



HAL
open science

Approche générique d'aide à la décision pour l'amélioration de la sécurité des transports

Ahmed Maalel

► **To cite this version:**

Ahmed Maalel. Approche générique d'aide à la décision pour l'amélioration de la sécurité des transports. Intelligence artificielle [cs.AI]. Ecole Nationale des Sciences de l'Informatique (ENSI), La Manouba (Tunisie), 2015. Français. NNT: . tel-02343916

HAL Id: tel-02343916

<https://hal.science/tel-02343916>

Submitted on 3 Nov 2019

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.

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
UNIVERSITÉ DE MANOUBA



THÈSE

Présentée en vue de l'obtention du diplôme de

DOCTEUR EN INFORMATIQUE

**Approche générique d'aide à la décision pour
l'amélioration de la sécurité des transports**

Par :

Ahmed MAALEL

Réalisé au sein de



Soutenue le 30 Avril 2015, devant le jury composé de :

Président	Khaled GHEDIRA	Professeur à l'ISG, Tunis
Rapporteurs	Chihab HANNACHI Sadok BOUAMAMA	Professeur à l'université Toulouse 1 Capitole Maître de conférences à l'ENSI, Tunis
Examinatrice	Narjès BELLAMINE	Maître de conférences à l'ISI, Tunis
Directrice de thèse	Henda BEN GHÉZALA	Professeure à l'ENSI, Tunis

Résumé

Ces dernières années, de plus en plus d'entreprises et de secteurs industriels prennent conscience de l'importance accrue de la concurrence imposée par les avancées technologiques constatées dans tous les domaines et l'évolution continue des nouvelles technologies de l'information et de la communication. Devant cette situation et pour maintenir leurs compétitivités, ces organisations ne peuvent s'échapper à l'adoption de ces technologies de plus en plus performantes en dépit du respect des contraintes et des exigences de la sécurité. Ceci, a donné naissance à de nombreux risques et a favorisé l'apparition de situations contradictoires avec le respect de la sécurité. Dans certaines situations, les échecs sont graves et coûteux, ils sont dus à des accidents causant des pertes humaines, surtout quand il s'agit d'un domaine à haut risque comme celui du transport.

L'objectif de ce travail de thèse est de proposer une nouvelle approche basée sur l'intelligence artificielle pour prédire les accidents et aider à la prise de décision minimisant les risques. Dans cette approche d'Aide à la Décision (AD), un couplage entre le Raisonnement à Partir de Cas (RàPC) et les ontologies a été opéré pour permettre, en tenant compte de certains critères, d'assister un décideur (experts certifieurs, experts analystes, les constructeurs, etc.) à bien comprendre une situation risquée donnée et de lui proposer des solutions possibles. L'approche mise au point dans ce travail, baptisée ADAST (Aide à la Décision pour l'Amélioration de la Sécurité des Transports) est une approche générique d'aide à la décision pour l'amélioration de la sécurité et a été appliquée dans le domaine du transport ferroviaire. L'intérêt d'une telle approche réside non seulement au niveau de la capitalisation des connaissances vécues (issues du retour d'expérience), mais aussi, dans le fait de pouvoir en tirer profit pour dégager un savoir-faire susceptible d'aider et d'assister les experts du domaine dans leur tâche d'analyse et d'évaluation de la sécurité.

Mots clés : Aide à la décision, raisonnement à partir de cas, ontologies, analyse de la sécurité.

Abstract:

In the recent years, more and more businesses and industries have become aware of the increasing importance of competition imposed by the technological advances found in all areas and the continuous development of new information and communication technologies. Due to this situation and in order to maintain their competitiveness, these organizations have to adopt these more efficient technologies without any regard to constraints and security requirements. Consequently, numerous risks have arisen and facilitated the emergence of conflicting situations regarding safety. In some situations, failures are serious and costly due to accidents causing life loss, especially when it is a high-risk area such as transport field.

The objective of this thesis is to propose a new approach based on artificial intelligence to predict accidents and help in taking the right decision to minimize risks. In this Decision-Aid (DA) approach, a coupling between the Case-Based Reasoning (CBR) and ontologies has been made to allow, taking into account certain criteria, assisting a decision maker (analyst, user, etc..) to understand a given risky situation and to propose to him possible solutions. The approach developed in this work, called ADAST (Aide à la Décision pour l'Amélioration de la Sécurité des Transports = decision aid to improve the transport safety) is a generic approach to decision aid for improving security analysis, which has been applied in the railroad transport. The significance of such an approach lies not only in the capitalization of knowledge (from experience feedback) but also in taking advantage of the know-how capable of stimulating and helping the domain experts in their task analysis and safety assessment.

Keywords: Decision aids, Case-Based Reasoning, Ontologies, Security analysis.

"La science ne connaît qu'une loi : la contribution scientifique", Bertolt Brecht

"Ce n'est pas en tournant le dos aux choses qu'on leur fait face", Pierre Dac

"Le difficile n'est pas de monter, mais, en montant, de rester soi", Jules Michelet

TABLE DES MATIERES

Introduction générale.....	1	
 Partie I : État de l'art		
Chapitre 1: Sûreté et sécurité des systèmes de transport automatisés	8	
Chapitre 2: Concepts de base : Représentation de connaissances et raisonnement.....	24	
Chapitre 3: Approches d'aide à la décision se basant sur le RàPC et les ontologies	61	
 Partie II : Approche proposée		
Chapitre 4: ADAST : une approche d'aide à la décision basée sur le RàPC.....	77	
Chapitre 5: Maquette de faisabilité d'ADAST	116	
Chapitre 6: Expérimentation et validation.....	140	
 Conclusion générale.....		165
Bibliographie.....	170	
Publications	185	

TABLE DES MATIERES

Introduction générale.....	1
A. Contexte	1
B. Problématique.....	3
a. Problématique applicative	3
b. Problématique théorique	3
C. Approche proposée.....	4
D. Organisation du rapport.....	5

Partie I : État de l'art

Chapitre 1: Sûreté et sécurité des systèmes de transport automatisés.....	8
1.1. Introduction.....	8
1.2. Sûreté et sécurité des systèmes industriels	9
1.2.1. Définition et concepts de base de la sûreté de fonctionnement	9
1.2.2. Sécurité.....	11
1.2.2.1. Principes	11
1.2.2.2. Techniques.....	12
1.2.2.3. Méthodes	12
1.2.2.4. Matrice Gravité/ Occurrence	12
1.2.3. Notion d'un scénario d'accident	13
1.2.3.1. Description générale d'un scénario d'accident	14
1.2.3.2. Processus d'aboutissement d'un scénario accident	14
1.4. Le retour d'expérience	16
1.4.1. Travaux existants en matière du Rex	17
1.4.2. Cadre législatif et réglementaire du Rex ferroviaire.....	18
1.4.3. Déroulement d'une enquête après un accident ferroviaire.....	19
1.4.4. Construction de la connaissance à partir de Rex.....	19
1.5. Carences et limites liées aux travaux d'analyse de la sécurité et de Rex	22
1.6. Nos choix méthodologiques.....	22
1.7. Conclusion	23

Chapitre 2 : Concepts de base : Représentation de connaissances et raisonnement.....	24
2.1. Introduction.....	24
2.2. Les ontologies	25
2.2.1. Présentation de la notion d'ontologie.....	25
2.2.2. Rôles de l'ontologie	25
2.2.3. Classification des ontologies.....	26
2.2.3.1. Classification de Psyché	26

2.2.3.1.1. Typologie selon l'objet de conceptualisation.....	27
2.2.3.1.2. Typologie selon le niveau de détail de l'ontologie	27
2.2.3.1.3. Typologie selon le niveau du formalisme	28
2.2.3.1.4. Typologie selon le niveau de complétude.....	28
2.2.3.2. Classification d'Oberle	29
2.2.4. Composantes d'une ontologie.....	29
2.2.5. Méthodologies de construction des ontologies	30
2.2.5.1. Méthodologie d'Uschold et Gruninger.....	30
2.2.5.2. Méthodologie Methontology	31
2.2.5.3. Méthodologie différentielle	32
2.2.5.4. Méthodologie <i>OntoSpec</i>	32
2.2.6. Modes de représentation de connaissances	33
2.2.6.1. Les réseaux sémantiques	33
2.2.6.2. Les langages de frames.....	34
2.2.6.3. Les logiques de description	34
2.2.6.4. Les graphes conceptuels	34
2.2.6.5. Les standards web pour la représentation des ontologies.....	35
2.2.7. Fondements pour la conception d'une ontologie	36
2.2.8. La similarité entre concepts dans une ontologie	37
2.2.8.1. Les méthodes de mesure de similarité.....	37
2.2.8.2. Méthodes structurelles de mesure de similarité au sein des ontologies	39
2.3. Le raisonnement à partir de cas (RàPC)	40
2.3.1. Origines et présentation du RàPC	40
2.3.2. Définition et structuration d'un cas.....	41
2.3.3. Modèles de RàPC.....	41
2.3.4. Tâches de RàPC	42
2.3.5. Cycle de RàPC	43
2.3.5.1. Élaboration.....	43
2.3.5.2. Remémoration	44
2.3.5.3. Adaptation	45
2.3.5.3.1. Approches d'adaptation	46
2.3.5.3.2. Approches d'acquisition des connaissances d'adaptation	49
2.3.5.4. Révision.....	50
2.3.5.5. Apprentissage	50
2.4. Fouille de données et extraction de connaissances	51
2.4.1. Les règles d'association	51
2.4.2. La méthode d'arbre de décision	53
2.4.3. Méthode d'Analyse des Concepts Formels.....	55
2.5. Synthèse et choix méthodologiques.....	57
2.5.1. Côté ontologique	57

2.5.2. Côté RàPC	58
2.5.3. Méthodes et techniques de fouille de données	58
2.5.3.1. Règles d'association	59
2.5.3.2. Règles de classification	59
2.5.3.3. Règles d'adaptation	59
2.6. Conclusion	60

Chapitre 3 : Approches d'aide à la décision se basant sur le RàPC et les ontologies 61

3.1. Introduction.....	61
3.2. Approches proposées dans la littérature	61
3.2.1. Travaux d'Aamodt	61
3.2.2. Travaux de Diaz-Agudo et Gonzalez-Calero	62
3.2.3. Travaux de Fush et Mille	62
3.2.4. Travaux de Recio Garcia et ses collègues.....	63
3.2.5. Travaux de Gómez-Gauchía et ses co-auteurs	64
3.2.6. Travaux de Chen	65
3.2.7. Travaux de Hadj-M'tir	66
3.2.8. Autre travaux.....	66
3.3. Approches proposées liées à la sécurité des transports et du retour d'expérience.....	67
3.3.1. Transport ferroviaire	67
3.3.2. Transport aérien	68
3.3.3. Transport maritime.....	68
3.3.4. Transport routier.....	69
3.3.5. Sécurité industrielle et analyse des risques	71
3.4. Bilan et synthèse	72
3.4.1. Comparaison des approches.....	72
3.4.2. Synthèse	74
3.5. Conclusion	74

Partie II : Approche proposée

Chapitre 4 : ADAST : une approche d'aide à la décision basée sur le RàPC..... 77

4.1. Introduction.....	77
4.2. Motivations	77
4.2.1. Objectifs attendus.....	78
4.2.2. Choix adoptés.....	78
4.3. Présentation d'ADAST	79
4.3.1. Architecture fonctionnelle d'ADAST	79
4.3.2. Processus de résolution de problème	80

4.4. Structuration et modélisation de connaissances.....	81
4.4.1. Modèle conceptuel	82
4.4.2. Modèle de domaine /de cas.....	85
4.5. Développement de l'ontologie.....	86
4.6. Raisonnement à partir de cas	91
4.6.1. Élaboration d'un cas cible.....	91
4.6.2. Phase de remémoration	93
4.6.2.1. Classification du cas cible	94
4.6.2.2. Extraction des cas similaires.....	95
4.6.2.2.1. Mesures de similarités.....	96
4.6.2.2.2. Exemple illustratif.....	100
4.6.3. Phase d'adaptation	102
4.6.3.1. Démarche d'adaptation.....	102
4.6.3.2. Acquisition des connaissances d'adaptation.....	103
4.6.3.2.1. Construction du contexte formel.....	104
4.6.3.2.2. Génération du treillis du concept	105
4.6.3.2.3. Génération des conditions d'adaptation.....	107
4.6.3.3. Phase de révision	109
4.6.3.4. Phase d'apprentissage.....	110
4.6.3.4.1. Démarche d'apprentissage proposée.....	110
4.6.3.4.2. Algorithme d'auto-incrémentation.....	111
4.6.3.4.3. Exemple illustratif.....	113
4.7. Conclusion	114
Chapitre 5 : Maquette de faisabilité d'ADAST	116
5.1. Introduction.....	116
5.2. Architecture générale de la maquette d'aide à la décision.....	116
5.3. Outils logiciels utilisés.....	116
5.3.1. Architecture en couches	118
5.3.2. La structure de la base de connaissances	119
5.3.2.1. Base ontologique	119
5.3.2.2. Base de règles	120
5.3.2.3. Base de cas.....	121
5.3.3. Architecture fonctionnelle de la maquette basée sur ADAST	124
5.4. Modules	125
5.4.1. Module d'acquisition de connaissances	126
5.4.2. Module d'aide à la décision	130
5.4.2.1. Préconception	130
5.4.2.2. Processus de raisonnement	131
5.5. Conclusion	139

Chapitre 6 : Expérimentation et validation	140
6.1. Introduction.....	140
6.2. Cas étudié.....	140
6.2.1. Échantillon d'apprentissage	140
6.2.2. Approche d'expérimentation.....	141
6.3. Construction et acquisition de la base de connaissances	142
6.3.1. Acquisition des règles d'association	142
6.3.2. Acquisition des règles de classification	144
6.3.3. Acquisition de connaissances d'adaptation.....	147
6.4. Scénario d'aide à la décision	151
6.4.1. Description du scénario d'accident injecté	151
6.4.2. Cycle de raisonnement	152
6.5. Évaluation du système	155
6.5.1. Étude comparative entre l'outil CLASCA et le module de classification d'ADAST	156
6.5.1.1. Interprétation des résultats de classification de l'outil CLASCA.....	156
6.5.1.2. Interprétation des résultats du module de classification d'ADAST	157
6.5.2. Interprétation des résultats de la démarche d'aide à la décision	159
6.6. Limites de la démarche d'aide à la décision	162
6.6.1. Problèmes de la qualité et de la quantité des cas d'accidents	162
6.6.2. La diversité des techniques et des méthodes appliquées.....	162
6.6.3. La validation globale.....	163
6.6.4. Complexité des algorithmes proposés.....	163
6.7. Conclusion	163
 Conclusion générale.....	 165
A. Contributions	166
B. Perspectives	167
a. Enrichir et consolider la base historique de scénarios.....	167
b. Intégrer et optimiser les algorithmes utilisés.....	168
c. Apprendre par le succès et l'échec pour améliorer la démarche d'aide à la décision	168
d. Prise en compte de la modélisation dynamique des cas d'accidents.....	168
e. Ontology Web Language- Services (OWL-S)	169
 Bibliographies	 170
Publications	185

TABLE DES FIGURES

Figure 1.1. Arbre détaillé de la sûreté de fonctionnement (Laprie J.C. et al, 1995).....	10
Figure 1.2. Interdépendance des paramètres de la Sdf (Cauffriez L. et al., 2004).....	11
Figure 1.3. Processus séquentiel d'un aboutissement d'accident (Desroches A. et al., 2003).	15
Figure 1.4. Cycle en V. de développement d'un système.....	20
Figure 1.5. Création de la connaissance à partir de l'expérience.....	21
Figure 1.6. Articulation des différentes étapes de processus Rex.....	21
Figure 2.1. Principaux rôles de l'ontologie (Kolli M, 2011).	26
Figure 2.2. Typologies d'ontologies selon quatre dimensions de classification (Psyché V. et al., 2003). 27	27
Figure 2.3. Typologies d'ontologies selon l'objet de conceptualisation (Psyché V. et al., 2003).....	27
Figure 2.4. Typologie selon le niveau de détail de l'ontologie (Psyché et al. 2003).....	28
Figure 2.5. Typologie selon le niveau du formalisme (Psyché et al. 2003).....	28
Figure 2.6. Typologie selon le niveau de complétude (Bachimont B., 2000).	29
Figure 2.7. Dimensions de classification d'ontologie (Oberle, 2006).	29
Figure 2.8. Composants d'une ontologie.	30
Figure 2.9. Méthodologie de construction de l'ontologie (Uschold et Gruninger, 1996).	31
Figure 2.10. Méthodologie Methontology (Fernandez M. et al, 1997).	32
Figure 2.11. Phases de développement d'ontologie (Kassel et Perpette 1999).	33
Figure 2.12. Exemple d'un réseau sémantique (Miller, 1990).	34
Figure 2.13. Exemple d'un graphe conceptuel (Bahloul, 2006).....	35
Figure 2.14. La pyramide des langages basés Web (Boussaid I., 2005).....	35
Figure 2.15. Similarité basée sur les arcs.....	40
Figure 2.16. Approches de RàPC (Kolodner, 1992).....	41
Figure 2.17. Modèles de RàPC (Bergmann R. et al., 2003), (Lamontagne L., 2004).	42
Figure 2.18. Classification des systèmes de RàPC (Althoff K.D. et al., 1995).	43
Figure 2.19. Types d'adaptation en RàPC (Wilke W. et Bergmann, R. 1998).....	46
Figure 2.20. Concept formel de l'attribut $Attr_6$	56
Figure 3.1 Les types des connaissances dans Creek (Aamodt A., 2004).....	62
Figure 3.2. Système de RàPC utilisant CBROnto (Diaz-Agudo B. et González-Calero P.A., 2000).	62
Figure 3.3. Modèles conceptuels (Fuchs B. et Mille A., 2005).	63
Figure 3.4. Les différents usages de l'ontologie selon (Recio-García J.A. et al, 2006).	64
Figure 3.5. Connecteurs de jColibri (Recio-García J.A. et al, 2006) (Diaz-Agudo B. et al, 2007).....	64
Figure 3.6. Les différents modèles de Gómez-Gauchía (Gómez-Gauchía et al, 2006).....	65
Figure 3.7. Partie de la mémoire de l'agent MemoPA (Chen K. et Barthès J.B., 2007).	65
Figure 3.8. Représentation d'un cas d'incident (Yanping C. et al, 2005).....	68
Figure 3.9. Architecture SAGEO (Arnaud V. et Napoli A., 2013).....	69

Figure 3.10. Architecture ROSAC (Capus L. et Tourigny N., 2000).	70
Figure 3.11. Architecture du système proposé (Capus L. et Tourigny N., 2000).	70
Figure 3.12. Architecture COBRA (Abou-Assali A., 2010).	72
Figure 4.1. Architecture fonctionnelle d'ADAST et ses interactions avec le module d'acquisition des connaissances	79
Figure 4.2. Articulations des différentes phases de cycle de RàPC retenu.	80
Figure 4.3. Un échantillon d'instances « Cause liée à l'environnement ».	82
Figure 4.4. Un échantillon d'instances pour les erreurs relatives à l'intervention des passagers (Hadj-Mabrouk A. and Hadj-Mabrouk H., 2003).	82
Figure 4.5. Exemple de classification d'un accident potentiel de type collision dans le transport ferroviaire (Hadj-Mabrouk A. and Hadj-Mabrouk H., 2003).	82
Figure 4.6. Les principales classes de l'ontologie noyau.	84
Figure 4.7. Diagramme de classe proposé.	84
Figure 4.8. Aperçu de la prise en compte de l'erreur humaine dans le diagramme de classe proposé.	85
Figure 4.9. Aperçu du modèle de cas.	86
Figure 4.10. Cycle de développement d'ontologie (Abou-Assali A. et al., 2008).	87
Figure 4.11. Représentation d'une relation d'agrégation.	89
Figure 4.12. Hiérarchie des classes et des sous-classes sous Protégé.	90
Figure 4.13. Une partie de l'ontologie de domaine avec le plugin OWL-Viz (Maalel A. et al, 2011).	91
Figure 4.14. Élaboration d'un cas cible.	92
Figure 4.15. Approche de remémoration retenue.	94
Figure 4.16. Démarche de génération et de représentation des règles de classification.	94
Figure 4.17. Aperçu des mesures de profondeurs des attributs de l'ontologie.	102
Figure 4.18. Approche de réutilisation adoptée.	103
Figure 4.19. Approche d'acquisition des connaissances d'adaptation retenue.	104
Figure 4.20. Représentation du contexte uni-valué dans l'outil Concept Explorer.	106
Figure 4.21. Génération de treillis de concept avec l'outil Concept Explorer.	106
Figure 4.22. Identification de l'extension et l'intension.	108
Figure 4.23. Déroulement de la phase de révision.	110
Figure 4.24. Déroulement de la phase d'apprentissage (Première hypothèse).	111
Figure 4.25. Déroulement de la phase d'apprentissage (Deuxième hypothèse).	111
Figure 5.1. Articulation entre les différents outils utilisés et la maquette ADAST.	117
Figure 5.2. Architecture en couches de la maquette d'aide à la décision.	118
Figure 5.3. Mapping entre les anciens et les nouveaux paramètres descriptifs.	123
Figure 5.4. Aperçu de la représentation du cas source « Cas_01 »	123
Figure 5.5. Architecture fonctionnelle de la maquette d'aide à la décision.	124
Figure 5.6. Écran d'authentification.	125
Figure 5.7. Menu principal de la maquette ADAST.	126
Figure 5.8. Explorateur de l'ontologie.	127

Figure 5.9. Ouverture de l'éditeur de cas.....	128
Figure 5.10. Écran d'acquisition de cas sources.....	128
Figure 5.11. Les éléments nécessaires pour acquérir un nouveau cas source.....	129
Figure 5.12. Ajout d'une instance à la description d'un cas source.	129
Figure 5.13. Écran de consultation de cas sources.....	130
Figure 5.14. Écran de préconception.	131
Figure 5.15. Écran dédié à l'élaboration du cas cible (Première méthode).	132
Figure 5.16. Écran dédié à l'élaboration du cas cible (Deuxième méthode).	133
Figure 5.17. Acquisition d'un nouveau cas cible.....	134
Figure 5.18. Écran dédié à la phase de remémoration.	134
Figure 5.19. Écran dédié à la phase de remémoration.	135
Figure 5.20. Écran dédié à la phase d'adaptation.	135
Figure 5.21. Écran dédié à la phase de révision.....	136
Figure 5.22. Message de validation de la solution proposée	136
Figure 5.23. Écran dédié à la phase d'apprentissage.	137
Figure 5.24. Message de confirmation de l'apprentissage automatique.....	137
Figure 5.25. Suite de messages relative à l'opération d'apprentissage.....	138
Figure 5.26. Aperçu du fichier journal d'ADAST.....	138
Figure 6.1. Représentation de l'échantillon d'apprentissage en format CSV.....	141
Figure 6.2. Évolution du nombre de règles générées en fonction de la confiance.	142
Figure 6.3. Évolution du nombre de règles générées en fonction du lift.	142
Figure 6.4. Évolution du nombre de règles générées en fonction du paramètre support.....	143
Figure 6.5. Schéma d'évaluation des modèles de classification.....	144
Figure 6.6. Courbe ROC du classifieur généré.....	146
Figure 6.7. Écran visualisation de l'arbre de décision généré via l'outil SIPINA.....	146
Figure 6.8. Partie de l'arbre de décision générée.....	147
Figure 6.9. Le treillis de concepts de la classe #CL1 Commutation de redondance.	149
Figure 6.10. Le treillis de concepts de la classe #CL8 Suivi de l'ordre.	149
Figure 6.11. Le treillis de concepts de la classe #CL10 Contrôle/Commande des aiguilles.	149
Figure 6.12. Le treillis de concepts de la classe #CL2 Séquence d'initialisation.....	150
Figure 6.13. Le treillis de concepts de la classe #CL11 Contrôle de vitesse.....	150
Figure 6.14. Synoptique de l'accident : Perte d'élément - pénétration sur un canton occupé (Dossier technique CR/94-16).....	151
Figure 6.15. Règles d'associations générées.....	153
Figure 6.16. Résultat de la classification.	153
Figure 6.17. Résultat de la remémoration.....	153
Figure 6.18. Le carré d'analogie pour le problème d'adaptation.....	154
Figure 6.19. Résultat de l'adaptation.	154
Figure 6.20. Résultat de l'apprentissage.....	155

Figure 6.21. Histogramme de la répartition des cas (apprentissage et test).....	156
Figure 6.22. Interprétation des résultats de classification obtenus par CLASCA.	157
Figure 6.23. Interprétation des résultats de classification obtenus par le système.	159
Figure 6.24. Interprétation des résultats obtenus par le système.	161

Liste des tableaux

Tableau 1.1. Tableau de fréquence des situations dangereuses selon la norme [EN 50126].....	12
Tableau 1.2. Catégories de gravité des situations dangereuses selon la norme [EN 50126]	13
Tableau 1.3. Niveau de risque selon la norme [EN 50126]	13
Tableau 1.4. Matrice gravité/occurrence selon la norme[EN 50126]	13
Tableau 2.1. Les techniques de mesures de similarité	38
Tableau 2.2. Comparaison des approches de calcul de similarité au sein des ontologies.....	39
Tableau 2.3. Énumération des approches d’adaptation.....	47
Tableau 2.4. Comparaison des approches d’adaptation	48
Tableau 2.5. Comparaison des approches d’acquisition de connaissances d’adaptation.....	49
Tableau 2.6. Algorithmes de génération de règles d’association.....	52
Tableau 2.7. Algorithmes de génération d’arbre de décision.....	54
Tableau 2.8. Exemple d’un contexte formel	56
Tableau 3.1. Travaux proposés pour la sécurité des transports et la sécurité industrielle.....	67
Tableau 3.2. Comparaison des approches étudiées	73
Tableau 4.1. Les différentes transformations du langage UML vers OWL (Hadj-M’tir R, 2010)	88
Tableau 4.2. Échantillon d’apprentissage Ω_a	95
Tableau 4.3. Description du cas cible.....	100
Tableau 4.4. Description du cas source_test	100
Tableau 4.5. Contexte formel multi-valué.....	105
Tableau 4.6. Contexte formel uni-valué.....	105
Tableau 4.7. Conditions d’adaptations trouvées	109
Tableau 4.8. Description du cas cible (adapté et révisé).....	114
Tableau 4.9. Description du cas source.....	114
Tableau 5.1. Distribution des cas selon la classe d’appartenance (Hadj-Mabrouk H, 1994).....	122
Tableau 6.1. Comparaison des ensembles de règles d’associations générées.....	143
Tableau 6.2. Comparaison des performances des algorithmes testés.....	145
Tableau 6.3. Répartition des solutions dans les classes de scénarios.....	148
Tableau 6.4. Description initiale du cas cible introduit.....	152
Tableau 6.5. Interprétation des résultats de classification des cas cible.....	158
Tableau 6.6. Interprétation des résultats de la démarche d’aide à la décision	160

Introduction générale

A. Contexte

Ces dernières années, de plus en plus d'entreprises et de secteurs industriels prennent conscience de l'importance accrue de la concurrence imposée par les avancées technologiques constatées dans tous les domaines et l'évolution continue des nouvelles technologies de l'information et de la communication. Devant cette situation et pour maintenir leurs compétitivités, ces organisations ne peuvent éviter l'adoption de ces technologies de plus en plus performantes en dépit du respect des contraintes et des exigences de la sécurité. Ceci, a donné naissance à de nombreux risques et a favorisé l'apparition de situations contradictoires avec le respect de la sécurité notamment dans le domaine des transports. Dans certaines situations, les échecs sont graves et coûteux, ils sont dus à des accidents causant des pertes humaines, surtout quand il s'agit d'un domaine à haut risque comme celui du transport ferroviaire.

En effet, le secteur du transport présente une importance cruciale dans notre vie. La sécurité ou encore l'analyse de risques dans les transports est une étape assez conséquente, complexe et interdisciplinaire par nature. Elle repose en grande partie sur une analyse profonde des risques résultant des scénarios d'accident. Un scénario d'accident est une succession adéquate d'événements pouvant être à l'origine d'un risque potentiel. En dépit de tous les efforts déployés pour proposer des méthodes et des techniques permettant de gérer, conduire et maintenir un niveau de risque acceptable, plusieurs insuffisances peuvent être décelées. D'une part, la plupart des méthodes souffrent de l'absence de définitions admises par tous et des ambiguïtés dues au manque, voire l'absence, de formalisme pour exprimer ou modéliser les connaissances relatives aux scénarios d'accidents. D'autre part, les méthodes actuelles d'analyse des risques souffrent de l'insuffisance au niveau d'exploitation de connaissances, du moment que les méthodes existantes n'abordent pas les spécificités liées à de tels systèmes complexes exigeant le recours à différents experts.

Pour apporter des éléments de réponse à ces limites, plusieurs méthodes, voire même des méthodologies ont été proposées dans la littérature, malgré leur intérêt indéniable, elles présentent quelques insuffisances. D'un côté, il n'y a pas suffisamment de travaux qui ont été consacrés à l'analyse des systèmes des transports ferroviaires automatisés. D'un autre côté, le déploiement des outils issus de méthodes d'analyse des risques dans des domaines autres que le domaine pour lesquels ils ont été conçus s'avère difficile, voire impossible.

Ce travail de recherche entre dans le cadre d'une collaboration entre le laboratoire RIADI et l'Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR). Ce contexte professionnel d'études et de recherche implique plusieurs acteurs (experts certifieurs, experts analystes, les représentants de collectivités locales, les constructeurs, les exploitants, etc.) qui concourent ensemble pour l'homologation (cas d'un système de transport jugé homologue à un modèle ancien) ou pour la certification (Cas d'un système de transport typiquement nouveau). L'homologation comme la certification sur le plan de sécurité, peuvent être apparentées à un processus de prise de décision assez complexe :

- D'une part, pour tous les intervenants dans ce processus d'étude du dossier de sécurité et surtout pour les experts de l'analyse de sécurité et aussi pour les constructeurs qui sont les premiers utilisateurs auxquels cette aide est adressée ;
- D'autre part, au niveau de chaque phase cruciale de l'analyse de sécurité.

L'objectif de ce travail de thèse est alors de proposer une nouvelle approche basée sur l'intelligence artificielle, pour aider à la prise de décision la plus informée. Cette prise de décision se manifeste à deux niveaux successifs :

- Le premier concerne l'octroi d'un agrément provisoire au système de transport automatisé sur le plan matériel et logiciel ;
- Le second concerne l'octroi d'un agrément définitif qui est seul permettant la mise sur site définitif du système de transport par l'exploitant (maître de l'ouvrage).

L'approche d'aide à la décision (AD) envisagée devrait permettre, à plusieurs échelles, de fournir un tableau de bord, sous forme d'indicateurs favorisant la prise de décision (homologation et/ou certification). En effet, ce processus d'aide à la décision est sollicité à plusieurs niveaux complémentaires et hiérarchiques du processus de construction de la sécurité: l'analyse préliminaire de risques (APR), l'analyse de la sécurité fonctionnelle (ASF), l'analyse des effets des erreurs du logiciel (AEEL), l'analyse des modes de défaillance, de leurs effets et de leur criticité (AMDEC), la méthode des combinaisons des pannes résumées (MCPR) et notamment la méthode de l'arbre des causes (MAC). Notre travail s'inscrit dans le cadre de l'ASF. L'une des difficultés de cette méthode de sécurité consiste à s'assurer de l'exhaustivité des risques et scénarios d'accidents non pris en compte lors de l'élaboration du dossier de sécurité par le constructeur. En effet, il convient d'examiner l'ASF avec le plus grand soin, tant la qualité de celle-ci a un impact sur la sécurité des usagers des systèmes de transport.

Notre approche d'AD, vise alors en tenant compte de certains critères, d'assister un décideur (expert, analyste, utilisateur etc.) à bien comprendre une situation risquée donnée. Cette approche préventive permettrait à mieux comprendre chaque situation à risque et de lui proposer les solutions possibles sous forme de mesures préventives et/ou correctives susceptibles d'éviter sa reproduction dans le futur. L'approche mise au point dans ce travail, baptisée ADAST¹, est une approche générique d'aide à la décision pour l'amélioration de la sécurité dans le domaine du transport. L'intérêt de proposer une telle approche réside non seulement au niveau d'acquiescer l'image la plus fidèle de la réalité (les accidents issus du terrain réel), mais aussi, dans le fait de pouvoir la capitaliser et d'en tirer profit pour dégager un savoir-faire susceptible d'aider et d'assister les experts du domaine dans leur tâche d'analyse et d'évaluation de la sécurité.

Dans notre approche, nous avons fait le choix de coupler le Raisonnement à Partir de Cas (RàPC) et les ontologies pour proposer une aide à la décision. Ceci est justifié par le fait que l'étude que nous avons menée sur la nature des connaissances disponibles, montre que ces dernières sont essentiellement, des scénarios d'accidents historiques, ce qui nous a fait pencher pour une approche fondée sur la réutilisation. Dans ce cadre, le RàPC, un paradigme de raisonnement par analogie dans l'apprentissage automatique, offrant des méthodes, des techniques et aussi des outils pour retrouver des solutions à de nouveaux problèmes résolus, s'adapte bien pour être appliqué comme un raisonnement. D'autre part, l'analyse plus profonde des scénarios d'accidents acquis a révélé l'existence de plusieurs divergences et aussi des incohérences dans la formalisation de ces scénarios. Ces divergences,

¹ADAST : Aide à la Décision pour l'Amélioration de la Sécurité des Transports

incomplétudes et incohérences sont surtout engendrées par la présence de plusieurs experts à plusieurs niveaux d'analyse de la sécurité. Le choix de l'utilisation des ontologies se justifie par le fait qu'elles permettent de proposer un mode de représentation de connaissances, fournissant l'élaboration d'un vocabulaire homogène et commun, applicable en particulier dans le contexte de l'ingénierie de connaissances et de l'AD. Ceci permet alors de réduire considérablement ces incohérences constatées en mettant en évidence la corrélation des faits.

ADAST propose ainsi de tirer profit des ontologies pour appréhender les limites et les insuffisances liées à la représentation des connaissances ainsi que du RàPC pour remédier au problème de capitalisation et d'exploitation de ces connaissances.

B. Problématique

Notre contexte d'étude fait apparaître plusieurs problématiques d'ordre applicatif et théorique :

a. Problématique applicative

Les mesures préventives et correctives à mettre en œuvre pour maintenir un niveau de risque acceptable, et par conséquent, pour éviter la reproduction de telles situations contraires à la sécurité, peuvent être liées à différents éléments. Il est difficile de connaître tous ces éléments, et donc de bien analyser les symptômes et les causes d'un cas d'insécurité.

Plusieurs défis sont ainsi énumérés :

- Manager des connaissances pour bien bâtir le mode de représentation adéquat pour les situations d'insécurité (les accidents) ;
- Rendre accessibles et exploitables ces connaissances dans le but de fournir une aide à la décision adéquate ;
- Assister et outiller les experts et les spécialistes du domaine dans leur tâche d'analyse et d'évaluation de la sécurité.

b. Problématique théorique

Plusieurs aspects principaux du RàPC ont constitué le cœur de notre réflexion dans ce travail de thèse, la représentation des cas au sein de la base de cas historique et notamment son processus de raisonnement :

- Nous avons constaté que les cas fournis par les experts ne partagent pas toujours les mêmes descripteurs et l'hétérogénéité des cas crée un certain nombre de problèmes. Comment alors homogénéiser et unifier la description de ces cas ?
- Quelle solution pouvons-nous donner pour compléter les données manquantes ou erronées lors de l'élaboration d'un cas cible ?
- Comment procéder pour calculer la similarité entre les cas déjà vécus présents dans la base de cas historique ?
- L'adaptation est l'étape la plus importante mais aussi la plus difficile dans un système de RàPC, parce qu'elle nécessite des connaissances spécifiques au domaine qui ne sont pas toujours faciles à acquérir. De ce fait, la question à laquelle nous souhaitons répondre est : Comment acquérir ces connaissances d'adaptation dans un système de RàPC ?

- Enfin, la convergence de la qualité des connaissances pérennisées dépend de la phase en aval de raisonnement, l'apprentissage. Comment conserver une bonne structure de la base de cas et assurer ainsi un apprentissage auto-incrémental ?

C. Approche proposée

Notre cadre théorique découle de notre problématique. Il est centré sur le RàPC et plus particulièrement sur les approches d'AD proposées se basant sur le RàPC. Nous avons pu à travers notre approche non seulement améliorer les méthodes usuelles d'analyse de la sécurité et de retour d'expérience, mais aussi, d'optimiser et de renforcer la complémentarité entre toutes les phases de raisonnement. Notre approche s'organise en trois niveaux complémentaires permettant d'aborder progressivement la résolution du problème d'acquisition, de capitalisation et d'exploitation des connaissances issues de terrain réel tout en conservant une forte interactivité avec l'expert du domaine (le spécialiste en matière de sécurité).

L'ensemble des contributions abouties peut être résumé dans les points suivants :

- **L'amélioration de la phase de l'élaboration du cas cible (nouveau cas d'accident) :** La représentation de ce dernier se fait moyennant l'ontologie de domaine. Des règles d'association sont utilisées également pour assister l'expert pour compléter et raffiner davantage la description initiale dans le cas où cette dernière est bruitée ou insuffisante ;
- **L'amélioration de la phase de remémoration :** Cette phase consiste d'abord, à repérer à partir de la base de cas (des cas anciens d'accidents déjà validés), la classe adéquate à laquelle l'accident courant est susceptible d'appartenir. Ceci, permet à l'expert de focaliser sur la nature et les connaissances impliquées dans cette nouvelle situation d'accident. Ceci est un deuxième degré d'assistance à la décision. Une méthode d'apprentissage supervisé nommée « Arbre de Décision » sera alors appliquée pour classer la nouvelle situation d'accident dans une famille connue d'accidents. Ensuite, une mesure de similarité sera donc opérée entre la nouvelle situation d'accident avec celles existantes dans la base. Ceci est très avantageux car il limite la procédure d'analyse experte aux seuls accidents d'une famille donnée ;
- **La proposition d'une technique d'acquisition de connaissances d'adaptation d'une solution historique pour résoudre le cas cible :** La situation d'accident historique la plus similaire au nouveau cas d'accident devrait être réadaptée à travers des transformations pour enfin proposer un remède à la nouvelle situation à risque. Ici, nous avons proposé une technique inspirée de l'ACF (Analyse des Concepts Formels) ;
- **L'amélioration de la phase d'apprentissage :** Ceci est fait à travers la proposition d'un algorithme d'apprentissage auto-incrémental. Cet algorithme permet de mémoriser dans la base de cas uniquement les cas cibles résolus ayant une description assez détaillée comparée à celle d'une ancienne situation d'accident historique très similaire ;
- **La proposition d'une maquette de faisabilité d'ADAST :** cette maquette est utilisée pour pérenniser les connaissances en matière d'accident afin de tirer profit en proposant pour les experts de domaine en matière de sécurité l'aide à la décision adéquate comme la classe de l'accident, les cas historiques les plus similaires au cas d'accident en question et notamment les mesures préventives et/ou correctives permettant d'éviter sa reproduction dans le futur.

D. Organisation du rapport

Le manuscrit de thèse proposé est composé d'une introduction générale, de six chapitres structurés en deux parties (Partie 1 : État de l'art ; Partie 2 : L'approche proposée), d'une conclusion générale et de trois annexes.

Suite à une introduction générale décrivant le contexte, la problématique ainsi que l'approche proposée, la première partie de l'état de l'art comporte trois chapitres. **Le premier chapitre**, présente les principaux concepts liés à la problématique et le contexte général de la recherche en particulier, la sûreté et la sécurité des systèmes de transport automatisés.

Le deuxième chapitre vise à décrire dans un premier temps, les concepts de base que nous avons adoptés dans notre approche. Il aborde dans sa première partie, les ontologies, ses aspects importants et plus particulièrement ceux que nous abordons dans notre travail de thèse. La deuxième partie est consacrée au RàPC, à travers l'étude de ses origines, ses aspects importants à savoir : la structuration des cas, les modèles de RàPC, les tâches de RàPC et notamment les différentes phases de son cycle de raisonnement comme la remémoration, l'adaptation et l'apprentissage. La troisième partie de l'étude est focalisée sur la fouille de données et les méthodes d'extraction de connaissances à partir de données, plus précisément ceux adoptés dans le cadre de notre approche à savoir, les règles d'association, la méthode d'arbre de décision et la technique d'Analyse des Concepts Formels (ACF). Cette étude théorique va déboucher enfin sur une synthèse et un bilan des choix méthodologiques et techniques effectués.

Dans **le troisième chapitre**, nous étudions tout d'abord les différentes approches de RàPC qui se sont basées d'une façon ou d'une autre sur les ontologies. Dans la deuxième partie de ce chapitre, nous explorons les travaux liés à notre domaine d'application, la sécurité des transports. Notre étude va déboucher au final sur une synthèse de ces travaux pour montrer certaines limites et carences dans le but de bien situer notre approche et notamment les éventuelles contributions attendues de toute la démarche d'aide à la décision envisagée.

Dans **le quatrième chapitre**, nous procédons à la présentation de notre approche ADAST (Aide à la Décision pour Améliorer la Sécurité des Transports). Nous présentons tout d'abord, les motivations à travers les objectifs attendus et nous rappelons nos choix méthodologiques adoptés. Nous présentons ensuite l'architecture fonctionnelle d'ADAST et notamment les modèles de connaissances élaborés à savoir ; le modèle conceptuel, le modèle de domaine et le modèle de cas. Nous détaillons, les articulations entre toutes les phases de raisonnement et la démarche d'aide à la décision retenue et nous positionnons également le choix et l'utilisation de chaque méthode/technique pour bien expliquer la nature et l'utilité de leur alliance par rapport à notre contexte d'étude.

Le cinquième chapitre est consacré à la présentation de la maquette de faisabilité d'ADAST. Nous présentons d'abord les différents outils logiciels utilisés, l'architecture en couche à travers ses différents composants en particulier la base ontologique et la base de connaissances, puis l'architecture fonctionnelle. Nous présentons enfin, les deux modules d'acquisition de connaissances et de raisonnement.

Enfin, **le sixième chapitre**, est prévu pour les expérimentations et la validation. Les expérimentations que nous avons réalisées avaient ainsi pour but d'évaluer tous les éléments qui constituent notre approche ADAST. L'approche d'expérimentation est scindée en trois volets : tout d'abord, nous évaluons les différentes techniques utilisées issues de fouille de données pour générer les règles de classification (la méthode d'arbre de décision), les règles

d'association et notamment notre approche d'acquisition de connaissances d'adaptation en ayant recours à quelques sessions d'acquisition. En second lieu, nous présentons une première validation globale de l'approche à travers un exemple de fonctionnement du système illustrant les différentes étapes de raisonnement. Enfin, nous effectuons l'évaluation globale de toute la démarche d'aide à la décision en ayant recours à un ensemble de cas de tests (une partie de l'échantillon d'apprentissage). Les limites de notre démarche d'AD sont présentées à la fin de ce chapitre.

La conclusion de ce travail et les perspectives de recherche sont présentées à la fin du manuscrit de thèse.

Partie I

État de l'art

- Chapitre 1 : Sûreté et sécurité des systèmes de transport automatisés8
- Chapitre 2 : Représentation de connaissances et raisonnement24
- Chapitre 3 : Approches d'aide à la décision se basant sur le RàPC et les ontologies.....62

Chapitre 1 : Sûreté et sécurité des systèmes de transport automatisés

1.1. Introduction

Les conséquences fatales et le coût terrible des accidents, la survenue des échecs graves et des nouvelles catastrophes malgré le progrès de la technologie, sont à la base de la mise en place d'un processus de retour d'expérience (souvent appelé Rex), comme étant l'une des méthodes essentielles à promouvoir l'amélioration nécessaire de la sécurité dans les domaines à haut risque. Le transport ferroviaire (notre cadre d'application dans le cadre de notre travail de thèse) n'échappe pas à cette règle.

Au Canada, selon le site de BST ¹ (Bureau de la Sécurité des Transports), en 2012, 1011 accidents ferroviaires ont été signalés. Ce nombre est proche au total enregistré en 2011 (1022) et en baisse de 10 % par rapport à la moyenne de 1128 durant la période de 2007 à 2011. En France, les investissements poussés dans la modernisation du réseau ferroviaire ont fait affaiblir de 40% le nombre d'accidents ferroviaires en 10 ans, selon un rapport de RFF ² (Réseau Ferré de France) publié récemment. Sur 5 millions de trains ayant circulé en 2012, nous relevons 136 «accidents significatifs» selon la même source. De manière générale, la sécurité ferroviaire n'a pas cessé de s'améliorer ces dix dernières années et les statistiques d'accidents montrent que le pourcentage d'accidents ne cesse de diminuer.

En dépit de tous ces progrès, nous sommes encore loin d'empêcher des catastrophes qui peuvent survenir. Les accidents prennent parfois des proportions spectaculaires et entraînent de véritables catastrophes susceptibles de frapper l'imagination. Rien qu'au mois de Juillet 2013 seulement, trois accidents catastrophiques sont survenus. L'accident de 6 juillet 2013 qui a eu lieu au Lac-Mégantic (Québec, Canada) décrit par l'explosion d'un train de pétrole ravage le centre de la ville, faisant 47 morts. Le 12 juillet 2013 à Brétigny-sur-Orge, France, un *Corail Intercités* reliant la gare de Paris Austerlitz à Limoges Bénédictins déraile en gare, l'accident fait 7 morts. En 24 juillet 2013 à Saint-Jacques-de-Compostelle, en Espagne, un train à grande vitesse Alvia effectuant la liaison Madrid-Ferrol qui venait d'emprunter la ligne à grande vitesse de Galice s'engage à 190 km/h dans une courbe limitée à 80 km/h et déraile violemment, l'accident fait au moins 79 morts. Récemment, le 30 décembre 2013, un train transportant du pétrole et un autre convoi ferroviaire se sont percutés, dans le Dakota du Nord, un État du Midwest des États-Unis. La collision n'a pas fait de victimes mais a déclenché une série d'explosions spectaculaires et a répandu une épaisse fumée noire sur Casselton, une bourgade enneigée à l'ouest de Fargo. Les autorités ont dû ordonner l'évacuation de cette petite ville de 2 300 habitants.

Dès son origine, le transport ferroviaire présente un fort potentiel de risques, souvent découverts à travers les accidents lorsqu'ils se réalisaient. Afin de les conjurer, toutes les autorités et tous les intervenants sont efforcés au fil du temps de tirer les leçons provenant de Rex à travers les accidents ferroviaires vécus. C'est pourquoi le Rex et plus globalement,

¹ <http://www.bst-tsb.gc.ca/>

² <http://www.rff.fr/fr>

la sécurité est devenue pour de nombreuses entreprises industrielles notamment pour le domaine du transport ferroviaire, une préoccupation majeure et un élément indispensable non seulement à la réussite de leurs projets mais également à leur développement, voire même à leur survie.

Pour assurer un niveau de risque acceptable, la définition d'un système efficace d'aide à l'analyse de la sécurité reste toujours un défi à relever. Pour aboutir à ces fins, l'utilisation des techniques de représentation de connaissances, d'apprentissage automatique et notamment des approches informatiques d'aide à la décision semble être très intéressante.

Ce chapitre présente donc notre contexte général de la recherche. Il est consacré à l'étude de la sécurité et de Rex. Il aborde dans sa première partie, après une introduction des concepts de base de la sûreté de fonctionnement, les soucis liés la sécurité, ses principes, ses méthodes et ses principales techniques. Nous étudions par la suite quelques travaux en matière d'aide à l'analyse de la sécurité. Ensuite, nous introduisons le concept de Rex, les travaux existants en matière de Rex, son cadre législatif et réglementaire et notamment le déroulement d'une enquête technique après un accident. Nous détaillons aussi les principes, les modalités de déroulement de Rex et l'acquisition de connaissances. Finalement, nous présentons nos orientations, nos choix méthodologiques, les principales motivations et nous débouchons sur le choix convenable des approches et des techniques informatiques mises en œuvre dans notre travail de thèse.

1.2. Sûreté et sécurité des systèmes industriels

Selon (Joffrey C., 2009), les premiers travaux sur la sûreté de fonctionnement (SdF) sont apparus dans les années 1950 dans des domaines à haut risque comme le nucléaire. Selon (Tittel L., 1992) dans (Joffrey C., 2009), les études statistiques sur les fréquences de pannes et accidents qui étaient alors menées, avaient pour but de renforcer la sécurité des systèmes par l'amélioration de la fiabilité des matériels jugés critiques où en ayant recours à la redondance des équipements.

Les années 1960 ont été marquées par l'apparition des nouvelles méthodes de sûreté telles que l'Analyse des Modes de Défaillances et de leurs Effets (AMDE), la Méthode du Diagramme de Succès (MDS), la méthode de l'Arbre de Défaillances (AdD) et la Méthode des Combinaisons de Pannes Résumées (MCPR). La prise en compte de l'évolution de l'informatique dans le domaine industriel et les facteurs humains a marqué ensuite les années 80 et 90, à travers l'apparition d'outils tels que les chaînes de Markov, les réseaux de Petri. Selon (Joffrey C., 2009), ces nouvelles techniques accentuent les efforts menés pour converger vers des systèmes plus sûrs.

1.2.1. Définition et concepts de base de la sûreté de fonctionnement

La SdF est une préoccupation permanente dans la plupart des secteurs industriels. Elle est d'autant plus essentielle dans le secteur des transports terrestres automatisés en raison de la gravité des accidents et de leur impact dans le public.

La littérature propose plusieurs définitions de la SdF. Laprie (Laprie J.C., 1995) par exemple, la définit par : « *la propriété d'un système permettant à ses utilisateurs de placer une confiance justifiée dans le service que délivre le système* ». Selon Krishnakumar et ses co-auteurs (Krishnakumar R. et al., 2005), l'objectif de la sûreté est d'apprendre et de maîtriser les risques de dysfonctionnement des systèmes complexes, notamment leur fiabilité en mettant en œuvre des méthodes prévisionnelles appropriées. Boulanger (Boulanger J.L., 2006) la définit à son tour comme suit : « *La sûreté de fonctionnement d'un système (SdF) est définie comme la qualité du service délivré par un système, qualité telle que les utilisateurs de ce service puissent placer une confiance justifiée dans le système* ».

qui le délivre ». Laprie et ses co-auteurs (Laprie J.C. et al., 1995) présentent la sûreté de fonctionnement comme un arbre dont les feuilles sont les différents éléments qui la créent (Joffrey C., 2009). Cet arbre est illustré par la figure 1.1.

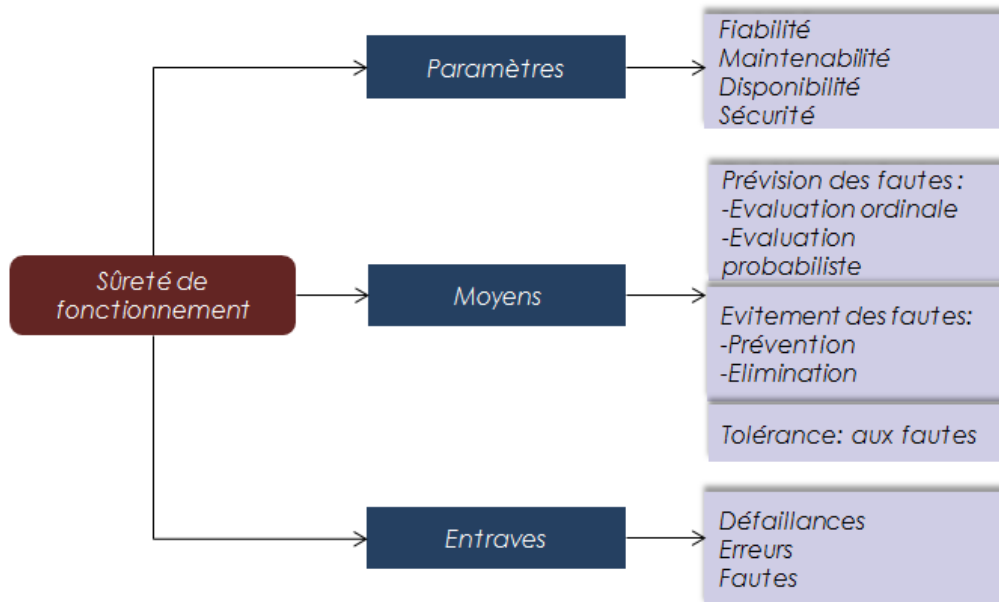


Figure 1.1. Arbre détaillé de la sûreté de fonctionnement (Laprie J.C. et al, 1995).

D'une manière générale, les paramètres de la sûreté de fonctionnement d'un système à risque s'articulent autour de quatre principales composantes, à savoir : la fiabilité, la disponibilité, la maintenabilité et la sécurité :

- **La fiabilité (Reliability)**, notée $R(t)$, selon (Joffrey C., 2009), est l'aptitude d'un système à rester régulièrement opérationnel sur l'intervalle $[0, t]$:

$$R(T) = P\{\text{Système non défaillant sur } [0, t]\}$$

- **La disponibilité (Availability)**, notée $A(t)$, selon (Joffrey C., 2009), est l'aptitude d'un système à être opérationnel au moment où il est souhaité. Selon (Chevalier M et al, 2004), une disponibilité élevée est compatible avec une fiabilité faible, pour peu que l'état de l'appareil puisse être restauré très rapidement :

$$A(T) = P\{\text{Système non défaillant à l'instant } t\}$$

- **La maintenabilité (Maintainability)**, notée $M(t)$, c'est l'aptitude d'un système à être restauré rapidement dans un état opérationnel. Ainsi selon (Chevalier M et al, 2004), les systèmes dont les composants sont aisément démontables peuvent bénéficier d'une meilleure maintenabilité que les autres :

$$M(T) = P\{\text{Système réparé au temps } t\}$$

- **La sécurité (Safety)**, notée $S(t)$, c'est l'aptitude d'un système à ne pas éprouver de pannes importantes ou critiques pendant une durée donnée :

$$S(T) = P\{\text{Système sans défaillance catastrophique sur } [0, t]\}$$

Selon (Cauffriez L. et al., 2004), ces quatre paramètres présentés précédemment sont fortement corrélés (Figure 1.2) :

Diminuer la fiabilité du système affaiblit la disponibilité et a un impact direct sur le niveau de sécurité du système. Selon (Cauffriez et al., 2004), une défaillance peut conduire

le système dans un état risqué. Une maintenabilité non parfaite peut nuire à la disponibilité du système et à sa sécurité. Selon les arguments de Hollnagel (Hollnagel E., 1999) et Cabeau (Cabeau E., 1999) ; « Augmenter le niveau de sécurité du système (par l'ajout de multiples éléments de sécurité) réduit sa disponibilité (le système stoppe intempestivement dès la première défaillance) tandis que l'augmentation du niveau de disponibilité est obtenue au détriment du niveau de sécurité ».

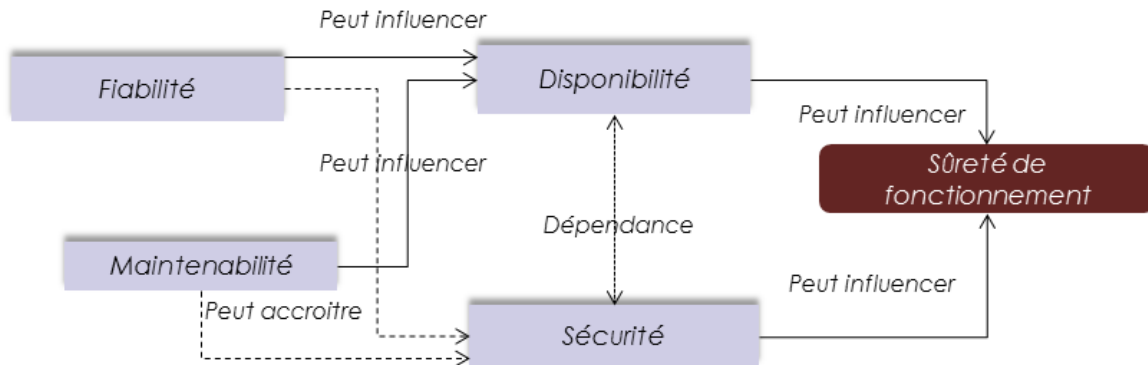


Figure 1.2. Interdépendance des paramètres de la Sdf (Cauffriez L. et al., 2004).

