe: "une description d'une classe à l'aide de paramètres statistiques permettant
            son évaluation";
            description_abstraite_classe: "une description abstraite de la classe facilitant son évaluation";
            critères-3: "liste de critères d'évaluation";
            critère-3: "critère choisi pour l'évaluation";
            valeur_critère-3: "valeur du critère";
            décision_pertinence-classe: "décider si oui ou non la classe est pertinente pour l'étude";
            décision_supprimer-3: "décider s'il faut écarter la classe de l'étude";
            violations-3: "déterminer pourquoi la classe n'est pas pertinente";
            liste_actions-3: "définir une liste d'action de modification à effectuer pour améliorer la
            pertinence de la classe";
            classe_modifiée: "description de la classe après sa modification";
            descr_classe_modifiée: "description de la classe après sa modification";
        CONTROL-STRUCTURE:
            obtenir-3 (classes -> description_classes);
            WHILE NOT EMPTY description_classes DO
                sélectionner-31 (description_classes -> description_classe);
                description_classes := description_classes DELETE description_classe;
                WHILE NEW-SOLUTION abstraire(description_classe -> description_abstraite_classe) DO
                    /* réutiliser la description abstraite comme input de l'inférence abstraire, jusqu'à ce
                    qu'il n'y a plus la possibilité de faire des abstractions */
                    description_classe := description_abstraite_classe;
                END WHILE
                spécifier-3 (description_abstraite_classe + liste-critères-3 -> critères-3);
            REPEAT
                sélectionner-3 (critères-3 -> critère-3);
                évaluer-3 (description_abstraite_classe + critère-3 -> valeur_critère-3);
                résultats_évaluation-3 := valeur_critère-3 ADD résultats_évaluation-3;
            UNTIL

```

```

HAS-SOLUTION combiner-3 (résultats_évaluation-3 + règles_décision_pertinence_classe ->
décision_pertinence-classe + violations-3);
END REPEAT
critiquer-3 (violations-3 + règles_modification_classe -> décision_supprimer-3 +
liste_actions-3);
IF décision_supprimer-3 == 'oui'
THEN
/*les classes supprimées ne seront pas utilisés pour construire les STA*/
classes := classes DELETE classe;
ELSE
modifier-3 (description_classe + liste_actions-3 -> classe_modifiée);
obtenir-31 (classe_modifiée -> descr_classe_modifiée);
/*après modification, les attributs seront réévalués*/
description_classes := descr_classe_modifiée ADD description_classes;
END IF
END WHILE
/* la liste des classes qui restent dans l'objet 'classes' est la liste de classes dont la
description synthétique représente les scénarios d'accidents recherchés*/
choisir-3 (préférences + attributs -> attributs_d-étude);

END TASK-METHOD abstraire_évaluer_modifier_attribut;

END TASK-KNOWLEDGE élaboration_scénarios;