Face à l'évolution technologique, tous les organismes et les structures de recherche dans ce domaine, se sont rendu compte de l'obligation de l'intégration de nouveaux concepts comme la qualité, les facteurs humains et l'ergonomie. Cependant, la sécurité demeure la composante cruciale de tout système industriel à risque. La sécurité mérite une attention particulière et doit être étudiée en amont de tout projet de développement d'un système industriel à risque. Nous consacrons le paragraphe suivant à l'étude de la sécurité.

1.2.2. Sécurité

Selon le petit Larousse, sécurité se dit « d'une situation où nous n'avons aucun danger à craindre ». Cette définition d'ordre général selon (Gayon A., 2001), se base sur le principe du risque zéro, elle ne peut pas être adaptée à des activités humaines et/ou industrielles, d'où le concept de risque acceptable.

Selon (Gayon A., 2001), se référant spécialement au domaine industriel, la sécurité peut être définie comme la possibilité d'un système à fonctionner en maintenant un niveau du risque acceptable pour les personnes, les biens et l'environnement. La sécurité industrielle, au sens général, consiste à assurer la sécurité des biens, des personnes et également la continuation de l'entreprise en réduisant les risques liés aux côtés : environnemental, social et économique. Selon la norme CENELEC 50 129¹, la sécurité est « l'absence de tout niveau de risque inacceptable ».

Hadj-Mabrouk (Hadj-Mabrouk H., 2010) préconise une séparation entre principe, technique et méthode de sécurité.

1.2.2.1. Principes

En Europe, il existe principalement trois principes de sécurité. L'Allemagne applique le principe MEM (*Minimum Endogenous Mortality*) qui prétend que nous ne poursuivons à améliorer le niveau de sécurité si seulement si le taux de mortalité dû aux faits technologiques (exogène à l'organisme) est inférieur au taux de mortalité endogène (dans un lieu et un espace de temps déterminé). Au Royaume-Uni, l'aspect est plutôt économique.

¹ NORME CENELEC EN 50129 (2000) : Application ferroviaire – Systèmes de signalisation, de télécommunications et de traitement à Systèmes électroniques de sécurité pour la signalisation.

En effet, le principe ALARP (*As Low As Reasonably Practicable*) appliqué exige selon (Hadj-Mabrouk H., 2007), « *qu'un niveau de risque soit approuvable si le coût dû à la réduction du risque n'est disproportionné vis-à-vis du gain en termes de l'amélioration de la sécurité* ». En France, GAME (*Globalement Au moins Équivalent*) est le principe retenu par le comité européen de sécurité. Ce principe exprime que le niveau de sécurité d'un nouveau système doit être au moins semblable à celui d'un système comparable déjà existant et réputé sûr.

1.2.2.2. Techniques

Généralement, il y'a trois techniques de sécurité : *intrinsèque, probabiliste et contrôlée*. En effet, certains systèmes ne peuvent tolérer aucun niveau de défaillances. Afin de répondre à ce besoin, la sécurité intrinsèque vise à atteindre un niveau de sécurité élevé. La sécurité probabiliste admet un niveau de risque admissible et définit un seuil préétabli. Pour la technique contrôlée, les défaillances sont détectées et leurs conséquences évitées par des équipements (Hadj-Mabrouk H., 2007).

1.2.2.3. Méthodes

Les méthodes de sécurité peuvent être classées en trois niveaux : *système, logiciel et matériel*. Côté logiciel, il y'a la méthode d'Analyse des Effets des Erreurs du Logiciel (AEEL) ainsi que sur les lectures critiques de code. Côté matériel, nous recensons plusieurs méthodes : l'Analyse des Modes de Défaillance, de leurs Effets et de leur Criticité (AMDEC), la Méthode des Combinaisons de Pannes Résumées (MCPR) et la Méthode de l'Arbre des Causes (MAC). Au niveau système, l'Analyse Préliminaire des Risques (APR). Elle vise à identifier les accidents potentiels liés au système et à ses interfaces afin d'évaluer leur probabilité d'occurrence ainsi que la gravité des dommages engendrée (Hadj-Mabrouk H., 2007). Et enfin, l'analyse de la sécurité fonctionnelle (ASF) qui permet à son tour de s'assurer de l'exhaustivité des risques et scénarios d'accidents non pris en compte lors de l'élaboration du dossier de sécurité par le constructeur.

1.2.2.4. Matrice Gravité/ Occurrence

La norme EN 50126¹ propose des probabilités quantitatives. Les deux tableaux 1.1 et 1.2 extraits de la norme EN 50 126 :2000, décrivent les critères qualitatifs relatifs à la probabilité d'occurrence et la gravité de dommages utilisés pour définir la sécurité.

<i>Catégorie</i>	<i>Description</i>
Fréquent (A)	Fréquent (A) : survient probablement souvent ($P > 10^{-3}$)
Probable (B)	Probable (B) : survient plusieurs fois ($10^{-3} > P > 10^{-4}$)
Occasionnel (C)	Occasionnel (C) : survient probablement plusieurs fois ($10^{-4} > P > 10^{-5}$)
Rare (D)	Rare (D) : survient probablement ($10^{-5} > P > 10^{-7}$)
Hautement improbable (H)	Extrêmement improbable ($10^{-9} > P$)

Tableau 1.1. Tableau de fréquence des situations dangereuses selon la norme [EN 50126]

La norme EN 50126 propose quatre niveaux de gravité des dommages concernant les personnes :

<i>Niveau de gravité</i>	<i>Conséquence pour les personnes ou l'environnement</i>
Catastrophique	Des blessés graves ou plusieurs morts, des dommages majeurs pour l'environnement.

¹ Norme CENELEC EN 50 126: 2000 Application ferroviaire – Spécification et démonstration de la fiabilité, de la disponibilité, de la maintenabilité et de la sécurité (FDMS).

Grave	Un blessé grave ou un mort
Signifiant	Un blessé léger
Insignifiant	Ni blessé, ni mort.

Tableau 1.2. Catégories de gravité des situations dangereuses selon la norme [EN 50126]

Le niveau de risque identifie le croisement du niveau de probabilité d'occurrence de l'accident potentiel avec le niveau de gravité des dommages engendrés par cet accident potentiel. Plusieurs classifications de niveau de risque sont présentes. La norme EN 50126 identifie 4 niveaux de risque :

<i>Niveau de risque</i>	<i>Description</i>
Risques intolérables	Doivent être exclus
Risques non souhaitables	Ne peuvent pas être acceptés que si le risque est impossible à réduire
Risques tolérables	Acceptables à travers des mesures appropriées
Risques négligeables	Acceptables

Tableau 1.3. Niveau de risque selon la norme [EN 50126]

Une matrice gravité/occurrence est identifiée également par la norme EN 50126 :

		<i>Niveau de gravité des dommages</i>			
		Catastrophique	Grave	Signifiant	Insignifiant
<i>Niveau de probabilité</i>	Fréquent				
	Probable				
	Occasionnel				
	Rare				
	Improbable				
	Hautement improbable				

	Risque intolérable
	Risque non souhaitable
	Risque tolérable
	Risque négligeable

Tableau 1.4. Matrice gravité/occurrence selon la norme [EN 50126]

La norme ISO 310001 : 2009 abandonne cette définition du risque et propose une nouvelle définition : « *le risque est l'effet de l'incertitude sur les objectifs* ». Cette norme propose donc une nouvelle approche Globale du Management des risques. Cette norme internationale ISO a été élaborée en Novembre 2009 puis adoptée comme norme Française en 2010 par l'AFNOR² sous le nom NF ISO 31000.

1.2.3. Notion d'un scénario d'accident

¹ ISO 31000:2009 - Management du risque (http://www.iso.org/iso/fr/catalogue_detail?csnumber=43170)

² www.afnor.org/

Selon (INERIS, 2004), « un scénario d'accident est une combinaison logique et chronologique d'un ensemble d'événements de défaillances d'éléments (équipements, procédures, actions humaines ou environnement) aboutissant à un événement redouté ou dangereux (l'accident) ». Nous présentons la description des paramètres généraux d'un scénario d'accident dans la section suivante.

1.2.3.1. Description générale d'un scénario d'accident

Pour décrire un scénario d'accident, il faut disposer à priori d'un ensemble de termes et de concepts. Plusieurs termes ont fait l'objet d'une normalisation comme par exemple :

- **Damage** : La norme CEN/CENELEC 50129 précise que le dommage comprend la mort, la blessure physique, l'atteinte à la santé ;
- **Accident** : Selon (Chopard G. et Hadj-Mabrouk H., 1996), l'accident est toute succession d'événements provoquant un ou plusieurs dommages. ;
- **Quasi-accident** : Selon la norme *NF F 00-101*, un quasi-accident est un événement ou succession d'événements ne provoquant aucun dommage (Chopard G. et Hadj-Mabrouk H., 1996) ;
- **Accident potentiel** : L'accident potentiel généralement défini par un accident ou un quasi-accident ;
- **Danger** : La norme *MIL-STD-882C*¹ spécifie que le danger est « *a condition that is a prerequisite to a mishap* ». Comme préconisé dans la norme EN 50126, le danger peut conduire soit à un accident ou à un quasi-accident, c'est-à-dire que le danger est une situation antérieure à un accident potentiel ;
- **Événement dangereux** : C'est un événement susceptible d'engendrer un danger. Selon (Desroches A. et al., 2003), « *quel que soit le domaine, la perte humaine ou matérielle peut être vue comme la conséquence d'un accident décrit par un séquençement d'évènement « présence d'un danger », « situation dangereuse» et « accident»* ».

1.2.3.2. Processus d'aboutissement d'un scénario accident

Le maillon initiateur d'un scénario d'accident est le danger, en présence d'autre d'événements, nous pouvons parler d'une situation dangereuse, le deuxième maillon d'un scénario d'accident, il correspond à un état fragile mais décidé pour aboutir à un accident.

Selon (Desroches A. et al., 2003), notre conduite personnelle obéit bien souvent à des motivations comme le stress, la colère, l'excès de confiance, orgueil... etc., qui peuvent nous faire nier le danger réel d'une situation

¹ MIL-STD-882, Standard practice for system safety, United States Military Standard.

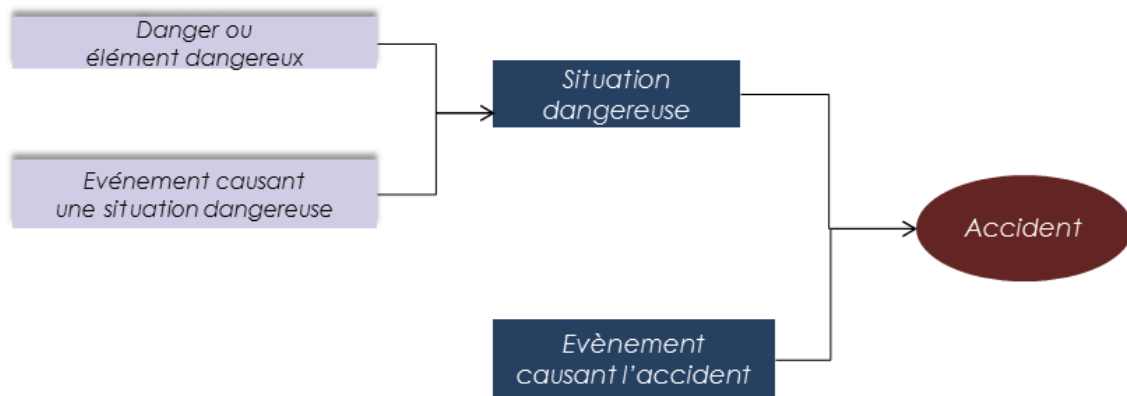


Figure 1.3. Processus séquentiel d'un aboutissement d'accident (Desroches A. et al., 2003).

1.3. Travaux d'aide à l'analyse de sécurité

Plusieurs travaux ont été proposés dans la littérature en matière d'aide à l'analyse de la sécurité. Certains travaux ont été soutenus à travers des développements et une mise en œuvre d'outils logiciels. Nous proposons dans le tableau 1.5 quelques systèmes concrétisés.

Référence	Projet	Présentation
(Hadj-Mabrouk H., 1992) (Hadj-Mabrouk H., 2007) (Mejri L., 1995) (Mejri L., 2009)	ACASYA	Un système d'Aide à la Certification par Apprentissage des SYstèmes de transport Automatisés basé sur les techniques d'acquisition et d'exploitation de connaissances.
(Hadj-Mabrouk H. et Darricau, 1996) (Hadj-Mabrouk H., 2007)	SAUTREL	Outil d'aide à l'analyse des effets des erreurs de logiciel basé sur le raisonnement à partir de cas.
(Mazouni M.H., 2008)	SIGAR	Outil logiciel d'analyse des risques qui se base sur trois niveaux : physique, conceptuel et opérationnel.
(Maalel A. et al., 2010)	REXCAS	Outil logiciel qui permet d'une part de pérenniser les connaissances issues de retour d'expérience ferroviaire (des accidents) et d'autre part les exploiter en appliquant le raisonnement à partir de cas.

Tableau 1.5. Travaux d'aide à l'analyse de la sécurité

Le système ACASYA d'Aide à la Certification par Apprentissage des SYstèmes de transport Automatisés (Hadj-Mabrouk H., 1992), (Mejri L., 1995), (Hadj-Mabrouk H., 2007) et (Mejri L., 2009) repose sur l'utilisation conjointe des techniques d'acquisition de connaissances et d'apprentissage automatique en vue d'assister les experts du domaine de la sécurité des systèmes de transports ferroviaires. Le système ACASYA est un système destiné à coopérer avec un expert pour l'assister dans sa décision, il possède deux caractéristiques ; la prise en compte de l'incrémentalité nécessaire à une évolution progressive des connaissances apprises et la coopération Homme/Machine.

ACASYA comporte 4 modules : le premier module de formalisation concerne l'acquisition et la représentation d'un scénario, il est du ressort de la phase d'acquisition de connaissances. Les trois autres modules, CLASCA, EVALSCA et GENESCA concernent respectivement les problèmes de classification, d'évaluation et d'aide à la génération d'autres scénarios potentiels. Le système ACASYA présente plusieurs apports : Il prend en compte

l'interactivité Expert/Système pour contrôler et compléter les connaissances produites ; Il permet de renforcer en aval les méthodes classiques d'analyse de sécurité ; il met à profit l'expérience acquise sur les anciens systèmes de transport automatisés pour aider à la validation d'un nouveau système ; il permet finalement d'évaluer la qualité des scénarios d'accidents proposés dans le dossier d'analyse de sécurité du constructeur.

En dépit de l'intérêt d'ACASYA, ce système présente deux principales limites : une limite qui concerne le **langage de description** utilisé dans les scénarios d'accident qui **manque une cohérence globale** et une deuxième qui concerne l'insuffisance de **la stratégie de simulation adoptée** à travers l'application de l'apprentissage automatique pour résoudre de tels problèmes.

L'outil « SAUTREL » d'aide à l'Analyse des Effets des Erreurs du Logiciel (AEEL), (Hadj-Mabrouk H. et Darricau M., 1996), (Hadj-Mabrouk H., 2007), considéré parmi les premiers travaux de recherche sur l'application du raisonnement à partir de cas aux AEEL. La maquette logicielle a été réalisée à l'aide de la plateforme « ReCall » de la société ISoft. SAUTREL est à ce jour une maquette qui montre l'intérêt de la démarche d'aide aux AEEL proposée et qui, de ce fait, requiert certaines améliorations et extensions. Ces améliorations selon (Hadj-Mabrouk H., 2007), portent notamment sur le choix des critères d'évaluation des nouveaux cas d'AEEL, l'amélioration des stratégies d'adaptation des solutions proposées par le système, l'enrichissement de la base de cas d'AEEL pour couvrir l'ensemble du problème et enfin l'amélioration et la validation du formalisme de représentation des AEEL élaboré.

L'outil d'aide à la décision « REXCAS » proposé dans (Maalel A. et Hadj-Mabrouk H., 2010) et (Maalel A. et al, 2011), permet d'une part de pérenniser les connaissances en matière d'accidents issues de domaine de transport ferroviaire et d'autre part de déboucher sur la proposition des mesures de préventions et/ou correctives afin de minimiser le niveau de risque d'une nouvelle situation d'accident. La méthode de raisonnement à partir de cas (RàPC) est choisie comme une méthode de raisonnement sur l'ensemble de connaissances déjà capitalisées dans une base de données conventionnelle. L'outil logiciel mis en œuvre est perfectible et nécessite plusieurs améliorations. Ces améliorations touchent essentiellement, la formalisation des connaissances (une utilisation d'une base de données conventionnelle) et l'amélioration du processus de raisonnement (qui ne couvre pas tout le cycle de RàPC).

Dans un autre volet, « SIGAR », un outil d'aide à l'analyse des risques ferroviaires fut proposé par (Mazouni M.H., 2008). L'architecture de SIGAR comprend les trois niveaux ; physique, conceptuel et opérationnel. L'outil doit permettre aux utilisateurs de pouvoir manipuler les données, insérer de nouveaux éléments, vérifier l'intégrité de l'ensemble de données. Il doit aussi éviter dans la mesure du possible d'informations redondantes, afin de préserver la cohérence et optimiser l'exploitation des ressources. Il assure aussi la vérification de l'intégrité et « la partageabilité » des données (permettre l'accès simultané à la base de données par plusieurs utilisateurs). Malgré ces avantages, l'outil SIGAR présente certaines limites qui concernent surtout le non prise en compte de la richesse des connaissances modélisées en ayant recours à une base de données conventionnelle et notamment la méthode de raisonnement sur les connaissances utilisée qui est perfectible.

1.4. Le retour d'expérience

Plusieurs définitions, approches, et pratiques de retour d'expérience (Rex) ont été proposées dans la littérature. Joing et Keravel (Joing M. et Keravel F., 1993) par exemple, le définit comme « *un examen approfondi des circonstances conduisant à la réalisation d'événements touchant à la sécurité* ». C'est une démarche qui vise à mettre en ampleur les dysfonctionnements et les incompatibilités du système de sécurité et à formuler des propositions susceptibles d'éviter de telles situations.

Plusieurs auteurs tel que (Joing M., 1991), (Dominati A. et al., 1996), (Ferrandez F., 1999) et (Wanner J.C., 2000), précisent deux objectifs du Rex ; il s'agit de tirer des enseignements pour définir les actions correctrices de sécurité efficaces à court terme, et notamment de faire accumuler les connaissances sur comportements humains et les matériels. Pour Amalberti et Barriquault (Amalberti R. et Barriquault C., 1999), il s'agit d'un outil pour gérer l'information sur les dysfonctionnements des situations de travail. Cette présentation du concept Rex est proche de celle décrite par Bergmann (Bergmann R., 2002), Holitiana (Holitiana R., 2004) et Hadj-Mabrouk (Hadj-Mabrouk H., 2004). En effet, Bergmann (Bergmann, 2002) considère le management de l'expérience comme étapes de collecte, de modélisation, de stockage, de réutilisation, d'évaluation et de mise à jour de l'expérience. Holitiana (Holitiana, 2004) le représente comme une démarche structurée de capitalisation et d'exploitation des informations issues de l'analyse d'événements positifs et/ou négatifs. Elle met en œuvre un ensemble de ressources humaines et technologiques qui doivent être managées pour contribuer à réduire les répétitions d'erreurs et à favoriser certaines pratiques performantes. Hadj-Mabrouk (Hadj-Mabrouk H., 2004) le considère comme étant « *un processus dynamique de collecte, de stockage, d'analyse et d'exploitation des données relatives à des situations contraires à la sécurité (accident ou incident)* ». Il consiste en une étude détaillée des différents facteurs impliqués dans la genèse des incidents ou accidents.

Dans un autre volet, Cédric (Cédric B., 2008) le représente comme étant un mécanisme de gestion des connaissances particulier, conceptuel et ascendant. Selon Clermont et ses co-auteurs, il s'agit d'une démarche structurée de capitalisation et d'exploitation des connaissances issues de l'analyse d'événements positifs et/ou négatifs. Elle met en œuvre un ensemble de ressources humaines et technologiques qui doivent être organisées pour contribuer à réduire les répétitions d'erreurs et à favoriser certaines pratiques performantes. Enfin, selon Wybo (Wybo J.L., 2012) : « *Le retour d'expérience (REX) peut être défini, au sens large, comme toute formalisation d'un événement passé. Sous cet angle, il s'agit d'un concept très ancien et chacun le pratique d'une manière informelle* ».

Nous pouvons conclure que le but de Rex est de tirer profit des enseignements de l'expérience passée pour améliorer le niveau de sécurité en mettant en œuvre les mesures préventives et correctives adéquates afin d'éviter la reproduction des accidents dans le futur. Nous présentons dans le paragraphe suivant, les principaux travaux menés en matière de Rex.

1.4.1. Travaux existants en matière du Rex

Plusieurs travaux en matière de Rex ont été proposés par la communauté scientifique. Ces travaux constituent pour nous un centre d'intérêt important.

Certains travaux mettent **l'accent sur les contributions du Rex** pour renforcer et maîtriser les risques (Tea C., 2009) et (Violaine B., 2012). Le travail de Frédérique CHAUDET (Frédérique C, 2002) par exemple, met l'accent sur l'apport du Rex à la maîtrise des risques relatifs à l'hygiène, la sécurité et l'environnement. Ce travail s'est basé sur l'industrie du traitement thermique comme champ d'application. Céline Tea (Tea C., 2009) traite le Rex et les données subjectives d'une manière générale. Elle met l'ampleur sur les

systèmes d'information pour la gestion des risques. Le cas de la sécurité des transports ferroviaire est considéré. Pour Violaine BRINGAUD (Bringaud V., 2012), il s'agit de mettre 20 recommandations corrélées et complémentaires pour organiser un dispositif de Rex.

D'autres travaux considèrent plus l'enrichissement **d'un processus Rex** à travers la prise en compte des facteurs humains (Hadj-Mabrouk A et Hadj-Mabrouk H, 2003), (Research FonCSI, 2008) et (Wybo J.L., 2009). Hadj-Mabrouk A et Hadj-Mabrouk H (Hadj-Mabrouk A et Hadj-Mabrouk H, 2003) présentent une nouvelle approche d'intégration de l'erreur humaine dans le Rex. Ils prennent le domaine de la sécurité des transports ferroviaires comme champ d'application pour leurs travaux. Le travail de (Research FonCSI, 2008), *Research teams from the FonCSI programme Operational Experience Feedback*, met l'ampleur sur le succès des facteurs socio-culturels dans l'opérationnalisation de Rex et notamment sur les facteurs clés pouvant réussir un processus Rex en général. Wybo (Wybo J.L, 2009) consolide la place du maillon humain au sein de l'organisation pour la gestion de crise. Il traite le Rex comme étant un processus d'acquisition de connaissances et d'apprentissage.

Les travaux de (Holitiana R, 2004), (Beller C., 2008) et (Bringaud V., 2012) par exemple, prennent en considération l'**intégration et l'organisation d'un processus Rex dans le domaine industriel**. Tous ces travaux considèrent le Rex comme un moyen indispensable pour améliorer la sécurité. Pour Holitiana Rakoto (Holitiana R., 2004), Il s'agit d'intégrer le Rex dans les processus industriels. Le cas d'Alstom Transport est pris comme application. Pour Cédric BELER (Beller C., 2008), il s'agit de la modélisation générique d'un Rex cognitif, il prend le domaine de la prévention des risques comme cadre de son application. Le cas de la sécurité des transports ferroviaires est traité plus particulièrement dans les travaux de (Hadj-Mabrouk A et Hadj-Mabrouk H, 2003), (Holitiana R., 2004) et (Tea C., 2009). La plupart de ces travaux proposent des modélisations purement théoriques à l'exception de travail de Jabrouni H. (Jabrouni H., 2012), qui vise la définition et la spécification d'un outil collaboratif, support de la résolution de problèmes et de Rex implémenté au sein d'Alstom.

1.4.2. Cadre législatif et réglementaire du Rex ferroviaire

Bien que nous considérons que le cadre législatif et réglementaire du Rex ferroviaire est très ancien en Tunisie, la tendance actuelle est de plus en plus vers l'instauration d'un processus Rex commun, plus fiable sur le plan européen. En France, cette tendance est consolidée par un ensemble de textes législatifs et réglementaires (*la directive 2004/49/CE* relative à la sécurité des chemins de fer, *la Loi du 03 janvier 2002* et relative à la sécurité des infrastructures et systèmes de transport, aux enquêtes techniques, *l'arrêté du 19 mars 2012* fixant les objectifs, les méthodes, les indicateurs de sécurité et la réglementation technique de sécurité et d'interopérabilité applicables sur le réseau ferré national. Récemment, la LOI n° 2014-872 du 4 août 2014 portant la réforme ferroviaire.

Ces textes réglementaires renforcent la nécessité de mettre en place un organisme d'enquête permanent, spécialisé et indépendant. Parmi les textes institutionnels du Bureau d'Enquêtes sur les Accidents de Transport Terrestre (BEA-TT¹), nous pouvons citer :

- Les articles L. 1621-1 à L. 1622-2 du code des transports issus de la loi n° 2002-3 du 3 janvier 2002 modifiée relative à la sécurité des infrastructures et systèmes de transport, aux enquêtes techniques et au stockage souterrain de gaz naturel,

¹ Le Bureau d'Enquêtes sur les Accidents de Transport Terrestre (<http://www.bea-tt.equipement.gouv.fr>)

d'hydrocarbures et de produits chimiques ;

- Le décret n° 2004-85 du 26 janvier 2004 relatif aux enquêtes techniques sur les événements de mer et les accidents de transport terrestre, modifié par les décrets n° 2006-1279 du 19 octobre 2006 et n° 2012-668 du 4 mai 2012.

Le BEA-TT a été créé en 2004, il a pour mission de réaliser, en toute indépendance, des enquêtes techniques sur les accidents ou incidents graves ou potentiellement graves de transport terrestre afin d'en arranger les circonstances, d'en identifier les causes certaines ou possibles et d'émettre des appuis de sécurité destinées à anticiper de futurs accidents similaires.

1.4.3. Déroulement d'une enquête après un accident ferroviaire

Chaque État s'assure à ce que les enquêtes soient effectuées par un organisme permanent et indépendant après la survenu d'un accident. Le BEA-TT (Bureau d'Enquêtes sur les Accidents de Transport Terrestre), pour le cas de la France. Selon (Hadj-Mabrouk A. et Hadj-Mabrouk H., 2003), les organismes d'enquêtes doivent communiquer pour bien élaborer de procédés d'enquêtes et des recommandations communes. L'agence ferroviaire européenne¹ soutiendra ces organismes d'enquête dans la réalisation de cette tâche. Chaque enquête sur un accident doit faire l'objet d'un rapport indiquant **l'objectif de l'enquête** et éventuellement **les mesures nécessaires sous forme de recommandations relatives à la sécurité**.

Pour mener efficacement leur travail, les enquêteurs techniques doivent accéder à plusieurs éléments, témoignages et informations utiles, même couverts par le secret professionnel ou le secret médical. La Loi du 03 janvier 2002 préconise dans son article 14, le devoir de réaliser une enquête technique après accident ou incident, son objet, son but ainsi que le statut de l'organisme d'enquête. La Directive 2004/49/CE du 29 avril 2004 définit les principes des enquêtes obligatoires sur les accidents et/ou les incidents dont la gravité dépasse un seuil prédéfini et prévoit la mise en place d'organismes d'enquête indépendants.

1.4.4. Construction de la connaissance à partir de Rex

L'enchaînement d'un ensemble d'activités mises en œuvre dans un ordre bien déterminé pour réaliser un système est appelé cycle de développement ou cycle de vie d'un système. Le cycle de vie du système est classiquement synthétisé par un diagramme en «V» comprenant une branche descendante pour la spécification et la conception du système et une branche ascendante pour l'intégration et la validation du système (Hadj-Mabrouk A. et Hadj-Mabrouk H., 2003). Cette méthodologie de développement suppose le passage par les cinq phases chronologiques suivantes (Figure 1.4) : spécification du système, conception du système, réalisation des équipements, intégration du système et validation fonctionnelle du système :

¹ L'Agence ferroviaire européenne (European Railway Agency (ERA)) est une agence communautaire créée le 29 avril 2004 par un règlement du Parlement européen et du Conseil.

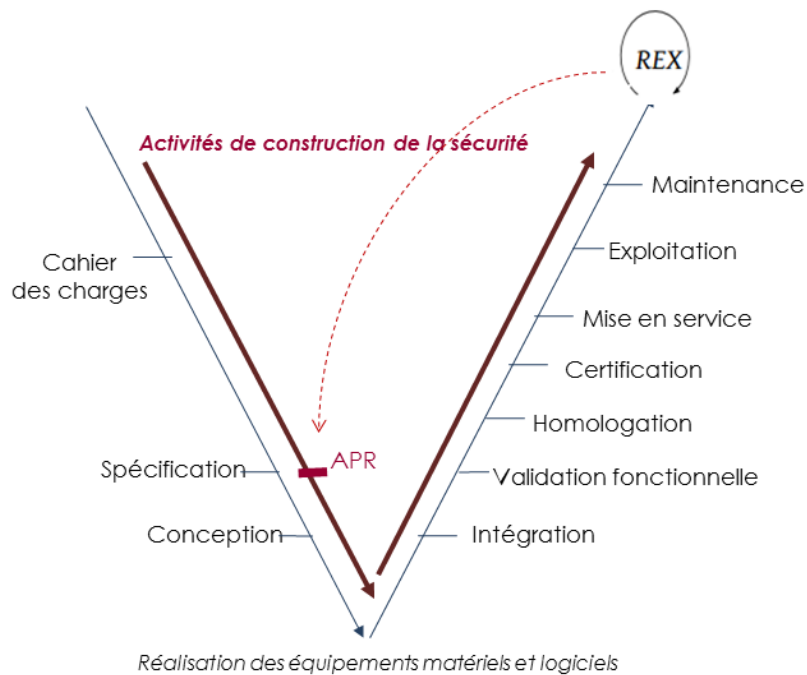


Figure 1.4. Cycle en V. de développement d'un système

En effet, le système après avoir été validé, homologué, certifié, mis en service par le ministre chargé des transports, exploité par le maître d'ouvrage et maintenu par les opérateurs de maintenance, la phase en aval correspond au Rex. Dans ce cadre, selon (Mbaye S., 2010), le Rex a pour but de collecter, de conserver, de transmettre, et de réutiliser les connaissances issues des expériences vécues.

Il existe un grand nombre de termes employés pour caractériser des expériences dans le domaine industriel. Le terme de Retour d'Expérience (Rex) est le plus utilisé en France. En Europe, le terme d'Expérience Management, en référence au Management des Connaissances (*Knowledge Management*), fait appel à la notion de gestion des connaissances liées à l'expérience. Aux États-Unis, c'est la notion de leçons acquises (*Lesson Learned*) qui est surtout employé. Pour la traduction littérale, le terme qui correspond au retour d'expérience est celui de « *Experience Feedback* ». Malgré la diversité des termes employés, c'est bien la gestion de l'expérience qui est au cœur de tous les travaux considérés.

Très schématiquement, un flux brut de données est lié à chaque expérience vécue. Une activité d'analyse et de traitement permet de transformer ce flux en un flux informationnel, puis, dans un deuxième temps en connaissances suite à une phase de compréhension. Ces connaissances résultantes de l'expérience passée pourront être pérennisées puis exploitées ultérieurement (Figure 1.5).

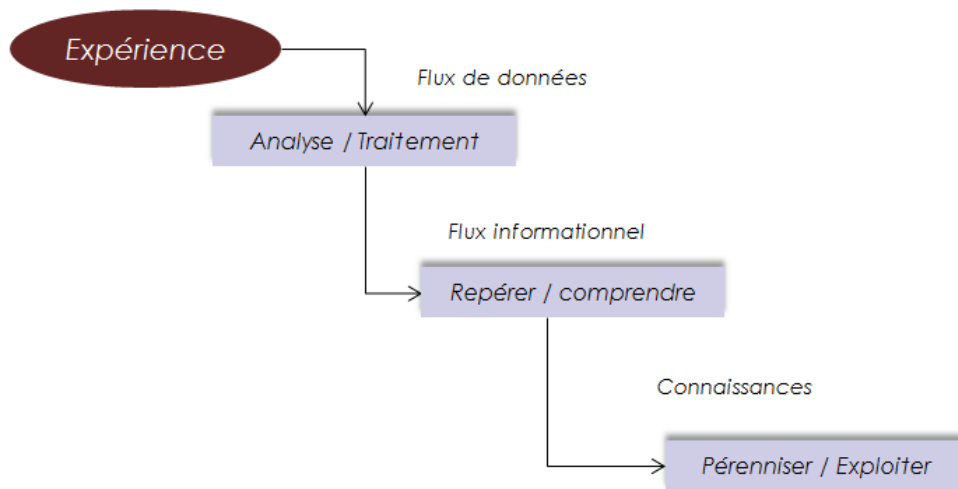


Figure 1.5. Création de la connaissance à partir de l'expérience.

Dans le cadre de notre travail nous considérons une articulation de processus Rex (Figure 1.6) se basant sur trois étapes complémentaires ; Analyse/Traitement, Compréhension et Capitalisation/Exploitation. Les trois étapes découlent respectivement de trois niveaux d'abstractions différentes : le niveau de donnée brut, le niveau informationnel et le niveau de connaissances résultantes.

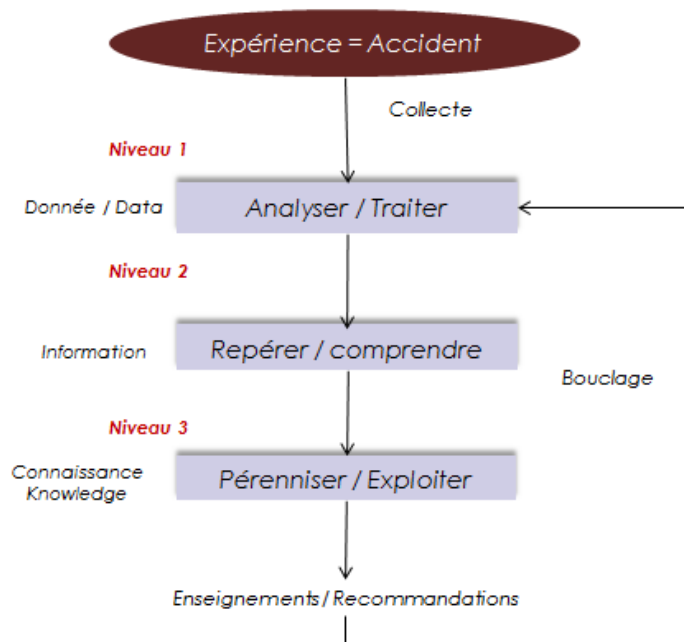


Figure 1.6. Articulation des différentes étapes de processus Rex.

La première phase du processus de construction du Rex s'intéresse au recueil maximal des données brutes. Ensuite, vient l'analyse et le traitement pour mieux cerner les événements affectant la sécurité. Donc, transformer les données brutes en un flux informationnel. La phase suivante est la formalisation, une phase aussi importante considérée comme un moyen de partage de l'expérience en favorisant la compréhension et la discussion à travers de représentations explicites. Elle présente toutefois certaines limites non négligeables comme la perte potentielle d'informations : tous les niveaux de compréhension ne sont pas toujours faciles à retranscrire. La dernière phase, concerne la

capitalisation et l'exploitation des connaissances générées grâce souvent à un outil informatique intelligent.

L'objectif essentiel du Rex n'est pas la réalisation du Rex lui-même, mais d'extraire l'événement réellement prédictif et de prédire et d'anticiper les futurs scénarios d'accidents ou événements non pris en compte, éclairer et identifier les mesures de préventions et de protections adéquates pour limiter la reproduction d'un événement d'insécurité dans le futur.

1.5. Carences et limites liées aux travaux d'analyse de la sécurité et de Rex

Nous avons insisté dans les sections précédentes sur la sûreté et la sécurité des systèmes industriels, en particulier le secteur de transport. Nous avons également montré l'importance cruciale occupée par le retour d'expérience (Rex) pour l'amélioration de la sécurité. En dépit des progrès accomplis en matière de sécurité, les situations d'insécurité, en particulier, les accidents dans le transport, avec leur coût humain et leur caractère parfois spectaculaire ou dramatique, nous rappellent que les hommes, les matériels et les organisations restent faillibles. Plusieurs travaux d'aide à l'analyse de la sécurité et de Rex ont été proposés dont quelques-uns sont présentés dans la section 1.3. Malgré leur intérêt incontestable, certaines difficultés sont décelées. Parmi ces limites, nous notons essentiellement :

- **L'hétérogénéité de sources de données** : Plusieurs sources de connaissances (rapports d'enquêtes techniques, études effectuées, connaissances expertes,...) sont disponibles pour les experts de domaine pour réaliser leur tâche d'analyse et d'évaluation de la sécurité. Ces sources sont malheureusement dans la plus part des cas mal formalisées, incohérentes et extrêmement ambiguës dans certains cas ;
- **Le vocabulaire utilisé et la représentation des connaissances** : Le vocabulaire utilisé souffre d'imprécisions, de confusions et de l'absence de définitions imparables admises par tous. L'absence d'une terminologie cohérente et une représentation standard des connaissances ne peut qu'affaiblir la communication et la collaboration entre les experts du domaine. En effet, l'analyse et l'évaluation de ces connaissances de sécurité sont généralement réparties entre plusieurs experts du domaine afin de traiter les différents aspects de la sécurité ; au niveau système, au niveau logiciel et au niveau matériel ;
- **Non prise en compte du facteur humain** : Les démarches existantes prennent d'une manière générale en compte le risque technique **en dépit du risque humain**. Elles n'abordent pas les spécificités liées à la composante humaine bien que celle-ci participe fortement à la genèse des risques d'accidents et des situations contraires à la sécurité.

Nous pouvons constater ainsi **le manque de méthodologies explicites d'analyse de la sécurité et de retour d'expérience**. Ces méthodologies standardisent, pour les experts de différents niveaux, non seulement la terminologie et le vocabulaire utilisés mais également le déroulement de Rex et les méthodes appliquées pour chaque étape de son processus.

1.6. Nos choix méthodologiques

Pour remédier à l'ensemble des limites déjà présentés et pour tenter des éléments de réponse à celles-ci, il est indispensable de se doter de systèmes d'aide à la décision permettant non seulement l'archivage des accidents historiques vécus (sur des systèmes des transports analogues homologués ou certifiés), mais surtout, d'assurer un retour d'expérience permettant d'exploiter ces scénarios d'accident, en vue de proposer et d'en

dégager un savoir-faire susceptible d'aider et d'assister les experts de domaine dans leur tâche d'analyse de la sécurité.

Dans notre travail de thèse, le **Raisonnement à Partir de Cas (RàPC)** (Case-Based Reasoning (CBR)), développé dans le domaine de l'intelligence artificielle (IA) et des sciences cognitives, au début des années 1980, considéré aussi comme méthode de raisonnement par analogie dans le domaine de l'apprentissage automatique, est choisi conjointement avec **les ontologies** pour surmonter ces défis. Tout d'abord, le RàPC est **une méthode parente du retour d'expérience**. Il est considéré comme étant un processus de résolution de problème. Dans ce cadre, l'expérience se définit comme un problème résolu ou « un cas », conservé dans une « base de cas » structurelle. Elle donne la possibilité de rechercher des solutions à un problème nouveau, par association aux cas résolus dans le passé. Dans ce même cadre, **les ontologies** jouent un rôle très important dans la formalisation et l'harmonisation des connaissances du Rex. Les ontologies se présentent alors comme un moyen efficace pour formaliser, structurer et stocker les connaissances. Elles permettent aussi d'établir un vocabulaire commun partagé pour la description des cas et la modélisation des connaissances nécessaires à l'indexation et à l'organisation des cas.

1.7. Conclusion

Nous avons introduit dans ce premier chapitre notre contexte général de la recherche. Après avoir présenté les concepts de base de la sûreté de fonctionnement notamment les principes, méthodes et techniques de la sécurité. Le présent chapitre a présenté quelques travaux en matière d'aide à l'analyse de la sécurité, a introduit aussi, le retour d'expérience Rex, les travaux existants en matière de Rex, son cadre législatif et réglementaire et le déroulement d'une enquête technique après un accident. Il a précisé par la suite les principes et les modalités de la construction de la connaissance à partir de Rex et notamment notre vision des différentes articulations de son déroulement.

Nous avons étudié par la suite les principales lacunes et obstacles dont souffrent les méthodes d'analyse de la sécurité et notamment le Rex ferroviaire. Ce sont essentiellement les limites relatives à la diversité des sources de données, à l'absence d'une terminologie homogène et une représentation standard des connaissances, au vocabulaire utilisé qui souffre d'imprécisions, de confusions et de l'absence de définitions rigoureuses admises par tous et aux démarches existantes qui prennent en compte le risque technique en dépit du risque humain. Cette partie a débouché sur le choix convenable des approches et des techniques informatiques mises en œuvre dans notre travail de thèse à savoir, le Raisonnement à Partir de Cas (RàPC) et les Ontologies. Notre démarche consiste à combiner et à tirer profit à la fois des ontologies (pour appréhender les limites et les insuffisances liées à la représentation des connaissances) et du RàPC (pour combler le problème de l'exploitation de la connaissance).

Il s'agit pour nous plus globalement de proposer des solutions pour étudier et animer efficacement un processus de Rex, pour anticiper les risques, de manager des connaissances pour bien bâtir le mode de représentation et de raisonnement adéquat, de rendre accessibles et exploitables ces connaissances dans le but de fournir une aide aux experts de sécurité dans leur tâche cruciale d'analyse de la sécurité et finalement pour concevoir et mettre en œuvre un système d'aide à la décision adéquat.

La suite de notre étude sera focalisée sur l'étude bibliographique autour des concepts de base que nous utilisons pour proposer notre approche d'aide à la décision.

Chapitre 2 : Concepts de base :

Représentation de connaissances et raisonnement

2.1. Introduction

Après avoir présenté dans le premier chapitre notre contexte général de la recherche, la sûreté et la sécurité des systèmes de transports et avoir débouché sur les limites d'en souffrir le retour d'expérience (Rex) et notamment les choix effectués pour proposer notre approche dans le cadre de ce travail de thèse, nous focalisons notre étude maintenant sur les concepts de base inhérents à la représentation de connaissances et aux méthodes de raisonnements.

Ce présent chapitre est consacré alors à l'étude théorique de ces concepts de base que nous utilisons pour élaborer notre démarche d'aide à la décision. Cette démarche est basée sur le raisonnement à partir de cas (RàPC), les ontologies et les techniques/méthodes issues de la fouille de données. **Le RàPC** est une méthode de raisonnement par analogie en apprentissage automatique. Il est considéré aussi comme étant une méthode parente du Rex permettant ainsi la résolution des problèmes. Débutant par une remémoration des situations similaires passées et se poursuivant par la réutilisation des informations et des connaissances disponibles sur ces situations, le RàPC élabore une nouvelle solution au problème. L'élaboration de la solution se fait généralement par adaptation des solutions passées retrouvées similaires sous l'hypothèse d'avoir des solutions proches lorsque les problèmes sont similaires.

Pour renforcer les étapes d'acquisition, de modélisation, de capitalisation et d'exploitation des connaissances de sécurité, il est primordial dans un premier temps, de se doter de moyens pour harmoniser, homogénéiser et assurer la cohérence du vocabulaire utilisé par les experts et les spécialistes de domaine. **Les ontologies** sont apparues pour pallier ces défis. La représentation des connaissances par les ontologies doit alors pouvoir s'appliquer aux connaissances d'un domaine particulier. Les ontologies se présentent comme des moyens efficaces pour formaliser, structurer les connaissances. Elles permettent d'établir un vocabulaire commun pour la description des problèmes ou pour modéliser les connaissances nécessaires à leur résolution. En outre, **la fouille de données** a pour objet l'extraction d'un savoir ou d'une connaissance à partir de grandes quantités de données, par des méthodes automatiques ou semi-automatiques et l'utilisation industrielle ou opérationnelle de ce savoir. Dans la fouille de données, il existe plusieurs techniques comme le clustering, les règles d'association, les réseaux de neurones, les arbres de décision etc.

Le présent chapitre aborde dans sa première partie, les ontologies, leurs aspects importants et plus particulièrement ceux que nous abordons dans notre travail de thèse à savoir : les rôles, les typologies, les composantes de base et les méthodologies de construction. Nous présentons à cet effet aussi, les différents modes de représentation de connaissances et notamment les standards web pour la représentation des ontologies. Nous présentons à la fin de cette partie, les fondements de base pour la conception des ontologies et les différentes approches pour évaluer la similarité sémantique entre deux objets au sein d'une ontologie. La deuxième partie est consacrée au RàPC, nous présentons ses origines, ses aspects importants à savoir : la structuration des cas, les modèles de RàPC, les tâches de RàPC et les différentes phases de son cycle de raisonnement comme la remémoration,

l'adaptation et l'apprentissage. La troisième partie de notre étude est focalisée sur la fouille de données et les méthodes d'extraction de connaissances à partir de données, plus précisément ceux que nous utilisons dans le cadre de notre approche à savoir, les règles d'association, la méthode d'arbre de décision et la technique d'Analyse des Concepts Formels (ACF). Notre étude théorique va déboucher enfin sur une synthèse comportant nos choix méthodologiques et techniques adoptés.

2.2. Les ontologies

2.2.1. Présentation de la notion d'ontologie

Dans la philosophie, le mot ontologie est construit à partir des racines grecques : *ontos* qui veut dire « ce qui existe », « l'existant », et *logos* pour « le discours », « l'étude ». En d'autres termes, ontologie signifie l'étude de ce qui existe, la science de l'être. Selon (Kolli M., 2011), l'ontologie au sens philosophique est une branche de la Métaphysique (une branche de la philosophie et de la théologie¹ qui porte sur la recherche des causes) qui s'intéresse à l'existence de l'être et aux catégories fondamentales de l'existant. Dameron (Dameron O., 2003) à son tour, la présente comme suit : « *l'ontologie au sens philosophique a été une source d'inspiration pour l'acquisition, la représentation théorique et le partage des connaissances* ».

D'un autre côté, l'ontologie a été abordée dans le domaine de l'intelligence artificielle (IA) pour la première fois par John McCarthy dans les années 80 selon (Psyché V. et al., 2003). Neeches et ses co-auteurs (Neeches R. et al., 1991) définissent l'ontologie comme suit : « *une ontologie définit les termes et les relations de base du vocabulaire d'un domaine ainsi que les règles qui indiquent comment combiner les termes et les relations de façon à pouvoir étendre le vocabulaire* ». Gruber (Gruber T.R., 1993) propose à son tour la définition suivante : « *spécification explicite d'une conceptualisation* » qui est jusqu'à présent la définition la plus citée dans la littérature en IA. Cette définition a été reprise par (Borst W., 1997) comme « *spécification formelle d'une conceptualisation partagée* ». Studer et ses co-auteurs (Studer et al., 1998) proposent la définition suivante « *spécification formelle et explicite d'une conceptualisation partagée* ».

- *Formelle* : l'ontologie doit être lisible par une machine, ce qui exclut le langage naturel ;
- *Explicite* : la définition explicite des concepts utilisés et des contraintes de leur utilisation ;
- *Conceptualisation* : le modèle abstrait d'un phénomène du monde réel par identification des concepts clefs de ce phénomène.

Sowa (Sowa J., 1999) la définit comme suit : « *l'ontologie en tant que domaine d'étude des catégories d'entités abstraites et concrètes qui existent ou peuvent exister* ». Noy et McGuinness (Noy N.F. et McGuinness D.L., 2001) présentent l'ontologie à leur tour comme une « *description formelle explicite des concepts dans un domaine du discours* » (concepts ou classes), des propriétés de chaque concept décrivant des caractéristiques et attributs du concept (attributs ou propriétés) et des restrictions sur les attributs (facettes).

2.2.2. Rôles de l'ontologie

Nous avons regroupé dans la figure 2.1 les quatre principaux rôles des ontologies au sein des Systèmes à Base de Connaissances (SBC) proposés dans (Kolli M., 2011) :

- *La communication* : Les ontologies interviennent dans la communication entre

¹ L'étude et l'exégèse de la religion, de Dieu, des textes sacrés ou des dogmes. <http://fr.wikipedia.org/wiki/>

personnes, organisations et logiciels à travers un vocabulaire conceptuel commun (Uschold M. et Grüninger M., 1996) ;

- *L'interopérabilité* : Le développement et l'implantation d'une représentation explicite d'une compréhension partagée dans un domaine donné peut améliorer la communication pour permettre ainsi un partage plus large et une interopérabilité ¹ plus étendue Uschold et Grüninger (Uschold M. et Grüninger M., 1996) ;
- *Les connaissances du domaine d'un SBC* : Les ontologies servent de squelette à la représentation des connaissances du domaine (Kolli M., 2011) ;
- *L'aide à la spécification de systèmes* : En effet, l'ontologie fournit une classification des objets que doit manipuler le système d'où une spécification explicite des connaissances modélisées (Kolli M., 2011).

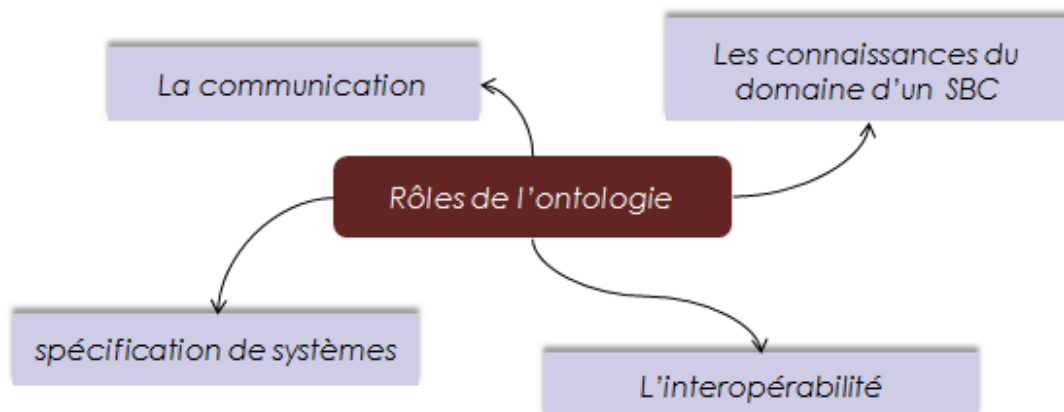


Figure 2.1. Principaux rôles de l'ontologie (Kolli M, 2011).

2.2.3. Classification des ontologies

Selon le domaine étudié et éventuellement les tâches pour lesquelles elles sont conçues, il existe plusieurs types d'ontologies. Pour notre approche de RàPC, il est intéressant de connaître ces typologies pour bien cibler le type d'ontologie adéquat que nous avons besoin.

Pour clarifier ces types d'ontologies construites, différentes classifications ont été proposées. Nous avons choisi de présenter dans ce paragraphe les deux classifications les plus citées dans la littérature, élaborées par (Psyché V. et al., 2003) et (Oberle D., 2006).

2.2.3.1. Classification de Psyché

Psyché (Psyché V. et al., 2003) présente plusieurs dimensions de classification illustrées dans la figure 2.2.

¹ L'interopérabilité est la capacité que possède un produit ou un système, dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre.

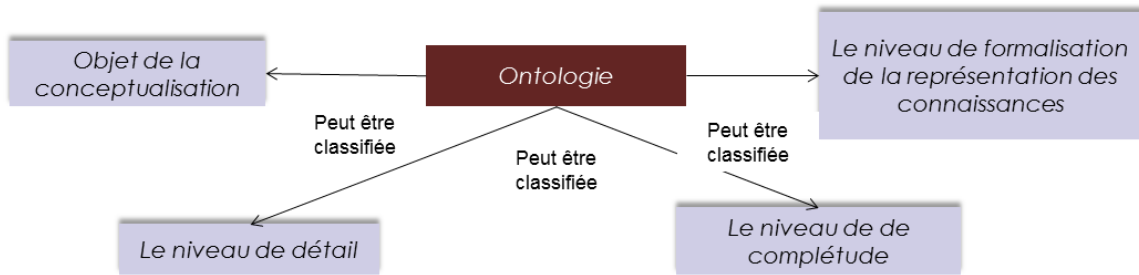


Figure 2.2. Typologies d'ontologies selon quatre dimensions de classification (Psyché V. et al., 2003).

2.2.3.1.1. Typologie selon l'objet de conceptualisation

Selon l'objet de conceptualisation, plusieurs typologies d'ontologie ont été proposées, nous pouvons distinguer cinq typologies (figure 2.3) :

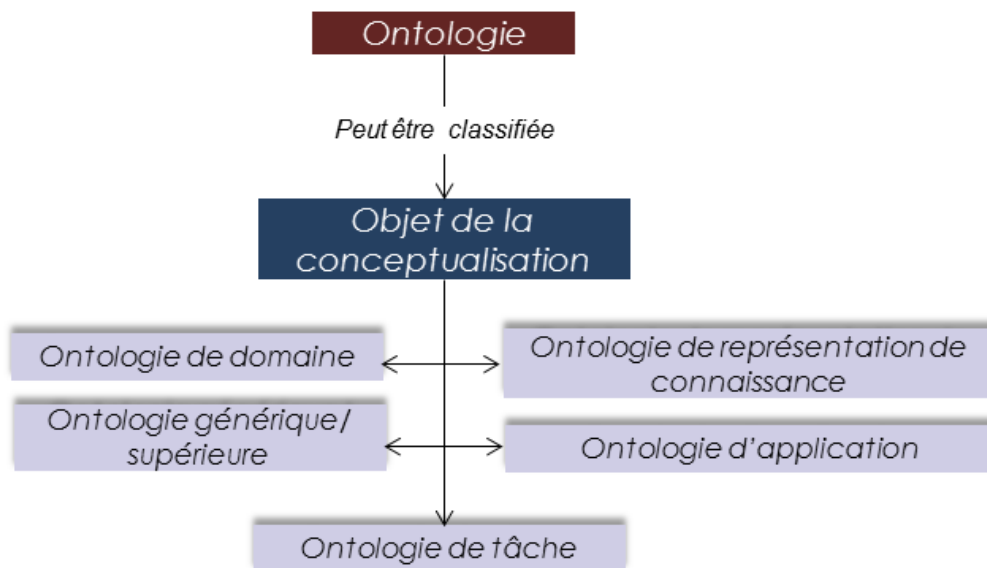


Figure 2.3. Typologies d'ontologies selon l'objet de conceptualisation (Psyché V. et al., 2003).

- **Ontologie de représentation de connaissances** Elle regroupe les concepts impliqués dans la formalisation des connaissances ;
- **Ontologie générique (*Generic Ontology*)** : Elle est appelée aussi ontologie de haut niveau ou bien ontologie supérieure. Elle traite des concepts généraux mais qui sont moins abstraits que ceux traités par l'ontologie de haut niveau. Cette généricité permet à ces ontologies d'être réutilisées dans plusieurs domaines d'études ;
- **Ontologie du domaine (*Domain Ontology*)** : Elle contient des concepts propres à un domaine précis. Une ontologie de domaine décrit le vocabulaire ayant trait à un domaine générique (l'enseignement, la médecine, la sécurité etc.), notamment en spécialisant les concepts d'une ontologie de haut niveau ;
- **Ontologie de tâches (*Task Ontology*)** : Elle fournit un ensemble de termes au moyen desquelles nous pouvons décrire généralement comment résoudre un type de tâche ;
- **Ontologie d'application (*Application Ontology*)** : Elle est utilisée pour modéliser les concepts d'un domaine particulier dans le cadre d'une application donnée.