END KNOWLEDGE-MODEL Scénarios_accidents;

```

Une représentation graphique de la tâche d'élaboration des STA est donnée dans la Figure A.1.0-4. La représentation graphique d'une tâche se fait à travers un rectangle dont les coins sont arrondis. La représentation d'une méthode de tâche se fait à l'aide d'un rectangle :

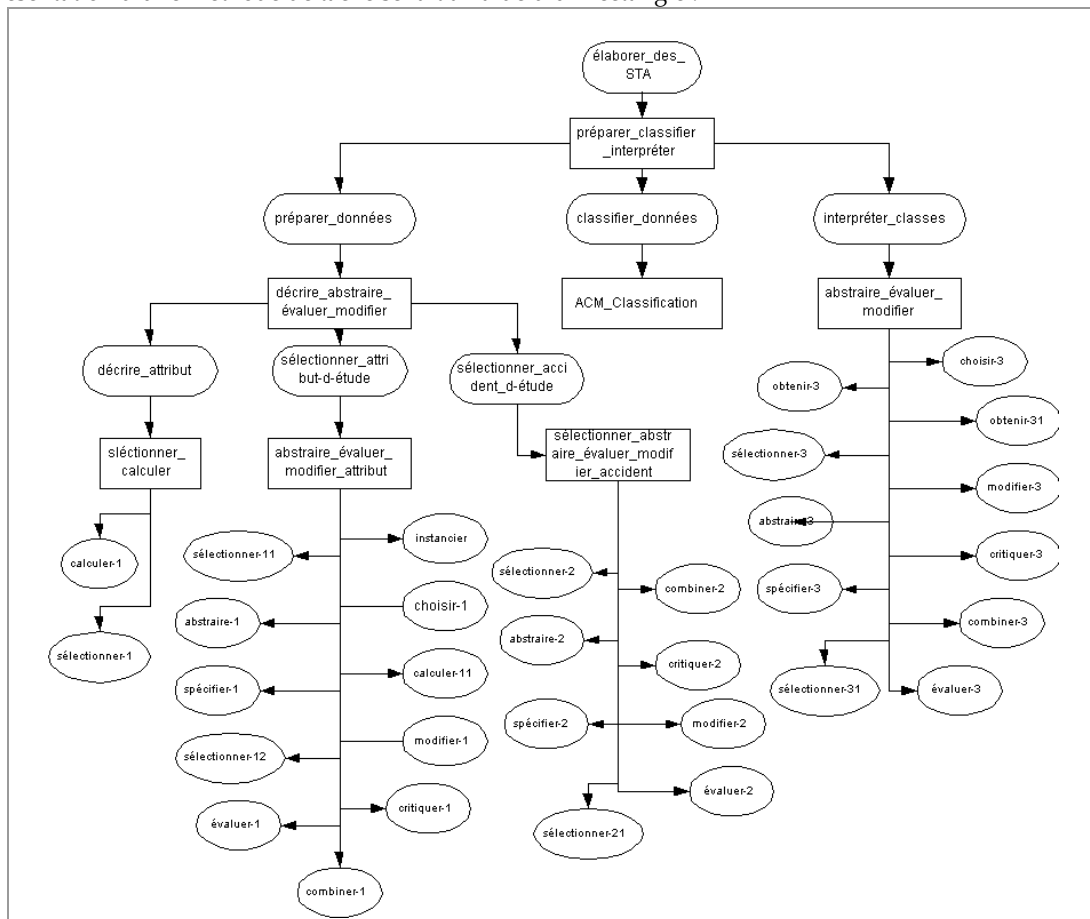


Figure A.1.0-4- Représentation graphique de la tâche d'élaboration des STA.

A-2. Enrichir la base EDA en utilisant ASMEC

description de la modalité	étape fonctionnelle
visibilité amoindrie(nuit/pluie)	perception
manœuvre autrui	
obstacle non éclairé	
visibilité limitée	
pression situationnelle induisant une manœuvre précipitée	
difficulté créneau de traversé ou d'insertion	
défaut aménagements	
visibilité limitée infra	
défaut signalisation	
gêne à la visibilité ponctuelle(soleil,autre VL,...)	
défaut d'éclairage de la zone	
présence d'un usager sans éclairage sur la chaussée	
identification risque	
tâche annexe	
pb directionnel	
contrainte tps situation	
statut prioritaire/sentiment prioritaire	
méconnaissance lieu	
faible expé conduite	
surexpé manœuvre	
surexpé trajet	
faible attention	
vigilance faible/fatigue	
confiance excessive dans les signaux	
handicap visuel	
défaut aménagements	
visibilité limitée infra	
comportement gênant d'un usager	
polysémie indices émis par autrui	
absence indices annonceurs d'une manœuvre	
tracé difficile(virage serré,en rupture)	
défaut présignalisation(insuffisante,peu visible)	
gêne à la visibilité ponctuelle(soleil,autre VL,...)	
aménagement de la zone incitant prise de vitesse	
défaut de pression des pneus	
identification risque	
pb directionnel	
contrainte tps situation	
méconnaissance lieu	
faible expé conduite	
surexpé manœuvre	
surexpé trajet	
faible attention	
vigilance faible/fatigue	
conduite à risque(ludique,test d'un VL,transgression)	
vitesse élevée	
contrainte temps	
distracted/discussion passager	
alcoolisation légère	
âge(lenteur de réaction)	
confiance excessive dans les signaux	pronostic
manœuvre autrui	
visibilité limitée infra	
absence indices annonceurs d'une manœuvre	
aménagement de la zone incitant prise de vitesse	
chaussée étroite	
statut prioritaire/sentiment prioritaire	
faible expé conduite	
surexpé trajet	
conduite à risque(ludique,test d'un VL,transgression)	
vitesse élevée	
contrainte temps	
banalisation situation	
confiance excessive dans les signaux	
visibilité limitée	
difficulté créneau de traversé ou d'insertion	
défaut aménagements	
visibilité limitée infra	

Figure A.2.1- ASMEC : approche systémique de modélisation multi-vues de connaissances (première approche dans SAF-Next)

A-3. Représentation XML de la projection d'un accident sur les méta-données multi-vues

A-3.1 Exemple d'utilisation de XML

Nous prenons un exemple simple donné [Van Lancker, 2004] pour expliquer l'utilisation d'XML pour la structuration des données. Il s'agit d'un fichier XML contenant des titres de chansons en mp3 (cf. Figure A.3.1)

```
<?xml version="1.0" ?>
- <compilation>
  = <mp3>
    <titre>Foule sentimentale</titre>
    <artiste>Alain Souchon</artiste>
  </mp3>
  = <mp3>
    <titre>Solaar pleure</titre>
    <artiste>MC Solaar</artiste>
  </mp3>
  = <mp3>
    <titre>Le baiser</titre>
    <artiste>Alain Souchon</artiste>
  </mp3>
  = <mp3>
    <titre>Pourtant</titre>
    <artiste>Vanessa Paradis</artiste>
  </mp3>
  = <mp3>
    <titre>Chambre avec vue</titre>
    <artiste>Henri Salvador</artiste>
  </mp3>
</compilation>
```

Figure A.3.1- Exemple XML

Comme le XML n'utilise pas des balises prédéfinies, le navigateur ne « comprend » pas les balises du XML et ne sait pas comment afficher un document XML. Pour le faire, il est nécessaire d'avoir un mécanisme pour décrire comment le document pourrait être affiché. Les feuilles de style classiques du Html (CSS) sont l'un de ces mécanismes. Mais, le XSL pour eXtensible Stylesheet Language est de loin un langage de feuille de style plus adapté au XML et donc plus performant.

Le XSL ne permet pas uniquement l'affichage de XML. Il permet aussi :

- De sélectionner une partie des éléments XML ;
- De trier des éléments XML ;
- De filtrer des éléments XML en fonction de certains critères ;
- De choisir des éléments ;
- De retenir des éléments par des tests conditionnels.

La figure A.3.2 présente le contenu du fichier XSL à associer avec le fichier XML de la Figure A.3.1). Ce fichier XSL permet de sélectionner et colorer en vert les titres de « Alain Souchon ».