2.2.3.1.2. Typologie selon le niveau de détail de l'ontologie

Psyché (Psyché V. et al., 2003) affirme que la classification peut s'effectuer en fonction du niveau de granularité, c'est-à-dire du niveau de détail des objets de la conceptualisation. Ainsi, selon l'objectif opérationnel de l'ontologie, une connaissance plus ou moins fine du domaine est nécessaire et des propriétés peuvent se révéler indispensables pour d'autres applications. Deux types de granularités peuvent être révélés : fine et large (Figure 2.4) :

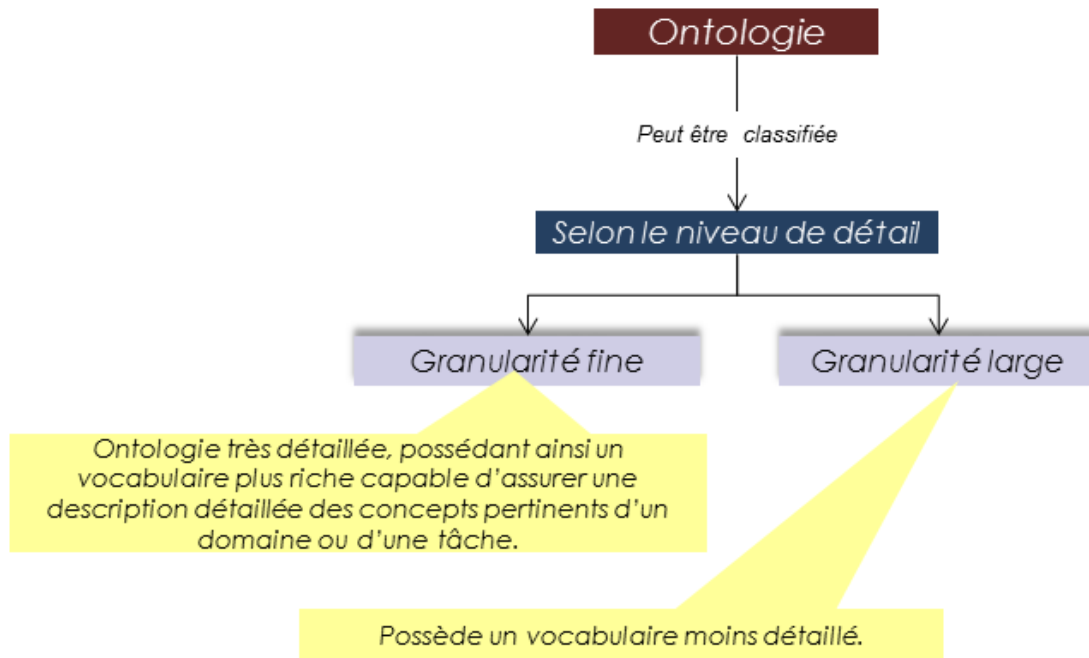


Figure 2.4. Typologie selon le niveau de détail de l'ontologie (Psyché et al. 2003).

2.2.3.1.3. Typologie selon le niveau du formalisme

La figure 2.5 présente une classification qui comprend quatre catégories par rapport au niveau du formalisme est proposée par (Uschold M. et Grüninger M., 1996) :

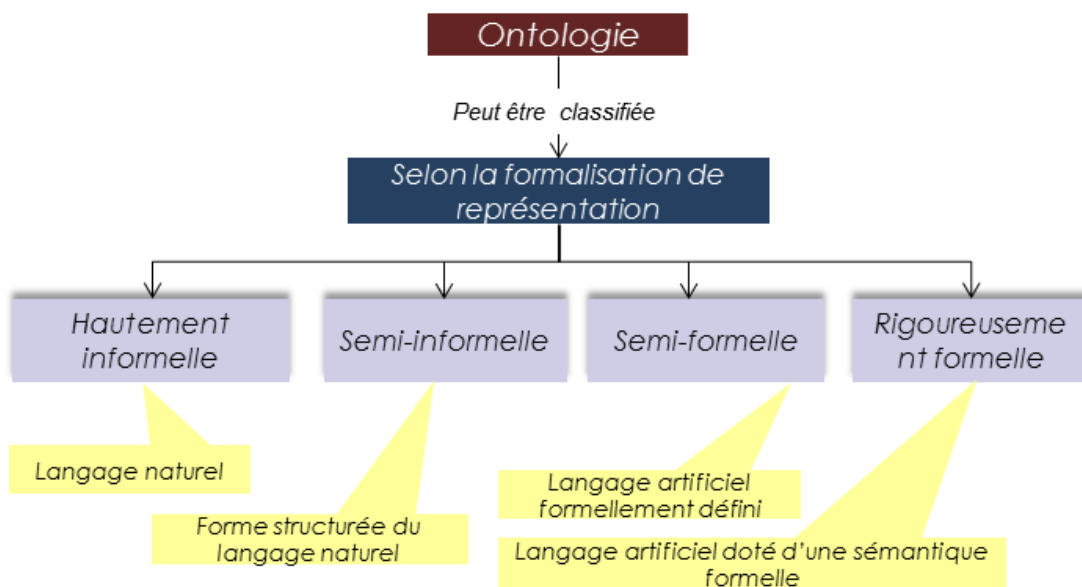


Figure 2.5. Typologie selon le niveau du formalisme (Psyché et al. 2003).

2.2.3.1.4. Typologie selon le niveau de complétude

A son tour, Bachimont (Bachimont B., 2000) propose une classification selon le niveau de complétude regroupant trois niveaux (Figure 2.6) :

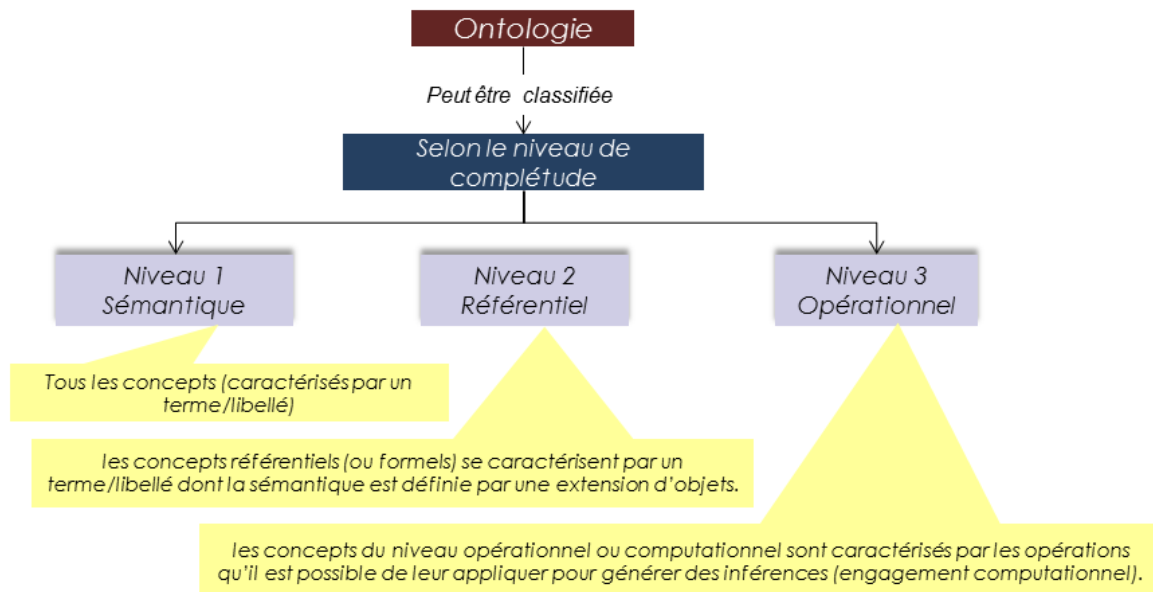


Figure 2.6. Typologie selon le niveau de complétude (Bachimont B., 2000).

2.2.3.2. Classification d'Oberle

Une autre classification est élaborée par Oberle (Oberle D., 2006). En effet, il propose une autre vision basée sur trois axes (le but, l'expressivité et la spécificité) (Figure 2.7) :

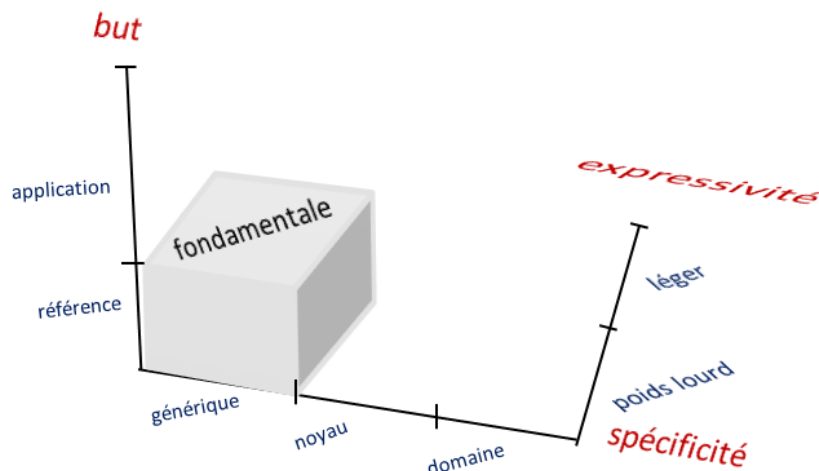


Figure 2.7. Dimensions de classification d'ontologie (Oberle, 2006).

- Selon **le but** : Les ontologies d'application qui sont destinées à une application donnée ou à une référence ;
- Selon **l'expressivité** : Les ontologies *light-weight* contiennent une simple hiérarchie de concepts des relations entre ces concepts. Les ontologies *heavy-weight* sont définies en déterminant des propriétés avancées sur les concepts ;
- Selon **la spécificité** : Les concepts généraux définis par des ontologies génériques. Les concepts communs à un ensemble de domaines sont définis par des ontologies noyaux. Les concepts propres à un domaine sont définis par des ontologies de domaine (Oberle, 2006).

2.2.4. Composantes d'une ontologie

Selon (Uschold M. et Grüninger M., 1996), une ontologie peut être vue comme un treillis de concepts et de relations entre ces concepts destinés à représenter les objets du monde sous une forme accessible et lisible par les hommes et par les machines. Une ontologie est constituée des concepts et des relations ainsi que des propriétés et des axiomes. Perez et Benjamins (Perez, A.G. et Benjamins V.R., 1999), considèrent que la connaissance au sein des ontologies est formalisée en utilisant cinq composants :

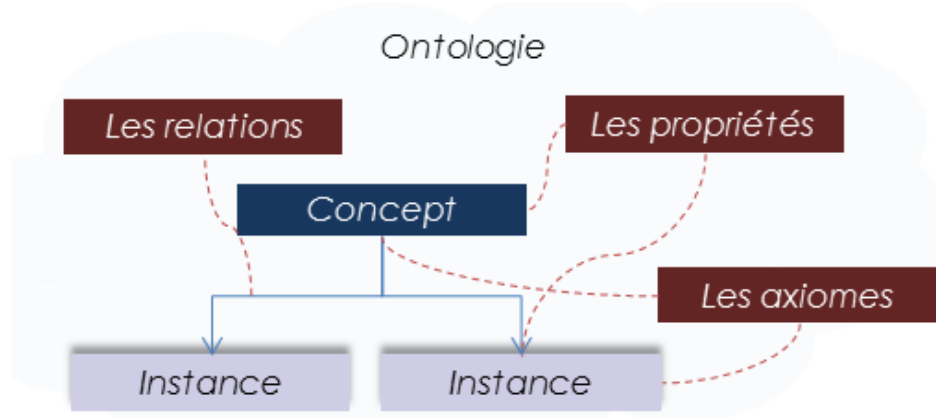


Figure 2.8. Composants d'une ontologie.

- *Les concepts* : sont des notions permettant la description d'une fonction ou d'une stratégie de raisonnement, etc. ;
- *Les relations* : sont les liens organisant les concepts de façon à représenter un type d'interaction entre les concepts d'un domaine. Nous pouvons citer les relations binaires comme : *sous-concept-de, connecté-à, sorte-de, etc.* ;
- *Les propriétés (ou attributs)* sont des caractéristiques liées aux concepts. Ces caractéristiques peuvent être évaluées ;
- *Les axiomes de l'ontologie* permettent de définir la sémantique des termes. Ils sont définis à l'aide de la logique du premier ordre en utilisant les prédicats de l'ontologie ;
- *Les instances* sont utilisées pour valoriser les concepts.

2.2.5. Méthodologies de construction des ontologies

La construction d'ontologie est un processus complet impliquant plusieurs acteurs, diverses connaissances et formalismes. Fernandez (Fernandez M., 1999) a proposé une bonne synthèse de ces méthodologies de construction. Psyché (Psyché V. et al., 2003) à son tour, a proposé un cadre référentiel qui permettra une analyse comparative des méthodologies recensées. Ces méthodologies recensées peuvent être analysées selon plusieurs critères ou dimensions, l'un d'entre eux étant le type du processus de construction :

- Construction de nouvelles ontologies à partir de zéro ;
- Réingénierie d'ontologies ;
- Fusion ou intégration d'ontologies ;
- Construction collaborative d'ontologies.

Nous présentons dans la suite un aperçu des méthodologies de construction d'ontologies les plus utilisées et citées dans la littérature.

2.2.5.1. Méthodologie d'Uschold et Gruninger

Quatre étapes pour développer une ontologie ont été spécifiées par (Uschold M. et Grüninger M., 1996) :

- L'identification du but et de la portée de l'ontologie ;
- La construction de l'ontologie qui se décompose en trois sous étapes : Capturer, coder dans un langage formel et intégrer les ontologies existantes ;
- L'évaluation de l'ontologie ;
- Le renseignement des choix adoptés au cours du développement afin de faciliter leur réutilisation plus tard.

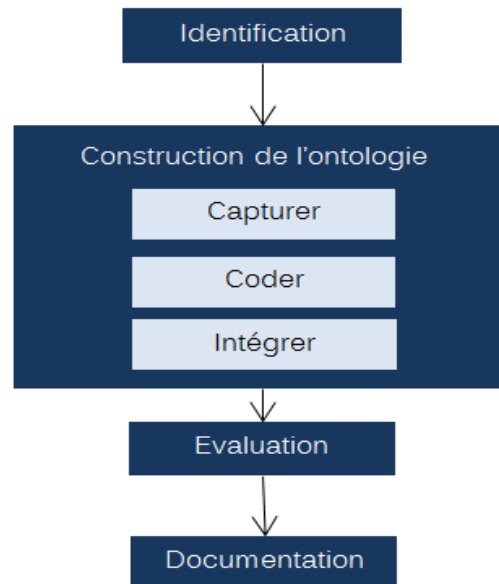


Figure 2.9. Méthodologie de construction de l'ontologie (Uschold et Gruninger, 1996).

2.2.5.2. Méthodologie Methontology

Methontology est une méthodologie qui fut proposée en 1997 au sein du laboratoire d'IA à l'université Polytechnique de Madrid. Selon (Fernandez M. et al., 1997) et (Nagypál G., 2005), cette approche décrit le processus de développement, le cycle de vie d'une ontologie, et les techniques utilisées dans chaque phase de développement. Son processus de développement comprend trois types d'activité (Figure 2.10) :

- les activités de **management de projet** (*planification, contrôle de qualité, etc.*) ;
- les activités de **développement** (*spécification, conceptualisation, formalisation, implémentation, maintenance*) ;
- les activités **intégrales** qui accompagnent toutes les phases de développement (*acquisition des connaissances, intégration d'autres ontologies, évaluation, documentation, etc.*).

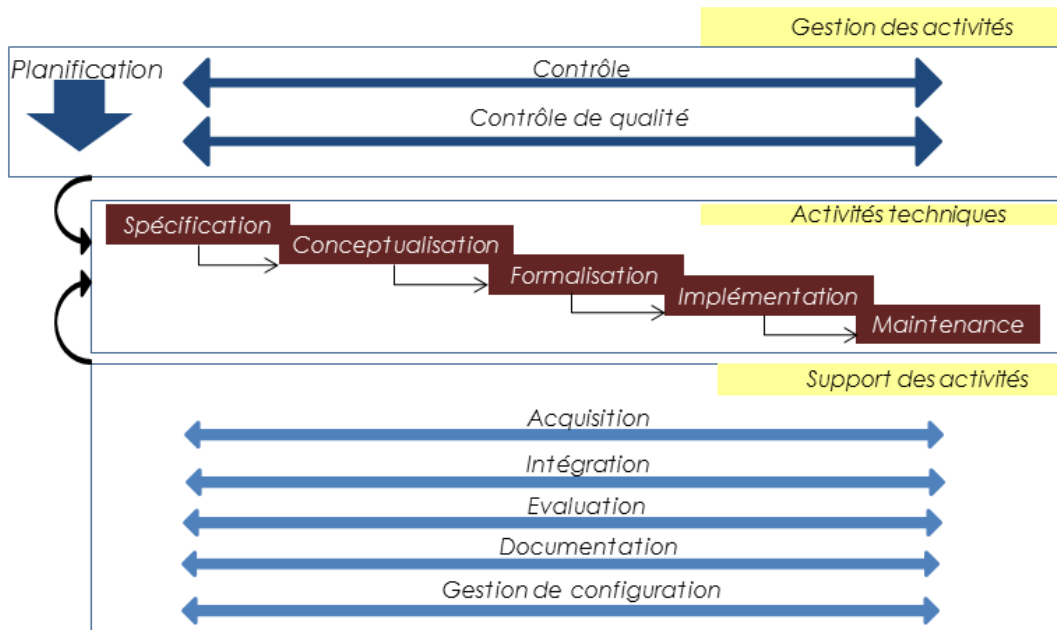


Figure 2.10. Méthodologie Methontology (Fernandez M. et al, 1997).

2.2.5.3. Méthodologie différentielle

Bachimont (Bachimont B., 2000) considère la définition et la mise en place d'une ontologie comme suit : « *définir une ontologie est une tâche de modélisation permettant de passer de l'expression linguistique des connaissances à une représentation formelle exploitable par l'ordinateur* ». La méthode de Bachimont propose d'élaborer successivement deux ontologies en faisant varier le cadre sémantique : une ontologie « sémantique », fondée sur une sémantique linguistique, puis une ontologie « référentielle » (ou « extensionnelle »). Cette modélisation s'effectue en trois phases d'engagements :

- **L'engagement sémantique** : consiste à déterminer les concepts (les primitives de représentation de connaissances) d'un domaine et d'une tâche précise ;
- **L'engagement ontologique** : consiste à amener la sémantique référentielle de tout concept. C'est-à-dire, l'ensemble de référents associés à chaque concept, pour permettre la définition d'une ontologie référentielle ;
- **L'engagement computationnel** : consiste à définir la sémantique computationnelle ¹ permettant de reconnaître la caractérisation informatique des opérations et règles applicables aux concepts. Cet engagement conduit à la définition d'une ontologie computationnelle (Bachimont, B. 2000).

2.2.5.4. Méthodologie *OntoSpec*

Kassel et Perpette (Kassel G. et Perpette S., 1999) considèrent deux étapes pour développer une ontologie (Figure 2.11) :

- *L'ontologisation* : correspond à un travail coopératif entre l'ingénieur de la connaissance, les spécialistes et/ou des utilisateurs de système, pour parvenir à une conception cohérente et consensuelle ;
- *L'opérationnalisation* : Codage de l'ontologie conceptuelle dans un langage de représentation adéquat.

¹ Computationnelle : de manière logico-algébrique

La méthodologie *OntoSpec* fut proposée dans (Kassel G., 2002). Elle a pour objectif de reconnaître les exigences nécessaires et suffisantes permettant de bien définir les concepts et les relations d'une ontologie.

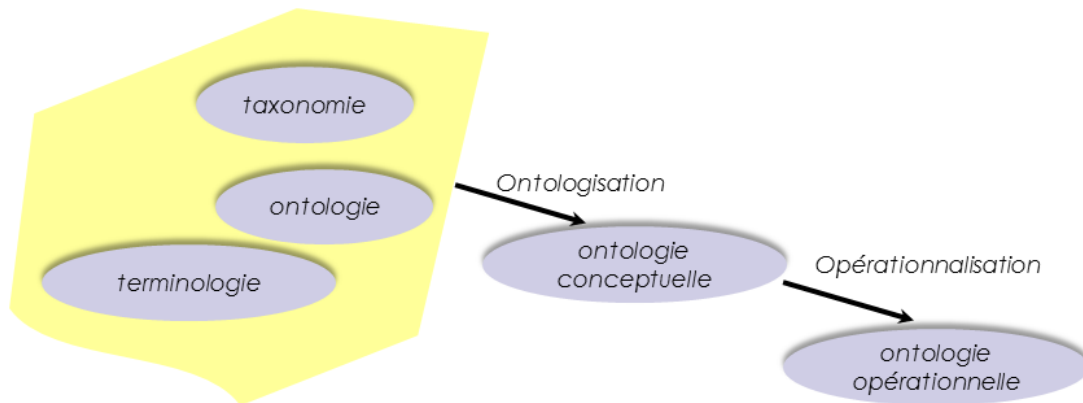


Figure 2.11. Phases de développement d'ontologie (Kassel et Perpette 1999).

2.2.6. Modes de représentation de connaissances

Très schématiquement, une représentation formelle exprime des connaissances sous forme logique. La représentation de connaissances permet de construire une première modélisation partiellement cohérente (semi-formalisée). Il est nécessaire de la représenter dans un langage compréhensible par un ordinateur pour la doter des extensions ayant le pouvoir de raisonner sur ces connaissances. Selon (Euzenat J. et al., 2000), « *représenter des connaissances propres à un domaine particulier consiste à décrire et à coder les entités de ce domaine de manière à ce qu'une machine puisse les manipuler afin de raisonner ou de résoudre des problèmes* ». Bahloul (Bahloul D., 2006) considère l'objectif étant de rendre l'ontologie opérationnelle à travers la manipulation des connaissances représentées selon l'objectif du système conçu. Il existe plusieurs niveaux d'abstraction et de représentation des connaissances et par conséquent différents modes de représentation de connaissances dans le domaine de l'IA. A cet effet, une variété de langages de représentation fut proposée. Le paragraphe suivant fera l'objet de cette étude.

2.2.6.1. Les réseaux sémantiques

Issus des modèles psychologiques de (Quillian R., 1968), les réseaux sémantiques sont des graphes dont les nœuds représentent des objets, événements ou concepts et les arcs représentent des relations tout type entre les nœuds (Figure 2.12). Dans la même lignée, un nouveau formalisme graphique, *les réseaux d'héritage structurés*, est implémenté dans le système KL-ONE (Brachman R. et Schmolze J.C., 1985). Cette représentation rend visible les diverses relations existantes entre les objets et les distances entre deux concepts. Selon (Miller, 1990), ce formalisme convient aux domaines où les concepts sont simples et fortement liés entre eux comme les phrases en langage naturel, c'est pourquoi il a été adopté pour produire l'ontologie Wordnet.

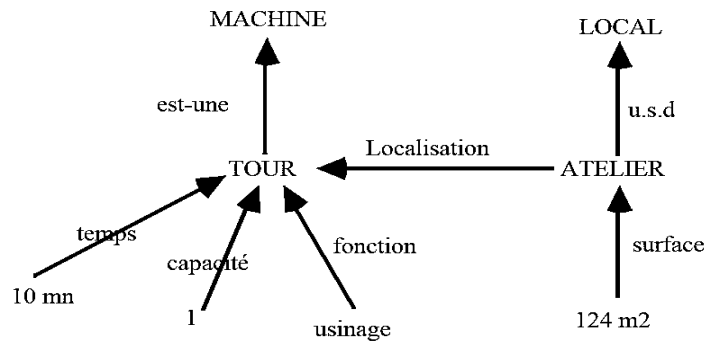


Figure 2.12. Exemple d'un réseau sémantique (Miller, 1990).

2.2.6.2. Les langages de frames

Le langage de frames ou schémas a été introduit par M. Minsky (Minsky M., 1975). Un des buts de ce mode est de représenter toutes les connaissances relatives à une situation dans un schéma plutôt que de répartir la connaissance à travers de axiomes. Les frames sont généralement organisés en réseaux. Les systèmes de frame sont très utiles pour modéliser les domaines où les objets sont complexes. Le regroupement de toutes les informations d'un objet dans une unité fournit une représentation concise et facilement exploitable. En 1995, Kifer et ses co-auteurs (Kifer et al., 1995) ont proposé de formaliser la structure des langages orientés objets et des langages de frame à travers la Logique de Frame (abrégiée en F-logic).

2.2.6.3. Les logiques de description

Le système « KL-ONE » (Brachman R. et Schmolze J.C., 1985), est en effet le premier système représentant des logiques de description. Son objectif était d'être un langage de représentation dont la sémantique formelle est externe à la représentation et aux algorithmes qui opèrent sur cette sémantique. Selon (Bahloul D., 2006), dans le cas où *une ontologie* est formalisée à l'aide des logiques de description, cette ontologie permet de décrire les concepts d'un domaine à travers des prédicats unaires spécifiant les objets du domaine et des rôles atomiques, décrivant les relations entre les objets. Une Logique de Description LD est composée de deux parties : un langage terminologique (appelé T-Box) et un langage assertionnel (appelé A-Box).

- *La T-Box (Terminological Box)* correspond au niveau descriptif qui permet de décrire les concepts en fonction d'autres concepts à partir des relations et des contraintes sur ces relations ;
- *La A-Box (Assertional Box)* correspond au niveau des assertions, elle est réservée à la description et la manipulation des individus et des instances.

Les LD fournissent les primitives qui comprennent trois composantes formelles : le concept, le rôle et l'individu.

2.2.6.4. Les graphes conceptuels

Développés par (Sowa, 1984), les graphes conceptuels se dérivent des réseaux sémantiques et représentent un modèle de représentation de connaissances du type réseaux sémantiques fondés sur la logique. Djida Bahloul (Bahloul D., 2006) présente un graphe conceptuel comme : « un graphe fini, connexe, biparti ; composé de deux sortes de nœuds : les nœuds conceptuels, appelés également sommets, et les nœuds relationnels (ou relations). Chacun de ces nœuds a une étiquette. Un nœud concept est étiqueté par un type correspondant à une classe sémantique, et un marqueur précisant une instance particulière de la classe ».

Nous présentons dans la figure 2.13, un graphe conceptuel présenté dans (Bahloul D., 2006). Le graphe peut être interprété par « Accident mortel » (accident dans lequel une personne est morte).

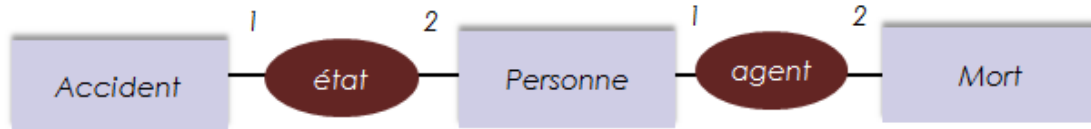


Figure 2.13. Exemple d'un graphe conceptuel (Bahloul, 2006).

2.2.6.5. Les standards web pour la représentation des ontologies

Il existe plusieurs langages d'ontologies qui ont été développés ces dernières années. Ils deviendront certainement des langages d'ontologie dans le contexte du Web sémantique. Certains de ces langages sont basés sur la syntaxe de XML (*Extended Markup Langage*), tels que XOL (*Ontology Exchange Langage*), SHOE (*Simple HTML Ontology Extension*) qui a été précédemment basé sur le HTML, OML (*Ontology Markup Langage*), RDF (*Resource Description Framework*), RDF Schéma. Trois langages additionnels sont établis sur RDF(S) pour améliorer ses caractéristiques : OIL (*Ontology Inference Layer*), DAML+OIL (*DARPA Agent Markup Langage + Ontology Inference Layer*) et enfin OWL (*Web Ontology Langage*).

La figure 2.14 représente les rapports principaux entre tous ces langages sous la forme d'une pyramide des langages du Web sémantique élaborée par (Boussaid I., 2005).

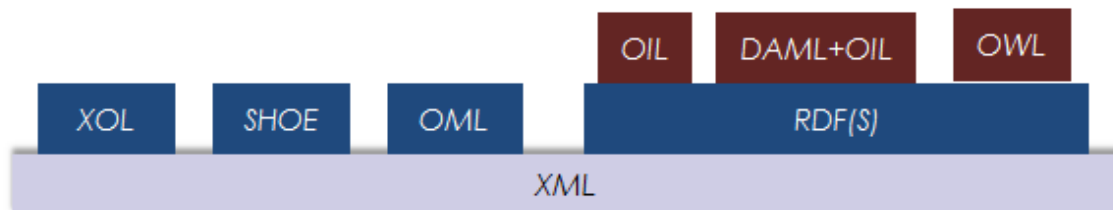


Figure 2.14. La pyramide des langages basés Web (Boussaid I., 2005).

- **RDF (Resource Description Framework)** : Le W3C ¹ (World Wide Web Consortium) a adopté le langage RDF comme formalisme standard de représentation qui repose sur la syntaxe XML, (constitué déjà un standard). Selon (Manola F et Miller E, 2004), RDF permet de décrire des ressources Web en termes de ressources, propriétés et valeurs :

- *Les ressources* : toutes choses décrites par des expressions RDF sont appelées des ressources. Elles sont identifiées par une URI (Universal Resource Identifier) ;
- *Les propriétés* : il s'agit de caractéristiques, d'attributs ou de relations utilisées pour décrire des ressources spécifiques ;
- *Les valeurs* : la valeur de la propriété peut être une autre ressource ou elle peut être littérale.

- **RDF Schema (RDF Vocabulary Description Language)** : Selon (Brickley D., 2004), l'extension à RDFS permet de déclarer les propriétés et le type des ressources.

¹ W3C : un organisme de normalisation à but non lucratif, fondé en octobre 1994 chargé de promouvoir la compatibilité des technologies du World Wide Web.

RDFS fournit les primitives de modélisation permettant de construire des ontologies simples. Dans une vision favorisant l'utilisation d'ontologies sur le Web, le langage RDF(S) a été enrichie par l'apport du langage OIL (Ontology Interchange Language) (Fensel D. et al., 2000) qui permet d'exprimer une sémantique à travers le modèle des frames tout en utilisant la syntaxe de RDF(S).

- **OIL (Ontology Inference Layer)** : Selon (Fensel D. et al., 2000), le langage OIL combine les primitives de modélisations utilisées dans le langage de frames et le raisonnement formel des logiques de description pour exprimer des ontologies sur le Web. Par rapport à RDFS, OIL reprend la séparation faite en logique de description entre concept primitif (décrit par des conditions nécessaires d'appartenance) et concept défini (décrit par des conditions nécessaires et suffisantes d'appartenance). OIL offre de nouveaux éléments permettant de définir des classes moyennant l'intersection de classes et l'union de classes par exemple.
- **DAML+OIL (DARPA Agent Markup Language)** : DAML+OIL est issu de l'union de deux langages : le langage OIL¹ et DAML-ONT² (Hendler et McGuinness, 2000). L'intégration de OIL rend possible les inférences compatibles avec les logiques de description. Ce nouveau langage supporte désormais les types de données primitifs, et la définition d'un certain nombre d'axiomes comme l'équivalence de classes ou de propriétés.
- **OWL (Ontology Web Language)** : Le langage OWL (Ontology Web Language) est le standard le plus utilisé. Il a acquis le statut de recommandation du W3C le 10 février 2004, fondé sur le standard RDF. Actuellement, OWL2 (Ontology Web Language 2) (W3C OWL Working Group, 2012), est un langage d'ontologie pour le Web sémantique d'un sens formellement défini. OWL2 regroupe des classes, des propriétés, des individus et des valeurs de données et sont tous stockées sous forme de documents Web sémantique. OWL2 peut être utilisé conjointement avec RDF et OWL2 et échangé comme les fichiers RDF (W3C OWL Working Group, 2012). Pour des raisons de modularité, OWL a été fractionné en trois langages distincts :
 - **OWL Lite** : contient un sous-ensemble des constructeurs. (Horrocks et Patel-Schneider, 2003) montrent ainsi que ce langage est dans la même classe de complexité que la logique de description SHIF(D). Il a une complexité formelle inférieure à celle d'OWL DL ;
 - **OWL DL** : ce langage utilise la logique de description DL. Il a été défini pour les utilisateurs qui sollicitent une expressivité maximale ;
 - **OWL Full** : sans aucune contrainte est la fusion de RDFS et d'OWL DL. OWL Full permet à une ontologie d'augmenter la signification du vocabulaire prédéfini (RDF ou OWL).

Nous venons de voir dans cette section les différents langages qui ont conduit à la définition d'un standard permettant de représenter des ontologies sur le Web. Nous étudions dans la section suivante les différents environnements et éditeurs pour les ontologies.

2.2.7. Fondements pour la conception d'une ontologie

¹ <http://www.ontoknowledge.org/oil>

² <http://www.daml.org/committee>

Le processus de construction d'ontologies, appelé ingénierie ontologique, peut être décrit selon les principes qui le gouvernent, et les méthodologies et les outils qui le soutiennent. Pour bien mener la construction d'une ontologie, il est souhaitable de tenir compte de quelques fondements établis par des experts. Gruber (Gruber T.R., 1993) et Arpirez (Arpirez J. et al., 2001) ont proposé huit critères pour guider le processus de conception d'ontologie :

- **Clarté** : l'ontologie devrait fournir des définitions claires pour les termes utilisés ;
- **Exhaustivité** : chaque définition doit être exprimée par une condition suffisante ;
- **Cohérence** : une ontologie cohérente devrait pouvoir permettre des inférences en conformité à ces définitions ;
- **Extensibilité monotone maximale** : les nouveaux termes, devraient être inclus dans l'ontologie sans entraîner de modifications dans les définitions existantes ;
- **Interventions ontologiques minimales** : l'ontologie devrait spécifier le moins possible la signification de ses termes ;
- **La diversification des hiérarchies** est utile afin d'optimiser la puissance dérivant des mécanismes d'héritage multiple ;
- **Minimisation de la distance sémantique entre des concepts frères** : les concepts proches sont regroupés et représentés dans des sous-classes d'une même classe ;
- **Normalisation des noms** : il est important de suivre des conventions pour définir les identifiants et les libellés des éléments de l'ontologie (classes, attributs, etc.).

2.2.8. La similarité entre concepts dans une ontologie

La similarité entre concepts dans une ontologie est un élément fondamental qui est utilisé par différentes techniques telles que la fouille de données, le web sémantique et en particulier, la recherche d'information. La similarité sémantique représente une notion définie entre deux concepts soit au sein d'une même hiérarchie ou appartenant à deux hiérarchies conceptuelles différentes. La similarité sémantique indique que ces deux concepts partagent des éléments en communs.

Dans cette section, nous allons exposer dans un premier temps, les différentes méthodes de mesure de similarité citées dans la littérature, puis nous justifions nos choix et nous enchainons une comparaison plus poussée des méthodes structurelles que nous utilisons dans le cadre de notre thèse.

2.2.8.1. Les méthodes de mesure de similarité

La littérature propose plusieurs méthodes de calcul de la similarité entre les concepts dans les ontologies. Un état de l'art complet est présenté par (Patwardham S., 2003) où différentes mesures ont été comparées par rapport à des évaluations faites sur des sujets humains et des classifications de celles-ci sont aussi données dans (Euzenat J., 2007), (Zghal S., 2010) et (Xavier A., 2011). Ces méthodes peuvent être ventilées en méthodes terminologiques, méthodes structurelles, méthodes extensionnelles et enfin, les méthodes sémantiques.

Le tableau 2.1 présente les différents types de méthodes de mesures de similarité :

<i>Méthode</i>	<i>Référence</i>	<i>Description</i>
<i>Les méthodes terminologiques</i>	(Levenshtein V., 1966) (Euzenat J., 2007).	Selon (Ziani M. et al., 2010), ces méthodes sont employées pour calculer la valeur de similitude des entités textuelles, telles que des noms, des méta-données sur les noms, des étiquettes, des commentaires. <i>Zghal Sami</i> (Zghal S., 2010), à son tour, les a séparées en méthodes syntaxiques et méthodes linguistique. Les méthodes syntaxiques permettent de calculer la valeur de la similarité des entités textuelles. Les méthodes linguistiques permettent de déterminer la similarité entre deux entités. Ces méthodes syntaxiques prennent en charge les propriétés expressives et productives du langage naturel (Zghal S., 2010).
<i>Les méthodes linguistiques</i>	(Sabou M. et al., 2006) (Safar S., 2007)	Ces méthodes ont eu recours à des ressources externes (dictionnaires, taxonomies,...), en particulier, en s'appuyant sur les liens qu'entretiennent les entités comparés avec d'autres entités repérées dans ces ressources linguistiques externes pour conclure la similarité entre deux entités donnés. (Ziani M. et al., 2010).
<i>Les méthodes structurelles</i>	(Madhavan J. et al., 2001)	Les méthodes structurelles déterminent la similarité entre deux entités en fonction des informations structurelles. Elles calculent la similarité entre deux concepts en exploitant les informations relatives à leur structure interne (valeur d'attributs, restrictions et cardinalités sur les attributs, valeurs des instances, etc.). Les entités sont reliées entre elles par des liens sémantiques ou syntaxiques. Ces liens forment ainsi une hiérarchie ou un graphe d'entités. Selon (Zghal S., 2010), les méthodes structurelles se subdivisent en deux familles : les méthodes structurelles internes et les méthodes structurelles externes . La première famille exploite uniquement les informations décrivant les attributs d'entités, tandis que la seconde famille considère des relations entre ces entités. Ces méthodes vont être davantage détaillées dans la section suivante.
<i>Les méthodes extensionnelles</i>	(Euzenat J., 2007)	Ces méthodes déduisent la similarité entre deux entités qui sont notamment des concepts ou des classes en analysant leurs extensions c'est à dire leurs couvertures en termes d'ensemble d'instances reconnues par une entité concept. Ces mesures calculent la similarité de deux entités en fonction de la similarité entre les deux ensembles de leurs instances.
<i>Les méthodes hybrides</i>	(Leacock C. et al., 1998)	Selon (Ziani M. et al., 2010), ces méthodes combinent plusieurs mesures lorsqu'une seule est insuffisante. Généralement, ces méthodes sont intégrées dans des outils permettant la mise en correspondance d'ontologies (Zghal S., 2010).

Tableau 2.1. Les techniques de mesures de similarité

Généralement, la définition d'une métrique de similarité peut changer selon les méthodes, selon les propriétés souhaitées et les objectifs sous-adjacents. Selon (Zghal S., 2010), la valeur d'une fonction de similarité est souvent normalisée, c'est à dire comprise entre 0 et 1, ce qui permet des possibilités d'interprétation probabiliste de la similarité.

Dans notre contexte de travail, partant des types de connaissances dont nous disposons, un cas d'accident comporte deux groupes d'attributs. Des attributs dits « simples » pour lesquels nous cherchons uniquement à comparer leurs valeurs respectives. D'autres attributs dits « complexes » pour lesquels nous exploitons les correspondances trouvés au niveau

des concepts supérieurs ou/et des concepts fils dans la hiérarchie des concepts de l'ontologie. Il s'agit donc, plus précisément de faire appel aux méthodes structurales. En effet, la similarité entre deux concepts quelconques de l'ontologie est perçue comme une évaluation de la similarité des concepts structurellement en liaison de parenté avec ceux-ci. Nous enchainons à cet effet, dans ce qui suit, avec une comparaison plus détaillée de ces méthodes structurales.

2.2.8.2. Méthodes structurales de mesure de similarité au sein des ontologies

Dans la littérature, plusieurs travaux sur les méthodes structurales de mesure de similarité ont été développés dans différents contextes. À partir de ces études effectuées, nous avons alors entamé une sorte de classification de ces méthodes en quatre familles de méthodes d'identification de la similarité sémantique (Tableau 2.2).

Référence	Contenu informationnel	Les arcs	Basée concept	Basée slot
	Quantité d'information partagée entre deux objets	Nombre d'arcs chemin entre les deux objets	Emplacement des objets dans l'ontologie	Valeurs des attributs communs de deux objets
(Rada R. et al., 1989)		×		
(Cover T.M. et Joy A.T., 1991)	×			
(Wu Z. et Palmer M., 1994)		×	×	
(Resnik P., 1999)	×			
(Jiang J., 1997)	×			
(Lin D., 1998)	×			
(Bergmann R. et Stahl A., 1998)				×
(Leacock C. et Chodorow M., 1998)			×	
(Ehrig M., 2004)		×		
(Zargayouna H. et Salotti S., 2004)		×	×	
(Zhang J. et al., 2006)				×
(d'Amato C. et al., 2008)			×	

Tableau 2.2. Comparaison des approches de calcul de similarité au sein des ontologies

Nous constatons la première famille d'approches basées sur **les nœuds** (Cover T.M. et Joy A.T., 1991), (Resnik P., 1999), (Jiang J., 1997) et (Lin D., 1998), utilisant le degré de la quantité d'information partagée entre des concepts et des mesures du contenu informationnel pour déterminer la similarité conceptuelle. Cette information partagée est égale au contenu informationnel du plus petit généralisant.

La deuxième famille d'approches repose uniquement sur **la hiérarchie ou sur les distances des arcs** (Rada R. et al, 1989), (Wu Z. et Palmer M., 1994) et (Ehrig M., 2004). L'ontologie est représentée par un graphe (Figure 2.15) dont les nœuds sont des concepts, et les arcs sont les liens entre concepts. Selon (Khatraoui M. et al., 2008), les mesures

reposant sur la distance considèrent que la similarité entre deux concepts peut être calculée à partir du nombre de liens qui séparent les deux concepts.

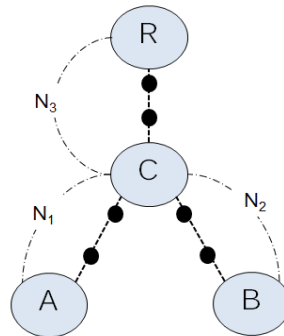


Figure 2.15. Similarité basée sur les arcs.

La troisième famille d'approches **combine les deux premières approches**, elle est basée concept (Leacock C. et Chodorow M., 1998). La mesure de similarité, notée *simcpt*, dépend de l'emplacement des objets dans l'ontologie. Elle s'inscrit ainsi dans la même catégorie que les approches basées sur les arcs (Wu Z. et Palmer M., 1994) et celle de (Zargayouna H. et Salotti S, 2004) qui propose une amélioration de la mesure (Wu et Palmer, 1994). Par exemple, (d'Amato C. et al., 2008) estiment que deux concepts peuvent être similaires sans avoir d'instances en commun. Ils proposent une nouvelle mesure qui évalue non pas l'intersection entre les ensembles d'instances de chaque concept, mais la variation de la cardinalité du plus petit subsumant en commun.

Finalement, la quatrième famille qui dépend **des valeurs des attributs communs des objets comparés** (Zhang J. et al, 2006), (Bergmann R. et Stahl A., 1998) et qui est basée sur une similarité basée-slot (ou similarité inter-classe).

2.3. Le raisonnement à partir de cas (RàPC)

Dans ce paragraphe, nous allons étudier les différents aspects du RàPC et particulièrement ceux que nous abordons dans notre travail.

2.3.1. Origines et présentation du RàPC

Les premiers travaux sur le RàPC sont ceux de (Schank R.C. et Abelson R.P., 1977). Selon ces travaux, nos connaissances génériques sur les situations rencontrées sont enregistrées sous forme de scripts nous permettant de faire des prévisions et de réaliser des inférences. Dans les années 80, Schank (Schank R.C., 1986) a continué à explorer le rôle de la mémoire dynamique des situations passées dans la résolution de problèmes et dans l'apprentissage. L'équipe de Schank (Riesbeck C.K. et Schank R.C., 1989) a pu rendre ce modèle opérationnel à travers l'implémentation des Memory Organization Packets (MOPs). Carbonell (Carbonell J.G., 1983) considère plutôt l'aspect résolution de problèmes par analogie : « *La résolution des problèmes analogues consiste à transférer les connaissances sur les étapes de résolution du problème passé vers le nouveau problème qui partage des aspects importants de l'expérience passé* ». Kolodner (Kolodner J., 1992) considère l'utilisation des cas antérieurs peut se faire dans le but d'adapter leur solutions aux nouvelles situations rencontrées, mais aussi d'expliquer, critiquer, ou interpréter ces situations. Selon (Aamodt A. et Plaza E., 1994), le RàPC est une approche de résolution de problèmes ayant pour objectif de résoudre un nouveau problème à l'aide d'un ensemble de problèmes déjà résolus. Autrement dit, raisonner à partir de cas c'est résoudre un nouveau problème en se

remémorant une situation précédente similaire et en réutilisant les connaissances de cette situation.

Kolodner (Kolodner J., 1992) sépare les approches de RàPC en deux catégories (Kolodner J., 1992) : les approches de résolution de problèmes, et les approches interprétatives (Figure 2.16).



Figure 2.16. Approches de RàPC (Kolodner, 1992).

2.3.2. Définition et structuration d'un cas

Selon (Cordier A. et al., 2009), un cas est habituellement formé d'une partie *problème* et d'une partie *solution*, chaque partie étant composée d'un ensemble de descripteurs, la plupart du temps définis dans une *ontologie* dédiée. (Haouchine M.K., 2009) le définit comme : « *Un cas est une expérience représentée par une connaissance. Selon le domaine d'application et les objectifs à atteindre, les informations contenues dans le cas varient* ». Fuchs (Fuchs B. et al., 2006) définit le cas comme étant la description informatique d'un épisode de résolution de problème.

Nous reprenons dans la suite la structure de cas présentée dans (Lieber J. et al., 2004) et (Renaud J. et al., 2007) : Dans une application de RàPC, nous considérons l'existence de deux espaces : l'espace de problèmes nommé *Problèmes* et l'espace de solutions nommé *Solutions*. Il existe une relation binaire, *Sol*, sur *Problèmes* \times *Solutions* signifiant "*a pour solution*". Résoudre un problème $pb \in$ *Problèmes* c'est donc trouver (ou construire) $sol \in$ *Solutions* tel que pb a pour solution sol ; nous avons, $Sol = Sol(pb)$. Un cas est défini par un couple $(pb, Sol(pb))$. Un raisonneur à partir de cas a pour but de résoudre un nouveau problème, appelé problème cible et noté *cible*, en faisant appel à un problème similaire déjà résolu, appelé problème source et noté *srce*, et à sa solution $Sol(srce)$. Généralement, la représentation structurée est la plus utilisée dans la majorité des travaux. Ainsi, le cas est souvent représenté sous la forme d'un ensemble de descripteurs. Un descripteur « *d* » est défini par une paire $d = (a, v)$ où « *a* » est un attribut et « *v* » est la valeur qui lui est associée.

2.3.3. Modèles de RàPC

Les modèles de RàPC sont regroupés en trois grandes familles : les modèles conversationnels, textuels et structurels (Bergmann R. et al., 2003), (Lamontagne L., 2004) :

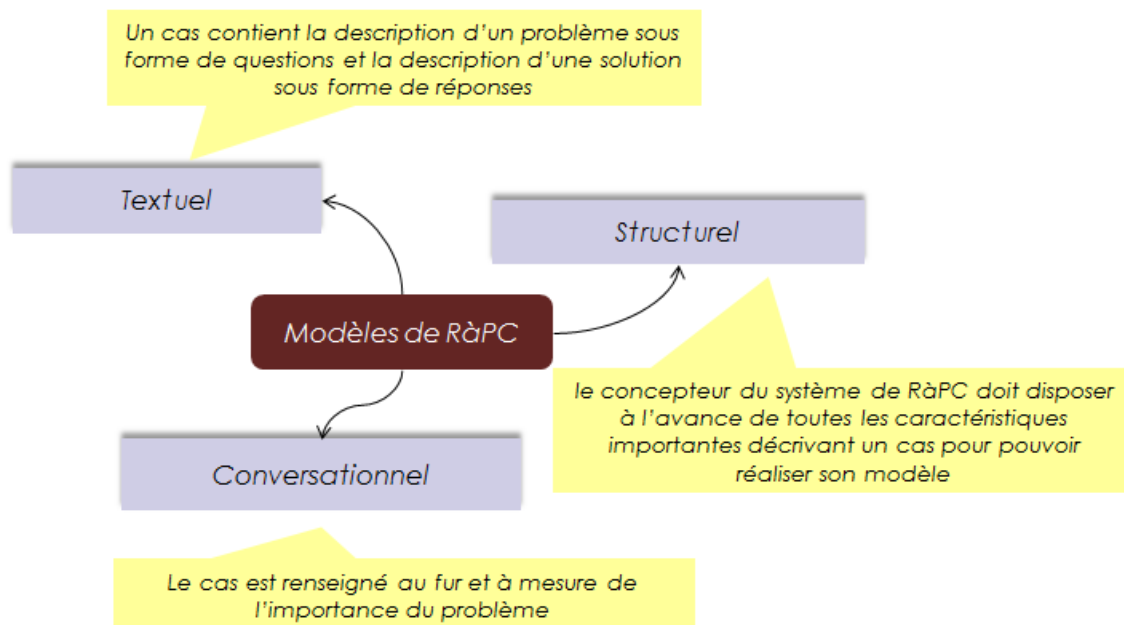


Figure 2.17. Modèles de RàPC (Bergmann R. et al., 2003), (Lamontagne L., 2004).

- **Le Modèle conversationnel :** Selon (Lamontagne L., 2004), le modèle conversationnel est mis sur l'interaction entre l'utilisateur et le système pour définir progressivement le problème à résoudre. Dans le schéma conversationnel, l'interaction entre le système et l'utilisateur se fait comme suit : L'utilisateur fournit au système une brève description textuelle du problème à résoudre. Le système propose une série de questions. Selon (Bergmann R. et al., 2003) (Lamontagne L., 2004), pour chaque réponse fournie, le système réévalue la similarité de chacun des cas ;
- **Le Modèle structurel :** Selon (Lamontagne L. et Lapalme G., 2002), dans ce modèle, « toutes les caractéristiques importantes pour décrire un cas sont déterminées à l'avance par le concepteur du système ». La similarité globale entre deux cas est habituellement évaluée par une somme pondérée de la similarité de chacun des attributs. A noter que tous les travaux sur l'adaptation de cas sont menés dans le cadre du modèle structurel ;
- **Le Modèle textuel :** Selon (Lamontagne L., 2001), les praticiens du domaine ont constaté les limites de l'approche structurale et ont proposé des modèles pour étendre son application à d'autres domaines. Dans cette approche, les cas textuels sont non-structurés si leur description est complètement "free-text".

2.3.4. Tâches de RàPC

Dans les travaux d'Althoff (Althoff K.D. et al., 1995), les systèmes de RàPC sont regroupés en deux grandes familles selon le type de problèmes qu'ils cherchent à résoudre (Figure 2.18) : les *systèmes de classification* et les *systèmes de synthèse*.

- **Systèmes d'analyse/classification :** Dans un problème d'analyse, le but est de placer un cas dans une catégorie prédéfinie. Selon (Althoff K.D. et al., 1995) : « il s'agit alors d'identifier les cas sources similaires à un nouveau problème en vue de déterminer le type de ce problème. La solution du cas le plus similaire dans cette "classe" est ensuite réutilisée pour résoudre le nouveau problème ».

- La prédiction, le diagnostic et le raisonnement juridique peuvent être pris comme exemple pour ces systèmes ;

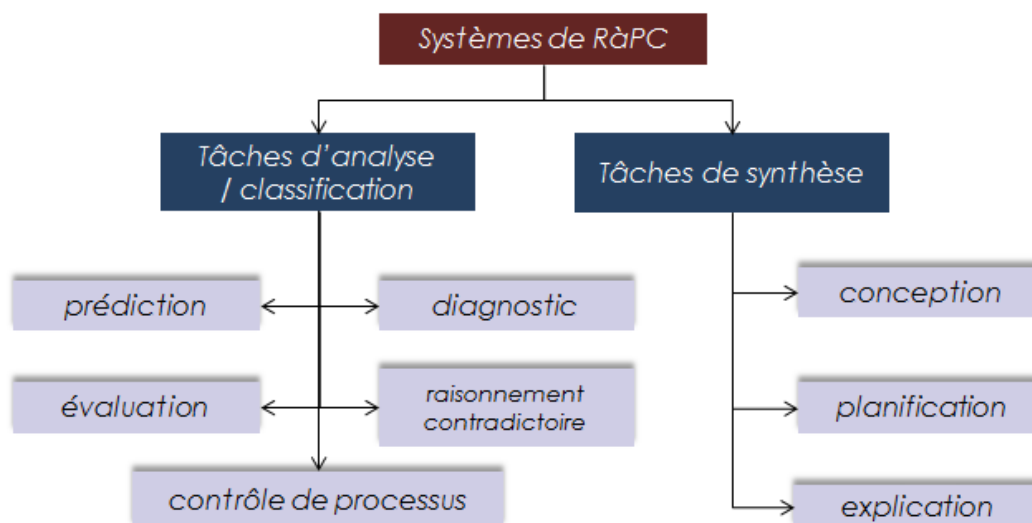


Figure 2.18. Classification des systèmes de Règle à Partir (Althoff K.D. et al., 1995).

- **Systèmes de synthèse** : Dans ce type d'application, le processus d'adaptation est de grande importance et requiert plus d'effort que dans les systèmes d'analyse/classification. Trois problématiques de synthèse peuvent être décrites :
 - o **La conception** : le raisonneur essaie de remémorer et de réutiliser d'anciens cas qui ont été créés avec des contraintes similaires à celles demandées actuellement ;
 - o **La planification** : selon (Renaud J. et al., 2007), « un problème est donné par un état initial, *init*, et un but à aboutir, *but*, et la solution sera un plan permettant de parcourir l'espace d'états de *init* jusqu'à un état réalisant le but ». A titre d'exemple, le système TAAABLE (Badra F. et al., 2008), un des systèmes destinés à produire une recette adaptée à une certaine demande à partir de recettes connues ;
 - o **L'explication** : il s'agit de l'identification de ce qui a été la cause d'un certain événement, généralement d'une anomalie. Selon (Schank R.C., 1986), expliquer un phénomène se fait en remémorant un phénomène similaire et en adaptant son explication.

2.3.5. Cycle de Règle à Partir

Le nombre de phases de Règle à Partir est différemment défini selon les chercheurs. Il peut être composé de trois, quatre ou cinq phases. Aamodt et Plaza (Aamodt A. et Plaza E., 1994) le composent de quatre phases : la remémoration (ou recherche du cas similaire), l'adaptation (ou la réutilisation du cas retrouvé), la validation (ou la révision du cas sélectionné) et la mémorisation (ou l'apprentissage). Fuchs (Fuchs et al, 2006) le composent de trois phases à savoir la remémoration, l'adaptation et la mémorisation. Quand à Mille (Mille A., 2006) et (Cordier A. et al., 2007) le déterminent sur cinq phases : La phase d'élaboration, la remémoration, l'adaptation, la révision et la phase de mémorisation.