```
<?xml version="1.0" ?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD-xsl">
  <xsl:template match="/">
    <html>
      <body>
        <table border="1" cellspacing="0" cellpadding="3">
          <tr bgcolor="#FFFF00">
            <td>Artiste</td>
            <td>Titre</td>
          </tr>
          <xsl:for-each select="compilation/mp3">
            <xsl:choose>
              <xsl:when test=".[artiste='Alain Souchon']">
                <tr bgcolor="#00FF00">
                  <td><xsl:value-of select="titre" /></td>
                  <td><xsl:value-of select="artiste" /></td>
                </tr>
              </xsl:when>
              <xsl:otherwise>
                <tr>
                  <td><xsl:value-of select="titre" /></td>
                  <td><xsl:value-of select="artiste" /></td>
                </tr></xsl:otherwise></xsl:choose>
            </xsl:for-each>
          </table>
        </body>
      </html>
    </xsl:template>
  </xsl:stylesheet>
```

Figure A.3.2- Exemple XSL

Le résultat obtenu sur le navigateur est donné dans la Figure A.3.3.

Artiste	Titre
Foule sentimentale	Alain Souchon
Solaar pleure	MC Solaar
Le baiser	Alain Souchon
Pourtant	Vanessa Paradis
Chambre avec vue	Henri Salvador

Figure A.3.3- Affichage du XML à l'aide du XSL

A-3.2 Projection sur les métadonnées transformationnelles

```

<?xml version="1.0"?>
<Base_de_donnees_daccidents>
<Accident>
<N_Accident> N 503 </N_Accident>
<Point2vue>
  <Nom_Point2vue> Vue_Transformationnelle </Nom_Point2vue>

  <Concept>
    <NomDuConcept>Situation_avant_la_conduite</NomDuConcept>
    <Attributs>
      <Age>51</Age><Situation_Professionnelle>Electricien</Situation_Professionnelle><Modele_d
        u_vehicule> Peugeot 206 </Modele_du_vehicule>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Situation_de_Conduite_normale </NomDuConcept>
    <Attributs>
      <Type_de_linfrastructure>
        Intersection</Type_de_linfrastructure><Conditions_de_surface>Sec
      </Conditions_de_surface><Vitesse_declaree>20</Vitesse_declaree>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Situation_de_Rupture </NomDuConcept>
    <Attributs>
      <Perception> Defaillance </Perception> <Difficulte_de_vision> Soleil
      </Difficulte_de_vision><Type_du_masque> Barriere anti-choc </Type_du_masque>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Situation_durgence </NomDuConcept>
    <Attributs>
      <Action_avant_le_crash>Pas de reaction</Action_avant_le_crash> <Action_apres_le_crash>
        Freinage </Action_apres_le_crash>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Situation_de_Choc </NomDuConcept>
    <Attributs>
      <Premier_impact> contre un vehicule </Premier_impact> <Deploiement_de_lairbag> Oui
      </Deploiement_de_lairbag><Cote_impacte> Frontal </Cote_impacte>
    </Attributs>
  </Concept>
</Point2vue>

```

A-3.3 Projection sur les méta-données ontologiques

```

<Point2vue>
  <Nom_Point2vue> Vue_ontologique </Nom_Point2vue>
  <Concept>
    <NomDuConcept>Conducteur</NomDuConcept>
    <Attributs>
      <Age> 51 </Age> <Situation_Professionnelle> Electricien </Situation_Professionnelle>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Vehicule</NomDuConcept>
    <Attributs>
      <Modele_du_vehicule> Ford Escort </Modele_du_vehicule>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Environnement</NomDuConcept>
    <Attributs>
      <Type_de_linfrastructure> Intersection </Type_de_linfrastructure> <Conditions_de_surface>
        Sec </Conditions_de_surface>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Cond_Veh</NomDuConcept>
    <Attributs>

```

```

    <Vitesse_declaree> 20 </Vitesse_declaree> <Action_avant_le_crash> Pas de reaction
    </Action_avant_le_crash><Action_apres_le_crash> Freinage </Action_apres_le_crash>
  </Attributs>
</Concept>
<Concept>
  <NomDuConcept> Cond_Envir</NomDuConcept>
  <Attributs>
    <Perception> Defaillance </Perception> <Difficulte_de_vision> Soleil
    </Difficulte_de_vision> <Type_du_masque> Barriere anti-choc </Type_du_masque>
  </Attributs>
</Concept>
<Concept>
  <Attributs>
    <Premier_impact> contre un vehicule </Premier_impact>
  </Attributs>
</Concept>
</Point2vue>

```

A-3.4 Projection sur les méta-données fonctionnelles

```

<Point2vue>
  <Nom_Point2vue> Vue_Fonctionnelle</Nom_Point2vue>
  <Concept>
    <NomDuConcept>Etape_de_Perception</NomDuConcept>
    <Attributs>
      <Action_perception>regarde a gauche</Action_perception><Defaillance_perception>
      Contr.visuel_inadapte </Defaillance_perception><Visibilite> Reduite
      </Visibilite><Difficulte_de_vision> Soleil </Difficulte_de_vision>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Etape_de_Diagnostic</NomDuConcept>
    <Attributs>
      <Estimation_de_danger> etat de securite </Estimation_de_danger><Vitesse_estimee> 20
      </Vitesse_estimee>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Etape_de_Pronostique</NomDuConcept>
    <Attributs>
      <Defaut_danticipation> Oui </Defaut_danticipation>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Etape_de_Decision</NomDuConcept>
    <Attributs>
      <Temps_de_reaction_necessaire> 2s </Temps_de_reaction_necessaire>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Etape_dAction</NomDuConcept>
    <Attributs>
      <Action_avant_crash> Pas de reaction </Action_avant_crash><Action_apres_crash> Freinage
      </Action_apres_crash>
    </Attributs>
  </Concept>
</Point2vue>

```

A-3.5 Projection sur les méta-données téléologiques

```

<Point2vue>
  <Nom_Point2vue> Vue_Teleologique</Nom_Point2vue>
  <Concept>
    <NomDuConcept>Navigation</NomDuConcept>
    <Attributs>
      <Action_perception>regarde a gauche</Action_perception><Defaillance_perception>
      Contr.visuel_inadapte </Defaillance_perception><Visibilite> Reduite
      </Visibilite><Difficulte_de_vision> Soleil </Difficulte_de_vision>
    </Attributs>
  </Concept>
  <Concept>
    <NomDuConcept> Contrôle_latéral</NomDuConcept>
    <Attributs>
      <Action_avant_le_crash> Pas de reaction</Action_avant_le_crash>
    </Attributs>
  </Concept>

```

```
</Attributs>
</Concept>
<Concept>
  <NomDuConcept> Contrôle_longitudinal</NomDuConcept>
  <Attributs>
    <Action_avant_crash> Pas de reaction </Action_avant_crash><Action_apres_crash> Freinage
    </Action_apres_crash>
  </Attributs>
</Concept>
<Concept>
  <NomDuConcept> Guidage_latéral</NomDuConcept>
  <Attributs>
    <Manœuvre_avant_accident> aborde un rond point </Manœuvre_avant_accident>
  </Attributs>
</Concept>
<Concept>
  <NomDuConcept> Guidage_longitudinal</NomDuConcept>
  <Attributs>
    <Manœuvre_avant_accident>suivre une file de voitures</Manœuvre_avant_accident>
  </Attributs>
</Concept>
</Point2vue>
</Accident>
</Base_de_donnees_daccidents> >
</Accident>
</Base_de_donnees_daccidents>
```

A-4. Traduction de l'objectif de l'étude sous forme d'une requête SQL

L'interface présentée dans la Figure 9-4, (page 125) permet à l'expert de traduire son objectif en fonction des modèles et des concepts du domaine identifiés et formalisés dans ASMEC. Cette interface traduit automatiquement la réponse de l'expert sous forme d'une requête SQL. La Figure A.4.1 en présente un exemple.

```

SELECT
Attribut_tabl.[systemic_aspect],
Attribut_tabl.[Components_interaction],
Attribut_tabl.[Functional_step], Attribut_tabl.[Transformational_step],

FROM Attribut_table
WHERE
(((Attribut_tabl.[systemic_aspect])="Functional") OR
(Attribut_tabl.[systemic_aspect])="Transformational") OR
(Attribut_tabl.[systemic_aspect])="Teleological" AND
(Attribut_tabl.[Components_interaction])="Driv/Vehicl" OR
(Attribut_tabl.[Components_interaction])="Driv/Envir" AND
(Attribut_tabl.[Functional_step])="Perception" OR
(Attribut_tabl.[Functional_step])="Diagnostic" OR
(Attribut_tabl.[Functional_step])="Prognostic" AND
(Attribut_tabl.[Transformational_step])="Accident_situation" OR
(Attribut_tabl.[Transformational_step])=" Emergency_situation"));

```

Figure A.4.1- Requête générée par l'interface AICEF présentée dans la Figure 9-4, p°125

Puisque les méta-données font le lien entre les concepts et les attributs, cette requête permet de sélectionner les attributs qui répondent au besoin de l'expert. Ainsi, les attributs sont sélectionnés selon le choix de l'utilisateur en combinant plusieurs points de vue existant dans le domaine et qui ont été identifiés par le méta-modèle (cf. §8.6.2.3 section 8.7) et formalisés par les experts (cf. section 8.8). Il s'agit donc d'un outil pour traduire, par l'expert, l'objectif de l'étude en fonction des concepts du domaine.

A-5. Une autre formalisation d'ASMEC

Dans la Figure 8-3 (page 102), nous avons présenté le modèle UML de l'approche ASMEC. Une autre illustration est donnée dans le modèle présenté dans la Figure A.5.1.