2.3.5.1. Élaboration

La première phase de raisonnement est l'élaboration d'un problème cible (cas cible) à partir d'une description initiale fournie par l'utilisateur. Rappelons que l'objectif de tout le cycle de raisonnement est l'attribution d'une solution convenable pour ce problème puis le mémoriser dans la base de cas si tout le processus s'est abouti correctement. Pratiquement, les cas cibles ne sont pas toujours bien fournis en termes de description. Nous parlons parfois d'une incomplétude qui désigne l'absence d'au moins un attribut ou encore une incohérence dans le cas d'une contradiction entre certains descripteurs. Des approches issues de la fouille de données peuvent être d'un grand apport pour remédier à des telles lacunes. Toutefois, dans la littérature, nous pouvons repérer différentes manières pour mener l'élaboration. D'après Fuchs et ses co-auteurs (Fuchs B. et al., 2006), la phase d'élaboration a pour objectif de préparer la remémoration en enrichissant la description du problème tel qu'il est saisi par l'utilisateur, afin d'obtenir le problème cible. L'élaboration est importante dans le RàPC car elle permet la facilité d'exécution et d'évaluation des différentes phases du cycle. Certains auteurs ont travaillé sur la phase d'élaboration comme Caïn et ses co-auteurs (Caïn T. et al., 1991), ils ont proposé une description des problèmes et des solutions par un ensemble de faits et un ensemble d'explications reliant le problème et la solution. Un fait est considéré comme descripteur problème s'il est intervenu comme règle d'une explication d'un problème passé. Bento et ses co-auteurs (Smyth B. et al., 1993) ont repris cette méthode pour tenir compte la qualité des explications. Les faits prétendants comme descripteurs des problèmes deviennent alors plus ou moins adéquats. Dans un autre volet, le système KRITIK fut proposé par (Goel V. et al., 2000) pour extraire de la base de cas un ensemble de cas selon une similarité de description de problème, puis pour évaluer le coût de l'adaptation sur les cas les plus proches du problème spécifié. Cette même idée est reprise par (Abou-Assali A., 2010) dans la phase de remémoration des cas similaires. Fuchs et ses co-auteurs (Fuchs B. et al., 2006), ont proposé à cet effet une première formalisation de l'élaboration en s'appuyant sur un exemple et sous l'hypothèse, assez générale, selon laquelle les problèmes et solutions peuvent être considérés comme des ensembles (ou conjonctions) de descripteurs.

2.3.5.2. Remémoration

Cette phase vise à récupérer les cas sources similaires au problème cible donné permettant de lui trouver une solution. Selon (Smyth B. et McKenna M.T., 2001), deux approches sont considérées dans cette phase, celles reposant **sur le calcul de la similarité entre le cas source et le cas cible et celles utilisant, en plus de la notion de similarité, la notion de diversité**. L'objectif de la deuxième approche est de remémorer des cas similaires au cas cible et de choisir, parmi ces cas, ceux qui ne sont pas très similaires entre eux. Quant à la première approche, l'objectif est de retrouver le cas similaire au problème actuel en mesurant leur degré d'appariement. Ainsi le degré de similarité représente la fonction *d'utilité/adaptabilité* de la solution.

Selon (Haouchine M.K., 2009), il existe plusieurs mesures de similarité génériques dans la littérature scientifique et dans plusieurs domaines tels que l'Analyse des Données (AD), la Reconnaissance des Formes (RF), l'Apprentissage Symbolique (AS), ou encore les Sciences Cognitives (SC). Pour les mesures de similarités, Rifqi (Rifqi M., 1996) a effectué une étude dans ce cadre et a présenté deux types de mesures : les similarités *locales* ou les similarités *globales*.

- *Similarités locales* :

Généralement, le calcul des similarités locales est basé sur la distance et dépend du type de descripteur :

- Les valeurs de descripteurs numériques :

$$sim(a, b) = \frac{|a - b|}{range}$$

- Les valeurs de descripteurs symboliques (mono-valeurs) :

$$sim(a, b) = \begin{cases} 1 & \text{pour } a = b \\ 0 & \text{pour } a \neq b \end{cases}$$

- Les valeurs de descripteurs symboliques (multi-valeurs) :

$$sim(a, b) = \frac{card(a \cap b)}{card(a \cup b)}$$

- Les valeurs de descripteurs taxonomiques :

$$sim(a, b) = \frac{h(commonnode(a, b))}{\min(h(a), h(b))}$$

- a et b : sont des valeurs des descripteurs ;
- $card$: est la cardinalité de l'ensemble ;
- $range$: est la valeur absolue de la différence entre la borne sup. et la borne inf. de l'ensemble des valeurs ;
- h : est le poids de l'arbre taxonomique.

- Similarités globales :

Elles sont calculées au niveau des cas ou des objets en intégrant les similarités locales. Plusieurs mesures de similarités globales ont été proposées, leur utilisation dépend étroitement du domaine étudié. Nous pouvons citer : *Weighted Block-City*

$$sim(A, B) = \sum_{i=1}^n w_i sim_i(a_i, b_i)$$

- la distance de *Minkowski*. Si $r = 2$ alors nous retrouvons la distance *Euclidienne* (*Weighted Block-City*). Si $r = 1$ alors il s'agira de la distance de *Manhattan* (s'explique par le nombre de pas verticaux et horizontaux pour atteindre le but en partant de l'état courant).

$$sim(A, B) = \left[\frac{1}{n} \sum_{i=1}^n sim_i(a_i, b_i)^r \right]^{\frac{1}{r}}$$

- *Maximum based*

$$sim(A, B) = \max_i w_i sim_i(a_i, b_i)$$

- n est le nombre d'attributs,
- w_i est le poids (évalué en fonction de l'importance) de l'attribut i ,
- sim_i est la similarité locale calculée pour l'attribut i .

Une fois la mesure de similarité établie pour un système donné, la remémoration des cas dans la base de cas se fait suivant un algorithme de recherche. Plusieurs types d'algorithmes peuvent être alors appliqués. L'algorithme *des K Plus Proches Voisins* « *KPPV* » plus connus en Anglais sous le nom *K-Nearest Neighbors* (K-NN) (Weiss & Kulikowski, 1991) est la méthode la plus utilisée.

2.3.5.3. Adaptation

Cette phase permet au système de réutiliser généralement une solution d'un cas source récupéré par la remémoration afin de résoudre le problème cible. Selon (Abou-Assali A., 2010), la réutilisation se fait en adaptant la solution du cas source aux spécificités du nouveau problème. Fuchs (Fuchs B. et al., 1999) considèrent l'adaptation comme un plan dont l'état initial est la solution de départ et l'état final est la solution adaptée. Lieber (Lieber J. et al., 2004) considèrent que la phase d'adaptation consiste à effectuer un raisonnement par analogie : « sachant que la solution du cas cible est à la solution du cas source ce que le cas cible est au cas source, connaissant le cas source et sa solution ainsi que le cas cible, que vaut la solution du cas cible ? ».

2.3.5.3.1. Approches d'adaptation

Wilke et Berglman (Wilke W. et Bergmann R., 1998) déterminent quatre types d'adaptation automatique (Figure 2.19) :

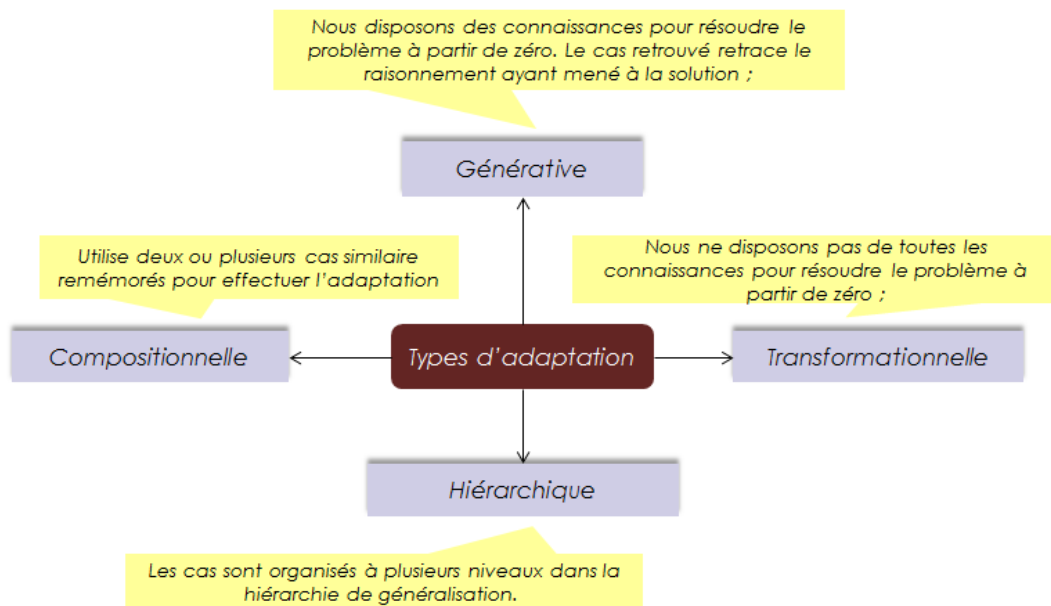


Figure 2.19. Types d'adaptation en RàPC (Wilke W. et Bergmann, R. 1998).

- **Adaptation par transformation** : Les premiers travaux ont été proposés par (Carbonell J.G., 1983) sur les transformations analogiques. Il s'agit de transformer les différences dans les problèmes en différences dans les solutions ;
- **Adaptation généralive** : Cette adaptation consiste à générer une nouvelle solution pour un problème donné. L'adaptation consiste à réutiliser la trace de raisonnement de la solution décrivant le processus qui a conduit à cette solution dans le passé (i.e. les décisions prises, les justifications, etc.) ;
- **Adaptation hiérarchique** : Selon (Abou-Assali A., 2010), les cas sont organisés dans une hiérarchie selon leur niveau d'abstraction. L'adaptation est réalisée d'une manière descendante : la solution est, dans un premier temps, adaptée au plus haut niveau de la hiérarchie, et ensuite affinée en descendant dans la hiérarchie où d'autres informations sont prises en compte ;
- **Adaptation par composition** : Cette adaptation utilise deux ou plusieurs cas similaire remémorés pour effectuer l'adaptation convenablement. Un ensemble de règles d'adaptation relatif à ces cas est alors utilisé.

La recherche a évolué depuis et la littérature de RàPC a proposé beaucoup d'autre d'approches d'adaptation. Une bonne synthèse de ces approches figure dans (Badra F.,

2009). Nous avons examiné, réorganisé et consolidé cette synthèse moyennant le tableau 2.3.

Référence	Approche d'adaptation	Transformationnelle et dérivationnelle	Par décomposition			Combinaison de cas	Interactive	A partir de cas	Tâche de recherche	Par satisfaction de contraintes			un jeu entre espaces
			Décomposer le problème	différentielle	reformulations					satisfaction de contraintes	un problème d'optimisation	Par la critique	
(Carbonell J.G., 1983)		×											
(Purvis L. et Pu P., 1998), (Aamodt A., 2004)	COMPOSER - CREEK				×								
(Leake D.B. et al., 1997), (Kinley A., 2001)	DIAL					×	×						
(Leake D.B. et Powell J.H., 2008), (Cordier A., 2008)	WebAdapt - FrakaS et IakA					×							
(Wiratunga et al., 2002) (Craw et al., 2006)							×						
(Hanney, 1997) (Lieber, 2002)								×					
(Smyth et Keane, 1993) (Lieber, 1999)												×	
(d'Aquin M. et al., 2005 ; 2007)	CABAMAKA DzCBR		×										
(Craw et al., 2006)			×										
(Fuchs et al., 2006)				×									
(Cordier A., 2008)	IakA-NF			×	×								
(Lieber J. et al., 2008)					×								
(Badra F. et al., 2008)	Taaable				×								
(Smyth, F.C. et al., 1995)	IDIOM								×				
(Purvis L. et Pu P., 1998)	COMPOSER								×				
(Schmidt R. et Vorobieva O., 2005)									×				
(Stahl R. et Bergmann O., 2000)										×			
(Cordier A. et al., 2007)	FrakaS										×		
(Abou-Assali A., 2010)	COBRA				×								

Tableau 2.3. Énumération des approches d'adaptation

Pour chaque approche d'adaptation figurant dans le tableau 2.3, nous présentons dans le tableau 2.4, brièvement son principe et éventuellement quelques constatations :

<i>Approche d'adaptation</i>	<i>Principe</i>	<i>Constatations</i>
<i>Transformationnelle et dérivationnelle</i>	<p>→ Transformation : consiste à déterminer une modification à appliquer à la solution du cas source mémorisé.</p> <p>→ Dérivation : consiste à adapter la méthode utilisée pour obtenir cette solution.</p>	<p>☺ Pas de connaissances d'adaptation à l'avance</p>
<i>Par décomposition</i>	<p>→ Décomposer un problème en sous-problèmes plus simples. Il existe 3 méthodes :</p> <ul style="list-style-type: none"> - décomposer le problème - différentielle - reformulations 	<p>☺ Pas de connaissances d'adaptation à l'avance</p> <p>☺ Plusieurs variantes</p>
<i>Combinaison de cas</i>	<p>→ Plusieurs cas sources sont mémorisés pour un problème d'adaptation donné puis une solution est composée à partir des solutions attribuées aux cas sources.</p>	<p>☺ Plusieurs systèmes ont été proposés</p> <p>☹ Difficile à mettre en œuvre</p>
<i>Interactive</i>	<p>→ Impliquer l'utilisateur dans l'étape d'adaptation.</p> <p>→ L'intervention de l'utilisateur peut être utilisée sous plusieurs façons.</p>	<p>☺ Intervention de l'utilisateur</p> <p>☺ Beaucoup de systèmes de RàPC sont proposés.</p> <p>☹ Pas toujours fiables</p>
<i>A partir de cas</i>	<p>→ Appliquer un processus de raisonnement à partir de cas à la tâche d'adaptation elle-même.</p> <p>→ L'adaptation est alors réalisée en raisonnant sur un ensemble d'expériences passées d'adaptation.</p>	<p>☹ Requiert à l'avance des connaissances d'adaptation</p>
<i>Tâche de recherche</i>	<p>→ Cette recherche s'effectue par l'application d'opérateurs d'adaptation, qui sont des transformations effectuées dans l'espace des solutions.</p>	<p>☹ Requiert la recherche des opérateurs d'adaptation à partir de la mémoire de système</p>
<i>Par satisfaction de contraintes</i>	<p>→ Peut être vue comme la recherche dans un espace des solutions d'une solution Sol (cible) qui satisfasse un certain nombre de contraintes.</p>	<p>☺ Plusieurs variantes</p>
<i>Un jeu entre espaces</i>	<p>→ L'espace des mémorisations désigne un ensemble de renseignements entre le problème cible et adaptations désigne un ensemble d'opérateurs qui peuvent être utilisés pour transformer la solution Sol (source) du problème source mémorisé pour donner Sol (cible)</p>	<p>☹ Requiert à l'avance des connaissances d'adaptation</p>

Tableau 2.4. Comparaison des approches d'adaptation

Nous avons vu tout au long de ce paragraphe, les différentes approches d'adaptation présentes dans la littérature de RàPC. Nous avons repris et réorganisé la décomposition de proposée par (Wilke W. et Bergmann R., 1998) qui comprend quatre approches et une deuxième plus récente élaborée par (Badra F., 2009). Cette dernière couvre la plupart

d'approches d'adaptation. Il est à noter que chaque approche a ses propres spécificités et aucune approche n'est parfaite pour des différents domaines d'études.

2.3.5.3.2. Approches d'acquisition des connaissances d'adaptation

Selon (Badra F. et al. 2009), l'étape d'adaptation est considérée comme le talon d'Achille du RàPC car elle nécessite des connaissances spécifiques au domaine d'application qui sont assez souvent difficiles à acquérir. En revanche, pour les systèmes RàPC de synthèses (le cas de notre système), l'étape d'adaptation est incontournable. Selon Abou-Assali (Abou-Assali A., 2010), « le fait d'abandonner l'étape d'adaptation dans la plupart des systèmes peut s'expliquer, du point de vue ingénierie des connaissances, par la difficulté de modéliser le processus et les connaissances mises en œuvre dans le cycle de raisonnement en général et dans la phase d'adaptation en particulier ».

Peu d'études ont été consacrées à ce jour à l'analyse des différentes approches d'acquisition de connaissances d'adaptation (ACA) en RàPC. Une bonne synthèse de ces travaux figure dans (Badra F., 2009). Le tableau 2.5 présente les six approches d'acquisition de connaissances d'adaptation les plus connus :

Approche	Référence	Moment d'acquisition		Méthode d'acquisition			Composantes				
		Hors-ligne	En-ligne	Manuelle	Semi-automatique	Automatique	Algorithme d'apprentissage	Connaissances	Ingénieur de connaissances	Expert	Utilisateur
Acquisition manuelle auprès d'experts	(Lieber J. et al., 2004) (Lieber J. et al., 2007)	×		×				×	×	×	
Automatiques en-ligne	(Hammond K.J., 1986) (Badra F. et Lieber J., 2007)		×			×	×	×			×
Automatiques hors-ligne	(Hanney K. et Keane M.T., 1996) (Craw et al., 2006)	×				×	×	×			×
Semi-automatiques auprès d'experts	(d'Aquin M. et al., 2007) (Abou Assali A., 2010)	×			×		×	×		×	
Interactives auprès d'utilisateurs	(Leake D.B. et al., 1996) (Cordier A. et al., 2007)		×	×				×			×
Semi-automatiques auprès d'utilisateurs	(Leake D.B. et Powell, 2008) (Badra F. et al., 2009)		×		×		×	×			×

Tableau 2.5. Comparaison des approches d'acquisition de connaissances d'adaptation

- Les approches d'acquisition manuelle auprès de l'expert : s'effectuent hors ligne en présence d'un expert du domaine au cours des sessions d'acquisition de connaissances ;
- Les deux approches automatiques en-ligne et hors-ligne : consistent à appliquer un algorithme d'apprentissage automatique (généralement issu de la fouille de

données) pour obtenir de connaissances d'adaptation à partir de connaissances disponibles. L'expert du domaine est impliqué dans ce processus ;

- *Les approches semi-automatiques auprès d'experts* : Selon (Badra F., 2009), ces approches utilisent des techniques d'extraction de connaissances à partir de données pour générer des éléments de connaissances à partir de sources de données différentes. L'apprentissage est réalisé hors ligne ;
- *Les approches interactives auprès d'utilisateurs* : un algorithme d'apprentissage est déclenché en ligne pour acquérir les connaissances d'adaptation nécessaires à la résolution d'un problème donné. L'apprentissage est guidé par un ensemble de paramètres liés au problème à résoudre et notamment l'interaction avec l'utilisateur au cours de session de résolution de problème.

2.3.5.4. Révision

Au cours de la phase de révision, la solution proposée à l'issue de la phase d'adaptation sera évaluée. La révision consiste à continuer éventuellement l'élaboration de la solution cible en cas de besoin. Cependant, Il n'existe pas encore de méthode standardisée pour rendre compte de l'étape de révision dans le RàPC. Selon Mille (Mille A., 1999), l'évaluation concerne plusieurs actions pouvant être employées :

- tester la solution proposée dans le monde réel ;
- faire une introspection dans la base de cas en utilisant l'ensemble des descripteurs de problème et de solution afin de vérifier que les cas similaires ont donné entière satisfaction ;
- utiliser une autre méthode d'évaluation de la solution (simulateur, système expert classique, etc.).

2.3.5.5. Apprentissage

Très schématiquement, à la fin de chaque session de raisonnement, le nouveau cas résolu est stocké dans la base de cas. Selon (Cordier A. et al., 2006), « *ceci en résulte une augmentation graduelle de taille de la base de cas, ce qui met donc en évidence le besoin d'organiser et de maintenir la base de cas tout au long de la vie du système* ». Pour répondre à ce besoin, plusieurs travaux proposent des techniques d'indexation ou de classification pour faciliter la remémoration des cas. D'autres approches s'appuient sur des stratégies par exemple pour ne retenir que les cas les plus pertinents et éviter l'explosion de la taille de la base de cas. Selon (Ganascia J.G., 1990, 1991), « *le comportement idéal d'un système d'apprentissage serait celui qui, bouclant sur lui-même, améliorerait continuellement ses connaissances au contact des expériences auxquelles il est confronté* ».

Cette caractéristique des systèmes qui apprennent est qualifiée de "dimension itérative" ou encore de "circularité" du savoir : une connaissance induite peut être utilisée en vue d'apprendre de nouvelles connaissances. Selon (Hadj-Mabrouk H., 1992), « *un système d'apprentissage est dit incrémental s'il est doté de la capacité de faire évoluer les connaissances apprises au cours d'un cycle antérieur, sans devoir à chaque fois retraiter la totalité des exemples collectés* ». D'autres travaux qui ont été consacrés au maintien de la cohérence de la base de cas, nous pouvons citer les travaux de Haouchine (Haouchine, 2009) sur la maintenance de la base de cas à travers son approche de maintenance de la base de cas pour remédier à ce problème d'évolutivité.

2.4. Fouille de données et extraction de connaissances

La fouille de données ou encore la découverte de connaissances à partir de données est un processus qui utilise des techniques d'apprentissage automatiques pour analyser et extraire des connaissances à partir de grandes quantités de données. Nous pouvons distinguer deux familles de méthodes d'apprentissage : les méthodes supervisées et les méthodes non supervisées.

D'après (Scharff C., 2004), l'apprentissage supervisé est représenté généralement par un modèle inductif où l'apprenant considère une base d'exemples et infère l'appartenance d'un objet à une classe en considérant les similarités entre l'objet et les éléments de la classe. Les classes sont étiquetées initialement. Les techniques de classification, d'estimation ou de prédiction utilisent l'apprentissage supervisé. Tandis que l'apprentissage non supervisé consiste à la construction d'un modèle et découverte des relations dans les données en ayant pas recours à d'autres données. La segmentation, le regroupement (cluster) et les associations sont des méthodes d'apprentissage non supervisées. Il est très intéressant pour nous dans le cadre de notre approche d'avoir recours à ces méthodes et techniques issues de la fouille de données. Nous présentons à cet effet quelques-unes que nous utilisons dans notre démarche d'aide à la décision, les règles d'association, la méthode d'arbre de décision et notamment l'Analyse des Concepts Formels (ACF).

2.4.1. Les règles d'association

Selon (Malek M., 2008), le but des règles d'association est de découvrir **des relations ayant un intérêt** pour un expert de domaine entre deux ou plusieurs variables stockées dans de bases de données volumineuses. La technique de règles d'association est qualifiée de « **méthode d'apprentissage non supervisée** ». Pasquier (Pasquier, 2000) présente cette technique comme suit : Pour un ensemble F d'items fréquents extraits, le nombre de règles d'association valides pouvant être générées à partir de F est $\sum_{f \in F} 2^{|f|} - 2$. Malgré son intérêt, cette méthode souffre d'une limite principale. Cette limite concerne le nombre d'itemsets fréquents extraits et leur taille qui est élevée dans la plupart des cas. Ce nombre important constitue un problème majeur pour la pertinence des règles d'association extraites. Agrawal (Agrawal R., 1993) fut le premier à introduire le problème de découverte des règles associatives. Le domaine d'étude était l'analyse du panier de la ménagère peut être décomposé en deux sous problèmes :

- Trouver tous les itemsets fréquents, qui apparaissent avec une fréquence supérieure ou égale à un seuil défini par l'utilisateur (Ben yahya S. et Engelbert M.N., 2004), appelé min support ;
- Générer l'ensemble des règles d'association, à partir de ces itemsets fréquents, ayant une mesure de confiance supérieure ou égale à un seuil défini aussi par l'utilisateur, appelé min confiance. Un seuil minimal de support minSupp est fixé à partir duquel un ensemble d'items est dit fréquent.

La littérature propose plusieurs algorithmes de règles d'association (Tableau 2.6). Ces algorithmes ont été proposés dans la plus part des cas pour optimiser la phase d'extraction des itemsets fréquents (la phase qui nécessite une particulière attention). Les algorithmes d'extraction des itemsets fréquents parcourent le treillis des itemsets fréquents « par niveau » d'une manière itérative : ils réalisent un parcours en largeur du treillis des itemsets fréquents, du bas vers le haut et déterminent dans chaque itération les itemsets fréquents pour un niveau. Les opérations d'entrée sortie sont généralement très coûteuses en temps d'exécution processeur CPU. Pasquier (Pasquier N., 2000) considère que ces balayages représentent une part importante des temps de l'extraction pour la génération des règles

d'association. Pour chaque itération k , un ensemble de k -itemsets fréquents candidats est généré et leur supports sont calculé ce qui permet de limiter le nombre total de balayage réalisés. L'efficacité et le temps de réponse de l'algorithme sont deux critères étroitement liés au nombre de balayage du jeu de données et du nombre d'itemsets fréquents considérés.

<i>Référence</i>	<i>Algorithme</i>	<i>Principe</i>
(Agrawal R. et Srikant R., 1993)	Apriori	Algorithme itératif de recherche des itemsets fréquents par niveaux.
(Agrawal R. et Srikant R., 1994)	AprioriTID	Il permet la réduction de temps d'extraction des itemsets fréquents.
(Savasere A. et al., 1995)	Partition	Après le premier balayage du contexte, il utilise des listes d'itemsets associées identifiants des objets en relation avec les itemsets afin de diminuer progressivement la taille du contexte.
(Taoivonen H., 1996)	Sampling	Utilise les techniques d'échantillonnage pour extraire les itemsets fréquents et vérifie leur supports en réalisant un balayage de l'ensemble du contexte.
(Brin S. et al., 1997)	DIC	Il permet d'améliorer l'efficacité de la recherche des itemsets fréquents par niveaux en diminuant le nombre de balayages du contexte réalisé.
(Lin Y.C., 1998)	Princer-Search	Extension de l'algorithme Apriori. La génération des itemsets candidats est étendue afin de générer un ensemble d'itemsets maximaux fréquents utilisés pour le parcours du haut vers le bas.
(Zaki M.J. et al., 1997)	MaxEclat	- MaxEclat est basé sur les classes d'équivalence des itemsets.
(Zaki M.J. et al., 1998)	MaxClique	- Maxclique utilise la même démarche mais basé sur les cliques maximales des hypergraphes uniformes.
(Bayardo R.J., 1998)	Max-Miner	Il effectue simultanément une recherche du bas vers le haut et une recherche par niveaux dont le résultat pour chaque itération k .
(Pasquier N. et al., 1999)	Close A-close	La génération des itemsets fréquents est étendue afin de générer un ensemble d'itemsets maximaux fréquents utilisés pour le parcours du haut vers le bas.
(Gasmi G. et al., 2005)	IGB	Un algorithme pour générer des règles d'association assez informatives et pertinentes pour l'utilisateur.
(Fournier P.V. et al., 2012)	TNR	Générer des règles d'association (top-k rules) non redondants.

Tableau 2.6. Algorithmes de génération de règles d'association

Les algorithmes **Apriori**, **AprioriTID**, **Partition**, **Sampling** et **DIC** sont développés pour des applications se basant sur des bases de données commerciales, dans la plus part des cas, ces données éparses et faiblement corrélés ce qui explique le temps d'exécution faible par rapport aux autres algorithmes. Par ailleurs, si le nombre d'itemsets (noté μ) augmente, les performances des algorithmes se dégradent considérablement. L'algorithme fondateur **Apriori** met μ itérations (balayages) pour extraire les itemsets fréquents. Il met un temps énorme si μ augmente davantage. Pour **AprioriTid**, le temps d'exécution est considéré important lors des premières itérations (car il sollicite le disque pour le stockage

des listes associées aux itemsets). Pour les deux algorithmes **Partition** et **Sampling**, le nombre de k -itemsets fréquents pour $1 < K < \mu$ est supérieur à celui d'**Apriori**. Ils présentent deux problèmes : le premier lié à la quantité de stockage (relativement supérieur à celle d'**Apriori**) et au temps de calcul qui est important parce qu'ils impliquent les itemsets inféquents (une partie considérée sans utilité). L'algorithme **DIC** présente aussi deux limites, une liée à la quantité de l'espace de stockage occupée pour l'ensemble d'itemsets fréquents traités simultanément et une autre au coût de calcul des supports des candidats considérée relativement supérieur à celui d'**Apriori**.

Généralement, les bases de données caractérisées par une valeur de μ élevée représentent une part importante dans les bases de données réelles de marché. Les données denses corrélés sont donc caractérisées par une valeur de μ élevée. Il est nécessaire de développer d'autres algorithmes qui tiennent en compte cette caractéristique et la valeur élevée de μ . Les algorithmes **Princer-Search**, **MaxEclat**, **MaxCliques**, **Max-Miner**, **Close/A-close** et plusieurs autres ont été développés pour tenir en compte ceci. En effet, ils réduisent le nombre total des itérations nécessaires à l'extraction des items fréquents quant μ est considérée importante. Ils réduisent également le temps de calcul processeur (par exemple : si un itemset max est identifié, ses sous-ensembles ne sont plus considérés).

L'algorithme **MaxMiner** est globalement plus performant que **Princer-Search**, **MaxEclat**, **Maxcliques** et **Close/A-close** en termes de temps d'exécution. **Princer-Search** met un temps de calcul processeur très important dans la génération des itemsets fréquents ce qui met en cause le temps de réponse de l'algorithme. **Max Eclat** présente une limite liée au manque de précision des itemsets fréquents maximaux qu'il utilise. Il doit faire davantage d'itérations que les deux algorithmes **MaxMiner** et **Princer-Search** pour résoudre ce problème. **Maxcliques** est soumis à un problème de performance dû aux temps de calcul requis pour la génération des itemsets fréquents maximaux d'où une dégradation de temps de réponse. **Max Mine** présente une alternative pour l'algorithme **Close/A-close**, il présente un avantage dans la réduction de temps d'extraction des items fréquents dans le cas où μ élevée mais il est soumis de même à deux problèmes : il nécessite une bonne quantité de mémoire et difficulté à extraire les règles d'association quand il s'agit de données denses et fortement corrélées. Ces dernières années ont été marquées par un effort algorithmique "monomaniaque" pour la réduction du temps de calcul de l'étape d'extraction des motifs intéressants en dépit de la qualité des résultats proposés.

L'algorithme **IGB** fut proposé en 2005 par Gasmi (Gasmi G. et al., 2005), son objectif est de créer une base de règles d'association en utilisant l'analyse formelle des concepts pour véhiculer le maximum d'informations utiles et éviter ainsi le problème de la non sémantique des résultats produits. Cet algorithme nécessite donc un prétraitement à l'avance ce qui met en évidence le temps global mis en œuvre pour générer les résultats.

De même pour l'algorithme **TNR** (*Top-k Non-redundant rules*) proposé par (Fournier, P.V. et Vincent, S.T., 2012), l'objectif est de conserver l'aspect sémantique des règles produites en ne gardant que celles qui présentent une utilité pour l'utilisateur. Cet algorithme récemment développé présente une solution pour résoudre les règles redondantes issues de l'algorithme **top-k** association rules proposé par le même auteur. L'expérimentation donnée par l'auteur indique de même que **DNR** est moins performant que **top-k rules** bien qu'il présente une sémantique non négligeable.

2.4.2. La méthode d'arbre de décision

Selon (Belacel N., 2000), l'arbre de décision représente une méthode d'apprentissage **supervisée** et une structure qui représente des ensembles de décisions. Ces décisions génèrent des règles pour la classification d'un ensemble de données. À nos jours, il existe

un très grand nombre d’algorithmes de génération d’arbre de décision (**ID3**, **C4.5**, **Exhaustive CHAID** etc.). Cependant, il n’existe pas de méthode qui soit dans la pratique automatiquement plus performante. Ces algorithmes se distinguent par le ou les critères de segmentation utilisés, par les méthodes d’élagages¹ implémentées. Nous avons recensé dans le tableau 2.7, les principaux algorithmes proposés pour la communauté scientifique.

Tableau 2.7. Algorithmes de génération d’arbre de décision

Référence	Algorithme	Principe
(Quinlan R., 1979)	ID3	L’algorithme ID3 a été développé à l’origine par Ross Quinlan. publié dans le livre « Machine Learning » en 1986. Il se base sur des exemples déjà classés dans un ensemble de classes pour déterminer un modèle de classification.
(Kass G., 1980)	CHAID	CHAID est la variante supervisée. Il se diffère de deux manières : le critère de segmentation ; les feuilles produites par un partitionnement peuvent être fusionnées
(Cestnik B. et al., 1987)	ASSISTANT 86	ASSISTANT 86 dérive de ID3. Il introduit un certain nombre d’améliorations destinées à surtout guider l’induction.
(Quinlan R., 1993)	C4.5	Il est basé sur l’algorithme ID3 auquel il apporte plusieurs améliorations. Il se base sur une mesure de l’entropie dans l’échantillon d’apprentissage pour produire le graphe d’induction.
(Rakotomalala R. et al., 1999)	Improved C4.5	C’est une généralisation de C4.5. Cet algorithme introduit les entropies généralisées de type beta pour le calcul du gain ratio.
(Chauchat J.H. et Rakotomalala R., 2001)	Cost sensitive C4.5	Cet algorithme représente une variante de C4.5. Il intègre explicitement la matrice de coûts de mauvais classement lors de l’exploration de l’espace de solutions.
(Rakotomalala R., 2005)	Improved CHAID	Cet algorithme est directement dérivé de CHAID. Il lui apporte quelques améliorations

En apprentissage automatique, la plupart des travaux s’appuient sur la théorie de l’information. Il est d’usage de citer l’algorithme **ID3** de Quinlan (Quinlan R., 1979). **ID3** choisit l’attribut de coupure S, qui conduit à une partition de D en s feuilles dont le Gain d’Information correspondant est maximum. Son approche a pris un tournant important dans les années 90 lorsqu’il a introduit la méthode **C4.5** qui est une référence incontournable. L’algorithme **C4.5** présente des nettes améliorations pour **ID3**, il a permis le traitement de données manquantes, la prise en charge des attributs continus, etc. **C4.5** est considéré comme référence au sein de la communauté. Par rapport aux autres algorithmes, **C4.5** à tendance à produire des arbres plus grands, plus profonds quand il s’agit d’un effectif important. Son inconvénient est que pour préserver l’efficacité de l’apprentissage et la pertinence du modèle produit, les variables continues doivent être discrétisées avant la mise en œuvre de l’algorithme.

L’algorithme **Assistant 86** proposé par (Cestnik B. et al., 1987) présente aussi quelques améliorations à l’algorithme **ID3**. L’arbre induit est nécessairement binaire. Cet algorithme cherche la combinaison qui maximise le gain d’entropie lors de la segmentation. Il présente un avantage car il introduit un nombre de paramètres pour contrôler la taille de l’arbre. L’algorithme **CHAID** (CHi-squared Automatic Interaction Detector) proposé par Kass

¹ L’élagage consiste à orienter ou limiter le développement d’un arbre

(Kass G., 1980), est une variante supervisée. Il évite l'élaboration d'arbres trop « larges » dans le processus de fusion. Ce point peut être considéré comme avantage. L'échantillonnage dans **CHAID** doit être suffisamment large de manière à ce que la taille de chaque groupe ne devienne pas trop petite, ce qui peut rendre l'analyse peu fiable. L'algorithme **Cost sensitive C4.5** proposé par (Chauchat J.H. et Rakotomalala R., 2001) permet de tenir en compte des coûts durant les deux phases : l'expansion de l'arbre (cette phase n'est pas déterminante dans **C4.5**) et le post élagage. L'algorithme **Improved CHAID** proposé par Rakotomalala (Rakotomalala R., 2005) basé sur **CHAID**, apporte quelques améliorations et introduit des paramètres supplémentaires pour mieux contrôler la taille de l'arbre.

2.4.3. Méthode d'Analyse des Concepts Formels

Nous introduisons dans ce paragraphe le formalisme de l'Analyse des Concepts Formels (ACF). Selon R. Wille (Wille R., 1982), l'ACF est une théorie basée sur la théorie des treillis (Birkhoff G., 1950) et des treillis de Galois (Barbut M. et al., 1970), qui permet de regrouper des individus qui partagent des propriétés communes. Les groupes d'individus et de propriétés mutuellement correspondants sont appelés des concepts formels (Birkhoff G., 1950). Ces concepts sont organisés en une hiérarchie, dite treillis de concepts reposant sur la relation d'inclusion ensembliste entre groupes d'individus et de propriétés.

Un concept formel regroupe deux parties :

- **L'extension** qui contient les individus appartenant au concept ;
- **L'intension** qui contient les propriétés partagées par les individus.

Ces concepts construisent un treillis de concepts. Un contexte formel (notion principale de l'ACF) est représenté par un triplet $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{J})$,

- \mathcal{G} est un ensemble fini d'objets ;
- \mathcal{M} est un ensemble d'attributs ;
- $\mathcal{J} \subseteq \mathcal{G} \times \mathcal{M}$: une relation binaire entre \mathcal{G} et \mathcal{M} . Un couple $(g, m) \in \mathcal{J}$ signifie que l'objet $g \in \mathcal{G}$ possède l'attribut $m \in \mathcal{M}$.

Les concepts formels d'un contexte formel \mathbb{K} sont identifiés. Pour tout $A \subseteq \mathcal{G}$ et $B \subseteq \mathcal{M}$, nous avons :

$$A' = \{m \in \mathcal{M} \mid \forall g \in A, (g, m) \in \mathcal{J}\}$$

$$B' = \{g \in \mathcal{G} \mid \forall m \in B, (g, m) \in \mathcal{J}\}$$

L'ensemble des attributs partagés par les objets de A est A' . B' est l'ensemble des objets ayant tous les attributs B . Pour tout $A_1, A_2, A \subseteq \mathcal{G}$ et $B_1, B_2, B \subseteq \mathcal{M}$:

$$1) A_1 \subseteq A_2 \Rightarrow A_2' \subseteq A_1' \qquad 1') B_1 \subseteq B_2 \Rightarrow B_2' \subseteq B_1'$$

$$2) A \subseteq A'' \Rightarrow A' \subseteq A''' \qquad 2') B \subseteq B'' \text{ et } B = B'''$$

$$A \subseteq B' \Leftrightarrow B \subseteq A'$$

Un couple $\mathbb{C}(A, B)$ est un concept formel dans un contexte formel $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{J})$, si et seulement si : $A \subseteq \mathcal{G}, B \subseteq \mathcal{M}, A' = B$ et $A = B'$ (i)

- $ext(\mathbb{C})$: contient tous les objets appartenant au concept
- $int(\mathbb{C})$: contient tous les attributs partagés par ces objets

Le support de l'intension B du concept formel est défini par :

$$support(B) = \frac{|B'|}{|\mathcal{G}|} = \frac{|A|}{|\mathcal{G}|}$$

Avec $|A|$ est la cardinalité de l'extension A . Les concepts formels de \mathbb{K} sont ordonnés partiellement par une relation de subsomption, notée \leq^c , telle que : $(A_1, B_1) \leq^c (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \ (\Leftrightarrow B_2 \subseteq B_1)$

L'ensemble ordonné de tous les concepts formels de $(\mathcal{G}, \mathcal{M}, \mathcal{I})$, est dénoté par $\mathfrak{B}(\mathcal{G}, \mathcal{M}, \mathcal{I})$, et appelé le treillis des concepts (ou le treillis de Galois) de $(\mathcal{G}, \mathcal{M}, \mathcal{I})$.

Exemple : Soit le contexte formel : $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$, avec $\mathcal{G} = \{Attr_1, Attr_2, Attr_3, Attr_4, Attr_5, Attr_6\}$ et $\mathcal{M} = \{Obj_1, Obj_2, Obj_3, Obj_4, Obj_5\}$

\mathcal{G}/\mathcal{M}	$Attr_1$	$Attr_2$	$Attr_3$	$Attr_4$	$Attr_5$	$Attr_6$
Obj_1	×	×		×		
Obj_2		×	×			×
Obj_3		×			×	×
Obj_4	×	×	×	×		
Obj_5	×	×	×		×	×

Tableau 2.8. Exemple d'un contexte formel

Dans le tableau 2.8, la présence d'une croix (×) indique qu'un objet possède un attribut. Le diagramme dans la figure 2.20 montre le treillis des concepts ordonnés par la relation de subsomption \leq^c . Le nœud étiqueté par l'attribut $Attr_6$ représente le concept formel $(\{Obj_2, Obj_3, Obj_5\}, \{Attr_2, Attr_6\})$.

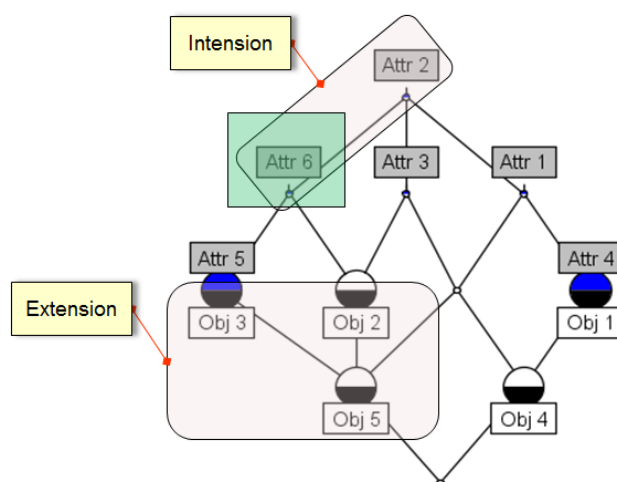


Figure 2.20. Concept formel de l'attribut $Attr_6$.

Le nœud étiqueté par l'attribut $Attr_2$ représente le concept formel $(\{Obj_1, Obj_2, Obj_3, Obj_4, Obj_5\}, \{Attr_2\})$.

Le contexte formel présenté sont des contextes uni-valués car les attributs prennent des valeurs binaires (1,0). Nous parlons d'un contexte formel multi-valué quand les valeurs d'attributs sont variées. Selon (Ganter B. et Wille R., 1997), pour construire le treillis d'un tel contexte, il faut d'abord le transformer en un contexte uni-valué. La transformation se fait par une mise à l'échelle conceptuelle : chaque attribut $m \in \mathcal{M}$ est remplacé par un ensemble d'attributs qui correspondent aux différentes valeurs de m . Ferré et Ridoux (Ferré et Ridoux, 1999) ont proposé une généralisation de l'ACF en transformant les attributs par des expressions logiques à travers leur extension : l'analyse des concepts logiques (ACL). Ferré et Ridoux (Ferré S. et Ridoux O., 1999) considèrent \mathcal{L} est une logique propositionnelle dont la relation de déduction est \models , et dont les opérations de disjonction et de conjonction

sont respectivement \vee et \wedge , i est une fonction de \mathcal{G} dans \mathcal{L} qui associe à chaque objet une formule qui décrit l'intension de l'objet.

2.5. Synthèse et choix méthodologiques

Après avoir étudié les différents concepts relatifs aux ontologies, le RàPC et notamment les méthodes/techniques issues de la fouille de données, il convient dans ce dernier paragraphe de synthétiser nos choix méthodologiques effectués.

Tout d'abord, le fait de coupler le RàPC et les ontologies et d'associer aussi les techniques issues de la fouille de données est d'un apport indéniable pour les problèmes liés à l'ingénierie, le management de connaissances et l'aide à la décision, surtout quant il s'agit de faire face à un domaine à haut risque, comme celui du transport ferroviaire.

Le RàPC est une méthode de raisonnement par analogie qui favorise la réutilisation des connaissances déjà capturées, formalisée et structurées. De ce point de vue, son exploitation pour le domaine de l'analyse des accidents ferroviaires ne pourrait être que bénéfique. En effet, la réutilisation des connaissances issues de ces accidents passés contribue à éviter l'apparition de nouveaux accidents et des nouvelles situations d'insécurité.

En termes plus précis, une ontologie est une structure hiérarchique pouvant décrire les connaissances issues de l'analyse des risques et de retour d'expérience (Rex) du transport ferroviaire (les scénarios d'accidents, en particulier). En effet, l'intérêt des ontologies provient de leur capacité à structurer, homogénéiser les terminologies utilisées. Ceci, pourrait remédier au problème relatif au recours à plusieurs intervenants (États, organismes, experts analystes, experts certifieurs etc.) pour l'analyse de la sécurité. Le RàPC à son tour, présente un processus complet de résolution de problèmes, pouvant s'appuyer sur cette structure solide offerte par l'ontologie. Ce processus devrait permettre, en tenant compte de certains critères, d'assister un décideur (expert de la sécurité, analyste, utilisateur etc.) à bien comprendre une situation risquée donnée et de le stimuler pour enfin proposer des solutions possibles aux nouvelles situations d'accidents. D'autre part, la fouille de données permet à son tour, de consolider le processus du RàPC en permettant la déduction d'autres connaissances implicites intéressantes issues de l'analyse des bases de données du Rex ferroviaire, par conséquent, améliorer tout le processus d'aide à la décision mené.

Notre objectif dans le cadre de ce travail de thèse, est la proposition d'une approche générique d'aide à la décision pour l'amélioration de la sécurité des transports. Pour atteindre cet objectif, nous caractérisons et choisissons plusieurs aspects des concepts que nous venons de présenter tout au long de ce présent chapitre. Les choix qui ont été faits sont résumés dans les sections suivantes.

2.5.1. Côté ontologique

Dans le cadre de notre contexte d'étude, nous avons convenu tout d'abord de choisir la mise en œuvre d'une ontologie noyau, qui regroupe les concepts fondamentaux pour **décrire un scénario d'accident. Cette ontologie servira comme un point d'accès pour l'ontologie de domaine que nous avons proposé. En effet, la formalisation des connaissances de sécurité manque énormément de cohérence et la terminologie utilisée dans la plupart des cas est ambiguë. L'utilisation d'une ontologie de domaine permettant de bien définir le vocabulaire, d'harmoniser le lexique et la description de connaissances de domaine s'annonce indispensable.**

Pour la construction de l'ontologie, nous avons convenu de recourir à l'approche proposée par (Abou-Assali A., 2010) qui s'inspire largement de *Methontology de* (Fernandez M., 1997). En effet, cette méthodologie est considérée comme la plus mature et

la plus détaillée parmi d'autres méthodologies présentées dans une étude réalisée par (Fernandez M., 1999) et plus tard par (Mendes O., 2003). Elle est aussi très utilisée dans les travaux présents dans la littérature.

Pour le calcul de la similarité, nous utilisons deux types de mesures : une similarité **basée-concept** et une **similarité basée-slot**. En effet, un cas d'accident comporte deux groupes d'attributs. Des attributs dits « simples » pour lesquels nous cherchons à comparer leur valeur uniquement d'où le choix de la mesure basée slot de (Zhang K. et al., 2006). D'autres attributs dits « complexes » pour lesquels nous extrayons les correspondances des concepts supérieurs ou et concepts fils dans la hiérarchie des concepts de l'ontologie. Il s'agit plus précisément de faire appel à un calcul de profondeur d'où le choix sollicité de la mesure basée-concept défini par (Wu Z. et Palmer M., 1994) et améliorée dans (Zargayouna H. et Salotti S., 2004).

2.5.2. Côté RàPC

Nous résumons les différents aspects de RàPC que nous retenons pour notre démarche d'aide à la décision proposée.

Dans notre approche, nous avons eu recours à un cycle de raisonnement de cinq phases (l'élaboration, la remémoration, l'adaptation, la révision et l'apprentissage). Le modèle qui correspond le mieux à notre cadre de travail est le **modèle structurel** partant de fait qu'une base d'exemple (un échantillon d'apprentissage) et des caractéristiques décrivant les cas d'insécurité sont disponibles à l'avance. Il s'agit pour nous de chercher les mesures préventives/correctives plus généralement la solution convenable à admettre pour un cas d'accident donné, pour éviter ainsi sa reproduction dans le futur. **Le système de RàPC que nous retenons est hybride. Il est un système de synthèse** dans la mesure où la solution d'un cas similaire au cas cible (un cas d'accident dont nous cherchons une solution) ne peut pas être toujours prise directement pour résoudre un problème courant.

L'étape d'adaptation étant une étape délicate, elle nécessite pour nous une attention particulière pour bien mener tout le processus de raisonnement vers ses fins utiles. **Notre système de RàPC peut être considéré aussi comme un système de classification** car nous avons affaire dans notre travail à une tâche d'analyse et de classification du cas cible avant de procéder l'adaptation.

Dans la phase d'adaptation, il s'agit d'utiliser l'ensemble des cas similaires au cas cible saisi pour adapter et proposer une éventuelle solution. Des connaissances d'adaptions liées à chaque cas remémoré vont être utilisées pour assurer ce processus d'adaptation. Notre approche d'acquisition de connaissances d'adaptation retenue donc, est **semi-automatique auprès d'experts**. Le choix de l'approche d'adaptation dépend de la complexité de l'application et la nature des connaissances modélisées en termes de tâche à réaliser, de structure des cas, de solutions espérées, etc. Dans notre travail, nous avons choisi **une adaptation par substitution** (un jeu entre espace), vu la nature des cas et des solutions que nous cherchons à fournir. Nous utilisons des règles d'adaptation, qui sont des transformations ou un enrichissement à effectuer dans la partie problème du cas cible pour cibler une solution. Toutefois, une présence d'un expert est souhaitable pour bien contrôler tout le processus d'adaptation. Il peut éliminer les règles d'adaptation non autorisées présentes et ceci dans le but de transférer son expertise pour améliorer la qualité des connaissances produites.

2.5.3. Méthodes et techniques de fouille de données

Dans cette section, nous décrivons et justifions les techniques et les algorithmes issus de la fouille de données que nous utilisons dans notre approche.

2.5.3.1. Règles d'association

Nous avons convenu d'utiliser les règles d'association dans notre démarche d'aide à la décision. Cette aide se présente au moment de la caractérisation du problème (le cas cible). Elle se présente sous forme de concepts associés qui peuvent être utiles pour l'utilisateur. Dans le cas où la base d'exemples (les scénarios accidents) est faible c'est à dire se situer dans la première famille d'algorithme où μ le nombre d'itemsets n'est pas assez élevé (voir section 2.4.1), l'algorithme Apriori peut être considéré le meilleur choix pour nous.

Comparé aux autres performances d'algorithmes comme Partition, Sampling et DIC, Apriori peut être considéré comme un algorithme efficace, il donne des temps de réponse acceptables dans la majorité des cas et consomme moins de temps processeur avec un faible espace de stockage occupé. Il est implémenté également dans la majorité des outils de fouilles de données proposés pour la communauté scientifique.

2.5.3.2. Règles de classification

L'étude effectuée sur les méthodes de classification nous a orienté à choisir la méthode **d'arbre de décision**. L'objectif de cette méthode est de produire des groupes d'individus les plus homogènes possibles ayant la même variable à prédire.

Nous utilisons cette méthode dans notre démarche d'aide à la décision pour détecter la classe d'appartenance du cas cible pour opérer ensuite la recherche des cas similaires uniquement dans la classe repérée. La méthode d'arbre de décision est arrivée à maturité. En effet, ses caractéristiques, ses points forts et ses points faibles sont maintenant bien connus. Nous pouvons citer quelques arguments qui ont sollicité notre choix :

- La simplicité et l'efficacité de cette méthode ;
- Les concepts que nous pouvons utiliser sont disjonctifs ;
- La représentation restrictive (attribut-valeur) ;
- Un résultat essentiellement visuel et facilement interprétable ;
- **Robuste** au bruit et aux valeurs manquantes ;
- Les attributs apparaissent dans l'ordre de **pertinence** ;
- **Outils disponibles** dans la plupart des environnements de fouille de données.

A l'instar de tous ces avantages, la méthode d'arbre de décision présente également quelques inconvénients qui concernent plus précisément, l'instabilité dans certains cas, la sensibilité de certains algorithmes au nombre de classes élevé causant ainsi une certaine dégradation des performances et également la nécessité de générer un nouveau classeur une fois les données en entrée ont subi des mises à jour ou une évolution dans le temps.

En résumé, nous pouvons conclure qu'en absence d'information a priori sur le problème et le domaine à traiter, il n'y a pas d'algorithme plus performant à un autre. Pour notre choix, nous avons convenu d'utiliser l'algorithme C4.5 proposé par (Quinlan R., 1993). Cet algorithme est implémenté également dans tous les outils de fouilles de données et ce choix sera consolidé dans l'expérimentation que nous allons effectuer.

2.5.3.3. Règles d'adaptation

Pour acquérir les connaissances d'adaptations (les règles d'adaptation), nous nous appuyons sur l'Analyse des Concepts Formels (ACF) appelée aussi Analyse Formelle de Concepts (AFC), et son extension l'Analyse des Concepts Logiques (ACL).

L'ACF est une méthode de classification symbolique qui vise à découvrir tous les regroupements possibles d'éléments ayant des points en commun. Ceci, peut être très bénéfique pour générer une sorte de règles de d'adaptation pouvant conduire à une solution

possible. Cette méthode a été utilisée dans plusieurs travaux et approches d'adaptation : Cabamaka (d'Aquin S. et al. 2007), (Badra F. et al. 2007) et COBRA (Abou-Assali A., 2010), (Abou-Assali A. et al., 2013).

Selon (Nizar M, 2009), l'ACF facilite la découverte des sources de données pertinentes et la structure du treillis obtenue (par application de l'ACF), facilite la visualisation des relations sous-jacentes entre les données et l'interprétation de ses relations. Ceci va nous aider à appliquer des règles d'adaptation pour converger vers une solution au cas cible (le problème dont nous cherchons la solution).

2.6. Conclusion

Tout au long du présent chapitre, nous avons étudié les concepts de base que nous utilisons pour élaborer notre approche d'aide à la décision pour l'amélioration de la sécurité des transports.

Nous avons essayé d'être exhaustif à travers cette étude bibliographique sur le RàPC, les ontologies et notamment les méthodes de la fouille de données que nous avons adoptées. En effet, l'apparition des ontologies comme une nouvelle tendance de modélisation de connaissances, s'est fait clairement dans le contexte de l'ingénierie de connaissances. Par ailleurs, il est aussi, à préciser qu'elle représente un intérêt indéniable dans notre contexte d'étude. Elle permet ainsi l'harmonisation de la terminologie et le lexique utilisé dans la représentation de connaissances de sécurité. Nous avons présenté les définitions de l'ontologie, les rôles, les typologies, les composantes de base et notamment les méthodologies de construction. Nous avons abordé par la suite, les modes de représentation de connaissances, les standards web pour la représentation des ontologies et les différents éditeurs et environnement de développement. Nous avons décrit finalement les fondements de base pour la conception des ontologies et les différentes approches d'évaluation de la similarité sémantique appropriées. Dans la deuxième partie de ce chapitre, nous avons exploré les principaux systèmes de RàPC, les tâches, les modèles, les architectures proposés dans la littérature et notamment les composantes essentielles (vocabulaire, base de cas, mesures de similarité, connaissances d'adaptation etc.). La troisième partie a été consacrée à l'étude des méthodes de la fouille de données à savoir, les règles d'associations, l'arbre de décision et l'analyse des concepts formels. Nous avons débouché à la fin de ce chapitre sur le choix des différents aspects de l'ontologie, du RàPC et notamment les algorithmes que nous utilisons pour élaborer toute la démarche d'aide à la décision.

Dans le prochain chapitre, nous allons étudier les différentes approches d'aide à la décision se basant sur le RàPC et les ontologies et notamment celles appliquées dans le domaine de la sécurité des transports.

Chapitre 3 : Approches d'aide à la décision se basant sur le RàPC et les ontologies

3.1. Introduction

De nos jours et même depuis quelques années, beaucoup de chercheurs ont ressenti la nécessité de faire émerger les méthodes d'acquisition de connaissances et d'apprentissage automatique, et ceci pour instaurer des démarches d'aide à la décision pouvant combler les limites et les lacunes rencontrées en ayant recours aux méthodes de résolution de problème conventionnelles. La littérature propose plusieurs approches qui combinent ces méthodes. Le raisonnement à partir de cas (RàPC) et les ontologies n'échappent pas à cette règle. En effet, les systèmes de RàPC se basant sur les ontologies ont montré leurs faisabilités et leurs apports incontournables pour cerner les problèmes de l'extraction, de la représentation et de l'exploitation de connaissances.

Dans le présent chapitre, nous étudions tout d'abord les différentes approches de RàPC qui se sont basées d'une façon ou d'une autre sur les ontologies. Nous examinons à cet effet, les développements récents survenus dans ce domaine. Dans la deuxième partie de ce chapitre, nous allons explorer les travaux liés à notre domaine d'application, la sécurité des transports, et plus globalement la sécurité industrielle. Notre étude va déboucher au final sur une synthèse de ces travaux pour montrer certaines limites et carences dans le but de bien situer notre approche et notamment les éventuelles contributions attendues de toute la démarche d'aide à la décision proposée.

3.2. Approches proposées dans la littérature

La littérature propose un nombre important d'approches de RàPC qui intègrent et utilisent les ontologies de différentes manières. Pour bien faire émerger aussi bien les invariants que les différences, nous allons étudier ces travaux pour prendre en compte les systèmes déjà proposés. Les différentes approches vont être présentées dans un ordre chronologique.

3.2.1. Travaux d'Aamodt

Aamodt (Aamodt A., 1991) a proposé une approche qualifiée de « *knowledge-intensive* ». Cette approche se base sur la base de cas et d'autres types de connaissances liées au domaine d'application. Aamodt (Aamodt A., 2004) présente également une comparaison entre les approches *knowledge-intensive*, à titre d'exemple le système Creek de (Aamodt, 1991). Les ontologies sont utilisées dans les systèmes *knowledge-intensive* pour structurer les connaissances. L'architecture de Creek se base sur un **modèle structurel de RàPC ayant une tâche de synthèse**, il propose un couplage assez fort entre les connaissances des cas et celles du domaine. Creek contient trois types de connaissances décrites par un réseau sémantique (Aamodt A., 2004) (Figure 3.1) :

- une ontologie générique ;
- une ontologie de domaine ;
- les cas antérieurs ;

Les connaissances des cas sont ainsi plus riches, car les descripteurs des cas correspondent à des nœuds (concepts ou instances) du réseau sémantique.

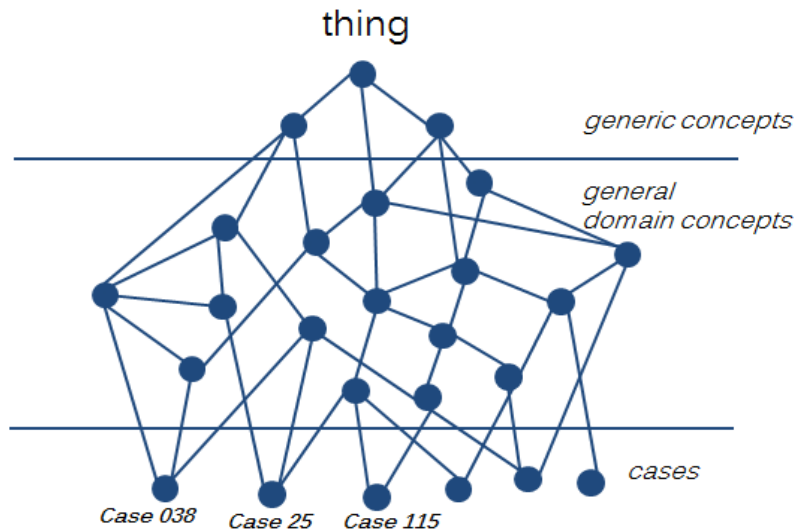


Figure 3.1 Les types des connaissances dans Creek (Aamodt A., 2004)

3.2.2. Travaux de Diaz-Agudo et González-Calero

Díaz-Agudo et González-Calero (Díaz-Agudo B. et González-Calero P.A., 2000) à leur tour ont présenté dans CBR_{Onto}, un système de RàPC qui utilise **une ontologie de tâche/méthode** qui fournit le vocabulaire nécessaire pour décrire les éléments impliqués dans **les processus de raisonnement**. Le système **CBR_{Onto}** présente un avantage car il permet d'intégrer le processus de raisonnement sur les connaissances d'un domaine sous forme d'**une ontologie**. Díaz-Agudo et González-Calero (Díaz-Agudo B. et González-Calero P.A., 2002) ont proposé une architecture **indépendante du domaine**, basée sur CBR_{Onto}, qui permet l'intégration d'ontologies dans les applications de RàPC se basant sur un **modèle Conversationnel et ayant une tâche de classification**. L'objectif consiste à élaborer des systèmes qui combinent des connaissances spécifiques aux cas avec des modèles génériques des connaissances du domaine. La figure 3.2 illustre la communication entre les différentes parties de cette architecture.

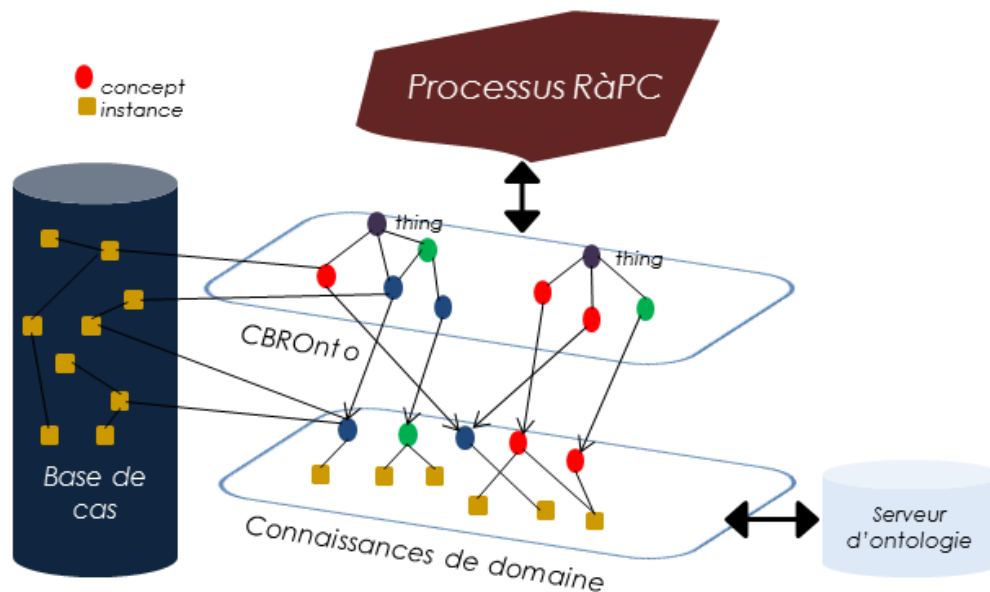


Figure 3.2. Système de RàPC utilisant CBR_{Onto} (Díaz-Agudo B. et González-Calero P.A., 2000).

3.2.3. Travaux de Fushs et Mille

Fuchs et Mille (Fuchs B. et Mille A., 2005) ont proposé quatre modèles de connaissances dans les systèmes de RàPC (se basant sur modèle structurel et une tâche de synthèse) (Figure 3.3) :

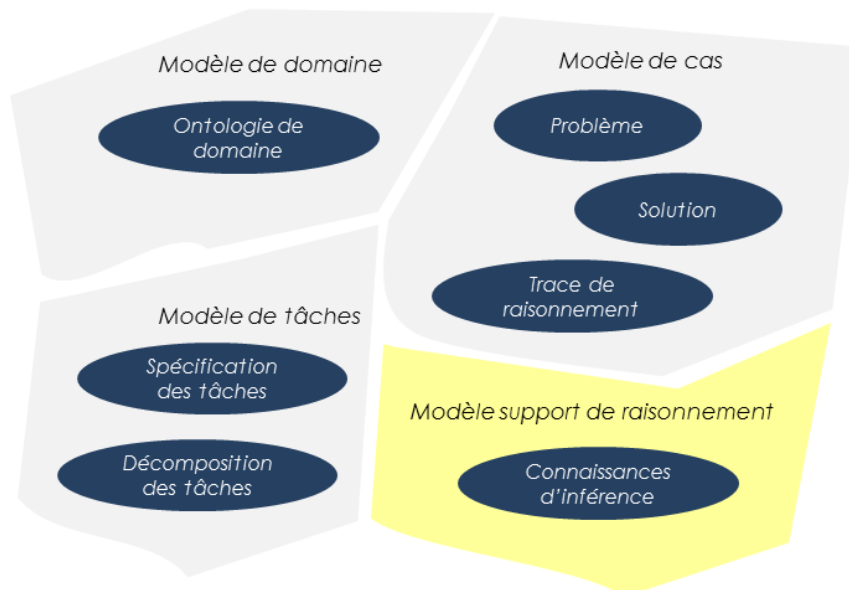


Figure 3.3. Modèles conceptuels (Fuchs B. et Mille A., 2005).

- **Le modèle conceptuel** du domaine décrivant les concepts utilisés pour décrire l'ontologie du domaine indépendamment du raisonnement ;
- **Le modèle de cas** qui sépare le cas en problème, solution, et trace de raisonnement ;
- **Les modèles de tâches** qui décrivent toutes les tâches de raisonnement (élaboration, remémoration, adaptation, etc.). Ces modèles sont représentés à l'aide de deux formalismes :
 - o Un formalisme de spécification de tâches permettant la description des connaissances traitées en entrée et en sortie de chaque tâche séparément ;
 - o Un formalisme de décomposition de tâches permettant la décomposition hiérarchique des tâches pour décrire le processus de raisonnement de manière détaillée ;
- **Les modèles supports du raisonnement** qui décrivent les connaissances nécessaires à la réalisation des inférences dans chaque tâche de raisonnement.

3.2.4. Travaux de Recio García et ses collègues

Recio García et ses collègues (Recio-García J.A. et al, 2006) à leur tour, préconisent différentes utilisations de l'ontologie au sein d'un système de RàPC. La figure 3.4 présente ces différentes utilisations.

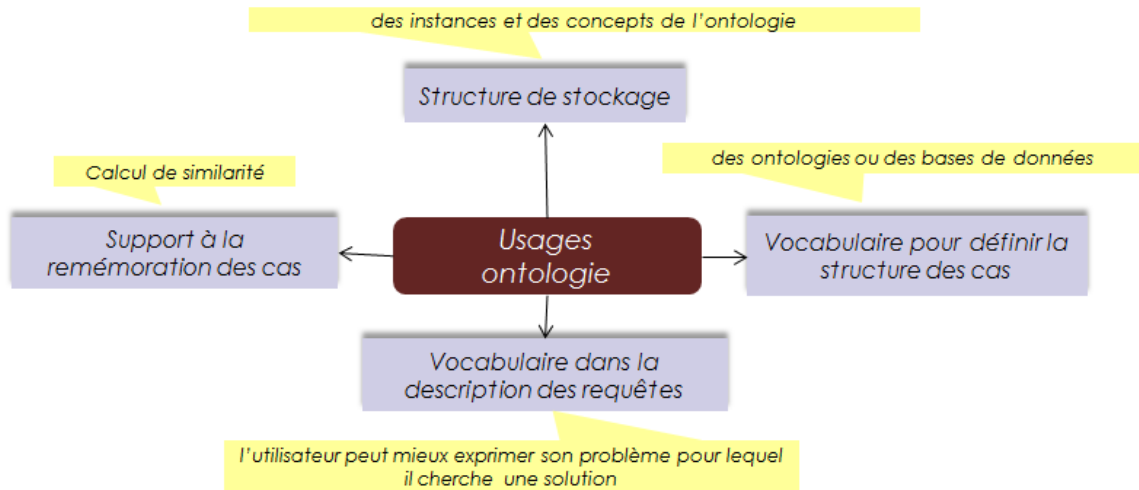


Figure 3.4. Les différents usages de l'ontologie selon (Recio-García J.A. et al, 2006).

Les travaux de (Recio-García J.A. et al, 2006), (Díaz-Agudo B. et al, 2007) ont conduit à la réalisation de jColibri, un Framework orienté objet écrit en Java pour la construction de systèmes de RàPC. Ce Framework s'est basé sur l'architecture CBRonto. Il a impliqué plusieurs algorithmes pour gérer la base de cas, la base de données, etc. Différents supports de stockage de cas (fichiers text/xml, ontologie, etc.) accessibles via des connecteurs spécifiques ont été proposés à cet effet (Figure 3.5). Le modèle de RàPC adopté est à la fois **conversationnel et structurel** et ayant une tâche de **synthèse**. La technique d'adaptation utilisée dans CBRonto est **transformationnelle** (voir section 2.3.5.3).

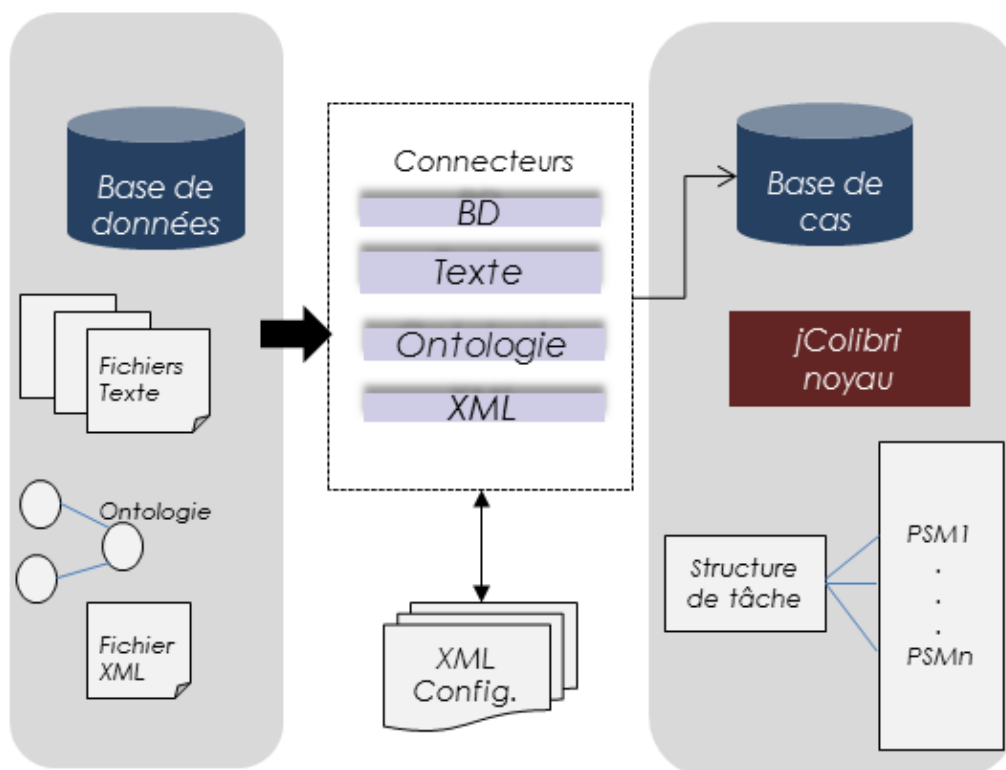


Figure 3.5. Connecteurs de jColibri (Recio-García J.A. et al, 2006) (Díaz-Agudo B. et al, 2007).

3.2.5. Travaux de Gómez-Gauchía et ses co-auteurs

Gómez-Gauchía et ses co-auteurs (Gómez-Gauchía H. et al, 2006) ont défini l'ontologie CCBROnto (Conversational CBROnto) comme une extension de CBROnto. Cette ontologie est destinée à la conception d'un système de RàPC conversationnel. Quatre modèles permettant d'organiser la base de cas ont été proposés par CCBROnto comme suit : Le modèle de domaine ; le modèle de cas ; le modèle de question (décrivant les types de questions possibles durant une session de conversation) ; le modèle d'utilisateur (permettant de personnaliser le système au profil des utilisateurs) (Figure 3.6).

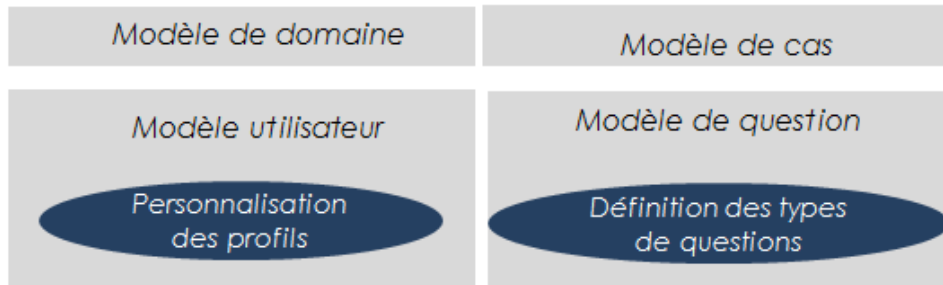


Figure 3.6. Les différents modèles de Gómez-Gauchía (Gómez-Gauchía et al, 2006).

3.2.6. Travaux de Chen

Chen (Chen K., 2008) a travaillé sur la conception et le développement d'une mémoire pour des agents assistants personnels (PA). Cette mémoire utilise un **mécanisme de RàPC**. Le modèle de RàPC adopté est textuel et ayant une tâche de **synthèse** et ayant la technique hiérarchique comme technique d'adaptation (voir section 2.3.5.3). Ce mécanisme a été testé en développant l'agent MemoPA (Chen K. et Barthès J.B., 2007) dans la plate-forme OMAS (Open Multi-Agent System). Un des modules clés de MemoPA est l'**ontologie** (Figure 3.7).

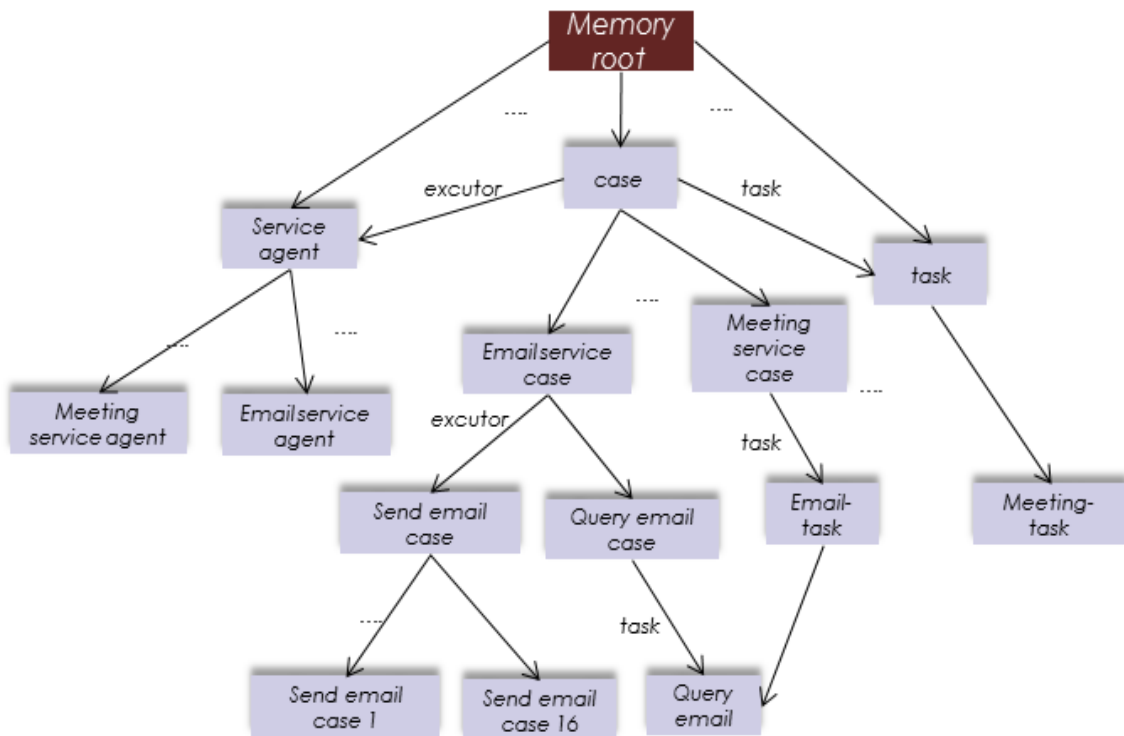


Figure 3.7. Partie de la mémoire de l'agent MemoPA (Chen K. et Barthès J.B., 2007).

L'ontologie est utilisée pour :

- Structurer la mémoire de l'agent ;

- Identifier les concepts qui peuvent apparaître dans les messages envoyés par d'autres agents ou par l'utilisateur ;
- Choisir un agent de service qui puisse répondre à la demande de l'utilisateur.

3.2.7. Travaux de Hadj-M'tir

Les travaux de (Hadj-M'tir R., 2010) réalisés au sein de RIADI (Laboratoire de Recherche en Génies Documentiel et Logiciel) à l'ENSI Tunis et LIRIS (Laboratoire d'InfoRmatique en Images et Systèmes d'Information) à l'INSA, ont permis de proposer une démarche de réutilisation des situations d'apprentissage dans les Systèmes de E-Learning.

L'approche proposée s'est basée conjointement sur le RàPC et les ontologies. Le RàPC permettant de s'appuyer sur une mémoire collective de situations d'apprentissage antérieures, pour offrir à l'apprenant courant la meilleure expérience adaptée à ses attentes, déjà validée par d'autres apprenants ayant des profils similaires d'apprentissage en prenant en compte les évaluations des situations antérieures d'un groupe d'apprenants. La modélisation ontologique des différentes connaissances manipulées par le système (profil de l'apprenant, le domaine, situation d'apprentissage) a été utilisée dans le but de garantir une réutilisation optimale des situations d'apprentissage.

Cette approche s'est concrétisée en développant un prototype baptisé PELSYS, pour "*Personalized E-Learning SYStem*", basé sur une architecture client-serveur permettant aux différents acteurs du système depuis un simple navigateur Web de se connecter et spécifier les connaissances du modèle du domaine, le profil de l'apprenant et les informations décrivant une stratégie d'apprentissage. PELSYS recourt à une approche de réutilisation basée sur le RàPC. Le modèle de RàPC adopté est **conversationnel** et ayant une tâche de **synthèse**. Il propose à cet effet une récapitulation des situations d'apprentissage afin de pouvoir les réutiliser.

3.2.8. Autre travaux

Valmi Dufour-Lussier (Valmi D.L. et al., 2010) a étudié l'impact de **raffinement de l'ontologie du domaine sur le processus de remémoration d'un système de RàPC**. En effet, l'enrichissement d'une ontologie améliore la remémoration et, plus globalement, la réponse finale du système. L'ontologie existante est enrichie en ajoutant automatiquement de nouveaux concepts qui vont affiner l'organisation initiale des concepts. Les nouveaux concepts introduits résultent d'un processus de fouille de données (l'analyse de concepts formels (ACF)) sur des données complémentaires relatives aux concepts de l'ontologie, collectées explicitement pour le processus de fouille. Les concepts formels créés par la fouille sont intégrés à l'ontologie, permettant ainsi de généraliser le problème cible de façon plus fine qu'avant cet ajout.

De leurs côtés, Essam A. et AbdEl-Badeeh (Essam A. et AbdEl-Badeeh M.S., 2010) ont proposé une approche qui se base sur l'utilisation conjointe de **RàPC et une ontologie de domaine de cancérologie**. L'objectif est de proposer un modèle de classification des cancers des seins pour décider s'il s'agit d'un cancer bénigne ou maligne. Une maquette baptisée « myCBR » basée sur le Framework jColibri fut proposée. Le système de RàPC utilisé est supposé conversationnel basé sur la tâche de classification. L'éditeur protégé est utilisé également pour construire l'ontologie de domaine.

Martin Kowalski et ses co-auteurs (Kowalski M. et al., 2012) ont proposé à leur tour, un système de RàPC baptisé « SCM Project Recommender » qui est piloté par une ontologie. Ce système permet de mesurer la similitude entre les collections de connaissances, qui sont écrites en langage naturel. Il permet la réutilisation des connaissances dans un contexte

complexe, en particulier les **projets logistiques internationaux**. L'outil prototype développé est basé de même que les travaux de (Essam A. and AbdEl-Badeeh M.S., 2010) sur le Framework jColibri.

3.3. Approches proposées liées à la sécurité des transports et du retour d'expérience

Nous avons regroupé quelques approches récentes d'aide à la décision proposées et appliquées dans le domaine de la sécurité des transports et notamment la sécurité industrielle dans le tableau 3.1. Nous allons consacrer cette section à l'étude de ces travaux.

Référence	Système proposé	Domaine d'application				
		Transport Ferroviaire	Transport Maritime	Transport Aérien	Transport Routier	Sécurité industrielle
(Yanping C. et al., 2005)		×				
(Arnaud V. et Napoli A., 2013)	SAGEO		×			
(Zubair, M. et Khan M.J., 2012)				×		
(Capus L. et Tourigny N., 2000)	ROSAC				×	
(Després S. et Ceausu V., 2004)	TERMINATE				×	
(Goh Y. et Chua D., 2010)						×
(Abou-Assali A., 2010)	COBRA					×

Tableau 3.1. Travaux proposés pour la sécurité des transports et la sécurité industrielle

3.3.1. Transport ferroviaire

Yanping Cui et ses co-auteurs (Yanping C. et al, 2005), ont proposé une approche d'aide à la décision dont le domaine d'application est identique au notre, le transport ferroviaire. L'approche proposée est basée sur le RàPC et le raisonnement à partir de règles pour la **prévention des accidents dans le transport ferroviaire**. Le système de RàPC utilisé est supposé conversationnel basé sur la tâche de classification.

Un cas dans le système représente des informations relatives à un incident de fonctionnement (Figure 3.8).

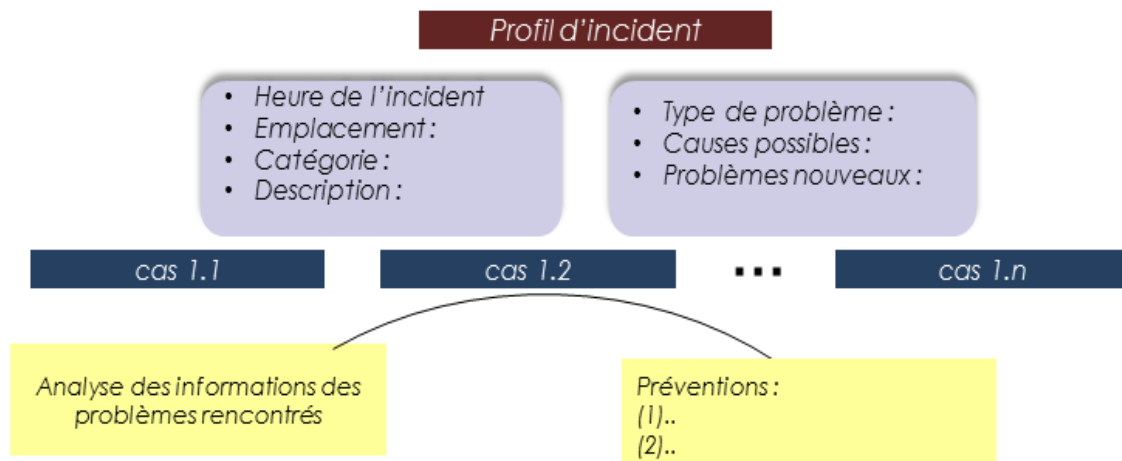


Figure 3.8. Représentation d'un cas d'incident (Yanping C. et al, 2005).

Il est défini comme un événement inattendu, qui a causé des pertes et des dégâts ou peut avoir des effets mauvais potentiels pour l'exploitation du train. Un accident ou un incident implique un ou plusieurs problèmes sous-jacents (par exemple, les erreurs humaines et les problèmes du système etc.).

Un problème sous-jacent peut être représenté par un certain nombre de caractéristiques qui peuvent être en outre liées à certaines causes possibles. Dans le système proposé, un cas d'incident se compose d'un certain nombre de sous-cas. Chaque sous-cas fournit des informations d'analyse et de préventions appropriées pour un problème sous-jacent. Les informations d'analyse de problèmes sous-jacents sont représentées par un certain nombre de paramètres : le type de problèmes, les caractéristiques de problèmes et les causes possibles aux problèmes.

3.3.2. Transport aérien

Zubair, M. et Khan, M.J. (Zubair, M. et Khan M.J., 2012) à leur tour ont proposé une approche basée sur le RàPC pour la prédiction des accidents aériens. Une base de cas est utilisée comme un référentiel pour les accidents antérieurs. Les auteurs ont proposé et mis en œuvre un algorithme qui intègre les différentes phases du cycle de RàPC. Le système de RàPC utilisé est de modèle textuel ayant une tâche de synthèse. Cinq fonctions de similarité pour la récupération des cas antérieurs ont été testées et une étude empirique de l'impact du voisinage le plus proche a été réalisée et présentée. Les auteurs préconisent que leur approche proposée peut produire jusqu'à 87 % des prévisions précises des accidents et des incidents aériens. Pour la préparation et la représentation de cas, ils ont caractérisé chaque instance d'un cas par divers paramètres disponibles d'un avion. Chaque cas est représenté comme un vecteur linéaire d'attributs, ce qui s'apparente bien au modèle structurel de cas que nous avons déjà décrit (section 2.3.3).

Les auteurs ont choisi cinq fonctions de distance incluant la distance de Hamming, la distance de Manhattan, la distance euclidienne, la distance de Canberra et celle de Bray-Curtis. Une fois le cas le plus proche récupéré, un mécanisme d'adaptation basé sur le vote à la majorité (qui consiste à un simple calcul de la moyenne arithmétique) est utilisé pour estimer la solution du cas cible (cas de test). Pour la faisabilité de leur approche, les auteurs ont utilisé des données extraites de la base de données de site de l'accident de l'aviation à travers un analyseur disponible sur <http://aviation-safety.net/database>.

3.3.3. Transport maritime

Arnaud Vandecasteele et Aldo Napoli (Arnaud V. et Napoli A., 2013) ont proposé un système SAGEO basé sur une modélisation ontologique (Figure 3.9) pour l'analyse de comportements de navires à risques.

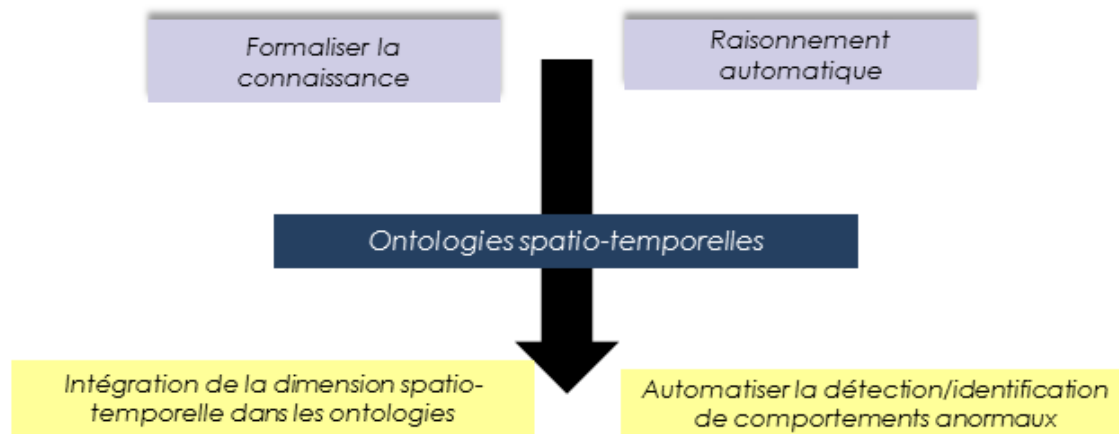


Figure 3.9. Architecture SAGEO (Arnaud V. et Napoli A., 2013).

Ils ont étudié les potentialités des ontologies spatiales à la fois comme objet de modélisation, de partage et d'inférence afin d'améliorer l'analyse de comportements des navires à risque. Les auteurs ont proposé un prototype de faisabilité pour la surveillance maritime composé de trois principales ontologies : une ontologie des trajectoires contenant les différents concepts spatio-temporels nécessaires à la description géométrique des trajectoires ; une ontologie géographique contenant les concepts spécifiques à la description du territoire (routes, ports, baies, etc.) et une ontologie du domaine qui contient les concepts permettant d'enrichir les informations initialement contenues et d'apporter ainsi une connaissance spécifique au sein de l'application.

Ils ont utilisé des règles d'inférence pour la détection des alertes (ces règles ont été définies par les experts du domaine maritime) puis un moteur de RàPC conversationnel permettant de déterminer automatiquement les scénarios potentiels issues de l'interprétation d'une situation à risque. Leur approche proposée a ensuite été mise en application au sein du système de surveillance maritime développé par le Centre de recherche sur les Risques et les Crises : FishEye.

3.3.4. Transport routier

Nous pouvons citer les travaux de Laurence Capus et Nicole Tourigny (Capus L. et Tourigny N., 2000) qui ont proposé une approche d'aide à la formation en analyse de sécurité routière à travers le système ROSAC (ROad Safety Analysis with Cases). ROSAC est un système de RàPC (utilisant un modèle conversationnel et ayant une tâche de synthèse), qui permet de capitaliser les études antérieures, de retrouver les études antérieures en fonction de la demande de l'utilisateur et notamment aider à la formation par des exemples.

L'architecture générale de ROSAC (Figure 3.10) comprend quatre parties principales :

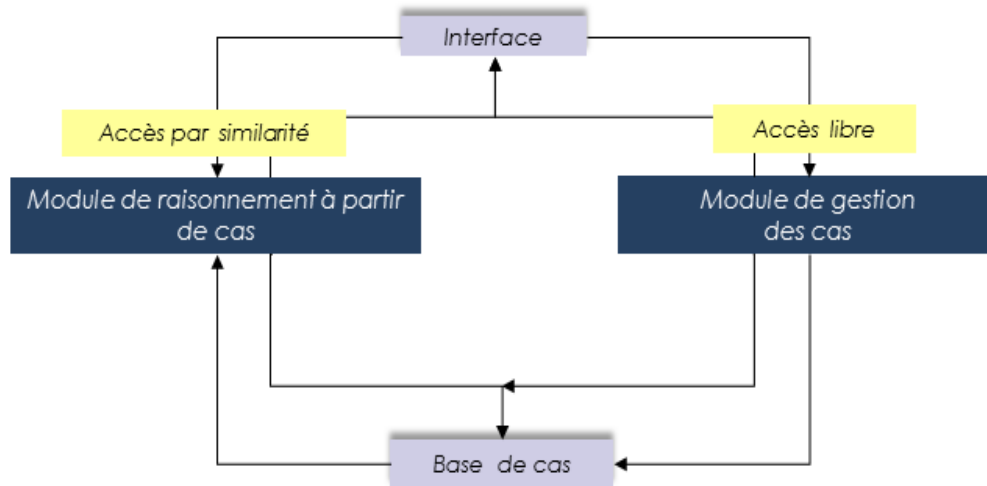


Figure 3.10. Architecture ROSAC (Capus L. et Tourigny N., 2000).

La base de cas, le module de gestion des cas, le module de RàPC et l'interface utilisateur. L'approche est basée sur la réutilisation d'expériences. Lorsqu'un nouveau cas est étudié par les analystes, ROSAC cherche des situations similaires déjà rencontrées afin de leur rappeler comment elles ont été résolues.

Nous pouvons citer également les travaux de (Després S. et Ceausu V., 2004) à travers leur approche qui se base sur le RàPC et une ontologie. Leur système proposé permet d'associer un profil de scénario d'accidents pour chaque groupe d'accidents de la route survenu dans un emplacement géographique particulier. Les auteurs utilisent une ontologie de l'accidentologie qui a été construite à partir de connaissances expertes (textes du domaine, scénarios d'accidents rédigés par les chercheurs du domaine et entretiens réalisés auprès des chercheurs). Elle est fondée sur une approche terminologique centrée sur les actions. L'architecture de l'approche proposée est scindée en deux modules : un premier module pour l'extraction de profils d'accidents à partir de la base de scénarios et un autre module pour la construction de la solution associée au profil (Figure 3.11).

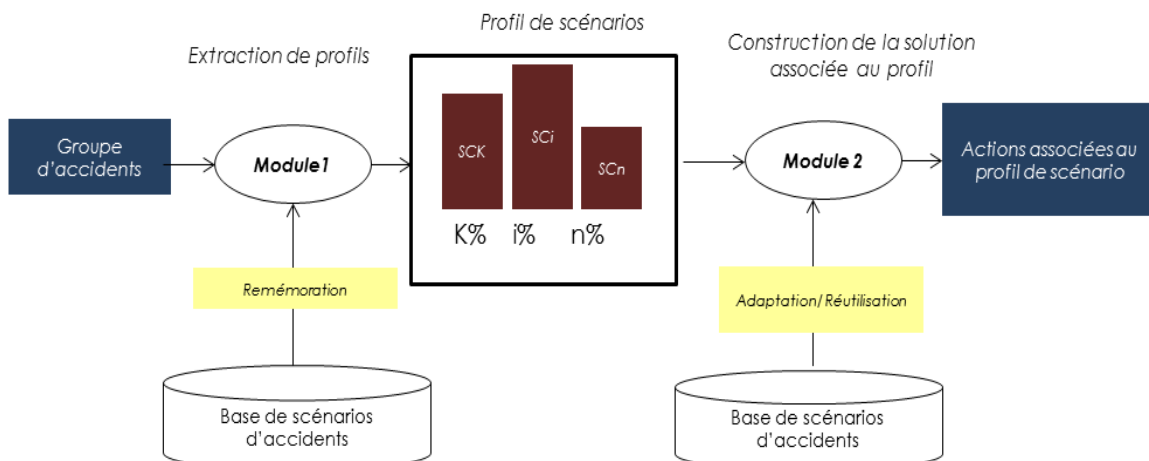


Figure 3.11. Architecture du système proposé (Capus L. et Tourigny N., 2000).

Le module d'extraction des profils traite les accidents afin de les apparier avec l'un des scénarios de la base de cas. A l'issue de ce traitement un profil est construit. Le filtrage des cas se fera sur les composants ; sur les relations caractérisant un scénario ; selon les niveaux de raisonnement où elles interviennent. Un profil constitue les données d'entrée du module de construction des propositions d'aménagement.

Les mêmes auteurs ont proposé dans (Ceausu V. et Després S., 2004), le système TERMINATE basée sur une approche mixte pour la construction d'une ressource terminologique. Le domaine d'application des expérimentations est celui de l'accidentologie routière. Leur but est d'analyser l'apport de techniques de fouille de données textuelles à une méthodologie de construction d'ontologie à partir de textes. Dans ce contexte, la méthodologie de construction de la ressource terminologique TERMINATE et les techniques de fouille de données textuelles sont utilisées simultanément. Les résultats issus de ces techniques sont utilisés pour contribuer à la construction de la ressource terminologique avec TERMINATE.

3.3.5. Sécurité industrielle et analyse des risques

Goh, Y. et Chua, D (Goh, Y. and Chua D., 2010) ont proposé à leur tour une approche de RàPC pour l'identification des risques. L'approche de RàPC utilisée vise à utiliser les connaissances passées sous forme de cas de danger pour améliorer l'efficacité et la qualité de l'identification des dangers. Pour un nouveau scénario de travail (Cas d'entrée ou cas cible), un arbre de l'identification des dangers les plus pertinents et un ensemble de cas d'incident sera récupéré pour faciliter l'identification des dangers. Pour la phase d'adaptation, elle est effectuée en trois étapes : (1) élagage de l'arbre de l'identification des dangers récupéré ; (2) élagage des cas incidents et ; (3) insertion de cas incidents dans l'arbre de l'identification des dangers.

Le processus d'adaptation est basé sur le calcul de similarité de partitions. Une étude de cas basée sur des identifications de dangers réels et les cas d'incidents, est utilisée pour valider la faisabilité de la technique d'adaptation proposée.

Inspiré de plusieurs travaux de recherche, Abou-Assali A. (Abou-Assali A., 2010) a proposé une approche d'acquisition des connaissances d'adaptation et de traitement de l'hétérogénéité de cas. Le système « COBRA » fut proposé par (Abou-Assali A., 2010) est considéré comme une plateforme de RàPC qui s'appuie sur une ontologie pour modéliser les connaissances de domaine ainsi que les connaissances relatives aux cas (Figure 3.12). Le système de RàPC est basée sur un modèle conversationnel et ayant une tâche de synthèse.

Le travail est mené en partenariat avec l'INERIS (Institut National de l'Environnement industriel et des RISques) dans le cadre du projet KMGR. Le système COBRA est dédié aux détecteurs de gaz, premiers maillons des barrières techniques de sécurité sur de nombreuses installations industrielles. Ces détecteurs ont pour fonction de signaler la fuite de certains gaz sur le site concerné.

La plate-forme COBRA est développée en tant qu'application Java basée sur Eclipse grâce au Framework RCP (Rich Client Platform). Son architecture est composée de deux modules principaux : le module d'acquisition des connaissances d'adaptation (ACA), et le module de résolution de problèmes. Ces deux modules font appel à un troisième module qui est le module d'alignement de cas.

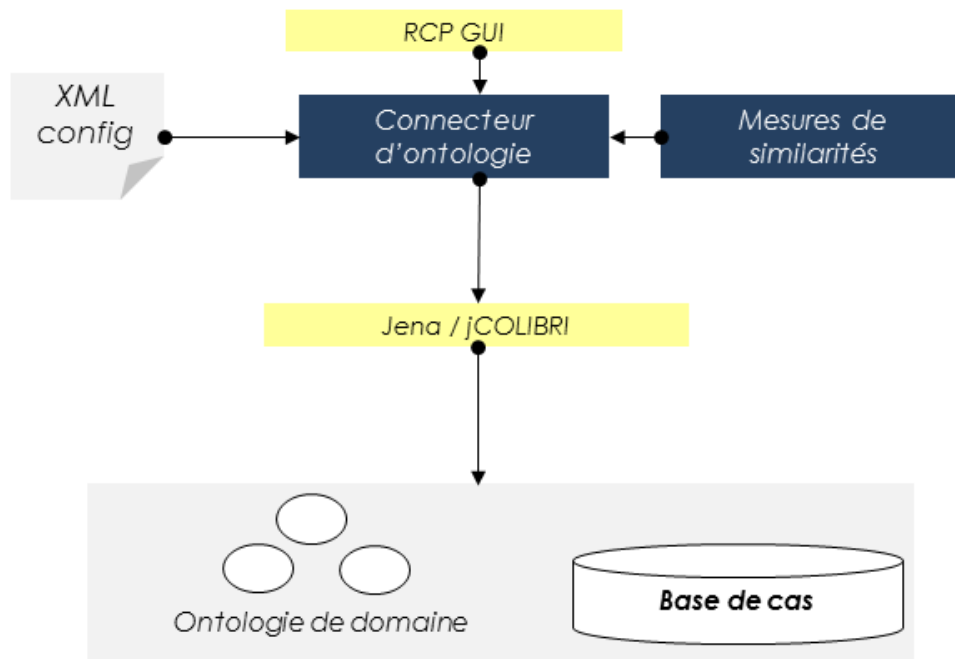


Figure 3.12. Architecture COBRA (Abou-Assali A., 2010).

3.4. Bilan et synthèse

3.4.1. Comparaison des approches

Malgré la divergence des environnements de développement et les techniques/méthodes mises en place pour concevoir et mettre en œuvre toutes les approches que nous venons de présenter, elles demeurent pour nous d'un apport non négligeable et peuvent nous être des bons points d'accès pour bien mener notre démarche de conception et de développement envisagée à ses fins utiles.

Pour pouvoir positionner et situer ces travaux, nous avons établi dans le tableau 3.2, une comparaison de ces travaux. Ce tableau n'est pas exhaustif et constitue une première phase de travail qui pourra par la suite être améliorée, par exemple, au moyen de questionnaires adressés directement aux auteurs.

Pour pouvoir comparer ces approches, nous avons fixé certains critères. Ces critères ont fait l'objet de notre étude effectuée dans le chapitre 2. Nous pouvons citer comme exemple ; le modèle de RàPC utilisé (conversationnel, textuel ou structurel), la tâche de RàPC (classification, synthèse), la technique d'adaptation utilisée (transformative, générative ou hiérarchique), le type de l'ontologie (de domaine, de tâche ou autre type), les mesures de similarité sémantique (basée sur le contenu informationnel ou bien sur les arcs/distances) et notamment le domaine d'application (particulier ou général).

A travers cette étude, nous avons ciblé les approches qui peuvent nous être utiles dans le cadre de notre travail quant aux différentes techniques et méthodes choisies (section 2.5).

Référence	Modèle de RàPC			Tâche de RàPC		Technique d'adaptation			Type d'ontologie		similarité entre concepts		Domaine d'application	
	Conversationalnel	Textuel	Structurel	Classification	Synthèse	Transformationnelle	Générative	Hiéarchique	Domaine	Autre type	Contenu informationnel	Arcs (distances)	Non spécifique	Spécifique
(Capus L. et Tourigny N., 2000)	x				x	?			x		?			x
(Díaz-A. et González-C. 2002)	x			x		x		x	x	x	x		x	
(Aamodt A., 2004)			x		x				x				x	
(Després S. et Ceausu V., 2004)		x			x	?				x	?			x
(Fuchs et Mille, 2005)			x		x				x	x			x	
(Yanping C. et al, 2005)	x				x	?			?		?			x
(Recio-Garcia, 2006)			x		x				x				x	
(Recio-G. et Díaz-A., 2007)	x				x	x				x	x		x	
(Chen K., 2008)		x			x			x		x	x			x
(Valmi D. et al., 2010)			x		x	x	x		x	x	x			x
(Essam A. et AbdEl-Badeeh, 2010)	x			x					x		x			x
(Abou-Assali A. 2010)	x			x		x	x		x			x		x
(Hadj-M'tir R., 2010)	x				x	?			x		x			x
(Kowalski M., 2012)	x			x					x		?			x
(Zubair M. et Khan M.J., 2012)		x			x		?		x			x	x	x
(Arnaud V. et Napoli A., 2013)	x				?				x	x		x	x	x

Tableau 3.2. Comparaison des approches étudiées

Nous retenons :

- x : Propriété acquise ;
- "blanc" : Propriété non acquise ;
- " ?" : Pas d'informations suffisantes pour prendre position.

3.4.2. Synthèse

Pour mieux synthétiser ces travaux, nous avons choisi volontairement de les classer en trois familles :

- *La première famille* concerne les **travaux qui ont été consacrés aux différents usages de l'ontologie dans les systèmes de RàPC**. Nous mentionnons à cet effet le système « CBROnto » de (Díaz-Agudo B. et González-Calero P., 2002) et (Recio García J.A. et al, 2006). De même pour Gómez-Gauchía (Gómez-Gauchía H. et al, 2006) qui a étendu CBROnto et défini l'ontologie CCBROnto (Conversational CBROnto). Les modèles de connaissances proposés par Fuchs et Mille (Fuchs et Mille, 2005) ayant pour but l'usage de l'ontologie dans les systèmes de RàPC. MemoPA de (Chen K., 2008) et (Chen K. et Barthès, 2007). Valmi Dufour-Lussier (Valmi D.L. et al, 2010) a étudié l'impact de raffinement de l'ontologie du domaine sur le processus de remémoration d'un système de RàPC ;
- *La deuxième famille* se **spécialise plutôt dans les méthodes de raisonnement précisément le RàPC** sans s'étaler beaucoup sur la représentation de connaissances. L'approche de (Recio-García J.A. et al, 2006), (Díaz-Agudo B. et al, 2007) qui a conduit à la réalisation de jColibri, qui a été utilisé également dans les travaux de (Abou-Assali A., 2010), (Essam A. et AbdEl-Badeeh M.S., 2010) et (Kowalski M. et al, 2012) ;
- *La troisième famille* regroupe les travaux qui **font partie d'un champ d'application réel**. Nous pouvons citer à cet effet, (Essam A. et AbdEl-Badeeh, 2010) qui a étudié le domaine de cancérologie. Les systèmes de e-learning pour (Hadj-M'tir R., 2010). La réutilisation des connaissances relatives aux projets logistiques internationaux dans SCM Project Recommender » par (Martin Kowalski et al, 2012). Pour le domaine de **la sécurité des transports**, nous pouvons citer ceux appliqués dans **le domaine du transport routier**, le système ROSAC (ROad Safety Analysis with Cases) de (Capus L. et Tourigny N., 2000) de même pour (Després S. et Ceausu V., 2004). **Le domaine ferroviaire** pour (Yanping C. et al, 2005) à travers la prédiction des accidents/incidents de chemin de fer. Le transport aérien pour (Zubair M. et Khan M.J., 2012). Le système SAGIO de (Vandecasteele A. et Napoli A., 2013) appliqué dans le domaine de **transport maritime**. Nous pouvons également mentionner les travaux appliqués dans le domaine de **l'analyse des risques et l'identification de dangers**, comme Abou-Assali Amjad (Abou-Assali A., 2010) et (Goh, Y. et Chua, D., 2010).

3.5. Conclusion

Nous avons consacré le présent chapitre à l'étude des approches d'aide à la décision, qui se sont basés sur le raisonnement à partir de cas (RàPC) et les ontologies, et qui ont été proposées dans la littérature. Nous avons examiné à cet effet dans la première partie les développements récents survenus dans ce domaine à travers les différentes architectures et les systèmes proposés. Nous avons exploré les travaux liés au domaine de la sécurité des transports (ferroviaire, routier, maritime et aérien) et notamment la sécurité industrielle (l'analyse des risques et l'identification des dangers). Nous avons pu constater ainsi que beaucoup d'approches présentent un intérêt indéniable, en l'occurrence, les travaux de (Fuchs B. et Mille A., 2005) (Gómez-Gauchía H. et al, 2006), (Hadj-M'tir R., 2010) et (Abou-Assali A., 2010) qui peuvent être très utiles pour élaborer notre approche.

Néanmoins, malgré les nombreux efforts effectués, et même si ces travaux apportent des contributions significatives, certaines limites peuvent être décelées :

- Une limite qui concerne **la généricité des approches proposées** : Malgré la diversité des systèmes proposés et les architectures retenues, il s'avère que leur déploiement dans d'autres domaines outre que le domaine pour lesquels ils ont été conçus est généralement difficile ;
- Une deuxième limite à-propos **les méthodes de raisonnement appliquées** : Nous révélons aussi que peu d'approches, permettant la couverture de cycle de vie de RàPC, ont été développées et qui ont permis d'optimiser toutes les phases du raisonnement. En outre, les travaux sont dans la plupart des cas tournés vers la phase d'adaptation au détriment des autres phases de raisonnement ;
- La troisième limite est liée à **l'insuffisance de travaux** qui ont été consacrés à l'analyse des systèmes des transports, en particulier, le transport ferroviaire.

Les points que nous venons de mentionner ont renforcé nos motivations et nos objectifs initialement fixés, pour proposer dans le cadre de notre travail de thèse une approche générique d'aide à la décision, basée conjointement sur le RàPC et les ontologies, permettant l'amélioration de la sécurité des transports. Pour aboutir à une approche réussie et efficace, notre démarche doit permettre d'optimiser et de renforcer éventuellement la complémentarité entre toutes les phases de raisonnement.

Le chapitre suivant présente exhaustivement notre approche d'aide à la décision baptisée **ADAST**.

Approche proposée

- Chapitre 4 : ADAST : une approche d'aide à la décision basée sur le RàPC
.....78
- Chapitre 5 : Maquette de faisabilité d'ADAST.....117
- Chapitre 6 : Expérimentation et validation.....141

Chapitre 4 : ADAST : une approche d'aide à la décision basée sur le RàPC

4.1. Introduction

Dans les chapitres précédents, nous avons dressé une étude bibliographique assez exhaustive concernant la sécurité, l'analyse des risques et notamment les problèmes particuliers liés à l'acquisition, la capitalisation et l'exploitation des connaissances issues de retour d'expérience (Rex). Par la suite, nous avons étudié les concepts de base que nous utilisons dans notre travail à savoir, le Raisonnement à Partir de Cas (RàPC), les ontologies, deux domaines de recherche en perpétuel évolution et aussi les méthodes et les techniques de la fouille de données qui seront exploitées pour élaborer toute la démarche d'aide à la décision. Nous avons examiné notamment les développements récents survenus dans le domaine de RàPC et en particulier les approches qui se sont basées sur les ontologies. Cette étude a débouché sur certaines limites des approches proposées. Ceci nous a permis dans un premier temps de consolider nos motivations et dans un second temps de bien situer notre approche par rapport aux travaux réalisés. Se basant sur cette étude, nous avons alors pu justifier nos choix et nos directives quant à la modélisation de connaissances et aussi quant à notre approche de résolution de problème.

Dans ce chapitre, nous allons alors procéder à la présentation d'une approche générique d'aide à la décision baptisée ADAST (Aide à la Décision pour Améliorer la Sécurité des Transports). Nous présentons tout d'abord, les motivations à travers les objectifs attendus et nous rappelons nos choix méthodologiques adoptés. Nous présentons ensuite l'architecture fonctionnelle d'ADAST et notamment les modèles de connaissances élaborés à savoir ; le modèle conceptuel, le modèle de domaine et le modèle de cas. Nous détaillons aussi les articulations entre toutes les phases de raisonnement et la démarche d'aide à la décision retenue et nous positionnons également le choix et l'utilisation de chaque méthode/technique pour bien expliquer l'utilité et la nature par rapport à notre contexte d'étude.

4.2. Motivations

Afin de rationaliser et de renforcer le processus d'acquisition et de management des connaissances issues de retour d'expérience (Rex), les scénarios d'accidents, dans le transport, il est nécessaire voire primordial dans un premier temps, d'harmoniser et d'homogénéiser la terminologie appliquée par les acteurs du domaine de la sécurité.

En dépit de tous les efforts déployés pour proposer des approches et des techniques permettant de gérer, conduire ce retour d'expérience, et malgré l'intérêt indéniable des approches et des méthodes existantes d'analyse et d'évaluation de la sécurité, la plupart des ces approches proposées dans la littérature souffrent de l'absence de définitions admises par tous et des ambiguïtés dues au manque, voire l'absence, de formalisme pour exprimer et modéliser les connaissances relatives aux scénarios d'accidents.

4.2.1. Objectifs attendus

Afin de mieux appréhender les limites que nous venons de mentionner dans les motivations, et dans le but de bâtir le mode de représentation, permettant de rendre accessibles et exploitables les connaissances (plus globalement l'expertise en matière de sécurité, qui est répartie entre les mains de plusieurs experts), nous avons eu recours aux techniques et aux méthodes de représentation de connaissances et de l'aide à la décision. Ces dernières ont montré leur intérêt et ont apporté des réponses pertinentes à des problématiques diverses.

Notre objectif dans ce cadre, est la proposition d'une approche d'aide à la décision permettant d'améliorer la sécurité des transports. Notre démarche d'aide à la décision envisagée permettra non seulement de structurer, de formaliser et de capitaliser les connaissances de sécurité, mais surtout de permettre à plusieurs échelles, de fournir un tableau de bord, sous forme d'indicateurs favorisant la prise de décision par les experts et les spécialistes de domaine.

4.2.2. Choix adoptés

Pour élaborer la démarche d'aide à la décision envisagée, **le Raisonnement à Partir de Cas (RàPC)** (Case-Based Reasoning (CBR)), **les ontologies** et notamment des méthodes et techniques de **fouille de données** ont été adoptées.

Tout d'abord, le choix du RàPC se justifie naturellement car il s'agit d'une méthode parente du Rex. Il est considéré comme étant un processus de résolution de problème débutant par une remémoration de situations similaires passées et se poursuivant par la réutilisation des informations et des connaissances disponibles sur ces situations afin d'élaborer une nouvelle solution au problème courant par adaptation des solutions passées retrouvées.

Pour justifier notre choix concernant l'utilisation **des ontologies**, nous nous basons sur trois définitions pour les ontologies proposées par Valéry Psyché (Psyché V., 2003), (Noy N.F. et McGuinness D. L., 2001) et (Grüninger M., 1995) (voir chapitre 2, section 2.2.1). A partir de ces trois définitions, il est possible de considérer les ontologies comme un ensemble de concepts et un ensemble de relations entre ces concepts, car ces dernières permettent de bien présenter des propriétés, des contraintes et des comportements. Ceci pourra être une solution pertinente pour bien définir un vocabulaire de description des cas. Rajoutons à ceci le fait que les ontologies permettent aussi de réduire considérablement l'effort de l'acquisition des connaissances dans les différentes phases du RàPC.