```

<A, PV1, PV2, R_PV1_PV2, MS, R_MS_PV, CS, R_CS_PV, At, R_At_CS, V, R_V_At>,
avec
- A : Un ensemble d'instances (e.g. accidents) ;
- PV1 : l'ensemble des points de vue instrumentant le premier niveau du méta-
modèle: point de vue « objet observé », point de vue « observateur » et
point de vue « contexte de l'observation » ;
- PV2 : l'ensemble des axes systémiques instrumentant chacun des points de
vue du niveau PV1. Il s'agit des axes : ontologique, fonctionnel,
transformationnel et téléologique ;
- R_PV1_PV2 : une relation d'« appartenance » instrumentant chacun des points
de vue du premier niveau du méta-modèle par les axes systémiques ;
- MS : L'ensemble de modèles du domaine pour représenter un accident;
- R_MS_PV/observateur/contexte_d'observation : MS ----> PV : relation
d'« appartenance » assignant à chacun des axes systémiques un ou plusieurs
modèles en tenant compte du point de vue observateur (Exemple : choisir les
modèles qui sont utilisés par les experts dans la tâche de construction des
STA) et du point de vue contexte (Exemple : choisir des modèles de
représentation de l'accident chez le concepteur. C'est une relation un-à-
plusieurs. Exemple : Point de vue ontologique  $\supseteq$  {modèle Conducteur-Véhicule-
Environnement, modèle du conducteur, modèle du véhicule, modèle de
l'infrastructure});
- CS : l'ensemble de concepts du domaine qui instrumentent chacun des
modèles. Exemple : { {conducteur, véhicule, infrastructure, trafic,
conditions ambiantes, interactions}, {conduite normale, rupture, urgence,
choc}, {perception, diagnostic, pronostique, décision, action}, [contrôle,
navigation, guidage]};
- R_CS_PV : CS ----> MS est une relation d'« appartenance » assignant chacun
des concepts dans CS au modèle du domaine correspondant MS. Exemple :
{conduite normale, rupture, urgence, choc}  $\subseteq$  Modèle séquentiel  $\in$  Axe
transformationnel  $\in$  Point de vue objet observé ;
- At : L'ensemble d'attributs caractérisant l'objet observé (e.g. accident)
dans la base de données utilisée. Exemple : age, type véhicule, température;
- R_At_CS : At ----> CS : une relation de « caractérisation » assignant chacun
des attributs à un ou plusieurs concepts. Autrement dit, c'est la projection
de chacun des attributs selon les différents modèles dans MS. Un seul
attribut peut caractériser plusieurs concepts appartenant des modèles
différents. C'est cette relation qui permet la construction des métadonnées;
- V : Un ensemble de valeurs. Exemple : les modalités des différents
attributs ;
- R_V_At : A, At ----> V est une relation d'« instanciation » qui assigne à
chacun des attributs et pour un accident donné une valeur.

```

Figure A.5.1- Modèle de représentation de l'approche ASMEC

Ce modèle résume l'architecture à plusieurs niveaux d'abstraction et multi-vues d'ASMEC que nous avons détaillée dans la section 8.4.

Les liens entre les différents points de vue du méta-modèle (niveaux 1 et 2) sont d'abord sémantiques et ils sont assurés par le fondement systémique des ces points de vue. Il est explicité dans l'architecture d'ASMEC présentée dans la section 8.4. Les trois relations « $R_{MS_PV} : MS \dashrightarrow PV$ », « $R_{CS_PV} : CS \dashrightarrow MS$ », et « $R_{At_CS} : At \dashrightarrow CS$ » formalisent ces liens. Par exemple, l'attribut « *coup de volant* » caractérise en même temps la « *phase d'urgence* » du *modèle séquentiel de l'axe transformationnel* du point de vue *objet observé* et la tâche de « *perception* » du *modèle de traitement d'information de l'axe fonctionnel* du point de vue *objet observé*. Ces relations permettent une lecture (interprétation) de l'accident selon les différents points de vue de la systémique. Nous avons illustré cela à travers un cas d'accident réel que nous avons projeté sur les différentes méta-données (cf. Figure 8-17, Figure 8-20, Figure 8-23 et Figure 8-26).

Liste des figures

Figure 1 - Une vue domaine de nos contributions : l'apport de SAFE-Next dans les domaines d'accidentologie, de conception, d'ingénierie des connaissances et d'extraction de connaissances de données	4
Figure 2- Guide de lecture du mémoire.....	5
Figure I.1- Partie I : Présentation de la problématique	8
Figure II.1- Partie II : Etat de l'art.	22
Figure 3-1- Réflexivité de l'approche systémique	34
Figure 4-1- Modèle CVE.	39
Figure 4-2- Modèle séquentiel de l'INRETS [Brenac, 1997].....	40
Figure 4-3- Modèle de traitement de l'information par le conducteur.....	43
Figure 4-4- Exemple de scénario d'accident présenté selon le modèle séquentiel [Brenac et Fleury, 1999].	46
Figure 6-1- Algorithme k-means.....	69
Figure 6-2- Algorithme k-means.....	70
Figure 6-3- Dendrogramme.....	71
Figure 6-4- Algorithme AGNES.....	71
Figure III-1- Partie III : Développement de SAFE-Next.....	81
Figure 7-1- Différents types de connaissances formalisés à l'aide de CommonKADS.....	83
Figure 7-2- Représentation du concept impliqué en CML.	84
Figure 7-3- Représentation d'une construction de type relation en CML	84
Figure 7-4- Schéma du domaine (sans les relations et les types-de-règle)	85
Figure 7-5- Exemple de règles d'évaluation de la pertinence d'un attribut.....	86
Figure 7-6- Description d'une classe à l'aide des modalités	87
Figure 7-7- Représentation graphique de l'inférence « opérationnaliser ».	89
Figure 7-8- Structure d'inférence générique pour l'évaluation d'un objet.....	90
Figure 7-9- Structure d'inférence générique pour la modification d'un objet	90
Figure 7-10- Structure d'inférence de sélection des attributs pertinents.	91
Figure 7-11- Illustrant de la méthode de classification mixte (Filière SPAD).	93
Figure 7-12- Tâche sélectionner_attribut_d'étude.....	94
Figure 7-13- Utilisation du rôle de connaissance d'abstraction des attributs dans la structure d'inférence.	96
Figure 7-14- L'une des règles abstraction des attributs dans la base de connaissances.	96