Nous utilisons également dans notre approche plusieurs techniques et méthodes issues de la fouille de données. Nous avons convenu d'utiliser, par exemple, **les règles d'association** pour aider à bien caractériser et décrire un cas cible (première phase de RàPC). La technique **d'arbre de décision est utilisée** pour déterminer les descripteurs clés ayant le pouvoir de classification des cas cibles saisis (Deuxième phase de RàPC). Nous avons aussi convenu de proposer une approche d'acquisition des connaissances d'adaptation basée sur **l'Analyse des Concepts Formels (ACF)**. Cette approche constitue un apport non négligeable pour remédier aux difficultés liées à l'extraction des connaissances d'adaptation (Troisième phase de RàPC).

4.3. Présentation d'ADAST

4.3.1. Architecture fonctionnelle d'ADAST

La littérature de RàPC propose plusieurs architectures. Ces architectures partagent plus ou moins les mêmes composants. Inspiré par ces travaux, notre système (Maalel A. et al., 2012b ; 2014) comprend deux processus principaux : *un processus hors-ligne et un processus en-ligne* (Figure 4.1).

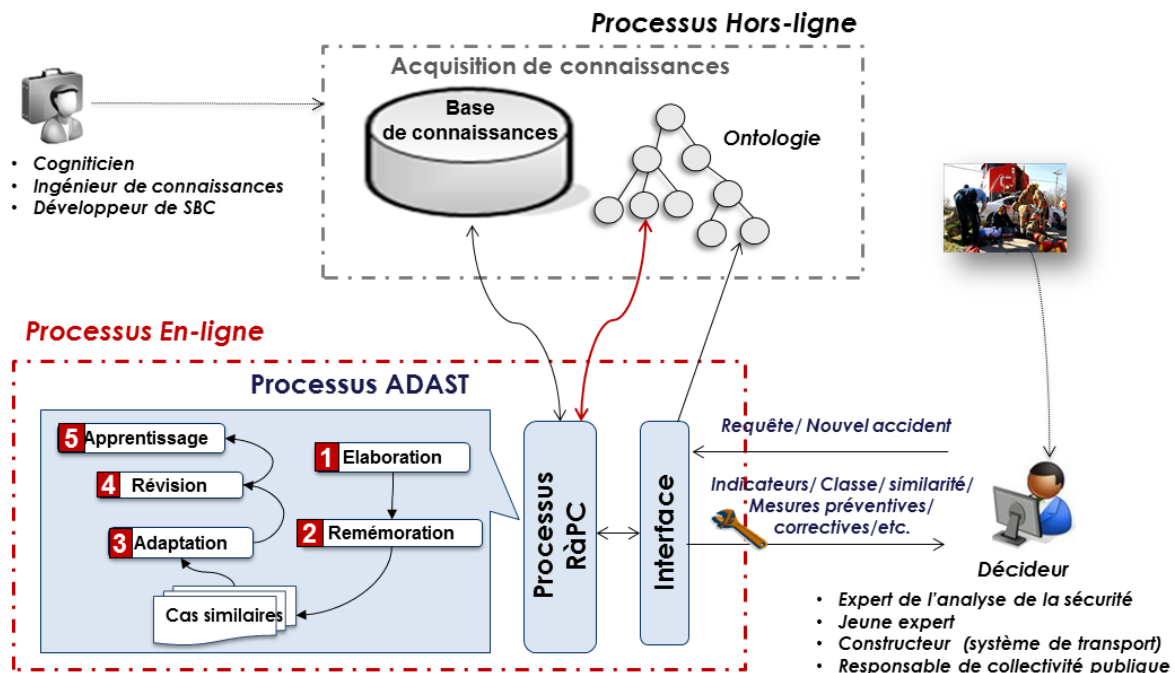


Figure 4.1. Architecture fonctionnelle d'ADAST et ses interactions avec le module d'acquisition des connaissances

Le *processus hors-ligne* concerne, dans un premier temps, la formalisation des connaissances et la mise en œuvre de l'ontologie de domaine à travers trois modèles de connaissances, le modèle conceptuel, le modèle de domaine et notamment le modèle de cas. Ensuite la construction de la base d'apprentissage (la base de cas). Enfin, la préparation et l'acquisition de connaissances relatives au raisonnement ; les règles d'association par application de l'algorithme Apriori, les règles de décision par application de l'algorithme C4.5 et enfin les règles d'adaptation issues de l'application de la démarche d'acquisition de connaissance d'adaptation basée sur l'ACF.

Le processus En-ligne concerne principalement le déroulement de la session d'analyse de sécurité à travers les différentes phases de RàPC. Rappelons que cette aide à la décision assez conséquente est adressée principalement à tous les intervenants dans le processus d'étude du dossier de sécurité et surtout pour les experts de l'analyse de sécurité.

Le décideur sera invité donc, dans chaque phase à interagir avec l'interface dédiée pour conduire toute la démarche d'aide à la décision. Chaque session débute par **l'introduction d'un problème** (un cas d'accident), nommé cas cible. Puis **la proposition de plusieurs indicateurs à plusieurs échelles sous forme d'un tableau de bord** (Instances pertinentes à ajouter, classe d'accident, seuils de similarité, pourcentages de cas similaires, etc.). La session se termine par **la proposition d'une solution**, souvent à travers des mesures préventives et/ou correctives adéquates pouvant éviter la reproduction de ce problème dans

le futur, et accompagnée par une **décision d'apprentissage** du cas courant associé avec sa nouvelle solution proposée.

Le processus En-ligne comprend également d'autres fonctions comme la gestion de la base de connaissances, l'enrichissement de l'ontologie et le paramétrage de raisonnement à travers l'interface de préconception. Cette interface va permettre, à l'ingénieur de connaissance l'ajustement de plusieurs paramètres liés au raisonnement et processus d'aide à la décision. L'utilisateur (jeune expert, quant à lui, pourrait exploiter le processus En-ligne donc ADAST en vue de se former progressivement au domaine de l'analyse de la sécurité ferroviaire

4.3.2. Processus de résolution de problème

Comme présenté dans la section 2.5.2, nous utilisons le cycle du RàPC de (Cordier et al, 2007). Ce cycle comprend cinq phases : (1) l'élaboration, (2) la remémoration, (3) l'adaptation, (4) la révision et (5) l'apprentissage (Figure 4.2). Dans cette section, nous allons décrire l'articulation entre les différentes phases de la démarche d'aide à la décision proposée.

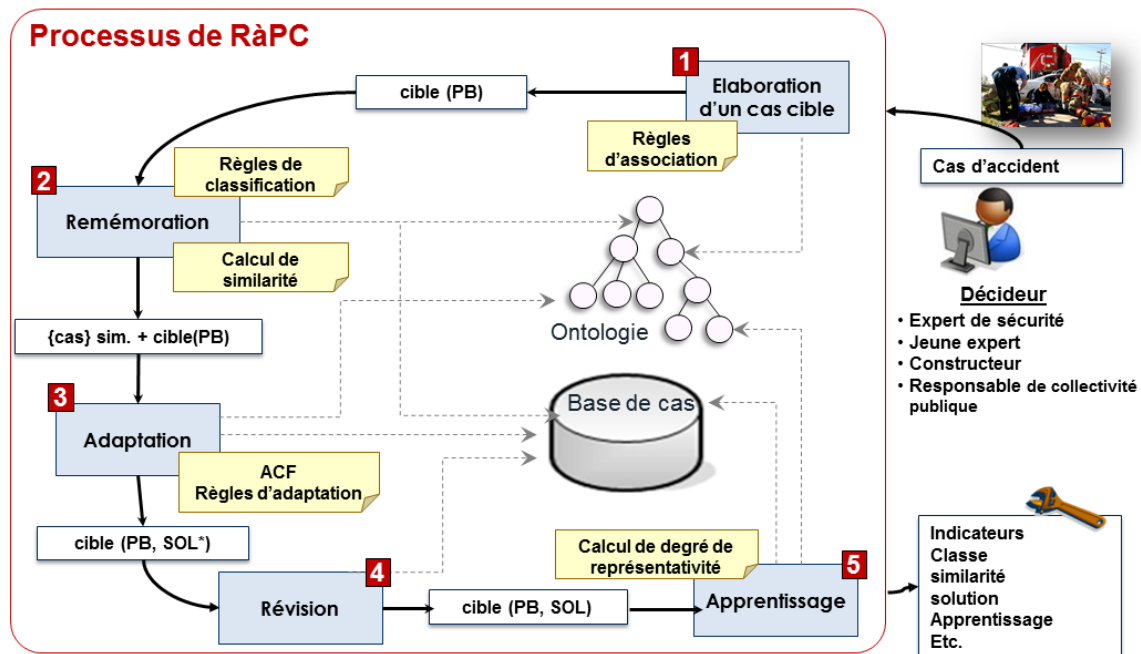


Figure 4.2. Articulations des différentes phases de cycle de RàPC retenu.

Un ensemble de cas sont capitalisés dans la base de connaissances (base de cas), les cas créés seront indexés selon leur classe d'appartenance prédéfinie lors de la phase d'acquisition (un cas source représente réellement une instance de l'ontologie de domaine et associé à une classe d'appartenance).

Nous déclenchons le processus de raisonnement à travers la première phase, **l'élaboration de cas cible**, l'ontologie de domaine sera instanciée et visualisée sous forme de formulaire à remplir par l'utilisateur qui sélectionne à son tour les concepts principaux de la partie problème aux quels il peut caractériser son cas cible. Dans cette phase, nous avons convenu de donner une aide sous forme d'extension permettant d'enrichir la description initiale fournie par l'utilisateur si celle-ci n'est pas suffisante ou bruitée. Nous avons utilisé alors **les règles d'association**. Cette technique issue de la fouille de données permet ainsi l'enrichissement de la description initiale. **La remémoration**, la deuxième

phase de raisonnement, permet généralement la recherche des cas jugés similaires au cas cible. Dans cette phase, nous avons eu recours à la méthode **d'arbre de décision**. L'arbre de décision assure la génération des règles de classification permettant la détection de la classe d'appartenance du cas cible. Notre objectif est de restreindre l'espace de recherche des cas sources aux seuls cas de la classe repérée. Nous procédons par la suite à la recherche des cas sources similaires. Des mesures de similarités sémantiques entre concepts au sein de l'ontologie sont utilisées et un algorithme de remémoration est proposé à cet effet. La phase suivante, **l'adaptation** se réfère à des règles d'adaptation générées en appliquant une approche semi-automatique d'acquisition de connaissances d'adaptation. Cette approche se base essentiellement sur **l'Analyse des Concepts Formels (ACF)**. Afin de permettre de tirer profit de(s) cas source(s) remémoré(es) et en vue de trouver des possibilités de substitutions ou un enrichissement matérialisé par une éventuelle solution au nouveau problème représenté par le nouveau cas « cible ». Au cours de la quatrième phase, **la révision**, la solution proposée à l'issue de la phase d'adaptation sera évaluée par l'utilisateur. La solution proposée peut ne pas lui plaire qui, à son tour, peut la modifier ou la refuser complètement. Dans la phase d'**apprentissage**, nous avons convenu de proposer **un algorithme d'auto-incrémentation**. Cet algorithme permet de conserver une bonne évolution de la base de cas du système en ne gardant que les cas les plus détaillés.

Avant de détailler notre approche ADAST, il convient dans un premier temps de présenter la structuration de connaissances ainsi que les différents modèles élaborés.

4.4. Structuration et modélisation de connaissances

Le travail de structuration de connaissances liées aux scénarios d'accidents (issus de retour d'expérience (Rex)) a permis de soulever des problèmes liés à un projet de grande envergure, dans un environnement industriel à haut risque.

Plusieurs sources de connaissances de sécurité (les experts, les chercheurs, les archives, les articles, ..., etc.) sont utilisées à cet effet. La plupart des connaissances relatives au domaine de la sécurité des systèmes de transport découlent principalement de l'analyse des situations contradictoires à la sécurité représentées sous forme de scénarios d'accidents potentiels, essentiellement issus des travaux antérieurs d'ACASYA : d'Aide à la Certification par Apprentissage des SYstèmes de transport Automatisés, développé dans le laboratoire LAMIH de l'université de Valenciennes et l'INRETS (devenu IFSTTAR) (Hadj-Mabrouk H., 1992 ; 1998 ; 2007) (Mejri L., 1995 ; 2009) (voir Annexe 1). Ces travaux ont permis de déboucher sur une variété de formes de modélisation complémentaires de scénarios d'accident. La plus sollicitée de ces formes est celle de la description statique d'un scénario. Les efforts dans le cadre de notre travail de thèse ont permis d'améliorer et d'homogénéiser sensiblement la description déjà proposée.

Un scénario d'accident ou encore un cas d'accident décrit un concours de circonstances qui peut conduire à une situation dangereuse. Il est caractérisé par un contexte et un ensemble de paramètres. Dans notre contexte, cette étape a débouché dans un premier temps sur l'élaboration d'un modèle « noyau » pour la représentation des scénarios d'accidents. Ce modèle est fondé sur l'identification de cinq paramètres généraux décrivant un scénario d'accident (le contexte de l'accident, les causes possibles, le niveau de risque de l'accident, l'accident potentiel et les solutions adoptées).

Avant de présenter les différents modèles que nous avons développés, il convient de recenser quelques connaissances de sécurité acquises. A cet effet, nous allons présenter dans la figure 4.3 un échantillon d'instances relatives à « Cause liée à l'environnement ». Une instance d'un concept est représentée par un nœud avec l'icône 1.

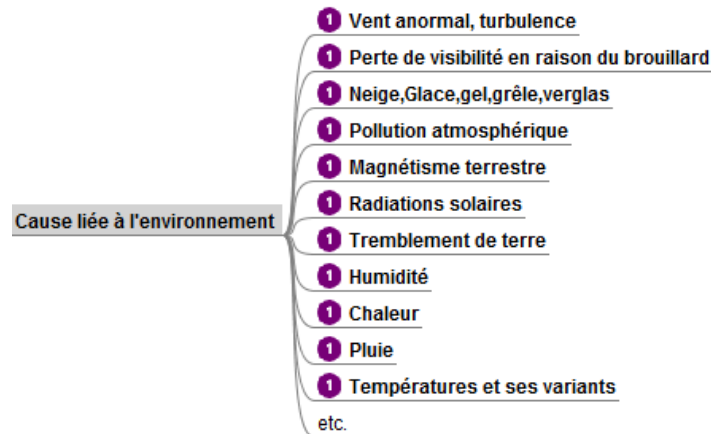


Figure 4.3. Un échantillon d'instances « Cause liée à l'environnement ».

Nous présentons à titre d'exemple, un aperçu d'une liste d'instances pour les erreurs relatives à l'intervention des passagers dans la figure 4.4 :

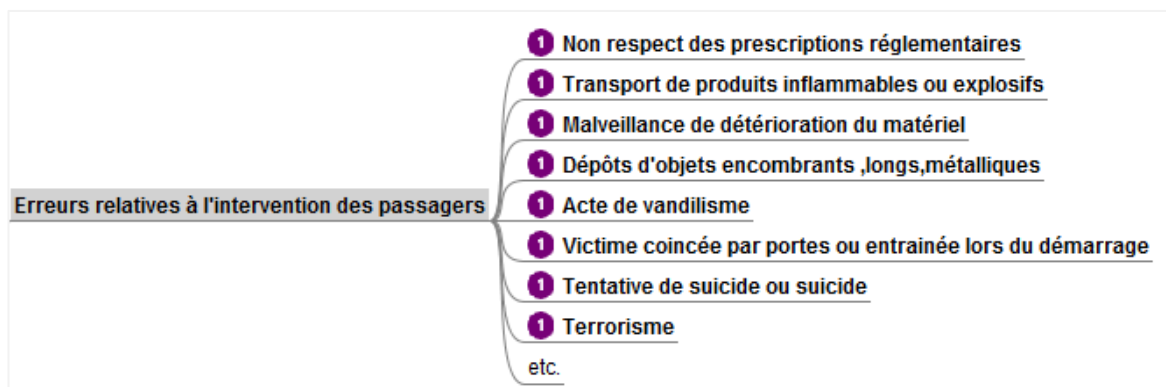


Figure 4.4. Un échantillon d'instances pour les erreurs relatives à l'intervention des passagers (Hadj-Mabrouk A. and Hadj-Mabrouk H., 2003).

Un exemple de classification d'un accident potentiel de type collision dans le transport ferroviaire est présent dans la figure 4.5 :

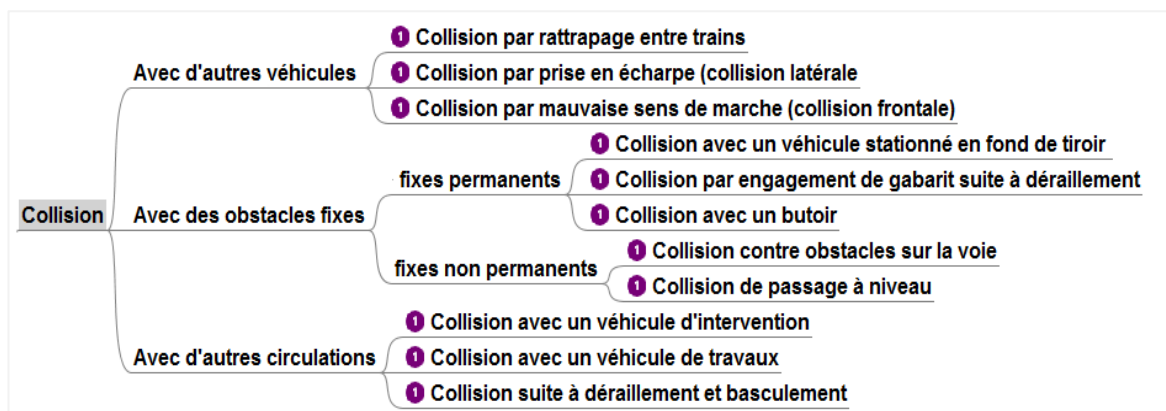


Figure 4.5. Exemple de classification d'un accident potentiel de type collision dans le transport ferroviaire (Hadj-Mabrouk A. and Hadj-Mabrouk H., 2003).

4.4.1. Modèle conceptuel

Pour établir le modèle conceptuel, il nous semble judicieux de choisir une notation conceptuelle qui permet de mettre en évidence tous les concepts, les relations et les contraintes formant l'ontologie. Pour ce faire, nous avons eu recours au langage UML (Unified Modeling Language) qui est une notation standard dans la modélisation orienté-objet. UML est un langage standard de modélisation des systèmes d'information.

Actuellement, l'utilisation d'UML est largement adoptée dans tous les domaines et sa propagation rapide sous entend qu'il fournit une méthode pertinente et évolutive pour la modélisation conceptuelle. Selon (Cranefield S. et Purvis, M., 1999) et (Bergenti F. et Poggi A., 2000), **UML peut être incontestablement considéré comme un langage approprié pour la modélisation d'ontologies**. Pour justifier notre choix concernant l'utilisation du langage UML en tant qu'un langage de modélisation d'ontologies, nous nous basons sur les points suivants :

- UML est un langage formel (ou semi formel) basé sur un méta modèle qui permet de définir : les concepts et les éléments de modélisation et notamment la sémantique de ces éléments ;
- La connaissance exprimée en UML est directement accessible, ce qui facilite d'un côté, sa compréhension par l'être humain ;
- Contrairement aux formalismes de la logique de description qui possèdent une syntaxe linéaire, il existe une représentation graphique standard pour les modèles exprimés en UML. Selon (Cranefield S. et Purvis, M., 1999), « *une représentation graphique possède une importance du fait qu'elle permet aux utilisateurs des systèmes d'information distribués de parcourir une ontologie et de découvrir les concepts qui peuvent apparaître dans leurs requêtes* » ;
- Des nouvelles connaissances peuvent être dérivées à partir de modèles UML par un raisonnement sur leurs contenus (Hadj-M'tir R., 2010).

En conclusion, UML, en particulier, le diagramme de classes d'UML **qui fournit une notation riche pour définir les classes, les concepts et les relations entre les classes**. Par conséquent, le modèle conceptuel que nous proposons va être ainsi exprimé sous forme de diagrammes de classes.

Nous avons donc élaboré un diagramme de classe (Figure 4. 6) qui servira comme base pour l'ontologie noyau. Ce diagramme contient les concepts génériques suivants :

- Contexte général d'apparition de l'accident ;
- Les causes possibles ;
- L'accident potentiel ou l'événement redouté ;
- Le niveau de risque ;
- Les solutions adoptées.

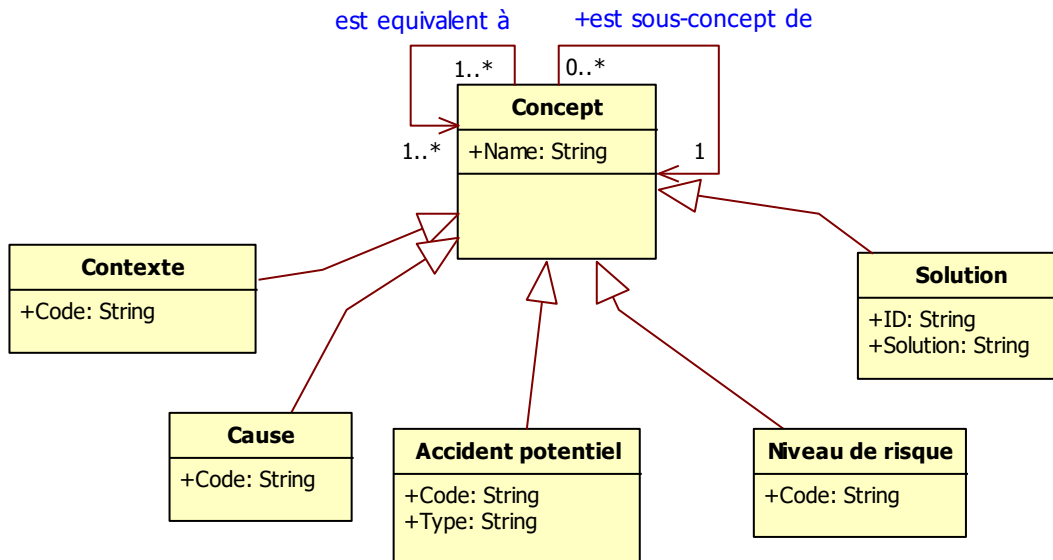


Figure 4.6. Les principales classes de l'ontologie noyau.

La figure 4.7 présente le digramme de classe qui servira comme point d'accès pour l'ontologie de domaine envisagée. Les classes présentes dans l'ontologie noyau seront étendues dans des sous-classes avec les cardinalités correspondantes.

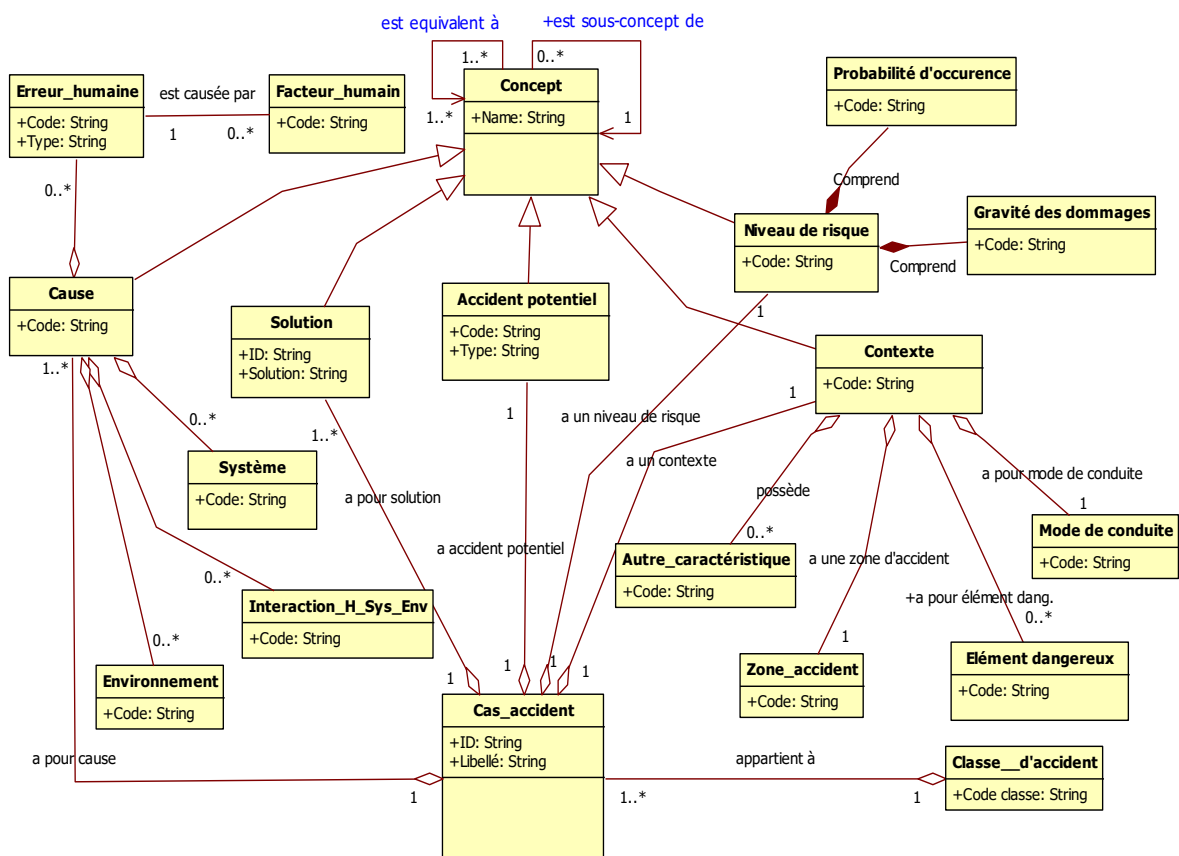


Figure 4.7. Diagramme de classe proposé.

Nous avons convenu également d'intégrer l'erreur humaine dans le modèle conceptuel. Celle-ci reste très présente dans les systèmes de transports et elle n'est pas prise en compte

de façon formelle dans les analyses de sécurité. En effet, selon (Hadj-Mabrouk H. et al., 2012), l'erreur humaine liée à une baisse de la vigilance est devenue un élément critique dans la fiabilité du système Homme-Machine. Les opérateurs humains dans le domaine du transport (les agents conducteurs, les de maintenance, etc.) sont soumis à des contraintes psychologiques, comportementales et physiques. L'erreur humaine est un symptôme révélateur d'une mauvaise organisation du travail, une formation insuffisante ou inadéquate, etc.

Un exemple de cette prise en considération est présent dans la figure 4.8 qui représente une partie de diagramme de classe proposé.

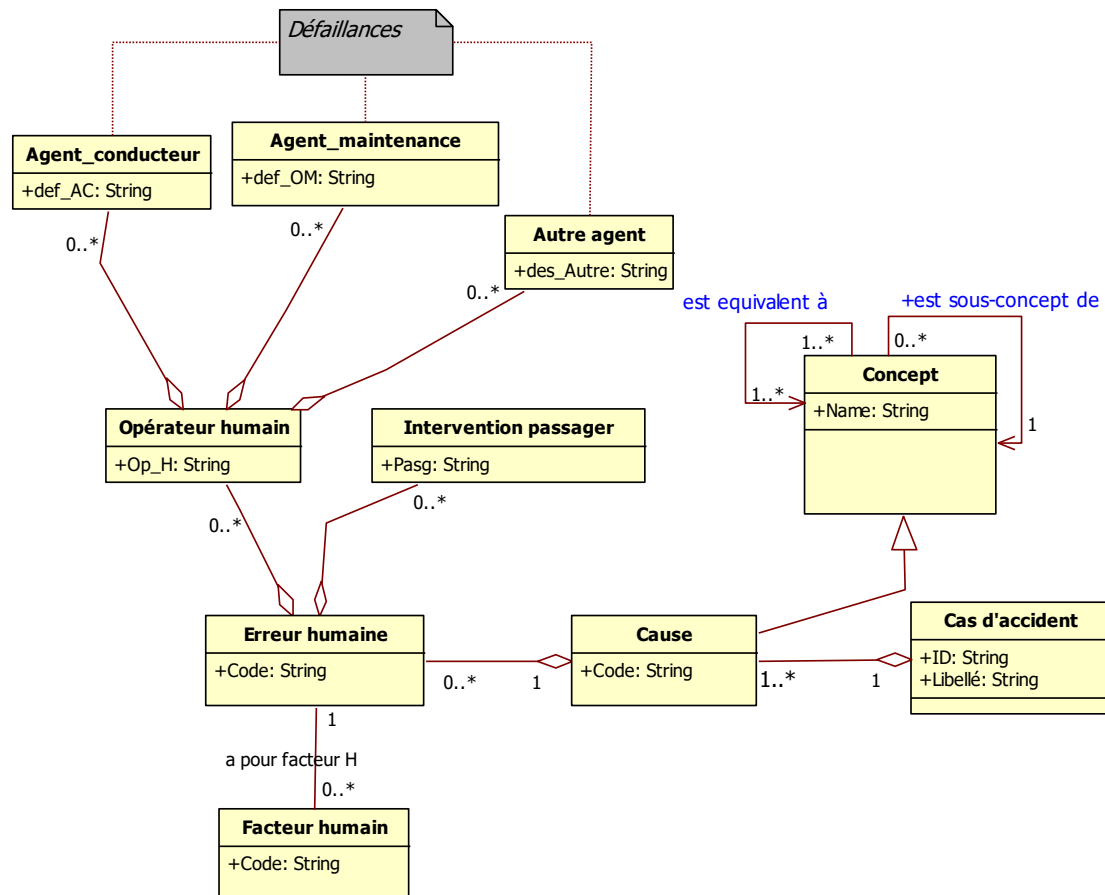


Figure 4.8. Aperçu de la prise en compte de l'erreur humaine dans le diagramme de classe proposé.

4.4.2. Modèle de domaine /de cas

Comme convenu, nous nous sommes basés dans notre travail sur les modèles proposés par Fuchs et Mille (Fuchs B. et Mille A., 2005) et (Abou-Assali A., 2010) (voir section 3.2.3) pour proposer le modèle de domaine/de cas.

Dans notre approche, ce modèle est représenté par une ontologie qui contient le vocabulaire permettant de décrire et de caractériser un scénario d'accident. Il se base essentiellement sur le modèle conceptuel en particulier le diagramme de classe de la figure 4.7 et celui de la figure 4.8.

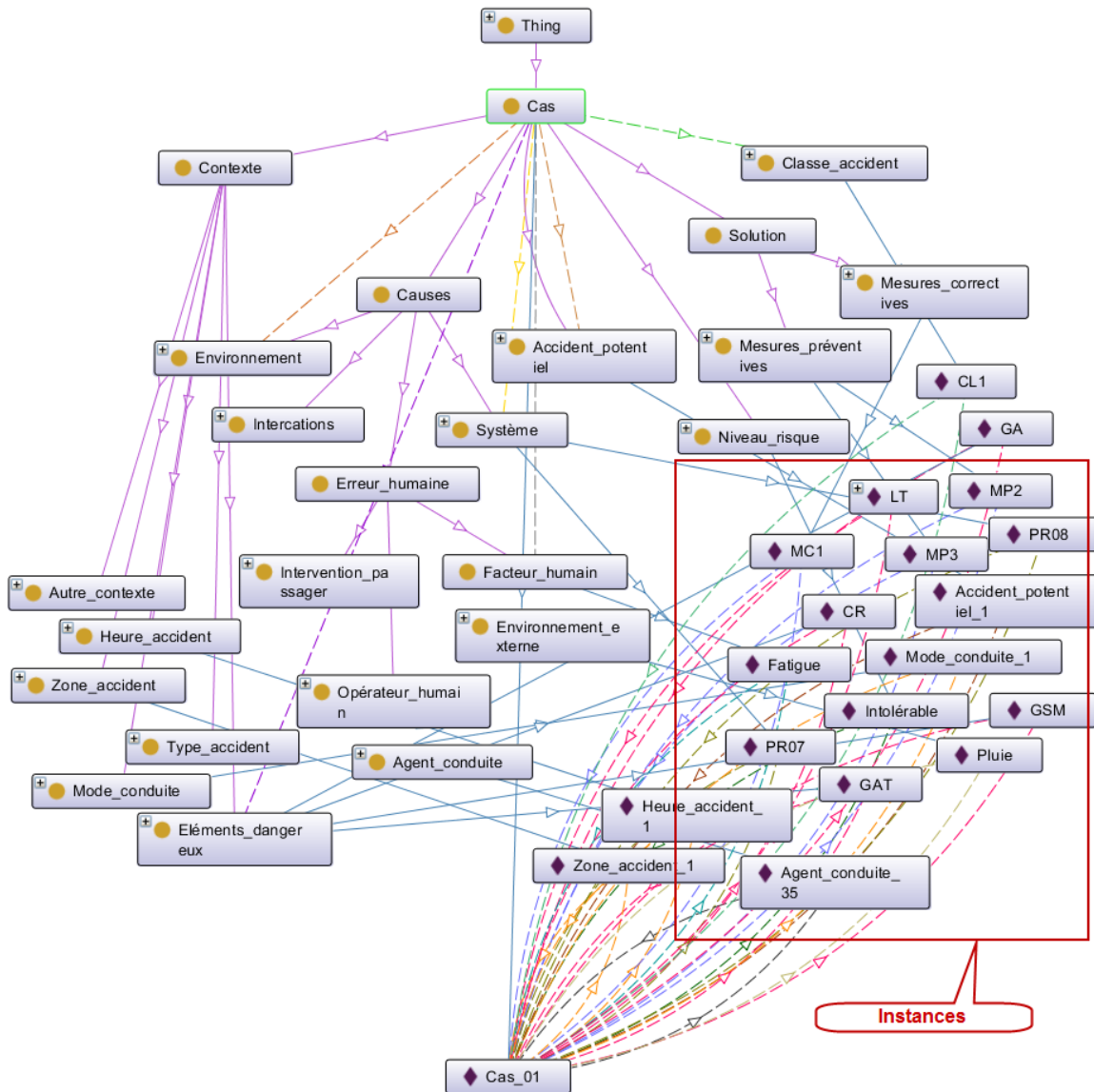


Figure 4.9. Aperçu du modèle de cas.

4.5. Développement de l'ontologie

Comme nous l'avons vu dans la section 2.2.5, la littérature propose plusieurs méthodologies de construction d'ontologies. Certains auteurs ont proposé des méthodologies inspirées de leur expérience de construction d'ontologies (Fernandez M. et al, 1997), (Uschold M. et Grüninger M., 1996) et (Nagypál G., 2005).

Dans le cadre de notre travail, nous utilisons comme convenu une approche utilisée dans (Abou-Assali A. et al., 2008) qui s'inspire largement de la méthodologie *Methontology* de (Fernandez M. et al. 1997).

Le cycle de développement adopté décrit également dans (Maalel A. et al., 2011), fait intervenir quatre principales étapes consécutives (Figure 4.10). L'étape de formalisation dans *Methontology* (voir section 2.2.5.2) fait partie désormais de l'étape de conceptualisation :

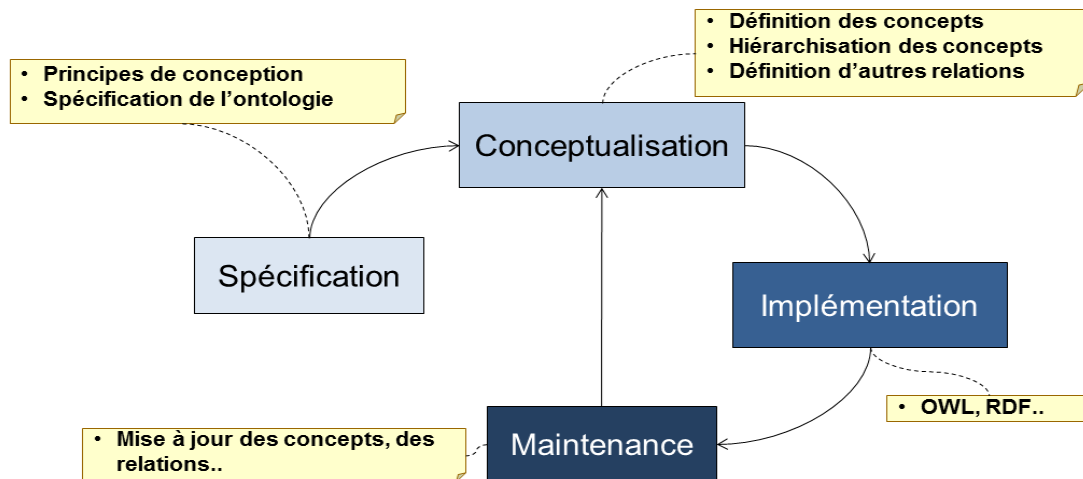


Figure 4.10. Cycle de développement d'ontologie (Abou-Assali A. et al., 2008).

Spécification

Cette phase consiste à donner une description générale de l'ontologie. Deux aspects sont pris en considération :

- **Les principes de conception de l'ontologie** qui tiennent compte de recommandations telles la normalisation des noms, etc. (voir section 2.2.8) ;
- **La spécification de l'ontologie** qui décrit le but, la portée, et les concepts essentiels de l'ontologie.

Conceptualisation

Cette phase consiste à produire le modèle conceptuel de l'ontologie qui contient les concepts du domaine et leurs propriétés (voir section 4.4.2 : le modèle conceptuel) :

- **La définition des concepts** : il s'agit d'identifier les concepts à partir des ressources qui ont été spécifiées au départ dans la phase de spécification. Ceci à pour but de recenser les connaissances de domaine de la sécurité ;
- **Hiérarchisation des concepts** : il s'agit d'organiser les concepts dans une hiérarchie qui exprime la relation de Spécialisation/Généralisation entre les concepts.

Implémentation

Cette phase consiste à passer du modèle conceptuel à un modèle implémenté dans l'un des langages d'ontologie comme l'OWL, RDF etc. L'éditeur d'ontologies *Protégé* est utilisé alors pour implémenter notre ontologie. Parmi les raisons qui ont motivé le choix de Protégé, nous citons :

- Protégé est un éditeur open source et gratuit ;
- Il permet d'importer et d'exporter des ontologies dans les différents langages d'implémentation d'ontologies (RDF-Schéma, OWL, DAML, OIL,...etc.) ;
- Il possède une interface modulaire, ce qui permet son enrichissement par des modules additionnels (plug-ins) ;
- Il permet l'édition et la visualisation d'ontologies ;
- Il permet le contrôle de la cohérence de l'ontologie par des vérifications de

contraintes.

Dans ce qui suit, nous allons utiliser quelques règles de transformation pour la traduction du modèle UML vers le modèle OWL. Ces règles ont été utilisées par Hadj-M'tir (Hadj-M'tir R., 2010) qui s'est basée sur l'article « *La Création d'Ontologies Web Sémantique avec Protégé-2000*¹ » :

- Une classe UML est traduite en une classe OWL ;
- Les attributs de classe UML sont traduits en propriétés OWL ;
- Les relations de généralisation UML sont traduites en relations de sous-classe OWL ;
- Pour les relations d'agrégation et de composition UML qui ne sont pas supportées par le langage OWL. Celles-ci pourront être traduites tout simplement en relations de sous-classe OWL de type « composé de » ou « partie de ».

Le tableau 4.1 présente les différentes transformations que nous venons de présenter.

UML		OWL
Classe		owl : Class
Nom de la classe		rdf : ID
Relation de généralisation		rdfs : subclassOf
Attribut de la classe		owl : DatatypeProperty
Type d'attribut		rdf : datatype
Relation binaire / relation unidirectionnelle		owl : ObjectProperty
Classe source d'une relation		rdfs : domain
Classe cible d'une relation		rdfs : range
Cardinalités	1	owl : cardinality 1
	0..*	owl : minCardinality 0
	0..1	owl : minCardinality 0 owl : maxCardinality 1
	N	owl : cardinality n
	M..N	owl : minCardinality m owl : maxCardinality n

Tableau 4.1. Les différentes transformations du langage UML vers OWL (Hadj-M'tir R, 2010)

Chaque classe dans le modèle conceptuel (section 4.4.2) de l'ontologie est traduite à une classe OWL portant la même identification. Si nous prenons par exemple la classe « *Concept* », elle est définie avec OWL comme suit :

¹ <http://www.cetic.be/>

```
<owl:Class rdf:ID="Concept">
```

Pour la relation d'agrégation/composition, nous proposons de la représenter tout simplement sous forme d'une relation (de type « composé de » ou « partie de ») entre classe et sous-classe. Selon (Hadj-M'tir R., 2010), la sous-classe joue le rôle d'un composant ou d'une partie de la classe.

Prenons par exemple la relation d'agrégation donnée par la Figure 4.11 qui montre qu'une cause d'un accident potentiel peut être formée par un ou plusieurs causes liées à l'environnement.

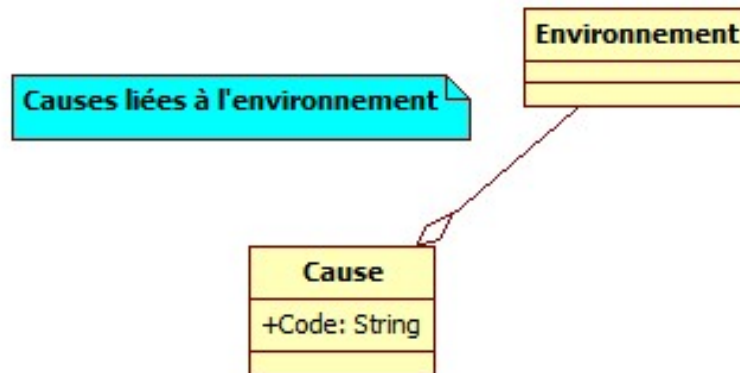


Figure 4.11. Représentation d'une relation d'agrégation.

Cette relation peut être représentée avec OWL par une relation nommée « composé_de » entre la classe « Cause » et la sous-classe « Environnement » avec une cardinalité de type 1..N de côté de la sous-classe « Environnement » :

```
<owl:Class rdf:ID="Cause"/>
<owl:Class rdf:ID="Environnement">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:minCardinality
rdf:datatype=http://www.w3.org/2001/XMLSchema#int>1</owl:minCardinality>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID="composé_de"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf rdf:resource="#Cause"/>
</owl:Class>
<owl:ObjectProperty rdf:about="#composé_de">
  <rdfs:domain rdf:resource="#Cause"/>
  <rdfs:range rdf:resource="#Environnement"/>
</owl:ObjectProperty>
```

Les figures 4.12 et 4.13 présentent successivement la hiérarchie des classes de l'ontologie dans l'éditeur protégé et à travers le plugin OWL-Viz :

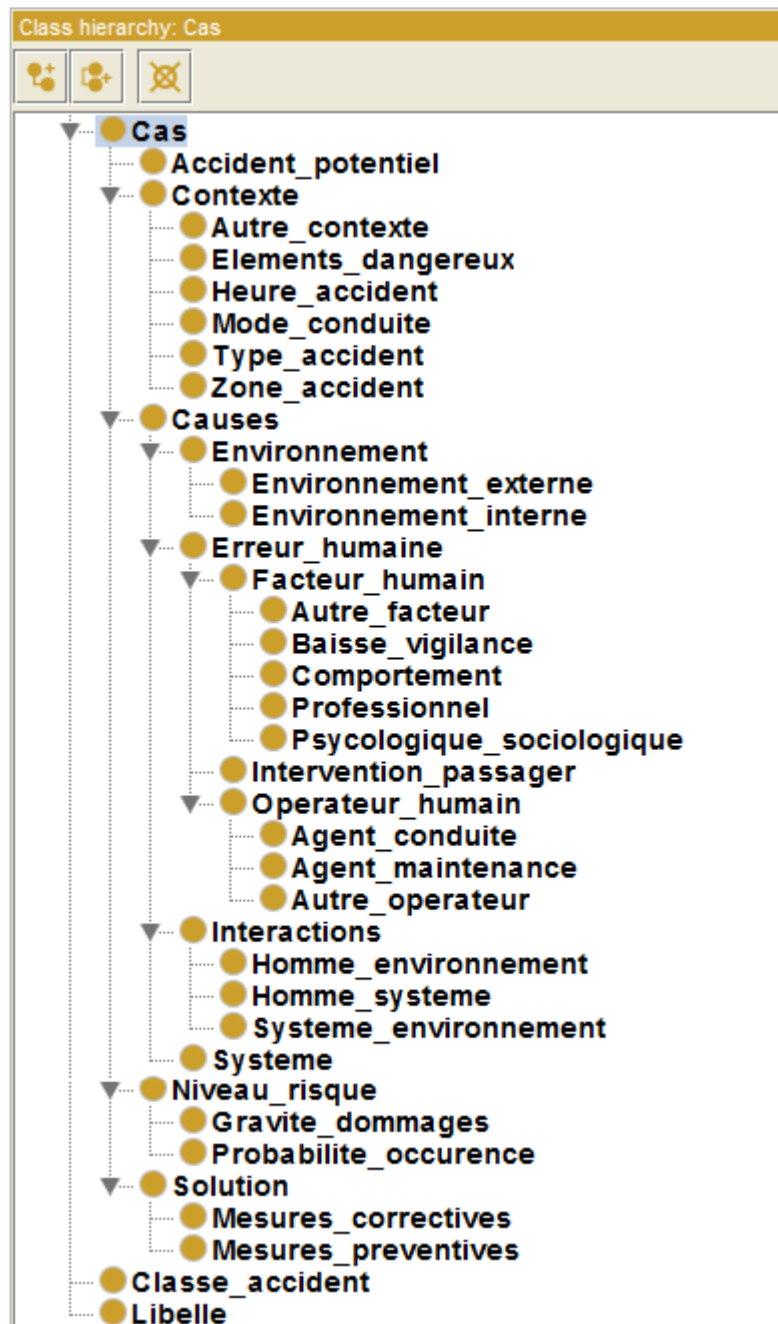


Figure 4.12. Hiérarchie des classes et des sous-classes sous Protégé.



Figure 4.13. Une partie de l'ontologie de domaine avec le plugin OWL-Viz (Maalel A. et al, 2011).

Maintenance :

Cette phase consiste à maintenir l'ontologie. Nous pouvons ainsi mettre à jour, ajouter et même supprimer des concepts. La maintenance désigne la modification apportée à l'ontologie après sa mise en œuvre, pour en corriger les fautes et améliorer son efficacité. Au cours de l'implémentation de l'ontologie, cet aspect de maintenance a été sollicité plusieurs fois ce qui a permis d'aboutir enfin à une ontologie exploitable.

4.6. Raisonnement à partir de cas

Dans ce paragraphe, nous détaillons les différentes phases de cycle de RàPC et nous mettons en œuvre les interactions entre les différentes méthodes et techniques adoptées issues de la fouille de données que nous utilisons.

4.6.1. Élaboration d'un cas cible

Comme nous l'avons déjà dit, cette phase a pour but d'élaborer un cas cible (un problème pour lequel nous cherchons une solution). Dans notre approche, l'élaboration se fait en utilisant des instances de l'ontologie de domaine. Dans un premier temps, l'utilisateur est censé fournir une description initiale pour caractériser son problème (Les valeurs des

attributs des concepts décrivant le scénario d'accident comme l'accident potentiel, les causes, le contexte d'apparition de l'accident...etc.)

Dans le cas où cette description n'est pas suffisante (présence des données manquantes) ou bien bruitée (entachée d'erreurs), l'utilisateur devra enrichir/affiner sa description à partir d'une liste d'instances enrichissantes.

Dans notre approche (Figure 4.14), nous utilisons **les règles d'association** pour élaborer cette liste. Nous avons effectué ce choix car cette technique de fouille de donnée permet de rechercher des relations d'association et de corrélation intéressantes parmi un grand ensemble de données, de découvrir des règles et des connaissances sans intervention de l'expert du domaine et notamment générer des nouvelles règles faciles à comprendre et à exploiter (des règles de type Antécédent → Conséquent). L'utilisateur pourra ainsi utiliser les règles produites pour mieux caractériser la description de son cas cible.

Cette aide présentée par le système aura des conséquences sur les autres phases de raisonnement.

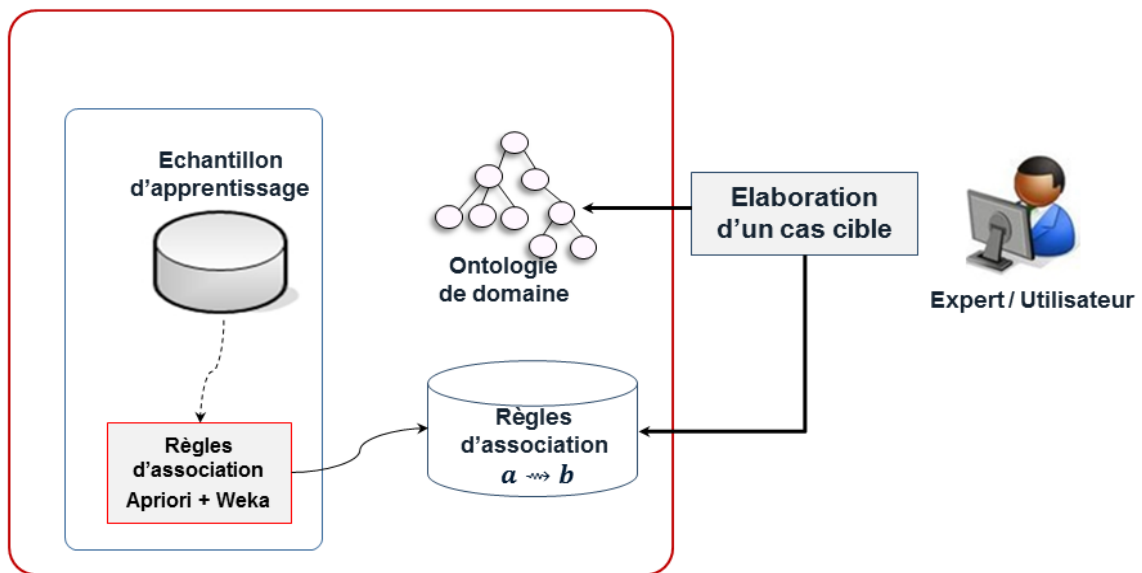


Figure 4.14. Élaboration d'un cas cible.

Par approche support-confiance, nous désignons les algorithmes d'extraction qui recherchent les règles d'association dont le support et la confiance dépassent des seuils fixés au préalable par l'utilisateur. Comme convenu, nous utilisons l'algorithme Apriori pour la génération des règles d'association.

Selon (Lallich S. et Teytaud O., 2004), les algorithmes du type Apriori, fondés sur le support et la confiance des règles, ont apporté une bonne solution au problème de l'extraction de règles. Cependant, ils produisent une grande masse de règles dont parfois, sans intérêt. Pour remédier à ceci, il faut disposer d'autres mesures venant compléter le support et la confiance.

Nous ne prétendons pas définir des nouvelles mesures, nous utilisons à cet effet une mesure très répandue, le lift de (Brin S. et al., 1997) qui vient de compléter les deux mesures utilisées par Apriori.

Soient $n(A)$ et $n(B)$ les nombres de transactions qui réalisent respectivement les items de A et de B , $n(AB)$ le nombre de celles qui réalisent à la fois A et B .

- **Le support** d'une règle est la proportion de transactions qui réalisent à la fois A et B :

$$Supp(A \rightarrow B) = P(A \cup B) = \frac{n(A \cup B)}{n}$$

- **La confidence** est la proportion de transactions qui réalisent B , parmi celles qui réalisent A , c'est-à-dire la fréquence relative conditionnelle de B sachant A :

$$Conf(A \rightarrow B) = \frac{P(A \cup B)}{P(A)} = \frac{n(A \cup B)}{n(A)} = 1 - \frac{n(A \cup \bar{B})}{n(A)}$$

- **Le lift** (Brin et al, 1997), défini par $lift(A \rightarrow B) = \frac{P(A \cup B)}{P(A)P(B)}$, s'interprète comme le quotient du nombre d'exemples observé par celui attendu sous l'hypothèse d'indépendance de A et B ;
- S'exprimant à partir des seuls exemples, il est symétrique, puisque les règles $A \rightarrow B$ et $B \rightarrow A$ ont les mêmes exemples. Le « lift » d'une règle $A \rightarrow B$ mesure l'amélioration apportée par la règle d'association par rapport à un jeu de transactions aléatoire (où A et B seraient indépendants). **Un « lift » supérieur à 1 explique une corrélation positive de A et B , et donc le caractère expressif de la règle d'association.**

4.6.2. Phase de remémoration

La phase de remémoration dans un processus de RàPC consiste à explorer et chercher dans toute la base de cas un ou plusieurs cas passés résolus jugés similaires au cas cible (Cordier et al, 2006). Dans notre approche, la phase de recherche est nettement améliorée parce qu'elle opère seulement sur des cas appartenant à une même classe (Figure 4.15).

La méthode de **l'arbre de décision** (issue de la fouille de données) qui permet d'aboutir à partir d'un ensemble d'exemples d'apprentissage à une structure arborescente est utilisée pour classifier un nouveau cas (cas cible). La classification permet de se pointer à la classe d'appartenance d'un cas avant de rechercher des cas sources similaires. Par conséquent, le calcul de similarité prend appui seulement sur l'ensemble de cas de la classe d'appartenance du cas cible.

La tâche de classification est une tâche cruciale car elle a une incidence sur les étapes de raisonnement en aval et sur la complémentarité souhaitée entre ces phases. Donc, d'une part, nous limitons l'utilisation des ressources et d'autre part, nous gagnons en termes de sémantique puisque les cas de la même classe partagent généralement les mêmes caractéristiques et sémantiquement très proches.

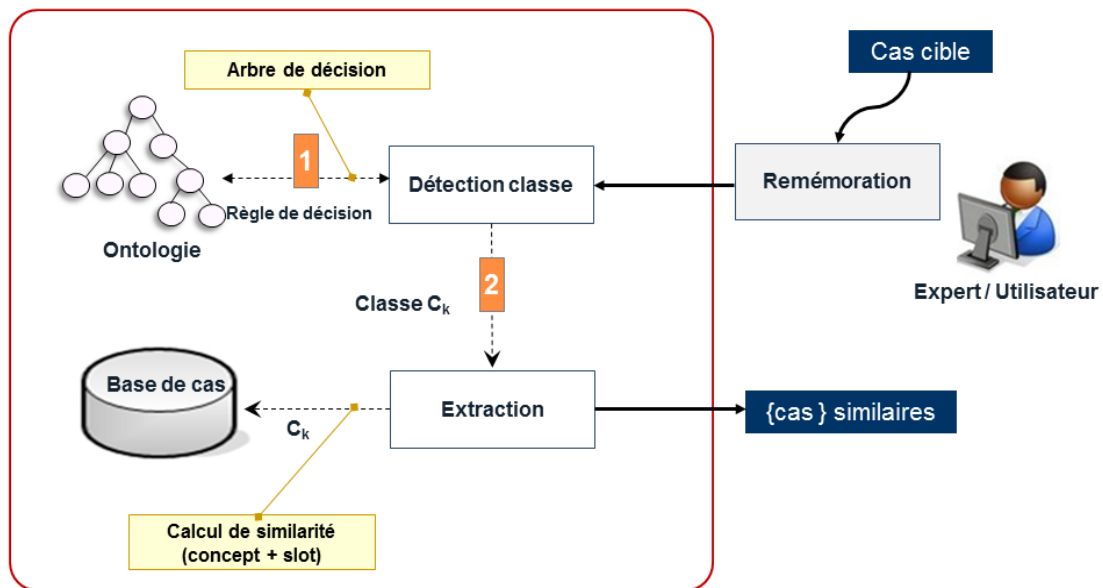


Figure 4.15. Approche de remémoration retenue.

Nous détaillons notre démarche de remémoration dans le paragraphe suivant.

4.6.2.1. Classification du cas cible

La figure 4.16 résume les différentes étapes que nous avons suivi pour d'une part (1) exploiter l'échantillon d'apprentissage à travers la technique d'arbre de décision (générer des règles de classification) et d'autre part, (2) filtrer et affiner ces règles pour enfin valider et formater les règles produites.

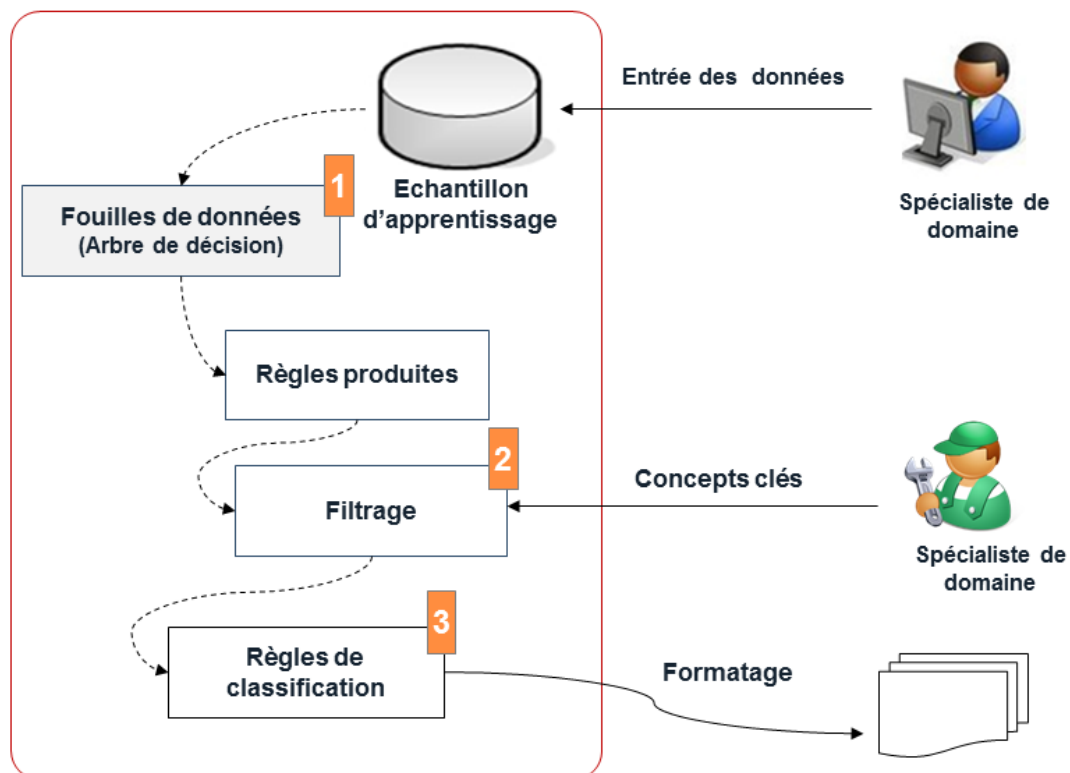


Figure 4.16. Démarche de génération et de représentation des règles de classification.

L'échantillon d'apprentissage constitue l'ensemble expérimental sur lequel va opérer le processus pour engendrer de nouvelles connaissances, des concepts, des caractérisations ou des règles. Pour aboutir à des règles consistantes, l'ensemble d'exemples doit être non bruité et suffisamment représentatif du domaine d'application. Un exemple est qualifié de positif s'il n'appartient pas à une classe particulière ou de négatif s'il n'appartient à cette classe.

Soit $\Omega = \{\omega_1, \omega_2, \dots, \omega_i\}$ l'échantillon d'apprentissage, c'est l'ensemble de descripteurs des cas qui va être utilisé pour la génération de l'arbre de décision. Chaque cas ω_i est décrit par une série de variables X_1, X_2, \dots, X_j dites variables descriptives. Pour illustrer un exemple simplifié, considérons notre problème de classification des cas d'accidents, ou nous précisons un exemple de 3 variables descriptives X_1, X_2, X_3 et auxquels est associée une décision Y : « Classe d'appartenance » ;

- X_1 : Accostage : $x_1^1 = OUI$ et $x_1^2 = NON$
- X_2 : Initialisation : $x_2^1 = OUI$ et $x_2^2 = NON$
- X_3 : Conduite manuelle : $x_3^1 = OUI$ et $x_3^2 = NON$

A partir de cet exemple, nous construisons un arbre dit de décision tel que :

- Chaque nœud correspond à un test sur la valeur $X_j(\omega)$ d'un attribut X_j qui possède plusieurs valeurs possibles ($(X_{j1}(\omega), X_{j2}(\omega) \dots X_{jn}(\omega))$) ;
- A chaque feuille est associée une valeur c_k de l'attribut cible Y .

Nous appliquons l'algorithme C4.5 intégré dans SIPINA (Zighed D. et Ricco R, 1996) sur l'échantillon d'apprentissage Ω_a (Tableau 4.2). Nous pouvons déduire plusieurs règles de classification $\{R_1, R_2, \dots, R_n\}$ qui sont de la forme :

SI (Condition) ALORS Conclusion.

ω	$X_1(\omega)$	$X_2(\omega)$	$X_3(\omega)$	$X_n(\omega)$	$Y(\omega)$
ω_1	<i>OUI</i>	<i>NON</i>	<i>NON</i>	...	<i>C5</i>
ω_2	<i>NON</i>	<i>OUI</i>	<i>NON</i>	...	<i>C2</i>
ω_3	<i>NON</i>	<i>NON</i>	<i>OUI</i>	...	<i>C9</i>
ω_i

Tableau 4.2. Échantillon d'apprentissage Ω_a

Nous donnons un exemple de deux règles de décision construites :

R1 : SI (Ligne = OUI) et (Conduite_manuelle = OUI)

ALORS = CL11_Controle_de_vitesse avec une probabilité = 0.66

R2 : SI (Ligne = OUI) et (Localisation_des_trains = OUI)

ALORS = CL3_Localisation_des_trains avec une probabilité = 1

4.6.2.2. Extraction des cas similaires

Dans ce paragraphe, nous allons présenter les mesures de similarités que nous utilisons et notamment l'algorithme de calcul de similarité proposé suivi d'un exemple illustratif.

4.6.2.2.1. Mesures de similarités

L'extraction de cas se base sur la notion de similarité, dans notre cadre ou nous utilisons une ontologie regroupant les différents concepts et instances de domaine, la similarité à mesurer est plutôt d'ordre sémantique et non pas seulement une similarité purement syntaxique, ce choix va être débattu dans ce présent paragraphe.

Nous avons étudié et mentionné dans la section 2.2.9, qu'il y'a quatre familles d'approches pour l'identification de la similarité sémantique. Nous ne prétendons pas définir de nouvelles mesures de similarité. Nous utilisons donc, les mesures qui nous conviennent le mieux. Nous nous basons dans notre approche conjointement sur les mesures de la similarité basée concept de (Wu J. et Palmer M., 1994) améliorée dans (Zargayouna H. et Salotti S, 2004) et celle de la similarité basée slot de (Zhang J. et al, 2006).

Soit C_{cible} un cas cible pour lequel nous cherchons des cas similaires, ou c_i est un attribut simple ou bien complexe :

$$C_{cible} = \{c_i : 1 \leq i \leq n, n \in \mathbb{N}^*\}$$

Soit Ω la sous base de cas relative à une classe C_k ,

$$\Omega = \{Source_j : 1 \leq j \leq k, k \in \mathbb{N}^*\}$$

Où $Source_j$ est un cas source appartenant à la classe C_k et s_{jl} est l'attribut l du cas $Source_j$.

$$Source_j = \{s_{jl} : 1 \leq l \leq m, m \in \mathbb{N}^*\}$$

→ **La similarité basée-concept, sim_{cpt}** est définie comme suit par (Wu J. et Palmer M., 1994) :

Pour chaque attribut complexe ; $c_i \in C_{cible}$ et $s_{jl} \in Source_j$

$$sim_{cpt}(c_i, s_{jl}) = w_{ci} \times \frac{2 \times Profondeur(LCS(c_i, s_{jl}))}{Profondeur(c_i) + Profondeur(s_{jl})} \quad (i)$$

Où w_{ci} est un poids associé à l'attribut c_i variant entre $\{0, 1, 2\}$. Profondeur est la profondeur d'un concept (ou d'une instance) dans l'ontologie, et LCS est le plus petit subsumant commun (*Least Common Subsumer*) de deux instances. Dans un cas particulier, quand c_i et s_{jl} représentent la même instance, nous avons :

$$Profondeur(LCS(c_i, s_{jl})) = Profondeur(c_i)$$

Selon Zargayouna et Salotti (Zargayouna H. et Salotti S, 2004), la mesure (i) de (Wu J. et Palmer M., 1994) est intéressante mais présente une limite car elle vise essentiellement à détecter la similarité entre deux concepts par rapport à leur distance de leur LCS.

Plus ce subsumant est général, moins ils sont similaires (et inversement). Dans (Zargayouna H. et Salotti S, 2004), les auteurs ont ajouté à la mesure (i) une

fonction $spec(c, s)$, qui calcule la spécificité de deux concepts par rapport au concept le plus bas de l'ontologie. Cette fonction servira à pénaliser les concepts qui ne sont pas dans la même lignée. Nous donnons :

$$sim_{cpt}(c_i, s_{jl}) = w_c \times \frac{2 \times Profondeur(LCS(c_i, s_{jl}))}{Profondeur(c_i) + Profondeur(s_{jl}) + spec(c_i, s_{jl})} \quad \text{(i)*}$$

Avec :

- $spec(c_i, s_{jl}) = profondeur(c_i) \times distance(LCS(c_i, s_{jl}), c_i) \times distance(LCS(c_i, s_{jl}), s_{jl})$
- $distance(LCS(c_i, s_{jl}), s_{jl})$: la distance en nombre d'arcs entre $LCS(c_i, s_{jl})$ et s_{jl}

Dans un cas particulier, quand c_i et s_{jl} représentent la même instance, nous avons :

$$spec(c_i, s_{jl}) = 0$$

→ **La similarité basée-slot sim_{slt} est définie pour tout attribut c_{is} simple de C_{cible} comme suit :**

$$sim_{slt}(c_{is}, s_{jl}) = \frac{\sum_{s \in CS} sim(c_{is}, s_{jl})}{|CS|} \quad \text{(ii)}$$

Où CS est l'ensemble des attributs simples en commun entre c_{is} et s_{jl} (Common Slots), $|CS|$ est sa cardinalité, c_{is} représente l'attribut simple de C_{cible} et $sim(c_{is}, s_{jl})$ est la similarité entre ces deux attributs.

Dans notre approche de calcul de similarité, nous ne considérons pas CS. Nous considérons uniquement les deux modes (Différent, Égal), et donc $sim_{slt}(c_{is}, s_{jl})$ serait définie comme suit :

$$sim_{slt}(c_{is}, s_{jl}) = \begin{cases} 1 & \text{si } (w_{c,s} = \text{Égal}) \wedge (v_{cis} = v_{s_{jl}}) \\ 0 & \text{sinon} \end{cases} \quad \text{(ii)*}$$

Où $w_{c,s}$ est le mode associé (Différent, Égal) à l'attribut c_{is} , v_{cis} est la valeur de cet attribut dans c_{is} et $v_{s_{jl}}$ est la valeur de cet attribut dans s_{jl} .

→ **La mesure globale de similarité** entre les deux attributs c_i et s_{jl} est définie par la formule suivante (Zhang J. et al, 2006) :

$$sim_{globale}(c_i, s_{jl}) = (1 - \alpha) \times sim_{cpt}(c_i, s_{jl}) + \alpha \times sim_{slt}(c_{is}, s_{jl}) \quad \text{(iii)}$$

Où α est un paramètre qui permet le contrôle de l'importance de la similarité basée-slot dans le calcul. Dans le cas où α prend la valeur 1, la similarité basée concept va être négligée et les attributs complexes n'entrent pas dans le calcul de la similarité.

Pour calculer la similarité basée-concept, chaque attribut complexe c_i du cas cible C_{cible} est comparé avec tous les attributs s_{jl} du cas source $Source_j$. Pour tout attribut supposé complexe $C_i \in C_{cible}$, nous considérons que l'attribut correspondant (le plus similaire) à C_i dans le cas $Source_j$ est celui qui a la valeur de similarité la plus élevée, nous écrivons :

$$sim_{cpt}(c_i, s_{jl}) = \max_{1 \leq l \leq n} (sim(c_i, s_{jl})), \quad s_{jl} \in Source, n \in \mathbb{N}^*$$

Pour extraire les cas similaires et en s'appuyant respectivement sur les formules **(i)***, **(ii)*** et **(iii)**, nous avons proposé un algorithme pour le calcul de la similarité. Nous donnons :

- **BC** : Base de cas ;
- **C_k** : Classe d'appartenance du cas cible c ;
- **nbcas**: Nombre de cas appartenant à la classe C_k ;
- **nb**: Nombre d'attributs/instances de cas cible C ;
- **nbs** : Nombre d'attributs de cas source courant ;
- **c** : Cas cible ;
- **S_i** : Cas source i
- **c_j** : j^{ième} attribut du cas cible ;
- **c_j.s**: j^{ième} attribut simple du cas cible ;
- **s_i^l** : l^{ième} attribut du cas source s_i ;
- **w_c^j**: Poids correspondant de chaque attribut du cas cible ;
- **P** : Somme des poids des attributs ;
- **α**: Paramètre permettant de contrôler l'importance de calcul de Sim_{slt} ;
- **T_{cpt}[]** : Tableau d'entiers pour mémoriser les valeurs de la similarité $sim_{cpt}(c_j, s_i^l)$;
- **T []** : Tableau d'entiers (le résultat retourné), pour mémoriser les valeurs de la similarité globale entre le cas cible c et l'ensemble des cas sources S_i.

Algorithme 4.1. Calcul_similarité (BC, C, classe C_k)

Entrée : C_k : classe de cas, BC : Base de cas (Ontologie_domaine), C : cas_cible

Sortie : T : Tableau (le résultat retourné) + Les cas les plus similaires au cas_cible C

0- Début

1- Obtenir nbcas //Nombre de cas appartenant à la classe C_k

2- Obtenir nb //Nombre d'attributs/instances de cas cible C_{cible}

3- Pour tout cas s_i (i=1..nbcas) faire // Parcours des cas sources appartenant à C_k

Obtenir nbs //nbs : Nombre d'attributs de cas source courant

$sim_{globale}(c, s_i) \leftarrow 0$ //initialisation de la similarité globale

$P \leftarrow 0$ //Initialisation de la variable servant au calcul global de la similarité

$somme_{slt} \leftarrow 0$ //Initialisation de la somme courante de la sim. basée slot du cas source si

Pour tout attribut c_j ($j=1..nbc$) faire //Parcours de tous les attributs de cas cible c

Récupérer w_c^j //le poids correspondant de l'attribut courant (attribut j du cas cible c)

Si c_j est un attribut simple alors

$P \leftarrow P + w_c^j \times \alpha$

$l \leftarrow 1$ //Initialisation du compteur

Tant que (attribut $c_j \neq$ attribut s_i^l) ou ($l = nbs$) faire

$l \leftarrow l+1$ //passer à l'attribut suivant du cas source

Fin Tant que

si (attribut $c_j =$ attribut s_i^l) alors

$somme_{slt} = somme_{slt} + (w_c^j \times sim_{slot}(c_j, s, s_i^l, s))$

Fin si

Si non // c_j est un attribut complexe

$P \leftarrow P + w_c^j \times (1 - \alpha)$

Pour tout attribut s_i^l ($l=1..nbs$) faire //Parcours de tous les attributs de cas source courant si

$$T_{cpt}[j] \leftarrow sim_{cpt}(c_j, s_i^l) = w_c^j \times \frac{2 \times Prof(LCS(c_j, s_i^l))}{Prof(c_j) + Prof(s_i^l) + spec(c_j, s_i^l)}$$

Fin pour

$$sim_{cpt}(c_j, s_i) = \max_{1 \leq l \leq nbs} (sim_{cpt}(c_j, s_i^l)) \text{ dans } T_{cpt}[j]$$

//Choisir la similarité sim_{cpt} ayant la valeur la plus élevée

Fin Si

$$sim_{globale}(c, s_i) = sim_{globale}(c, s_i) + (1 - \alpha) \times sim_{cpt}(c_j, s_i^l) + \alpha \times sim_{slt}(c_j, s_i^l)$$

//Déterminer la similarité globale

Fin pour

$$similarité(c, s_i) \leftarrow sim_{globale}(c, s_i) \times 100 / P$$

$T[j] \leftarrow similarité(c, s_i)$ //Mémoriser $sim_{globale}$ entre c et le cas source courant si

Fin Pour

4- Retourner T et le cas le plus similaire dans T

5- Fin

La complexité algorithmique est un concept primordial, elle permet de prouver si un algorithme A est meilleur qu'un algorithme B et s'il est optimal ou s'il ne doit pas être utilisé. Elle représente la mesure du nombre d'opérations essentielles qu'il effectue sur un jeu de données. Elle est exprimée en fonction de la taille n des données.

Pour calculer la complexité de l'algorithme que nous venons de présenter, nous devons calculer tout d'abord, la complexité des traitements élémentaires :

La complexité du traitement **conditionnel** relatif au calcul de la similarité est de $O(1)$, de ce fait, la complexité de la boucle **Pour tout attribut c_j ($j=1..nbc$)** est $O(1) \times O(nbc) = O(nbc)$. Pour la boucle **Pour tout attribut s_i^l ($l=1..nbs$)**, la complexité vaut $O(nbs \times nbc)$, elle est donc $O(nbcas \times nbs \times nbc)$ pour la boucle **Pour tout cas**

s_i ($i=1..nbcas$). En ignorant les complexités des traitements élémentaires simples, la recherche du cas ayant la valeur de similarité la plus élevée dans le tableau $T[i]$ est de complexité $O(nbcas)$.

En conclusion, la complexité de l'algorithme est alors $O(\max(O(nbcas \times nbs \times nbc), O(nbcas))) = O(nbcas \times nbs \times nbc)$.

4.6.2.2.2. Exemple illustratif

Pour illustrer le calcul de la similarité, nous avons eu recours à un exemple volontairement simplifié. Nous présentons à cet effet les différentes itérations de calcul de la similarité entre un cas source particulier (cas_source_test) et un autre cas supposé cible (cas_cible_test).

Les deux tableaux 4.3 et 4.4 présentent respectivement les instances présentes, leur type et leur poids respectifs.

CAS_cible_test Nombre d'instances : 11											
Attr.	1	2	3	4	5	6	7	8	9	10	11
	LT	GSM	CR	GA	Station	PR7	PR8	PR14	Pluie	Accident potentiel1	FH Psychologique
Type	S	S	S	S	S	S	S	S	S	S	C
Poids	2	1	1	1	1	2	2	2	1	2	2

Tableau 4.3. Description du cas cible

Cas_source_test Nombre d'instances : 12											
1	2	3	4	5	6	7	8	9	10	11	12
LT	GSM	GAT	GR	GA	Station	PR7	PR8	Défaillance conducteur_1	Pluie	Accident potentiel1	Fatigue

Tableau 4.4. Description du cas source_test

Pour illustrer le calcul de la similarité, nous procédons à une simulation d'une itération de calcul de la similarité entre un cas cible et un cas source disposants respectivement de 11 attributs (dont un est complexe) et de 12 pour le cas source.

A noter que ce calcul sera effectué de la même façon sur l'ensemble de cas sources de la classe d'appartenance du cas cible. Un aperçu des différentes valeurs de calcul que nous avons utilisé pour la similarité basée-concept est présent dans la figure 4.17. Les différentes itérations avec les mesures de similarités basées slot et basées concept correspondants sont présentées ci-après :

Nous fixons α à 0,8 et nous avons initialement $nbcas= 1$ et $nbc= 11$:

i=1 (Première itération- Cas source test)

Nous avons : $nbs= 11$ | $sim_globale=0$ | $P=0$ | $somme_slt=0$

■ $j=1$ (Premier attribut du cas cible) :

$w=2$ | $P=1,6$ | $Attribut_simple=OUI$ | $l=1$ $somme_slt=0+2=2$ | $sim_globale=0+0,8 \times 1,6=2$

■ $j=2$ | $w=1$ | $P=2,4$ | $Attr_simple=OUI$ | $l=1$ $l=2$ | $somme_slt=2+1=3$ | $sim_globale=2,4$

■ $j=3$ | $w=1$ | $P=3,2$ | $Attr_simple=OUI$ | $l=1$ $l=2$... $l=11$ | $somme_slt=3+0=3$ | $sim_globale=2,4$

- $j=4|w=1|P=4|Attr_simple=OUI| l=1 l=2\dots l=5| somme_slt=3+1=4| sim_globale=3,2$
- $j=5|w=1|P=4,8|Attr_simple=OUI| l=1 l=2\dots l=11| somme_slt=4+0=4| sim_globale=3,2$
- $j=6|w=2|P=6,4|Attr_simple=OUI| l=1 l=2\dots l=6| somme_slt=4+2=6| sim_globale=4,8$
- $j=7|w=2|P=8|Attr_simple=OUI| l=1 l=2\dots l=7| somme_slt=6+2=8| sim_globale=6,4$
- $j=8|w=2|P=9,6|Attr_simple=OUI| l=1 l=2\dots l=11|somme_slt=8+0=8| sim_globale=6,4$
- $j=9|w=1|P=10,4|Attr_simple=OUI| l=1 l=2\dots l=9|somme_slt=8+1=9| sim_globale=7,2$
- $j=10|w=2|P=12|Attr_simple=OUI|l=1 l=2\dots l=10|somme_slt=9+2=11|sim_globale=8,8$
- $j=nb=11|w=2|P=12+ 0,2\times 2= 12,4| Attr_simple=NON (Attribut Complexe):$

Calcul de la similarité basée concept :

$spec(LT, FH_Psyco)= 80 \quad simcpt (LT, FH_Psyco)= 1 \times 2 \times 1 / 4+5+ 80 =2/89 =0,022$

■ $simecpt (GSM, FH_Psyco.) = 1 \times 2 \times 1 / 4+5+ 80 =2/89 =0,022$

■ $simecpt (GAT, FH_Psyco.) = 0,022$

■ $simecpt (CR, FH_Psyco.) = 0,022$

■ $simecpt (GA, FH_Psyco.) = 0,022$

■ $simecpt (Station, FH_Psyco.) = 0,022$

■ $simecpt (PR7, FH_Psyco.) = 1 \times 2 \times 2 / 4+5+ 30 =4/39 =0,10$

■ $simecpt (PR8, FH_Psyco.) = 0,10$

■ $simecpt (Déf.C1, FH_Psyco.) = 1 \times 2 \times 3 / 5+5+20 =6/30 =0,20$

■ $simecpt (Pluie, FH_Psyco.) = 1 \times 2 \times 2 / 5+4+30 =4/30 =0,13$

■ $simecpt (Accident_potentiell, FH_Psyco.) = 1 \times 2 \times 1 / 5+3+40 =2/48 =0,041$

■ $simecpt (Fatigue, FH_Psyco.) = 1 \times 2 \times 4 / 5+5+ 5=8/15 =0,53$

➔ $Max (simecpt) = 0,53 \rightarrow Sim_globale = 8,8 + 0,2 \times 0,53 + 0,8 \times 0 = 8,9$

➔ **Similarité (cas_cible_test, cas_source_test)= $8,9 \times 100 / 12,4 = 71,77\%$**

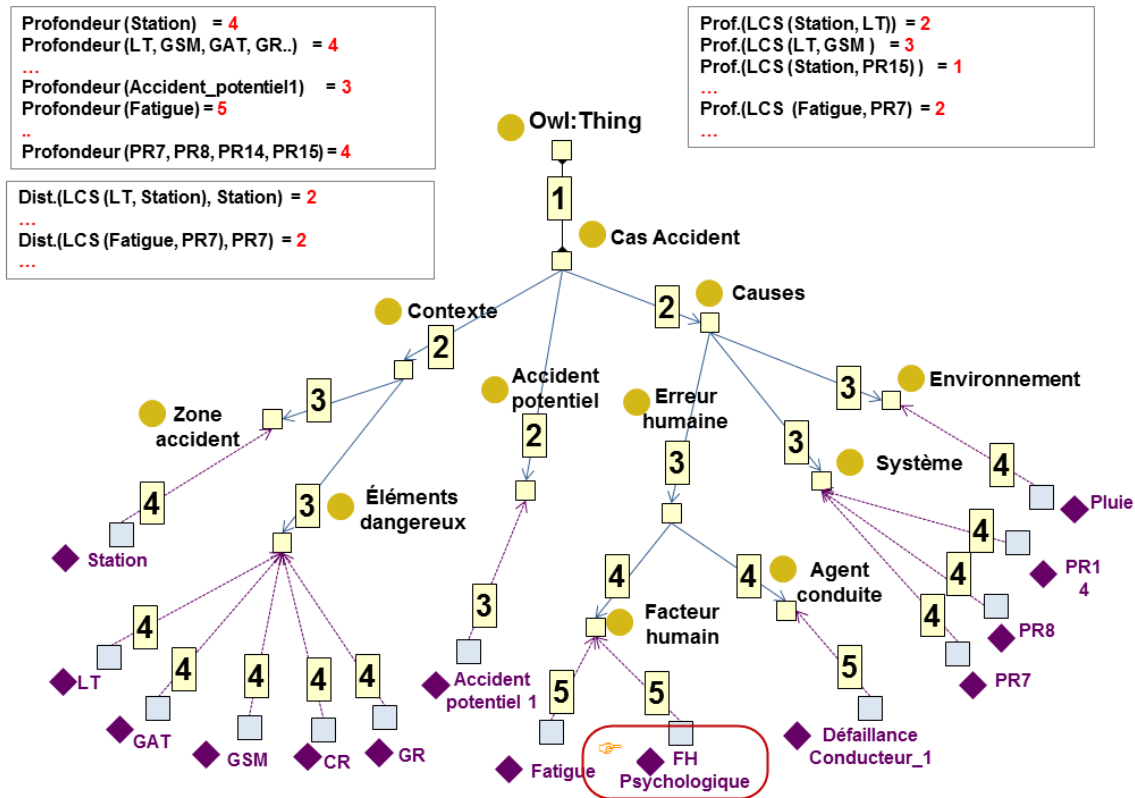


Figure 4.17. Aperçu des mesures de profondeurs des attributs de l'ontologie.

Comme nous l'avons déjà mentionné, ce calcul va être appliqué sur l'ensemble de cas sources provenant de la même classe d'appartenance du cas cible. Une fonction de recherche va être appliquée à la fin pour extraire la mesure de similarité globale ayant la valeur la plus élevée. Le cas source ayant la valeur de similarité la plus élevée sera considéré le plus similaire au cas cible.

4.6.3. Phase d'adaptation

Comme déjà mentionné dans la section 2.3.5.3, cette phase consiste à réutiliser une solution d'un cas source mémorisé afin de résoudre le problème cible.

Nous présentons notre démarche d'adaptation et notre approche d'acquisition de connaissances d'adaptation dans ce qui suit.

4.6.3.1. Démarche d'adaptation

L'approche d'adaptation que nous avons retenue comporte deux étapes (Figure 4.18). Une première étape, dans laquelle nous allons acquérir les connaissances d'adaptation, nous utilisons comme convenu l'Analyse des Concepts Formels (ACF). Durant la deuxième étape, le système infère sur ces connaissances d'adaptation (généralement sous forme des règles d'adaptation) pour solliciter une solution convenable au cas cible.

Nous détaillons ces deux étapes dans le paragraphe suivant.

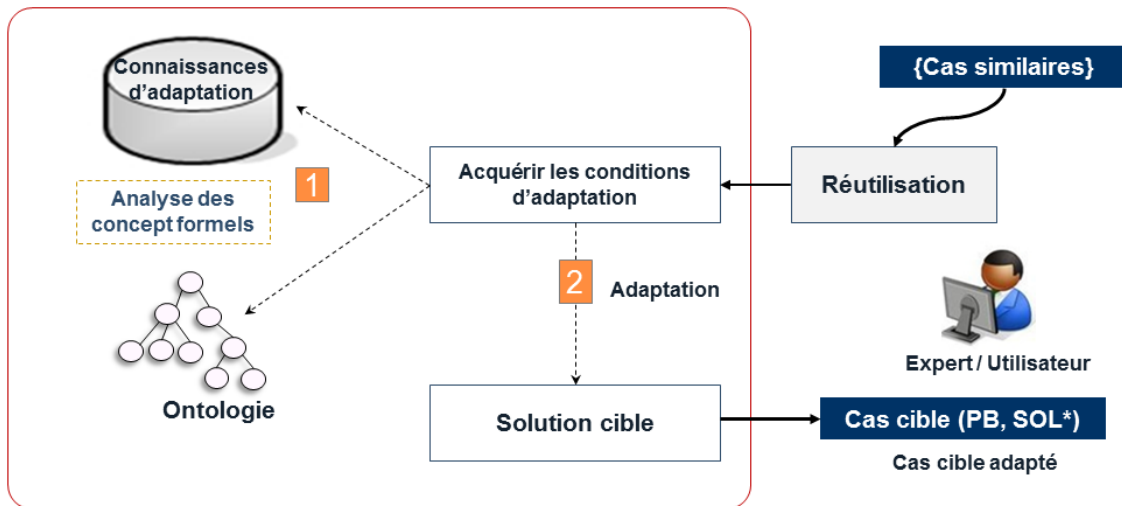


Figure 4.18. Approche de réutilisation adoptée.

4.6.3.2. Acquisition des connaissances d'adaptation

Notre approche d'acquisition des connaissances d'adaptation est considérée proche des travaux réalisés dans (d'Aquin M. et al, 2007), (Badra F. et al, 2007) et elle s'inspire notamment des travaux réalisés dans (Abou-Assali A., 2010) et (Abou-Assali A. et al, 2013).

Pour rappeler le contexte de notre application, un problème représente une situation d'accident et une solution qui représente les mesures préventives et/ou correctives à mettre en œuvre pour minimiser le niveau de risque de l'accident.

Soient $pb1$ et $pb2$ deux problèmes dont les solutions respectivement sont $Sol(pb1)$ et $Sol(pb2)$. Nous basons notre travail d'acquisition sur cette hypothèse :

$$pb1 \text{ et } pb2 \text{ sont similaires} \Leftrightarrow Sol(pb1) \text{ et } Sol(pb2) \text{ sont similaires}$$

En se basant sur cette hypothèse, notre premier objectif, est de découvrir les attributs en commun entre les problèmes conduisant à des solutions similaires. Nous séparons donc les cas dans la base de cas en sous-ensembles en fonction de leur **classe de solution**. Les classes de solutions ont été repérées et identifiées initialement hors-ligne en ayant recours aux connaissances de sécurité acquises pour mieux ajuster la classification. Un cas peut avoir une ou plusieurs solutions, et donc peut appartenir à plusieurs classes de solution.

Notre approche d'acquisition des connaissances d'adaptation se fait hors-ligne et comporte trois étapes successives (Figure 4.19) :

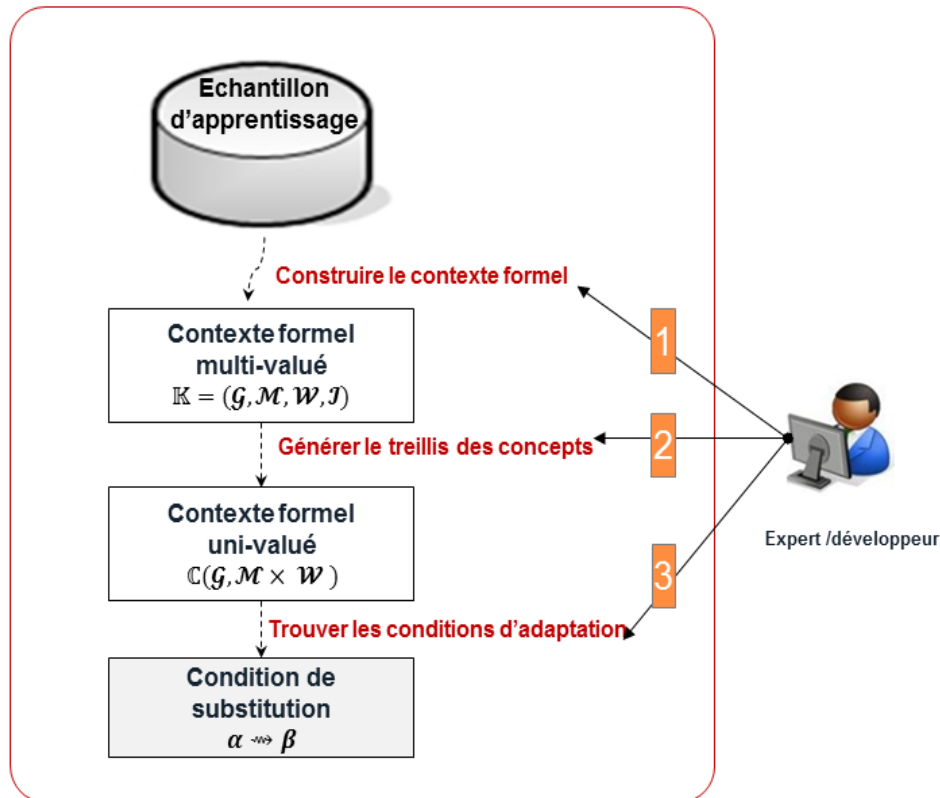


Figure 4.19. Approche d'acquisition des connaissances d'adaptation retenue.

Nous présentons les trois phases de la démarche d'acquisition de connaissances d'adaptation retenue et nous illustrons à cet effet, également un exemple volontairement simplifié.

4.6.3.2.1. Construction du contexte formel

Cette étape consiste à construire le contexte formel de chaque ensemble de cas appartenant à une classe de solution spécifique. Nous définissons le contexte formel multi-valué $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{W}, \mathcal{J})$, où \mathcal{G} est un ensemble de cas ayant la même classe de solution, \mathcal{M} est un ensemble d'attributs multi valués, \mathcal{W} est un ensemble de valeurs d'attributs, et $\mathcal{J} \subseteq \mathcal{G} \times \mathcal{M} \times \mathcal{W}$ est une relation entre \mathcal{G} , \mathcal{M} et \mathcal{W} telle que : $g, m, w \in \mathcal{I}$ signifie que la partie problème du cas g possède l'attribut m avec la valeur w . L'ensemble \mathcal{W} contient de valeurs d'attributs qui correspondent à des instances de l'ontologie $V \subseteq \mathcal{W}$.

Chaque classe de solution regroupe les cas ayant des solutions qui partagent les mêmes traits et/ou qui relatent à une fonction de sécurité bien précise. Pour illustrer un exemple, nous allons étudier trois cas de test (trois exemples de scénarios d'accident issus de transport ferroviaire).

Nous utilisons les acronymes suivants pour dresser le tableau de contexte formel :

- GCA : Gestion de conduite automatique ;
- LOC : Localisation des trains ;
- GMA : Gestion de sens de marche ;
- GAT : Gestion arrêt-train ;
- CR : Commutation de redondance ;

- GA : Gestion des alarmes ;
- C.E/S : Contrôle d'entrée/sortie ;
- ST : Suivi des trains ;
- PR07 : Panne résumée 7 (Maintien d'un itinéraire à tort par le PA en attente) ;
- PR08 : Panne résumée 8 (Commutation à tort du PA fixe par le PCC) ;
- PR14 : Panne résumée 14 (Commutation de redondance de PA) ;
- PR15 : Panne résumée 15 (Défaut d'enregistrement) ;
- PR36 : Panne résumée 36 (Défaut de transmission inter PA).

Pour chaque ensemble de cas, un contexte formel est construit et les concepts formels sont déterminés. Les conditions d'adaptation sont ensuite obtenues en **interprétant les intensions de chaque concept formel**. Chaque cas source est ainsi doté de plusieurs ensembles de conditions d'adaptation variées selon leurs capacités aux substitutions ou l'enrichissement trouvé.

<i>N°</i>	<i>1</i>	<i>2</i>	<i>3</i>
C/A	Éléments dangereux	Zone d'accident	Causes liées au système
Cas_Test01	LOC, GSM, GAT, CR, GA	Terminus	PR07, PR08
Cas_Test02	GCA, C.E/S, ST, CR	Limite tronçon	PR15, PR36
Cas_Test03	GCA, C.E/S, ST, CR	Terminus	PR14, PR15, PR36

Tableau 4.5. Contexte formel multi-valué

Cette étape consiste à transformer chaque contexte en un contexte uni-valué et calculer ses concepts formels afin de produire le treillis des concepts. En transformant ce contexte à un contexte uni-valué, nous obtenons le tableau 4.6.

<i>N°</i>	<i>1</i>								<i>2</i>		<i>3</i>				
Attribut	Éléments dangereux								Zone d'accident		Causes liées au système				
<i>C/A</i>	LOC	GSM	GAT	CR	GA	GCA	C.E/S	ST	Limite tronçon	Terminus	PR7	PR8	PR14	PR15	PR36
Cas_Test01	x	x	x	x	x					x	x	x			
Cas_Test02				x		x	x	x	x					x	x
Cas_Test03						x	x	x		x			x	x	x

Tableau 4.6. Contexte formel uni-valué

4.6.3.2.2. Génération du treillis du concept

Les concepts formels représentent les regroupements maximaux de cas ayant les mêmes valeurs d'attributs. Un concept formel est défini par un couple $\mathbb{C}(\mathcal{G}, \mathcal{M} \times \mathcal{W})$ dont l'extension \mathcal{G} (un ensemble de cas) et l'intension $\mathcal{M} \times \mathcal{W}$ (est une relation entre $m \subseteq \mathcal{M}$ et $w \subseteq \mathcal{W}$ telle que : $\forall g \in \mathcal{G}, \forall (m, w) \in \mathcal{M} \times \mathcal{W}, g.m = w$).

De ce fait, tous les cas qui appartiennent à l'extension du concept possèdent l'attribut m avec la valeur w .

Dans notre exemple, nous utilisons l'outil Concept Explorer¹ pour représenter dans un premier temps le contexte uni-valué (Figure 4.20) puis générer le treillis de concept correspondant (treillis de Galois) (Figure 4.21).

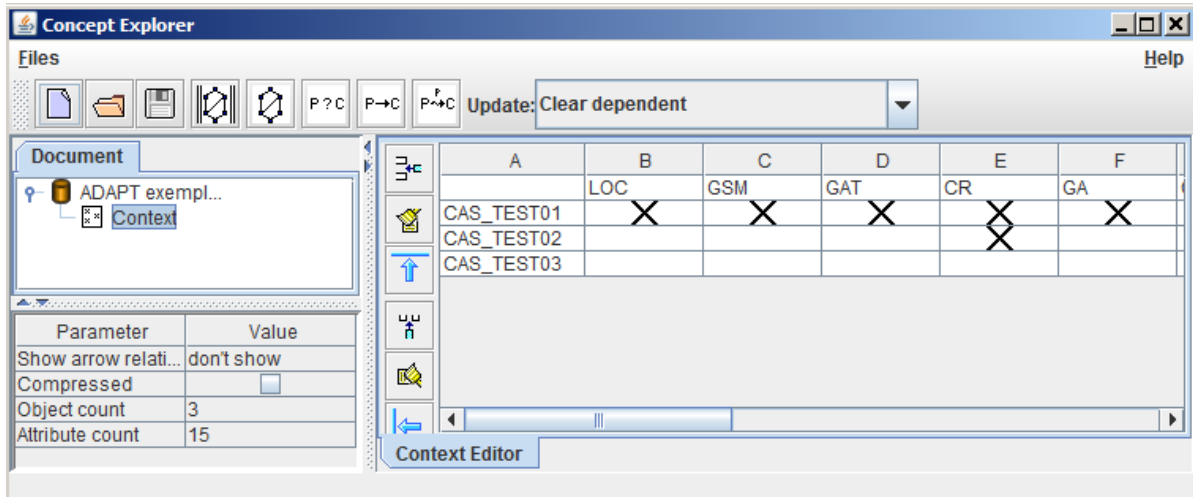


Figure 4.20. Représentation du contexte uni-valué dans l'outil Concept Explorer.

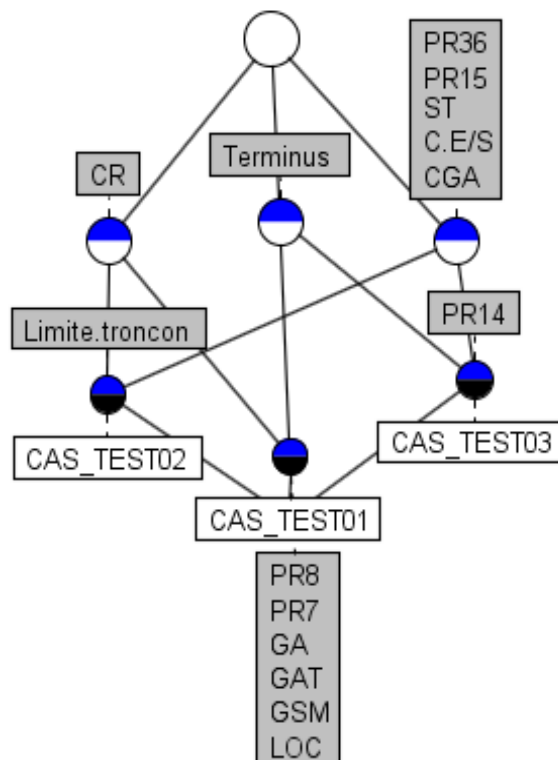


Figure 4.21. Génération de treillis de concept avec l'outil Concept Explorer.

¹ Concept Explorer : Outil libre programmé en java pour la construction des treillis de Galois.

4.6.3.2.3. Génération des conditions d'adaptation

Dans cette étape, nous cherchons les conditions d'adaptation spécifiques à chaque cas. Tout concept formel conduit à un ensemble de conditions noté CA, applicables aux cas appartenant à son extension¹.

Soit $\mathbb{C}(\mathcal{G}, \mathcal{M} \times \mathcal{W})$ un concept formel. Pour $\text{tout}(\mathbf{m}, \mathbf{w}) \in \mathcal{M} \times \mathcal{W}$, une condition d'adaptation est obtenue en suivant les règles suivantes (Abou-Assali A. et al., 2013) :

- **Condition d'égalité** : $w \in \mathcal{V} \Rightarrow$ toute substitution $w \rightsquigarrow \gamma$, telle que $\gamma \in \mathcal{V}$, est liée à une condition d'égalité ; γ doit être égal à w pour que l'adaptation soit acceptée ;
- **Condition sur les propriétés** : $w \in \mathcal{P} \Rightarrow$ toute substitution $\alpha \rightsquigarrow \beta$, telle que $\alpha, \beta \in \mathcal{V}$ et $\alpha.a = w$, est liée à une condition sur les propriétés. β doit avoir la propriété a avec la valeur w pour que l'adaptation soit acceptée ;
- **Condition d'appartenance** : Nous nous basons sur l'approche de l'ACL proposée par (Ferré et Ridoux, 1999) pour pouvoir obtenir ce type de condition. Supposons que l'**extension** du concept \mathbb{C} contient les cas :

$$\text{ext}(\mathbb{C}) = \{ \text{cas}_i(\text{srce}_i, \text{sol}(\text{srce}_i)) \mid i \in \mathbb{N}^*, i \leq n \}$$

Nous écrivons les parties *problème* des cas à l'aide de la **logique propositionnelle**. Les variables de cette logique correspondent à des éléments de l'ensemble \mathcal{V} (des instances de l'ontologie).

Cette logique utilise uniquement le connecteur de conjonction (\wedge). Nous retenons :

$$\text{srce}_1 = a_1 \wedge b_1 \wedge \dots$$

$$\text{srce}_2 = a_2 \wedge b_2 \wedge \dots$$

$$\text{srce}_n = a_n \wedge b_n \wedge \dots$$

L'Analyse des Concepts Logiques (ACL) préconise que l'**intension** du concept, $\text{int}(\mathbb{C})$, soit une formule propositionnelle qui correspond à la disjonction des parties problème des cas. Les conjonctions de disjonctions de cette formule donnent :

$$\begin{aligned} \text{int}(\mathbb{C}) &= \text{srce}_1 \vee \text{srce}_2 \vee \dots \vee \text{srce}_n \\ &= (a_1 \wedge b_1 \wedge \dots) \vee (a_2 \wedge b_2 \wedge \dots) \vee (a_n \wedge b_n \wedge \dots) \\ &= (a_1 \vee a_2 \vee \dots \vee a_n) \wedge (b_1 \vee b_2 \vee \dots \vee b_n) \wedge \dots \end{aligned}$$

Chaque bloc disjonctif peut conduire à une condition d'appartenance s'il satisfait les règles suivantes :

- **Règle 1** : Les variables du bloc sont des valeurs d'un même attribut de l'ensemble \mathcal{M} ;
- **Règle 2** : Supposons que le bloc $(a_1 \vee a_2 \vee \dots \vee a_n)$ permet de remplir la première règle. Soit $O \in \mathcal{O}$ le plus petit subsumant commun (qui généralise) des instances (a_1, a_2, \dots, a_n) et soit $V \subseteq \mathcal{V}$ l'ensemble de toutes les instances (directes ou non) du concept O . Nous disons que ce bloc est **généralisable par le concept O** si : $\frac{m}{|V|} \geq \mu$;

¹ L'extension : contient les individus appartenant au concept

Avec :

- $m \leq n$ est le nombre d'instances distinctes dans (a_1, a_2, \dots, a_n) ;
- $|V|$ est la cardinalité de V ;
- μ est un certain seuil à définir après les tests.

Cette règle signifie que l'analyse des instances d'un certain concept dans un cas particulier nous permet de déduire l'observation potentielle des autres instances dans la même situation.

La réalisation de ces deux règles donne lieu à **la condition d'adaptation** suivante : toute substitution $a_i \rightsquigarrow \alpha$, telle que $1 \leq i \leq n$, est acceptée si et seulement si α est une instance du concept O .

Généralement, un cas peut appartenir à plusieurs concepts formels. Donc plusieurs ensembles de conditions d'adaptation peuvent être décelées pour un cas donné. Le choix adéquat sollicite le concept formel **ayant le plus de tolérance au substituions et l'enrichissement**.

Soient \mathbb{C}_1 et \mathbb{C}_2 deux concepts formels, et \mathbb{CA}_1 et \mathbb{CA}_2 leurs ensembles de conditions d'adaptation respectivement.

$$\text{si } \mathbb{C}_1 \leq^c \mathbb{C}_2 \Rightarrow \mathbb{CA}_2 \text{ est plus fiable que } \mathbb{CA}_1$$

A partir du treillis de concept généré (Figure 4.21), nous pouvons illustrer donc les concepts suivants (Figure 4.22) :

$$\mathbb{C}_1 = (\{cas_test01, cas_test03\}, \{Terminus\})$$

$$\mathbb{C}_2 = (\{cas_test02, cas_test03\}, \{PR15, PR36, ST, C.E/S, GCA\})$$

$$\mathbb{C}_3 = (\{cas_test01, cas_test02\}, \{CR\})$$

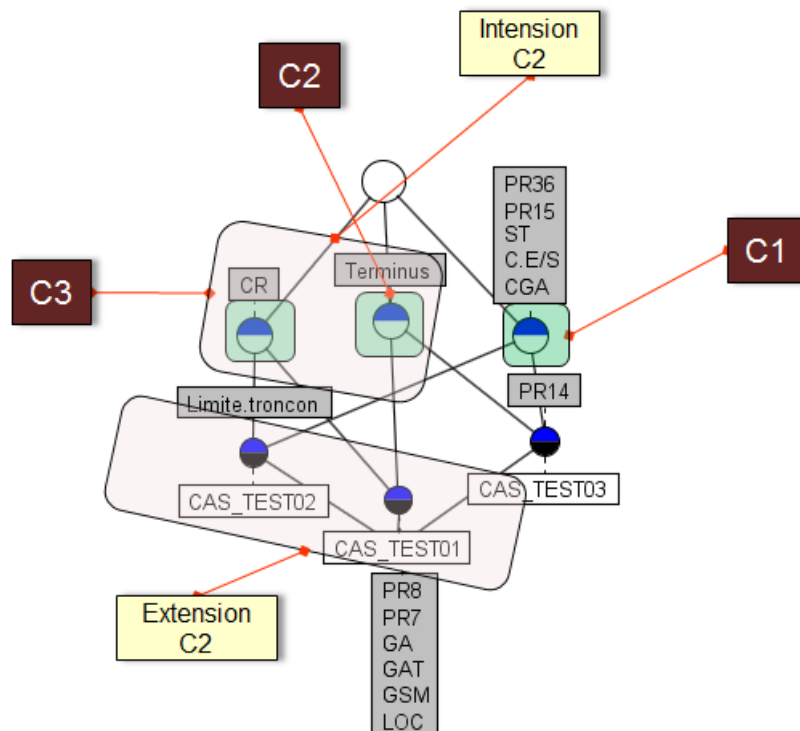


Figure 4.22. Identification de l'extension et l'intension.

Selon l'approche de l'analyse de concept logique, l'**intension** du concept $int(C)$ est une formule propositionnelle qui correspond à la disjonction des parties problème des cas. Nous écrivons à titre d'exemple les **intensions** du cas_test03 (C_1 et C_2) en forme normale conjonctive (des conjonctions de disjonctions) :

$$\begin{aligned} \text{Nous avons } int(C_1) &= cas_test01 \vee cas_test03 \\ &= (LOC_01 \wedge GSM_01 \wedge GAT_01 \wedge GA_01 \wedge PR7_01 \wedge PR8_01 \wedge Terminus_01) \vee \\ &\quad (PR14_03 \wedge Terminus_03). \end{aligned}$$

En se basant sur distributivité de « ou » par rapport à « et » : $(P \vee (Q \wedge R)) \Leftrightarrow ((P \vee Q) \wedge (P \vee R))$, nous allons trouver un ensemble de conditions (voir annexe 2 pour la démonstration).

Les conditions qui correspondent à C_1 et C_2 sont présentées dans le tableau 4.7. Nous constatons l'apparition de plusieurs règles d'adaptation liées aux conditions d'appartenance.

	<i>Condition de substitution</i>	<i>Substitution ($\alpha \in \mathcal{V}$)</i>
$\mathbb{C}A_1$	Condition d'égalité • <i>Zone d'accident</i>	$Terminus \rightsquigarrow \alpha$
	Appartenance au concept • <i>Causes liées au système</i> ▪ <i>Pilotage automatique</i>	$PR7 \rightsquigarrow \alpha$ $PR8 \rightsquigarrow \alpha$
$\mathbb{C}A_2$	Appartenance au concept • <i>Éléments dangereux</i> ▪ <i>Conduite automatisée :</i> ▪ <i>Conduite manuelle :</i> • <i>Causes liées au système</i> ▪ <i>Pilotage automatique :</i>	$GCA \rightsquigarrow \alpha$ $C.ES \rightsquigarrow \alpha$ $ST \rightsquigarrow \alpha$ $PR14, PR15, PR36 \rightsquigarrow \alpha$

Tableau 4.7. Conditions d'adaptations trouvées

Nous admettons aussi que l'ensemble $\mathbb{C}A_2$ est plus tolérant aux substitutions que $\mathbb{C}A_1$ par conséquent l'ensemble $\mathbb{C}A_2$ sera retenu pour cas_test03.

Par exemple :

- CGA, LOC \rightarrow sous concept « conduite automatisée » ;
- C.E/S et ST \rightarrow sous concept « conduite manuelle » ;
- PR14, PR15, PR36 \rightarrow sous concept « pilotage automatique ».

4.6.3.3. Phase de révision

Dans cette phase, la solution proposée à l'issue de la phase d'adaptation sera évaluée manuellement par l'utilisateur (Figure 4.23).

Il se peut que la solution proposée par le système ne convienne pas à l'utilisateur ou bien qu'elle ne soit pas en mesure de résoudre le problème, L'utilisateur a donc, pendant la phase de révision, l'opportunité de modifier ou choisir une autre solution à partir de l'ontologie.

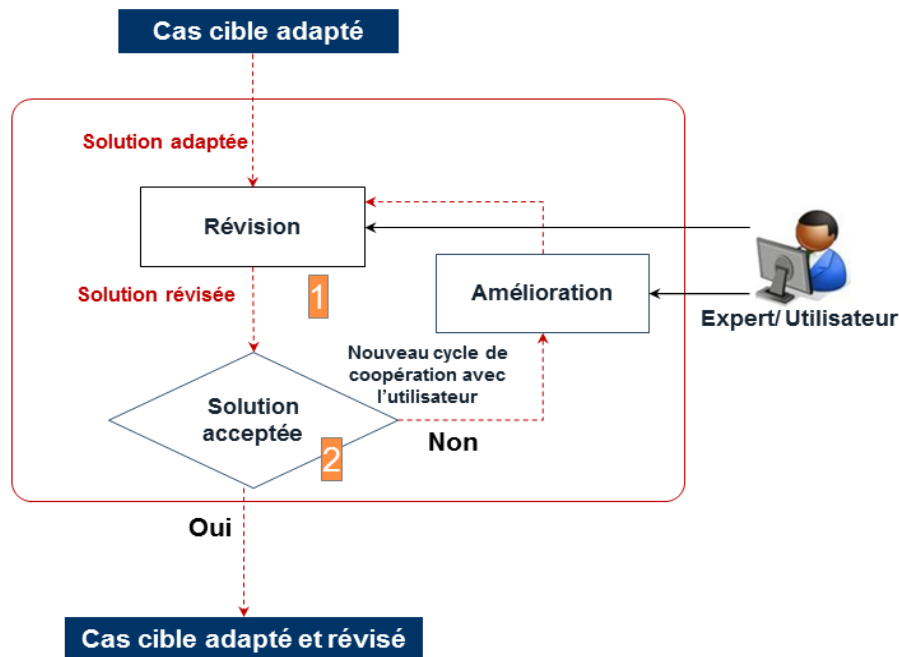


Figure 4.23. Déroulement de la phase de révision.

4.6.3.4. Phase d'apprentissage

Comme convenu, dans cette phase les connaissances présentes dans la base de cas vont être enrichies par des nouvelles connaissances pendant le déroulement des sessions de résolution de problème. La mémorisation du nouveau cas résolu (cas cible adapté et révisé) fait enrichir la base de connaissances par conséquent augmente l'expérience du système. La base de cas évolue dans le temps au fur et à mesure de l'ajout de nouveaux cas. Néanmoins, ceci pourrait perturber l'organisation et la structuration de la base de cas d'une façon générale. Donc, un contrôle de la cohérence du nouveau appris est indispensable pour ne pas nuire l'aspect incrémental de l'apprentissage en général.

Nous présentons alors notre démarche d'apprentissage retenue ainsi que l'algorithme d'auto incrémentation proposé.

4.6.3.4.1. Démarche d'apprentissage proposée

L'approche d'apprentissage que nous proposons est inspirée de travaux de (Haouchine M.K., 2009) sur la maintenance de la base de cas. Une méthode d'auto-incrémentation a été mise en place dans ses travaux conjointement avec une méthode de structuration des cas. Dans notre approche, nous donnons une autre optique qui met en considération plus l'aspect sémantique des connaissances.

Nous éliminons la méthode de structuration des cas, puisque notre base de cas regroupe un ensemble de classes de cas, d'où nous affirmons que nous possédons **une structuration initiale suffisamment préinstallée**, et nous affinons plus la méthode d'auto-incrémentation afin d'améliorer encore l'aspect sémantique de la phase d'apprentissage.