Figure 7-15- Utilisation du rôle de connaissances statique « modèles_multivues_domaine.....	97
Figure 8-1- ASMEC : approche systémique de modélisation multi-vues et multi-niveaux de granularité de connaissances (première approche dans SAFE-Next).....	98
Figure 8-2- Modélisation dirigée par un méta-modèle général (d'après [Le Moigne, 1999]).	101
Figure 8-3- Modèle UML de l'architecture de modélisation d'ASMEC.	102
Figure 8-4- Modèle UML du premier niveau du méta-modèle.	103
Figure 8-5- Modèle UML du deuxième niveau du méta-modèle.....	104
Figure 8-6- Modèle UML du méta-modèle dans ASMEC.....	104
Figure 8-7- Instrumentation des axes systémiques par des modèles.....	105

Figure 8-8- Développement du point de vue « objet observé » en tenant compte des points de vue « observateur » et « contexte d'observation ».....	110
Figure 8-9- Instrumentation des modèles par des concepts du domaine.....	111
Figure 8-10- Instrumentation des modèles par les concepts et les concepts par les attributs.....	112
Figure 8-11- Caractérisation des concepts par des attributs.....	113
Figure 8-12- Description textuelle d'un cas réel d'accident.....	114
Figure 8-13- Patron de l'élaboration des métadonnées transformationnelles.....	115
Figure 8-14- Comparaison des deux versions de classifications (selon le modèle transformationnel) des attributs effectuées par les deux équipes d'experts.....	115
Figure 8-15- Distribution des attributs selon le point de vue transformationnel.....	115
Figure 8-16- Représentation de l'accident n°503 selon le point de vue transformationnel.....	116
Figure 8-17- Projection de l'accident n°503 sur les métadonnées transformationnelles.....	116
Figure 8-18- Patron de l'élaboration des métadonnées transformationnelles.....	117
Figure 8-19- Distribution des attributs selon le point de vue ontologique.....	117
Figure 8-20- Représentation de l'accident n°503 selon le point de vue ontologique.....	118
Figure 8-21- Patron de l'élaboration des métadonnées transformationnelles.....	118
Figure 8-22- Distribution des attributs selon le point de vue fonctionnel.....	118
Figure 8-23- Représentation de l'accident n°503 selon le point de vue fonctionnel.....	119
Figure 8-24- Patron de l'élaboration des métadonnées transformationnelles.....	120
Figure 8-25- Distribution des attributs selon le point de vue téléologique.....	120
Figure 8-26- Projection d'un accident sur les métadonnées téléologique.....	120
Figure 8-27- Diagramme d'objets :exemples d'instanciation du diagramme de classes d'ASMEC.....	121
Figure 9-1- AICEF : approche d'incorporation des connaissances dans le processus d'ECD (deuxième approche dans SAFE-Next).....	123
Figure 9-2- Lien entre le modèle de connaissances dans ASMEC et la base de données.....	124
Figure 9-3- Exemple de requête de sélection d'attributs pour l'élaboration de STA.....	124
Figure 9-4- Interface d'AICEF aidant l'utilisateur à traduire son objectif en fonction des concepts du domaine et selon différents points de vue : sélection multi-vues des attributs.....	125
Figure 9-5- Description d'une classe à l'aide des modalités.....	127
Figure 9-6- Exemple de règle pour la sélection de modalités discriminantes caractérisantes.....	128
Figure 9-7- Exemple de règle pour assurer l'homogénéité des classes.....	128
Figure 9-8- Lien entre le modèle de connaissances dans ASMEC et la description des classes.....	128
Figure 9-9- Grille systémique : projection automatique des résultats de classification sur les métadonnées.....	130
Figure 9-10- La matrice binaire décrivant la relation I du contexte $C = (O, A, I)$	132
Figure 9-11- Le treillis de Galois $L(C)$ du contexte précédent.....	132
Figure 9-12- Treillis de Galois $L(C)$ correspondant à la classification C.....	134
Figure 9-13- Exemple d'utilisation du treillis pour l'interprétation des classes au niveau des modalités.....	135
Figure 9-14- Interprétation des classes au niveau des concepts systémiques.....	137
Figure 9-15- Grille systémique : projection automatique des résultats de classification sur les métadonnées au niveau des attributs.....	138
Figure 9-16- Interprétation des classes au niveau des attributs.....	139
Figure 9-17- Exemple d'utilisation du treillis pour l'interprétation des classes au niveau des attributs.....	140
Figure 9-18- Lecture multi-niveaux d'abstraction des classes.....	141

Figure 10-1- ASAIC : utilisation des méta-données pour l'analyse d'impact de changement (troisième approche dans SAFE-Next).....	142
Figure 10-2- Processus d'analyse d'impact de changement (adaptée de [Ma et al., 2003]).....	144
Figure 10-3- Un modèle d'AIC sur le système CVE.....	146
Figure 10-4- Structure d'inférence dans l'analyseur d'impact.....	147
Figure 10-5- Extrait des résultats de l'application d'ASAIC pour l'analyse de l'impact de l'ESP	149
Figure 10-6- Un modèle d'analyse d'impact de changement sur le système de conception.....	151
Figure 10-7- Un modèle d'analyse d'impact de changement sur le métier de l'accidentologie	152
Figure 10-8- Un modèle intégré pour l'AIC en sécurité routière.....	153
Figure 11-1- ASEM : approche d'évaluation de modèles (quatrième approche dans SAFE-Next)	154
Figure 11-2- La tâche de modélisation	155
Figure 11-3- Une vue générale sur l'évaluation d'un modèle.....	157
Figure IV.1- ASMEC : approche systémique de modélisation multi-vues de connaissances.....	183
Figure IV.2- Modèle UML de l'architecture d'ASMEC : modélisation dirigée par un méta-modèle multi-vues et multi-niveaux d'abstraction.	184
Figure IV.3- AICEF : approche d'incorporation des connaissances dans le processus d'ECD.....	184
Figure IV.4- ASAIC : utilisation des méta-données pour l'analyse d'impact de changement	185
Figure IV.5- ASEM : approche d'évaluation de modèles	186
Figure IV.6- Synthèse de l'ensemble des approches dans SAFE-Next.	187
Figure IV.7- Une synthèse des contribution de SAFE-Next en KM, IC, ECD, conception et accidentologie.	189
Figure A.1.0-1- Structure d'inférence de sélection des attributs.....	224
Figure A.1.0-2- Structure d'inférence de sélection des accidents	227
Figure A.1.0-3- Structure d'inférence d'interprétation des classes.....	230
Figure A.1.0-4- Représentation graphique de la tâche d'élaboration des STA.....	235
Figure A.2.1- ASMEC : approche systémique de modélisation multi-vues de connaissances.....	236
Figure A.3.1- Exemple XML.....	237
Figure A.3.2- Exemple XSL.....	238
Figure A.3.3- Affichage du XML à l'aide du XSL.....	238
Figure A.4.1- Requête générée par l'interface AICEF présentée dans la Figure 9-4, p°4	242
Figure A.5.1- Modèle de représentation de l'approche ASMEC	243

Liste des définitions

<p>ABS (Antiblockier system): système antiblocage des freins, 3</p> <p>Accidentologie, 14</p> <p>Activité, 43</p> <p>Analyse d'Impact de Changement (AIC), 3</p> <p>Attribut catégoriel, 70</p> <p>Classement, 47</p> <p>Classification, 47</p> <p>Classification automatique, 69</p> <p>Composante active dans une classe d'accidents, 133</p> <p>Concept, 83, 132</p> <p>Conjonction systémique, 30</p> <p>Connaissance, 51</p> <p>Contexte, 131</p> <p>Contexte binaire, 132</p> <p>Correspondance de Galois, 132</p> <p>Cybernétique, 27</p> <p>ECD, 17</p> <p>ESP (Electronic Stability Program) : Programme de Contrôle de Stabilité, 3</p> <p>Evolution d'un modèle, 162</p> <p>Extraction de Connaissances de Données, 17</p> <p><i>Fonctionnement d'un modèle</i>, 161</p> <p>Forme forte, 70</p> <p>Groupement stable, 70</p> <p>Holisme, 28</p> <p>IC, 18</p> <p>Inférences, 88</p> <p>Ingénierie des connaissances, 18</p> <p>Macro-accidentologie, 14</p> <p>Méthodes d'ingénierie des connaissances, 53</p> <p>Micro-accidentologie, 15</p> <p>Modèle, 156</p> <p>Modèle conceptuel des connaissances, 54</p> <p>Modèle de la tâche, 44</p> <p>Modèle du domaine, 18</p>	<p>Modèle opérationnel des connaissances, 54</p> <p>Modèle séquentielle, 40</p> <p>Modélisation, 32</p> <p>Modélisation des connaissances, 54</p> <p>MRP: méthode de résolution du problème, 103</p> <p>Ontologie d'un modèle, 159</p> <p>Problématique transversale en accidentologie, 139</p> <p>Processus d'Extraction de Connaissances de Données, 17</p> <p>Rôles de connaissance, 88</p> <p>SBC, 53</p> <p>Scénario-Type d'Accidents, 45</p> <p>Sécurité active, 37</p> <p>Sécurité passive, 37</p> <p>Sécurité primaire, 36</p> <p>Sécurité secondaire, 36</p> <p>Sécurité tertiaire, 37</p> <p>Sous-concept/Sur-concept, 132</p> <p>STA, 45</p> <p>Structure d'inférence, 89</p> <p>Système à Base de Connaissances, 53</p> <p>Système compliqué, 28</p> <p>Système cybernétique, 28</p> <p>Systèmes complexes, 28</p> <p>Tâche, 43</p> <p>Téléologie d'un modèle, 163</p> <p>Treillis de Galois, 132</p>
---	---

Liste des abréviations

- ABS** : Antiblockier system : système antiblocage des freins.
- ACACIA** : Acquisition des Connaissances pour l'Assistance à la Conception par Interaction entre Agents.
- AIC**: Analyse d'Impact de Changement.
- AICEF** : Approche d'Incorporation des Connaissances Expertes dans la Fouille de données.
- ASAIC** : Approche Systémique d'Analyse d'Impact de Changement.
- ASEM** : Approche Systémique d'Evaluation de Modèles.
- ASMEC** : Approche Systémique de ModElisation des Connaissances.
- BD** : Bases de Données.
- CEESAR** : Centre Européen d'Etude de Sécurité et Analyse des Risques.
- CML2** : Conceptual Modeling Langage.
- CVE** : Conducteur-Véhicule-Environnement.
- ECD** : Extraction de Connaissances à partir de Données d'accidents.
- EDA** : Etudes Détaillées d'Accidents.
- ESP** : Electronic Stability Program : Système de Contrôle de Stabilité.
- IAIC** : Ingénierie d'Analyse d'Impact de Changement.
- INRETS** : Institut National de Recherche sur les Transports et leur Sécurité.
- KM** : Knowledge Management.
- K-means**: algorithme de classification.
- KOD** : Knowledge Oriented Design.
- LAB** : Laboratoire d'Accidentologie, de Biomécanique et du Comportement Humain.
- MC** : Modèle conceptuel des connaissances.
- MIKE** : Model-based and Incremental Knowledge Engineering.
- MKSM** : Methodology for Knowledge System Management.
- MRP** : Méthode de résolution du problème.
- QFD** : Query Functional Deployment.
- SAFE-Next** : Systemic Approach For Enhanced kNowledge EXTraction.
- SBC** : Système à Base de Connaissances.
- SE** : Systèmes Experts.
- SPAD** : logiciel d'analyse de données.
- SS** : Systèmes de Sécurité.
- SSE** : Systèmes de Sécurité Embarqués.
- STA** : Scénario Type d'Accidents.

Résumé

Aujourd'hui, l'informatisation des saisies de données et la puissance des systèmes de collecte conduisent à la construction de grandes Bases de Données (BD). L'exploitation de ces millions de données en accidentologie et dans beaucoup d'autres domaines (e.g. management, marketing, etc.) fait appel à des techniques d'Extraction de Connaissances de Données (ECD). La complexité des données, du domaine d'application et des connaissances recherchées rendent fondamentale l'intégration des connaissances expertes dans le processus d'ECD. Cela nécessite la mise en place d'approches, méthodes et techniques d'identification, de représentation et d'opérationnalisation de ces connaissances.

Dans ce travail de thèse, nous proposons une nouvelle approche, appelée *SAFE-Next* (Systemic Approach For Enhanced kNowledge EXTraction, approche systémique pour l'extraction des connaissances) qui intègre les quatre approches suivantes : La première est appelée *ASMEC* (Approche Systémique de ModElisation des Connaissances). Elle consiste en une méthode de modélisation des connaissances multi-vues et selon une architecture à plusieurs niveaux d'abstraction. La deuxième approche, *AICEF* (Approche d'Incorporation des Connaissances Expertes dans la Fouille de données), propose l'élaboration et l'utilisation de *méta-données multi-vues* comme un moyen pour l'incorporation des connaissances formalisées par *ASMEC* dans le processus d'ECD. La troisième approche, *ASAIC* (Approche Systémique d'Analyse d'Impact de Changement), utilise le modèle de connaissances d'*ASMEC* pour une analyse interactive et multi-vues de l'impact d'un changement sur un système. La quatrième approche, *ASEM* (Approche Systémique d'Evaluation de Modèles), fournit un *modèle général d'évaluation de modèles de connaissances*.

Les fondements épistémologiques et méthodologiques de nos travaux sont respectivement le *constructivisme* et la *systémique* (ou *cybernétique*). En se basant sur ces fondements, nos travaux de recherche ont conduit à des contributions réparties en quatre domaines : En accidentologie, *SAFE-Next* fournit un outil efficace pour l'élaboration des STA permettant une meilleure analyse et compréhension de l'accident. Elle fournit aussi un moyen de capitalisation des connaissances offrant une vision synthétique des différents types de connaissances du domaine de l'accidentologie. En Ingénierie des Connaissances (IC), *SAFE-Next* propose un modèle général multi-vues et multi-niveaux d'abstraction de modélisation des connaissances pour le développement des Systèmes à Base de Connaissances (SBC). Elle permet aussi de guider l'élicitation des connaissances selon un modèle multi-vues. En ECD, *SAFE-Next* propose l'utilisation des métadonnées multi-vues pour l'incorporation des connaissances expertes du domaine dans la première et la dernière phase du processus d'ECD (i.e. préparation des données et interprétation des résultats). En conception de nouveaux systèmes, *SAFE-Next* fournit à travers les STA un moyen de communication entre les accidentologues et les concepteurs des systèmes de sécurité embarqués dans les véhicules. Cette interface entre les deux métiers (i.e. conception et accidentologie) permet la construction de l'espace de conception pour développer et évaluer les systèmes de sécurité. Elle offre aussi un moyen d'analyse de l'impact d'un changement (e.g. introduction d'un nouveau système de sécurité) sur le comportement du système Conducteur-Véhicule-Environnement.

Mots clefs : Ingénierie des Connaissances, Extraction de Connaissances de Données, Modélisation des connaissances, Modélisation des systèmes complexes, Cybernétique, Systémique, Analyse de changement, Conception de nouveaux systèmes, Evaluation de modèles

Abstract

Nowadays, given the automation of data collection, very large databases are constructed. The exploitation of these data in accidentology and several others fields (e.g. marketing, engineering, etc.) requires automatic techniques of Knowledge Discovery in Databases (KDD). Incorporating expert knowledge in the KDD process is fundamental to handle the complexity of data, domain and knowledge. This necessitates the development of approaches, methods and techniques intended to identify, represent and operationalize expert knowledge.

In this dissertation, we propose a new approach, *SAFE-Next* (Systemic Approach For Enhanced kNnowledge EXtraction), which integrates the following four approaches: the first one, *ASMEC* (Approche Systémique de ModElisation des Connaissances), allows knowledge modeling according to multiple viewpoints and granularity levels. The second approach, *AICEF* (Approche d'Incorporation des Connaissances Expertes dans la Fouille de données), uses the *ASMEC* knowledge model to elaborate multi-view *metadata*. It then uses these metadata as a tool for incorporating expert knowledge into the KDD process. The third approach, *ASAIC* (Approche Systémique d'Analyse d'Impact de Changement), uses the *ASMEC* knowledge model to carry out multi-view change impact analysis. The fourth approach, *ASEM* (Approche Systémique d'Evaluation de Modèles), provides an assessment framework for knowledge models.

The epistemological and methodological foundations of our work are constructivism and systemic approach (or cybernetics). Based on these backgrounds, our research contributions concern several disciplines, ranging from Accidentology, Knowledge Engineering, Knowledge Discovery in Databases and Design. In accidentology, *SAFE-Next* provides experts with an efficient tool for knowledge management. It enables the elaboration of multi-view accident scenarios, which are a powerful tool for understanding accident mechanisms in order to develop safety counter-measures. Furthermore, *SAFE-Next* provides a knowledge capitalization tool. In knowledge engineering, *SAFE-Next* supplies, via *ASMEC*, a multi-view knowledge model and thereby allows the integration of different viewpoints stemming from different users. Furthermore, it provides a multi-granularity knowledge model and in that way addresses the difficulty of knowledge identification and formalization. At the same time, *SAFE-Next* permits, via *ASEM*, the evaluation of knowledge models, an issue rarely addressed in literature. In Knowledge Discovery in Database, *SAFE-Next* enables, via *AICEF*, the incorporation of domain knowledge in the data preprocessing step (i.e. the first step in a KDD process) and more specifically in the attribute selection task. Likewise, the multi-view metadata enables the incorporation of domain knowledge in the interpretation step (i.e. the last step in a KDD process). In design, *SAFE-Next* provides safety system developers with an efficient tool to construct the design space. Scenarios enable them to understand complex behaviors and thereby to define solutions and alternatives. *SAFE-Next* also provides, via *ASAIC*, an approach for multi-view change impact analysis. Moreover, it proposes an extension of the change impact analysis to the use process of a given product as well as the evaluation process instead of limiting it to the design process.

Keywords: Knowledge Engineering, Knowledge Discovery in Databases, Knowledge Modeling, Complex System Modeling, Cybernetics, Systemic Approach, Change Impact Analysis, New System Design, Model Assessment.