Nous considérons un cas cible avec sa partie solution *renseignée et révisée* cas_{cible}^* . Deux hypothèses sont ainsi considérées :

Hypothèse 1 :

Il n'y a aucune solution dans la base de cas qui est similaire à la solution attribuée au cas cible* (l'espace solution cible* n'est recouvert par aucun espace de solution des cas sources appartenant à la classe C_k), ceci implique la mémorisation du cas dans la base de cas (Figure 4.24) :

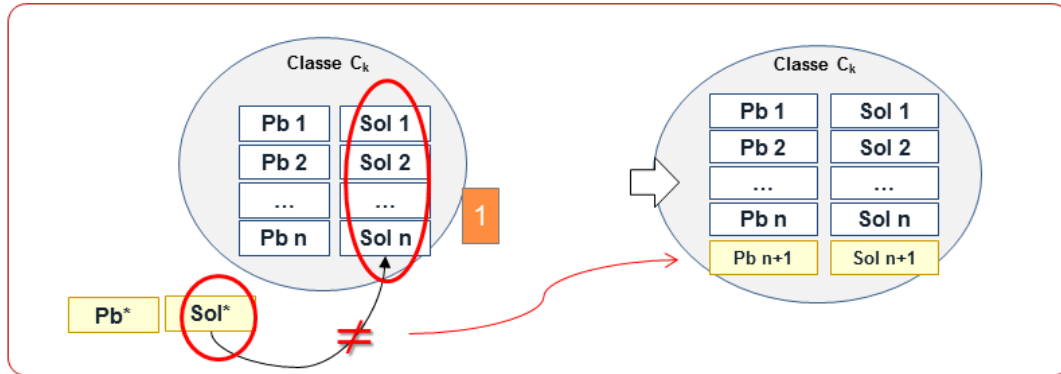


Figure 4.24. Déroulement de la phase d'apprentissage (Première hypothèse).

Hypothèse 2.

Il y'a au moins une solution attribuée à un cas \in classe C_k qui est similaire à la solution du cas_{cible}^* , nous sélectionnons donc, le(s) cas ayant la même solution.

Le cas ou le sous-ensemble de cas choisi sera utilisé pour calculer un score qui représente le degré de sémantique noté D_S . Étant donné que chaque cas de la base de cas présente une instance de l'ontologie de domaine, nous procédons alors à une comparaison des parties problèmes : le(s) cas source(s) choisis et le cas_{cible}^* .

Nous calculons le(s) score(s) en appliquant un calcul de similarité basé concept. Nous partons de l'hypothèse suivante : *Plus le cas est représenté par un nombre important d'ancêtres de l'ontologie plus le cas est assez représentatif et détaillé.* Par conséquent nous pouvons considérer qu'il présente un degré de représentativité élevé. Dans le cas ou le score du cas_{cible}^* est supérieur à l'un des cas sources sélectionnés, nous procédons l'apprentissage automatiquement et nous supprimons le cas source qui présente le score inférieur. Dans le cas contraire, le cas ne sera pas retenu, il va servir pour résoudre le problème courant uniquement (Figure 4.25).

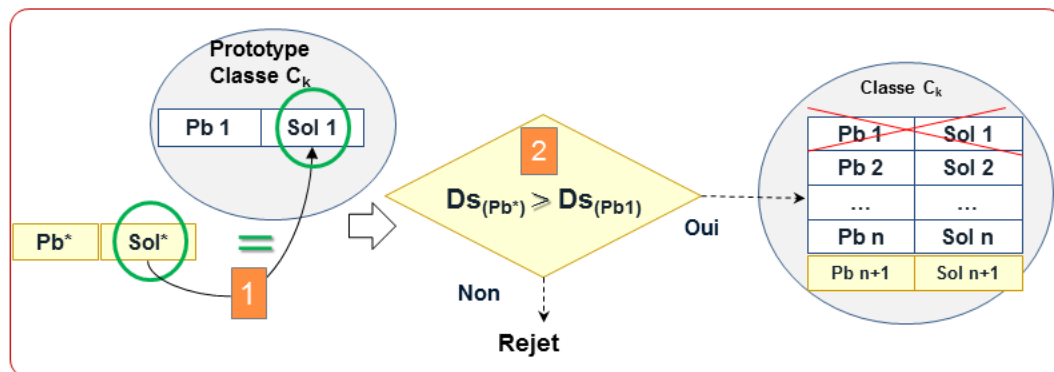


Figure 4.25. Déroulement de la phase d'apprentissage (Deuxième hypothèse).

4.6.3.4.2. Algorithme d'auto-incrémentation

Nous avons développé un algorithme pour mettre en œuvre la méthode d'auto-incrémentation proposée. La mise en œuvre de cet algorithme nous a permis non seulement d'assurer la phase d'apprentissage mais de bien maintenir la base de cas puisque nous cherchons toujours à mémoriser uniquement les cas les plus représentatifs pour assurer ainsi un apprentissage incrémental.

Algorithme 4.2. Apprentissage (BC, C_k , cas_{cible}^*)

Entrée : Base de cas BC + Classe C_k + cas_{cible}^* :cas cible adapté et révisé

Sortie : Décision : booléen (décision de l'apprentissage (Mémorisation du cas cible adapté et révisé))

0- Début

1- Cas cible $\leftarrow cas_{cible}^*$

2- Décision \leftarrow Non

3- Obtenir Nb // Nombre de cas ayant la même solution que le cas cible

4- Obtenir Nbc // Nombre d'attributs du cas cible

5- Obtenir Vas // Un entier qui prend la valeur 0 si sol^* n'est pas atteignable par une solution d'un cas prototype ou un ensemble de cas de la classe C_k . Il prend la valeur 1 dans le cas contraire.

5- Si $V_{as} = 0$ **Alors**

Décision \leftarrow OUI

BC_{ck} \leftarrow BC_{ck} \cup Cas cible // Mémoriser le cas cible dans BC_{ck}

Sinon // Calculer le degré de représentativité

DS_{Pb*} \leftarrow 0

Pour tout Attr_n (n=1..nbc) faire

DS_{Pb*} \leftarrow DS_{Pb*} + Profondeur (Attr_n)

Fin pour

Pour tout cas_i (i=1..Nb) faire

// Calculer le degré de représentativité des cas sources ayant la même solution

Obtenir Nbcas_i // Nbre d'attributs du cas source courant

D_sⁱ \leftarrow 0

Pour tout Attr_j (n=j..Nbcas_i) faire

D_sⁱ \leftarrow D_sⁱ + Profondeur (Attr_j)

Fin pour

T[j] \leftarrow D_sⁱ

Fin pour

// Comparer DS_{Pb*} et D_sⁱ pour décider d'apprendre le cas ou non

Si DS_{Pb*} > Max (T[j]) **Alors**

Décision \leftarrow OUI ; BC \leftarrow BC \cup Cas cible

Sinon

Décision \leftarrow Non ; Ne pas mémoriser cas_{cible}^* dans BC BC_{ck}

Fin SI

Fin Si

6- Fin

Nous donnons :

- **BC** : la base de cas ;
- **C_k** : la classe d'appartenance du cas cible ;
- **Nb** : Nombre de cas ayant la même solution que le cas cible ;
- **Nbc**: Nombre d'attributs/instances de cas cible ;
- **Vas** : Un entier qui prend la valeur 0 si la solution du cas cible n'est pas atteignable par une solution d'un cas source. Il prend la valeur 1 dans le cas contraire ;
- **DS_{pb*}** : Degré de représentativité du cas cible ;
- **D_sⁱ** : Degré de représentativité de l'ⁱème cas source ;
- **Attr_n** : n^{ème} attribut du cas cible ;
- **cas_i** : i^{ème} cas source ayant la même solution que cas cible ;
- **Nbcas_i** : Nombre d'attributs du n^{ème} cas source ;
- **T []** : Tableau d'entiers qui sert à la mémorisation des scores.

Pour calculer la complexité de cet algorithme, nous devons appliquer la formule conditionnelle :

Si <Condition> O(h(n)) Alors O(f(n)) sinon O(g(n)) fin si

La complexité d'une Si conditionnelle est : **Max{O(h(n)), O(f(n)), O(g(n))}**. Nous avons une complexité **O(1)** que ce soit pour le test **Si Vas = 0** ou l'instruction de mémorisation du cas dans la base de cas. Par contre pour le traitement de sinon, la complexité vaut en total : **O(nbc × nb × nbas)**, nous considérons cette dernière comme complexité générale de l'algorithme d'auto-incrémentation.

4.6.3.4.3. Exemple illustratif

Pour illustrer la phase d'apprentissage, nous avons eu recours aussi à un exemple volontairement simplifié.

Les deux tableaux 4.8 et 4.9 présentent respectivement les instances présentes pour le cas cible (adapté et révisé) et un cas source ayant la même solution attribué au cas cible. Les différentes mesures de profondeur utilisées sont représentées déjà dans la figure 4.17.

CAS_cible_adapté_révisé Nombre d'instances : 10										Solution
1	2	3	4	5	6	7	8	9	10	SA_T
LT	GSM	CR	GA	Station	PR7	PR8	Pluie	Accident potentiel1	Fatigue	

Tableau 4.8. Description du cas cible (adapté et révisé)

Cas_source Nombre d'instances : 10										Solution
1	2	3	4	5	6	7	8	9	10	SA_T
LT	GSM	GA	Station	PR7	PR8	Défaillance conducteur_1	Pluie	Accident potentiel1	Fatigue	

Tableau 4.9. Description du cas source

Pour illustrer le calcul, nous procédons à une simulation d'une itération de calcul de la similarité entre un cas cible adapté et révisé et un cas source disposants les deux de la même solution :

$$\underline{NB=1 | Nbc=10 | Vas=1}$$

$$DS_{pb^*}=0 | N=1 | DS_{pb^*} = 4 |$$

$$N=2 | DS_{pb^*} = 8 | N=3 | DS_{pb^*} = 12 | N=4 | DS_{pb^*} = 16 |$$

$$N=5 | DS_{pb^*} = 20 | N=6 | DS_{pb^*} = 24 | N=7 | DS_{pb^*} = 28 |$$

$$N=8 | DS_{pb^*} = 32 | N=9 | DS_{pb^*} = 35 | N=10 | \boxed{DS_{pb^*} = 40} |$$

$$\underline{Nbcas=10 |}$$

$$D_s^i =0 | i=1 | D_s^1 = 4 |$$

$$j=2 | D_s^1 = 8 | j=3 | D_s^1 = 12 | j=4 | D_s^1 = 16 |$$

$$j=5 | D_s^1 = 20 | j=6 | D_s^1 = 24 | j=7 | D_s^1 = 29 |$$

$$j=8 | D_s^1 = 33 | j=9 | D_s^1 = 36 | j=10 | \boxed{D_s^1 = 41} |$$

$$\rightarrow DS_{pb^*} < D_s^1$$

$$\rightarrow \text{Ne pas mémoriser cas}_{cible}^* \text{ dans la base de cas BC}_{ck}$$

4.7. Conclusion

Nous avons présenté tout au long de ce chapitre notre approche générique d'aide à la décision baptisée ADAST (Aide à la Décision pour Améliorer la Sécurité des Transports).

Nous avons décrit tout d'abord les motivations ainsi que les objectifs attendus et nous avons rappelé nos choix méthodologiques adoptés. Ces choix qui se sont basés sur le RàPC, les ontologies et un ensemble de techniques et de méthodes de la fouille de données. Nous avons par la suite présenté l'architecture fonctionnelle d'ADAST, les différentes articulations entre les phases de raisonnement et notamment les modèles de connaissances proposés à savoir, le modèle conceptuel, le modèle de domaine et le modèle de cas. La dernière section de ce chapitre a détaillé notre approche de raisonnement à partir de cas.

Au contraire d'une démarche de diagnostic, notre approche d'aide à la décision s'organise en trois niveaux complémentaires permettant d'aborder progressivement la résolution du

problème d'acquisition, de capitalisation et d'exploitation des connaissances issues de retour d'expérience Rex (les scénarios d'accidents) tout en conservant une forte interactivité future avec les spécialistes de domaine.

Le chapitre suivant présente les premières phases de validation de notre approche à travers la maquette de faisabilité d'ADAST.

Chapitre 5 : Maquette de faisabilité d'ADAST

5.1. Introduction

Le travail développé dans cette thèse relève d'une recherche prospective. Nos préoccupations concernent essentiellement la validation des hypothèses avancées et l'évaluation de la faisabilité et de la pertinence de toute l'approche d'aide à la décision proposée. Après avoir exposé dans le chapitre précédent les différentes articulations de notre approche ADAST, nous présentons dans ce chapitre notre démarche de validation du travail élaboré. Afin de concrétiser notre proposition nous avons développé une maquette de faisabilité d'ADAST. Cette maquette est utilisée dans un premier temps pour pérenniser les connaissances en matière d'accident dans le transport ferroviaire (notre cas d'étude actuel) puis proposer une aide à la décision à plusieurs échelles sous forme d'un tableau de bord pouvant assister convenablement l'expert de la sécurité face à des nouvelles situations d'insécurité.

Dans ce chapitre, nous présentons les spécifications générales de notre système à travers l'architecture générale qui comprend l'architecture en couches et ses différentes composantes (la base ontologique, la base de règles et la base de cas) et notamment l'architecture fonctionnelle. Nous présentons ensuite les deux modules de la maquette à savoir ; le module d'acquisition de cas et le module d'aide à la décision tout en illustrant les principaux écrans de la maquette.

5.2. Architecture générale de la maquette d'aide à la décision

Avant de procéder la présentation de l'architecture générale de la maquette d'aide à la décision, il convient dans un premier temps d'énumérer les différents outils logiciels utilisés pour mettre en œuvre ce premier prototype de faisabilité.

5.3. Outils logiciels utilisés

Les principaux outils logiciels utilisés sont énumérés ci-après :

- L'environnement de développement **Netbeans**¹ dans sa version 7.2. Il s'agit d'une plateforme complète de développement d'applications Java ;
- Une trentaine de **bibliothèques** destinées à l'interfaçage avec l'ontologie tel que *protege.jar*, *owlsyntax.jar*, etc. Ces bibliothèques sont indispensables pour faire fonctionner et mettre en œuvre tout le travail ;
- **Framework Jena**² pour la manipulation des contenus RDF, OWL etc. ;
- L'éditeur d'ontologies **Protégé** (v3.4 et v4.3) avec des différents plugins comme *OwlViz*, *OntoGraph* etc. Le recours à une version antérieure (v 3.4) était dans le but d'utiliser certains built-ins et des bibliothèques disponibles seulement dans cette version de Protégé ;
- D'autres outils issus de la fouille de données ont été également sollicités. Le schéma

¹ Netbeans : <https://fr.netbeans.org/>

² Jena : <http://jena.sourceforge.net/>

décrivant le processus d'enchaînement de ces différents outils et leur articulation avec la maquette ADAST est présenté dans la figure 5.1.

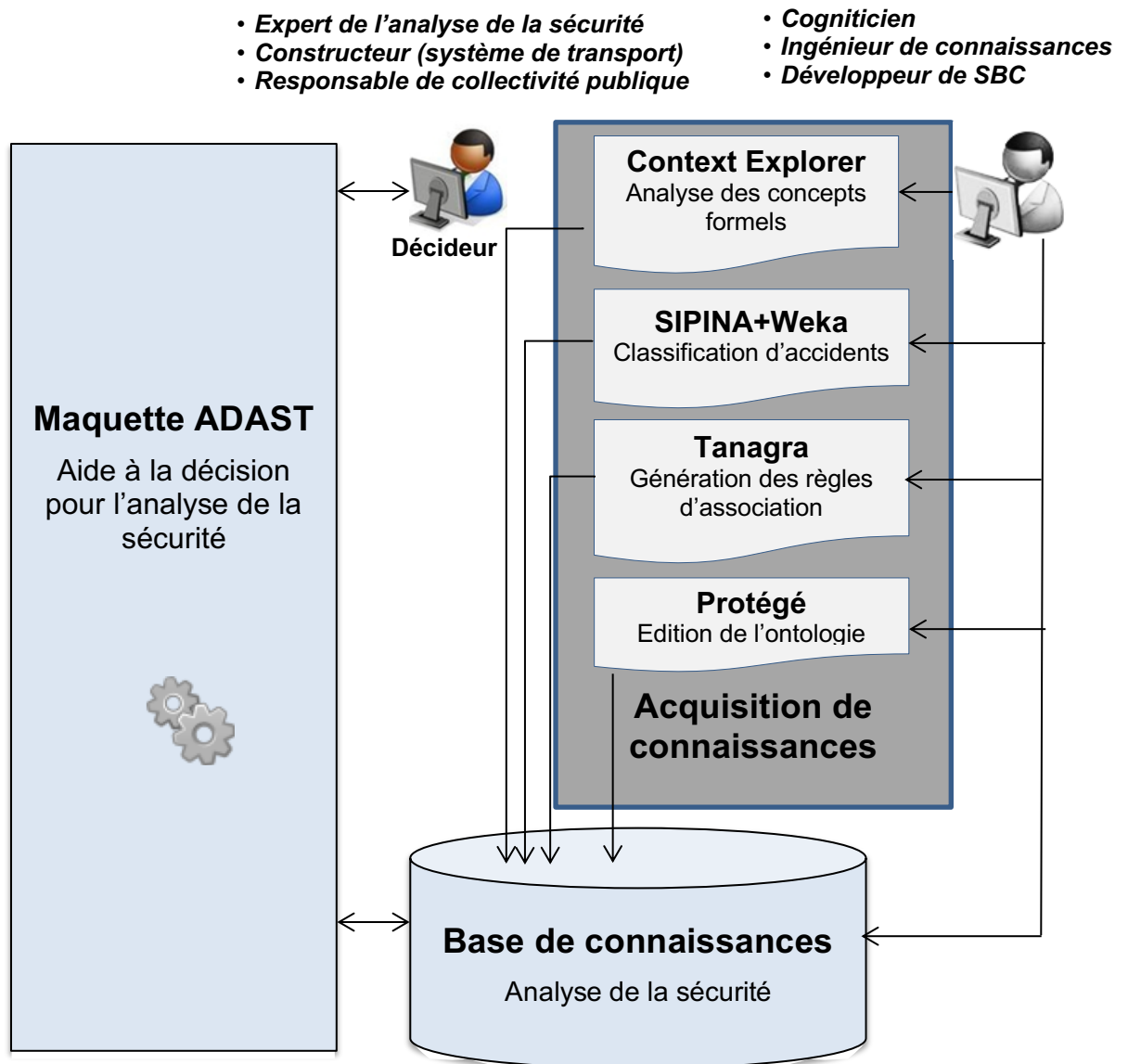


Figure 5.1. Articulation entre les différents outils utilisés et la maquette ADAST.

Dans cette articulation, nous trouvons les outils suivants :

- L'outil **Context Explorer**¹ pour la génération des treillis de concepts à travers l'ACF. Il s'agit d'un outil Java riche de fonctionnalités et qui permet également l'induction de règles de type (Prémisse → Conclusion) ;
- D'autres outils de fouille de données comme **Sipina**², **Tanagra**³ et **Weka**⁴ pour la production des règles d'association et de classification. Ces trois outils sont assez riches et implémentent la majorité des algorithmes de fouilles de données connus.

¹ ContextExplorer : <http://conexp.sourceforge.net/>

² Sipina : eric.univ-lyon2.fr/~ricco/sipina.html

³ Tanagra : <http://eric.univ-lyon2.fr/~ricco/tanagra/fr/tanagra.html>

⁴ Weka : www.cs.waikato.ac.nz/ml/weka/

Ces outils logiciels que nous venons de mentionner, font partie de la phase d'acquisition de connaissances sous le contrôle et la supervision de l'ingénieur de connaissance. Cette phase constitue de processus hors ligne qui est mis à contribution par la maquette ADAST. Vient juste après l'interaction et l'inférence entre la maquette de faisabilité d'ADAST et la base de connaissances pour conclure tout le processus d'aide à la décision.

5.3.1. Architecture en couches

L'architecture en couche de la maquette de faisabilité d'ADAST est composée de trois couches (Figure 5.2) :

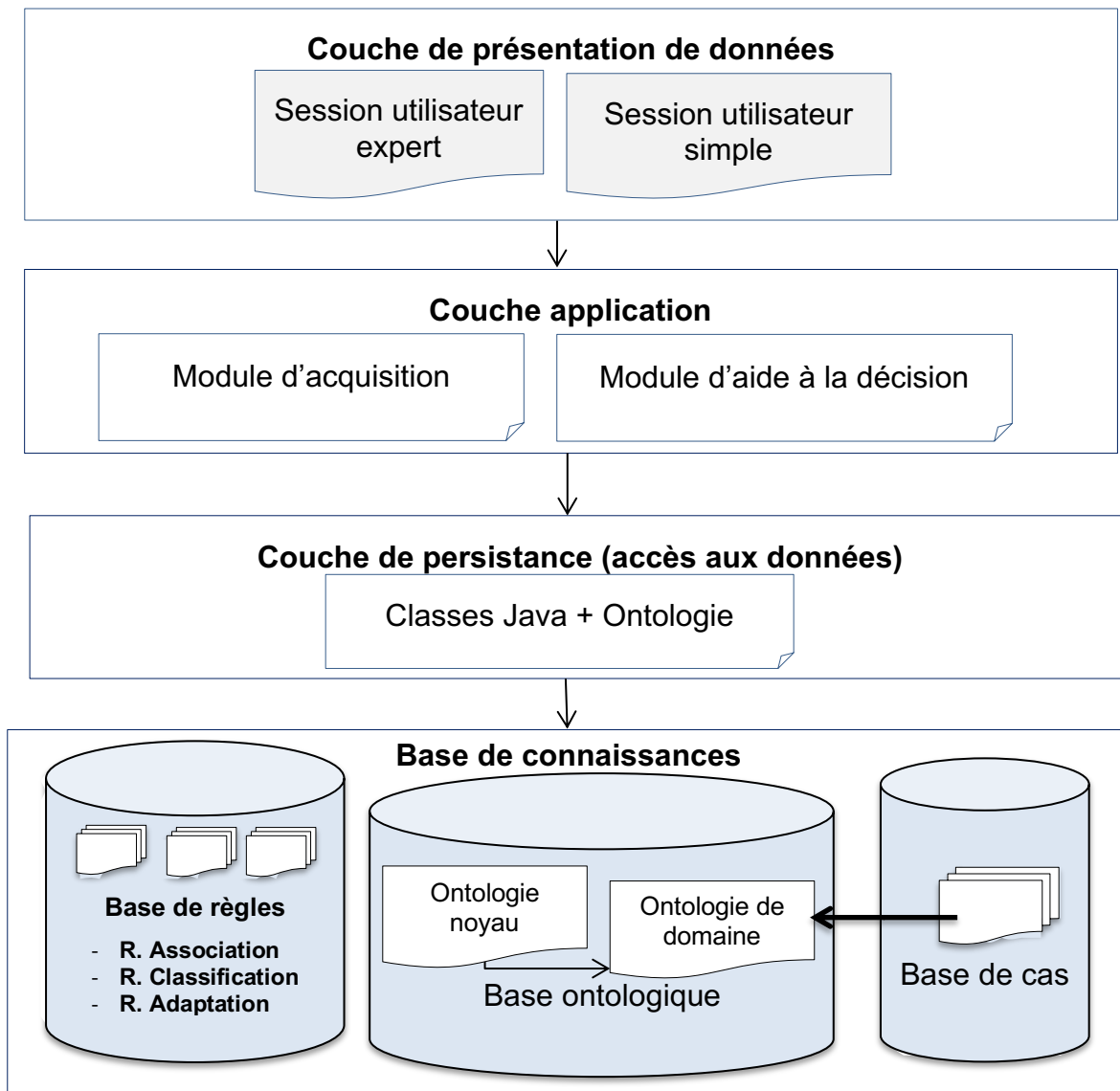


Figure 5.2. Architecture en couches de la maquette d'aide à la décision.

- Couche de présentation : Cette couche permet d'établir le lien et d'assurer l'interfaçage avec les différents types d'utilisateurs (expert, utilisateur, etc.). A cet effet, l'accès à toutes les fonctionnalités de la maquette est réservé uniquement à l'expert quant à l'utilisateur, il pourra uniquement visualiser et consulter les connaissances capitalisées. Cette couche permet donc, l'instanciation des modules applicatifs fournis par la couche application ;

- Couche applicative : Cette couche regroupe les deux modules principaux de la démarche d'aide à la décision proposée :
 - Le module d'acquisition des cas : Ce module permet l'interfaçage avec la base ontologique pour fournir le vocabulaire nécessaire pour l'acquisition et la capitalisation des cas sources ;
 - Le module d'aide à la décision : Ce module permet l'interfaçage avec la base de connaissances entière du système et les utilisateurs à travers les différents écrans dédiés aux phases de raisonnement à partir de cas. Il permet aussi le paramétrage du raisonnement.
- Couche de persistance : C'est la couche d'accès aux données. Elle fournit les objets nécessaires pour l'interfaçage avec la base ontologique, la base de cas et les différentes bases de règles.

5.3.2. La structure de la base de connaissances

La base de connaissances dans le cadre de notre maquette d'aide à la décision actuelle est structurée selon trois bases distinctes :

- Une base ontologique relative à l'ontologie de domaine ;
- Une base de cas qui collecte un ensemble de classes de cas. Chaque classe regroupe un nombre de cas (Accidents historiques) représentés en ayant recours à l'ontologie de domaine ;
- Une base de règles qui regroupe trois sous bases relatives aux différents types de règles exploitées par le module de raisonnement.

5.3.2.1. Base ontologique

La modélisation ontologique d'une situation d'insécurité (un scénario d'accident) est jugée comme une étape importante dans notre approche. Le but de cette étape est de mettre en évidence toutes les caractéristiques pouvant décrire le scénario d'accident ainsi que les relations sémantiques qui peuvent y exister. Cette étape prépare à l'intégration du raisonnement ontologique au sein du RàPC.

La base ontologique comme nous l'avons déjà vu dans notre approche ADAST (section 4.4.), comprend une ontologie noyau décrivant les principales classes d'un scénario d'accident et l'ontologie de domaine. Celle-ci servira comme référentiel pour l'acquisition des cas sources et notamment les cas cibles (pour lesquels nous cherchons des solutions). L'ontologie est alors représentée par un document XML, un aperçu présenté ci-après :

```
<?xml version="1.0"?>
<rdf:RDF
...
<owl:ObjectProperty rdf:ID="a_cause_interaction">
  <rdfs:domain rdf:resource="#Cas"/>
  <rdfs:range>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#Interactions"/>
        <owl:Class rdf:about="#Homme_système"/>
        <owl:Class rdf:about="#Homme_environment"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:range>
</owl:ObjectProperty>
```



```

    <owl:Class rdf:about="#Système_environnement"/>
  </owl:unionOf>
</owl:Class>
</rdfs:range>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="a_solution">
  <rdfs:range>
    <owl:Class>
      <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#Solution"/>
        <owl:Class rdf:about="#Mesures_correctives"/>
        <owl:Class rdf:about="#Mesures_préventives"/>
      </owl:unionOf>
    </owl:Class>
  </rdfs:range>
</owl:ObjectProperty>
...

```

5.3.2.2. Base de règles

La base de règles représente en effet une partie de la base de connaissances exploitée par le processus de raisonnement. Elle est répertoriée en trois sous bases :

- ***La sous base de règles d'association :***

Comme convenu dans notre approche, les règles générées sont issues de l'application de l'algorithme Apriori sur l'échantillon d'apprentissage dont nous disposons. Chaque règle est alors structurée en une suite de trois lignes du fichier, précédée d'un signe « @ » et comprend une conjonction d'attributs séparés par des « ; ».

Les trois paramètres (le lift, le support et la confiance) que nous avons déjà mentionnés (voir section 4.6.1) sont présents dans la dernière ligne de chaque règle. Un algorithme d'un moteur d'inférence est intégré dans la maquette va procéder en chainage avant pour faire apparaître les règles d'association en relation avec les descripteurs saisis dans le cas cible. Ces règles d'association vont permettre à l'utilisateur de la maquette de bien ajuster/compléter sa description initiale du cas cible à traiter. Un aperçu de quelques règles est présenté ci-après :

Exemples de règles d'association

~~~~~

```

@R_A
"Ligne=OUI"
"Communication__redondance=NON" ; "contrôle_E/S=NON"
1,35 57,35 97,50

```

```

@R_A
"Suivi_des_trains=OUI"
"Station=NON" ; "Gestion_des_alarms=NON"
1,25 32,35 100,00

```

```

@R_A
"Terminus=OUI"
"Sécurité_Quai-voie=NON" ; "contrôle_E/S=NON"
1,23 38,23 92,85

```

- ***La sous base de règles de classification***

Les règles de classifications sont générées en appliquant la méthode d'arbre de décision à travers l'algorithme C4.5 sur l'échantillon d'apprentissage dont nous disposons (voir

section 4.6.2). Chaque règle là aussi, occupe trois lignes, précédée d'un signe « @ » et comprend une conjonction d'attributs séparés par des espaces. La dernière ligne comporte la valeur de la probabilité associée à la règle.

Un aperçu de quelques règles est présenté ci-après :

#### **Exemple de règles de classification**

```
@R_C
Commutation_redondance
CL1_Commuation_de_redondance
0.8
```

```
@R_C
Controle_E-S
CL7_Controle_Entree_Sortie
0.8
```

```
@R_C
Ligne Localisation_des_trains
CL3_Localisation_des_trains
1.0
```

#### **- La sous base de règles d'adaptation**

Les règles d'adaptation sont issues de notre démarche d'acquisition de connaissances d'adaptation qui s'est basée sur l'analyse des concepts formels (voir section 4.6.3.2).

Chaque règle occupe de la même manière trois lignes, précédée d'un signe « @ ». Pour chaque cas, les instances séparés par des espaces représentent les attributs substituables avec ceux du cas cible. Les instances non précédés par de # peuvent enrichir la description du cas cible, elles présentent des instances clés.

#### **Exemple de règles d'adaptation**

```
@CASE_47
"Gestion_conduite_automatique" "Contrôle_entrée_sortie" "Suivi_trains"
"PR14" "PR15" "PR36"
```

```
@CAS_46
#ED
"Gestion_conduite_automatique" "Contrôle_entrée_sortie" "Suivi_trains"
#CS
"PR15" "PR36"
#ZA
"Limite_troncon"
```

#### **5.3.2.3. Base de cas**

La base de cas regroupe quant à elle, l'ensemble des cas sources. Pour élaborer la base de cas, nous avons eu recours à l'échantillon d'apprentissage (Hadj-Mabrouk H. et Bied-Charreton D., 1993), (Hadj-Mabrouk H., 1994, 1998) (Mejri L., 1995), (Mejri L. et al., 2009). Cet échantillon est relatif à un dossier technique comprenant soixante-dix scénarios d'accidents issus du terrain réel, dont 64 sont relatifs exclusivement au risque de collision.

En effet, la sûreté de fonctionnement des systèmes de transport automatisés requiert la prise en compte de tous les accidents potentiels (collision, déraillement, etc.) auxquels peuvent être associés de nombreux scénarios. Dans le cadre d'une étude de faisabilité, cet échantillon s'est limité à un seul accident potentiel : la collision. Les scénarios acquis sont modélisés et représentés **initialement** selon une fiche descriptive (voir Annexe 1). Ils ont été répertoriés en 12 classes (11 classes du risque de collision et une classe relative à tout autre type de risque).

La désignation  $C_i$  correspond au cas d'accident numéro  $i$  dans la base de cas,  $i$  varie de 1 à 70 tandis que  $CL_i$  correspond à la classe d'appartenance numéro  $i$  (Tableau 5.1).

| Cas d'accidents                                | Identification de la classe              |
|------------------------------------------------|------------------------------------------|
| C1, C8, C29, C30, C47                          | CL1 : Commutation de redondance          |
| C4, C5, C6, C21, C22, C23, C34, C66            | CL2 : Séquence d'initialisation          |
| C3, C7, C17, C19, C25, C38, C45, C53, C63, C65 | CL3 : Localisation des trains            |
| C9, C10, C20, C24, C50, C64                    | CL4 : Gestion du freinage d'urgence      |
| C13, C14, C15, C32, C33, C35, C40              | CL5 : Accostage                          |
| C2, C12, C16, C18, C55                         | CL6 : Gestion de sens de marche          |
| C11, C31, C36, C37, C39, C44, C46, C48         | CL7 : Contrôle d'Entrée / Sortie         |
| C26, C41, C42, C54                             | CL8 : Suivi de l'ordre des trains        |
| C27, C28, C43, C49, C51, C60                   | CL9 : Conduite manuelle                  |
| C52, C67                                       | CL10 : Contrôle / Commande des aiguilles |
| C68, C69, C70                                  | CL11 : Contrôle de vitesse               |
| C46, C48, C56, C57, C58, C59, C61, C62         | CL12 : Autre risque                      |

**Tableau 5.1. Distribution des cas selon la classe d'appartenance (Hadj-Mabrouk H, 1994)**

Nous avons repris cet échantillon d'apprentissage et nous avons harmonisé la représentation de tous les concepts figurants dans les scénarios selon les modèles de connaissances que nous avons développées dans notre approche ADAST (voir chapitre 4, section 4.4).

La figure 5.3 montre alors le « Mapping » entre les anciens paramètres descriptifs utilisés et les nouveaux descripteurs issus de l'emploi de l'ontologie :

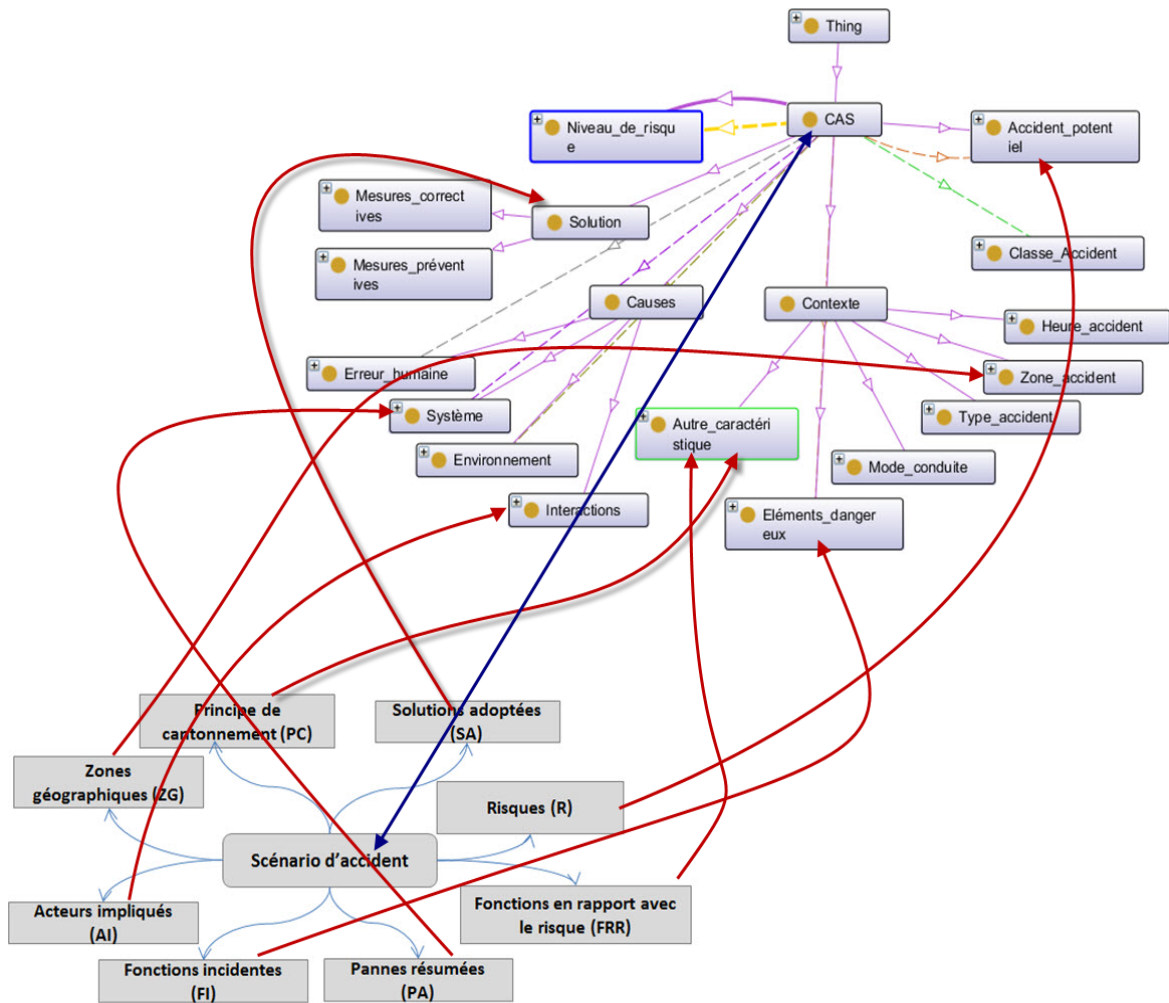


Figure 5.3. Mapping entre les anciens et les nouveaux paramètres descriptifs.

La figure 5.4 présente quant à elle un aperçu de la représentation du cas source « Cas\_01 » en langage Owl (Ontology Web Language).

```

<Cas rdf:ID="CAS_01">
  <a_niveau_risque rdf:resource="#Risque-intolérable"/>
  <a_libellé rdf:resource="#Scénario_1"/>
  <a_éléments_dangereux rdf:resource="#Gestion_sens_marche"/>
  <a_cause_système rdf:resource="#PR07"/>
  <a_éléments_dangereux rdf:resource="#Gestion_arrêt_train"/>
  <a_cause_env rdf:resource="#Pluie"/>
  <a_cause_système rdf:resource="#PR08"/>
  <a_cause_erreur_humaine rdf:resource="#Défaillance_conducteur_1"/>
  <a_contexte rdf:resource="#Terminus"/>
  <a_solution rdf:resource="#SA1"/>
  <appartient_classe_accident rdf:resource="#CL1"/>
  <a_éléments_dangereux rdf:resource="#Communication_redondance"/>
  <a_accident_potentiel rdf:resource="#Collision"/>
  <a_éléments_dangereux rdf:resource="#Localisation_des_trains"/>
</Cas>

```

Figure 5.4. Aperçu de la représentation du cas source « Cas\_01 »

### 5.3.3. Architecture fonctionnelle de la maquette basée sur ADAST

L'architecture de la maquette de faisabilité comprend deux modules principaux : le module d'acquisition des cas et le module de raisonnement. Ces deux modules sont en interaction avec d'autres composants faisant le rôle d'interface avec la base de connaissances et l'ontologie de domaine. Il est à signaler que les deux modules, d'acquisition de cas et d'aide à la décision, illustrent les traitements faites en-ligne.

La figure 5.5 présente l'architecture fonctionnelle de la maquette proposée. Les principaux composants de l'architecture fonctionnelle sont les suivants :

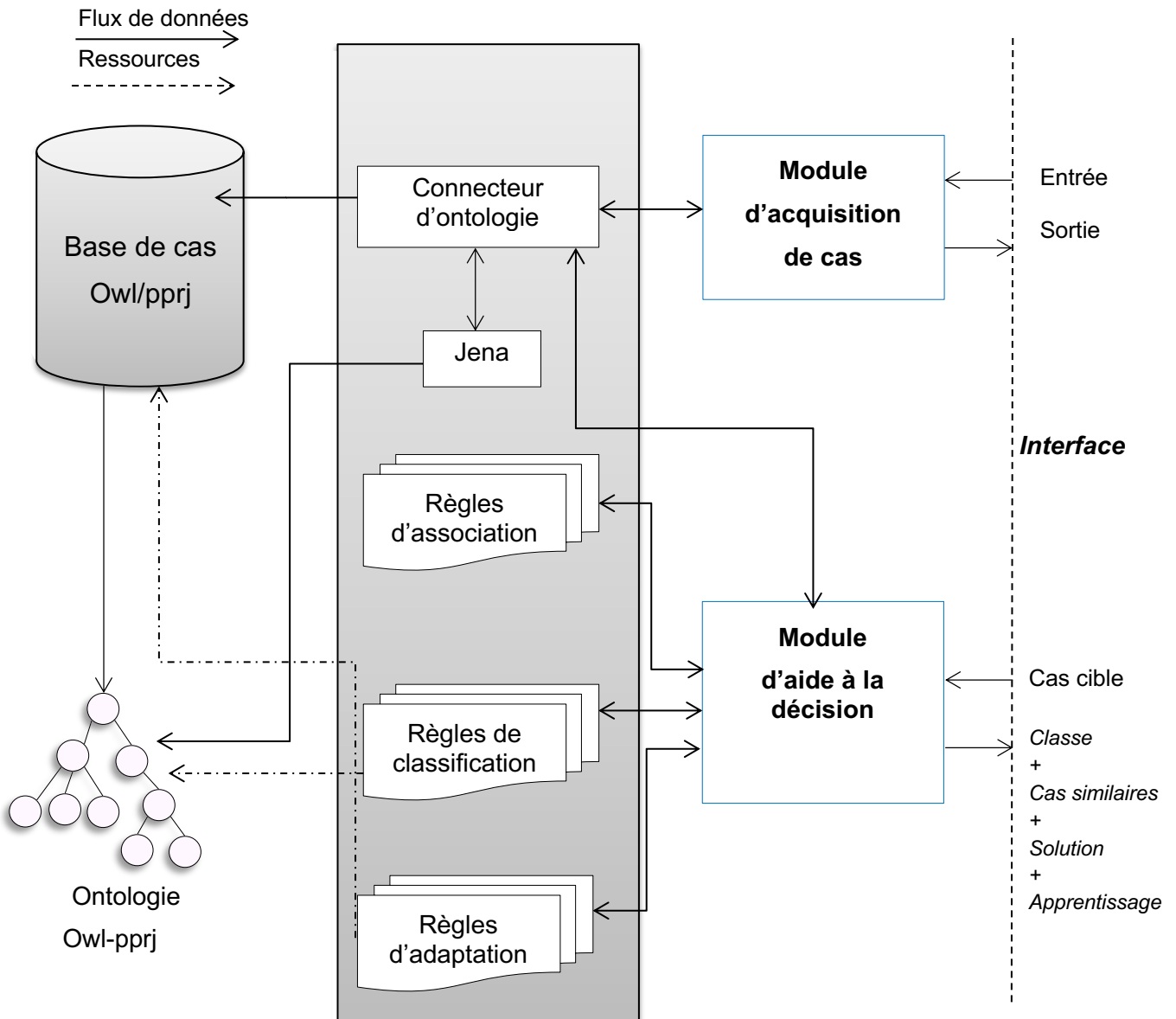


Figure 5.5. Architecture fonctionnelle de la maquette d'aide à la décision.

- **Jena** : C'est un Framework écrit en Java fournissant un environnement qui facilite le développement d'applications pour le web sémantique. Jena manipule des ontologies sous différents formats tel que RDFS et OWL. Jena permet aussi le raisonnement sur l'ontologie à travers des moteurs d'inférence. Dans notre application, Jena est utilisée :

- Par le connecteur d'ontologie lors de la récupération des cas :
  - « *plugins/edu.stanford.smi.protege.owl/jena.jar* »
  - owlModel = (OWLModel) ProtegeOWL.createJenaOWLModelFromURI(uri);*
- Pour récupérer des informations sur les concepts et les instances de l'ontologie lors du calcul de similarité et notamment les différentes opérations relatives à la gestion de l'ontologie.
- **Règles d'association** : Ces règles d'association vont permettre à l'utilisateur expert de bien ajuster/compléter la description initiale introduite du cas cible à traiter ;
- **Règles de classification** : Ces règles servent à la détection de la classe d'appartenance d'un nouveau cas cible. Elles se présentent comme des ressources pour l'ontologie ;
- **Règles d'adaptation** : Ces règles se trouvent dans des fichiers externes. Le module de raisonnement les utilise pour appliquer les substitutions ou l'enrichissement convenable permettant la proposition d'une solution possible au cas cible ;
- **Base de cas** : La base de cas regroupe l'ensemble de cas représentés en ayant recours à l'ontologie de domaine ;
- **Ontologie** : L'ontologie est le formalisme de base pour la description des cas sources.
- **Connecteur d'ontologie** : Il assure la connexion à l'ontologie pour pouvoir manipuler les connaissances emmagasinées dedans.

*String uri = "file:///c:/0accident.owl";*

*owlModel = (OWLModel) ProtegeOWL.createJenaOWLModelFromURI(uri);*

#### 5.4. Modules

La maquette de faisabilité d'ADAST est en effet une application Java. L'environnement de développement utilisé est Netbeans. De nombreuses fonctionnalités sont donc offertes que nous pouvons directement exploiter. Dans l'état actuel des travaux, les deux modules, le module d'acquisition de connaissances et le module de raisonnement à partir de cas, sont implémentés et constituent une première étape sur laquelle porte l'étude de faisabilité du système d'aide à la décision proposé.

L'utilisateur aura accès au premier écran d'authentification (Figure 5.6) dans lequel il est amené à saisir ses paramètres d'accès (le nom d'utilisateur et le mot de passe). Une fois la connexion est réussie, le menu principal sera affiché (Figure 5.7).

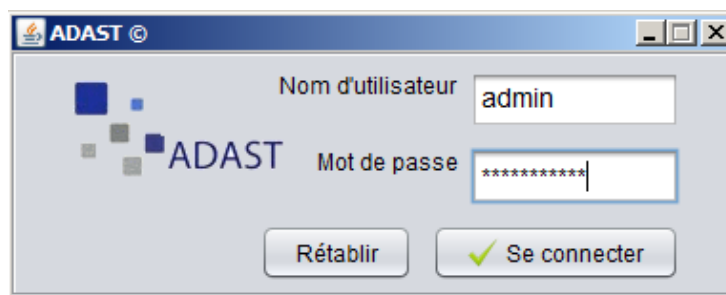


Figure 5.6. Écran d'authentification.

L'utilisateur peut interagir avec les deux modules principaux de la maquette. Il pourra dans un premier temps accéder à la base de connaissances du système qui comprend en effet la base de cas, l'ontologie de domaine et notamment les différentes bases de règles utilisées dans le processus de résolution de problème.

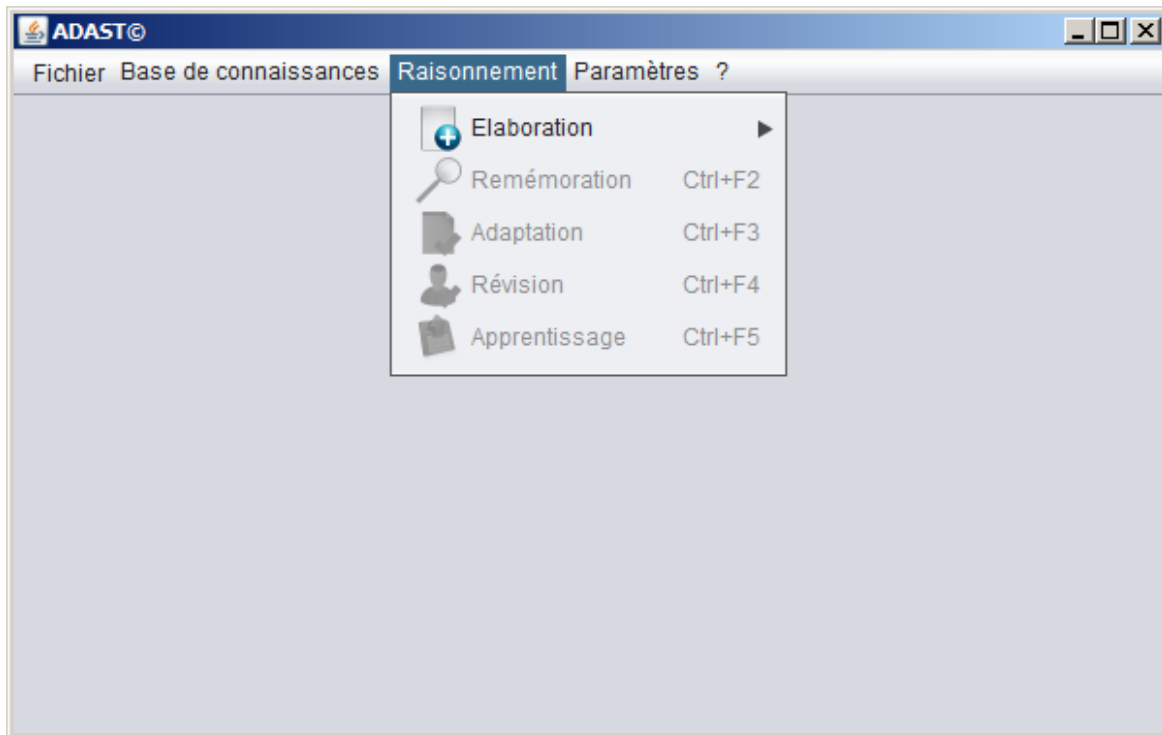


Figure 5.7. Menu principal de la maquette ADAST.

#### 5.4.1. Module d'acquisition de connaissances

Ce module présente un intérêt crucial vu qu'il représente un chantier de familiarisation avec les différents langages et formats de fichiers comme RDF et OWL. Il comprend des diverses fonctions permettant de gérer les connaissances capitalisées dans l'ontologie de domaine et notamment la base de cas.

Nous allons présenter successivement dans ce paragraphe les écrans suivants :

- L'explorateur de l'ontologie ;
  - L'écran d'acquisition des cas ;
  - L'écran d'ajout d'instances ;
  - L'écran dédié à la consultation et à la suppression des cas.
- **Explorateur d'ontologie**

Nous présentons dans la figure 5.8, l'explorateur d'ontologie. Cet explorateur va permettre la visualisation de la hiérarchie des concepts et de sous-concepts du modèle de domaine (les ontologies noyau et de domaine).

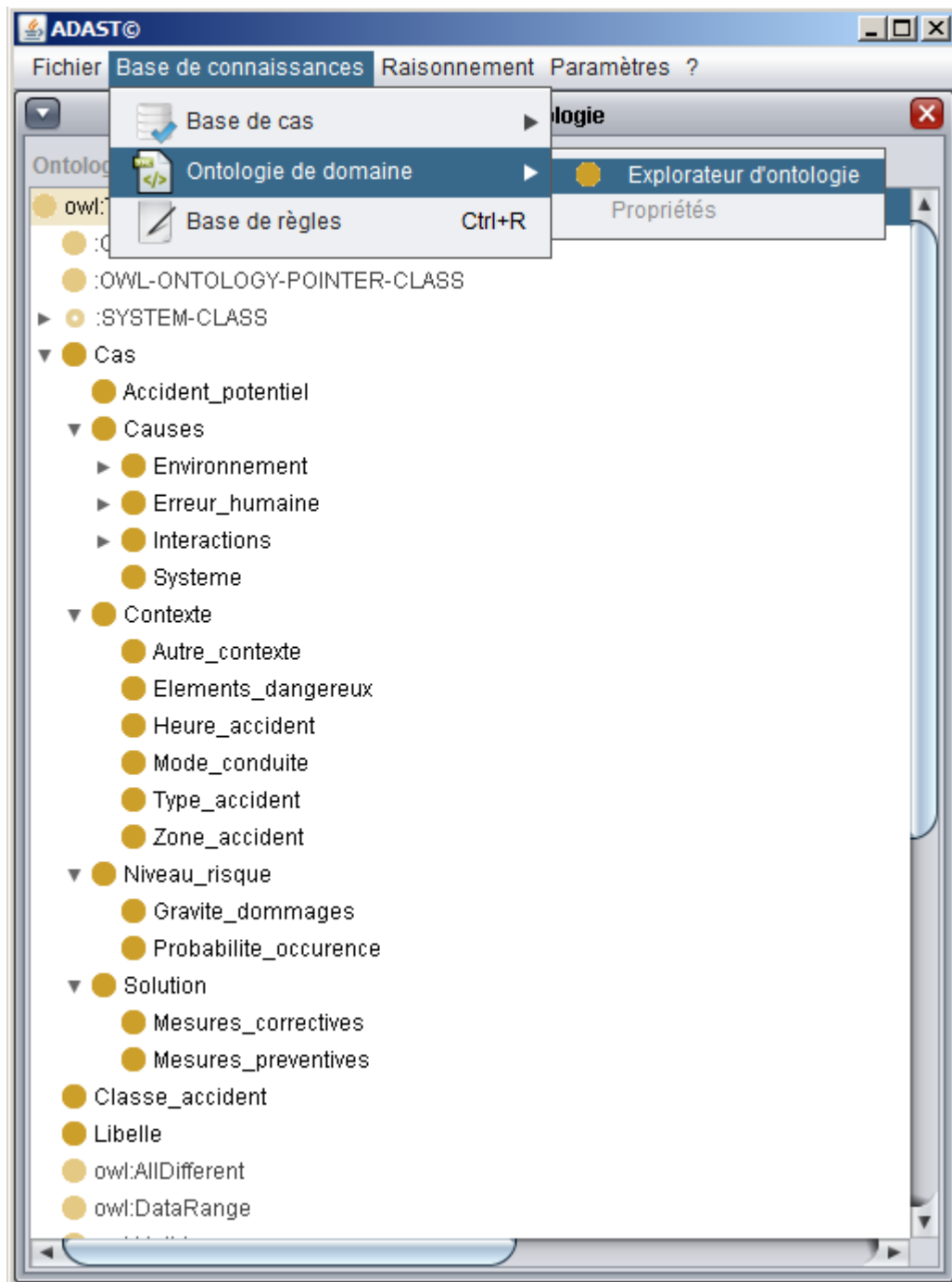


Figure 5.8. Explorateur de l'ontologie.

#### - Acquisition des cas sources

La figure 5.9 présente quant à elle l'écran d'acquisition de cas. C'est une fonctionnalité cruciale qui permet de construire et introduire un nouveau cas d'accident (cas source) à la base de cas. La construction de chaque nouveau cas source passe par les trois étapes essentielles suivantes :

1. Activer la commande « Acquisition » dans le menu « Base de cas » (Figure 5.9) pour lancer une nouvelle instance de l'éditeur de cas (Figure 5.10) :



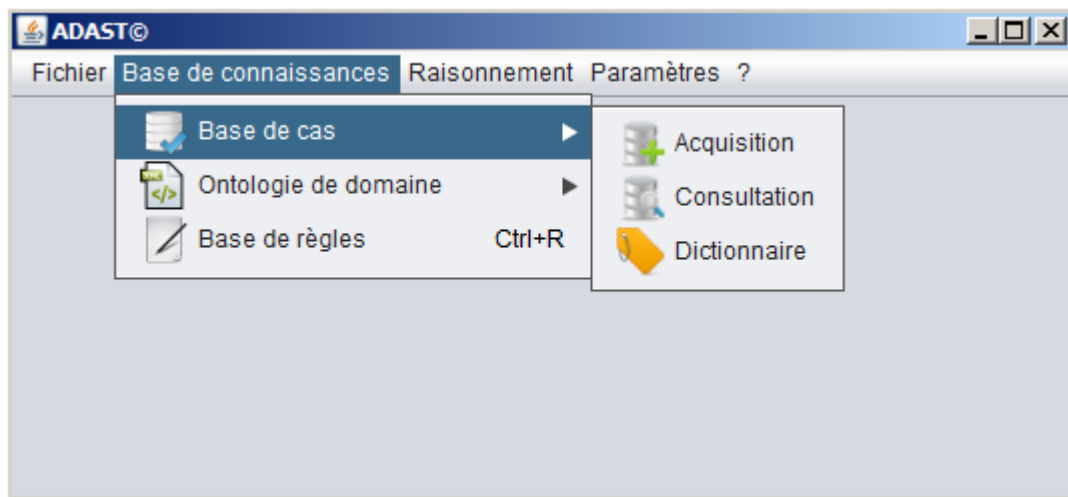


Figure 5.9. Ouverture de l'éditeur de cas.

2. Définir les attributs de la description du cas source qui correspondent à des instances extraites à partir de l'ontologie (Figure 5.10) :
  - a. Le système visualise les différents attributs et notamment les instances sous forme de formulaire à remplir. En ce sens, les différentes valeurs d'attributs sont regroupées par classe d'attributs selon les niveaux dans la hiérarchie ontologique ;

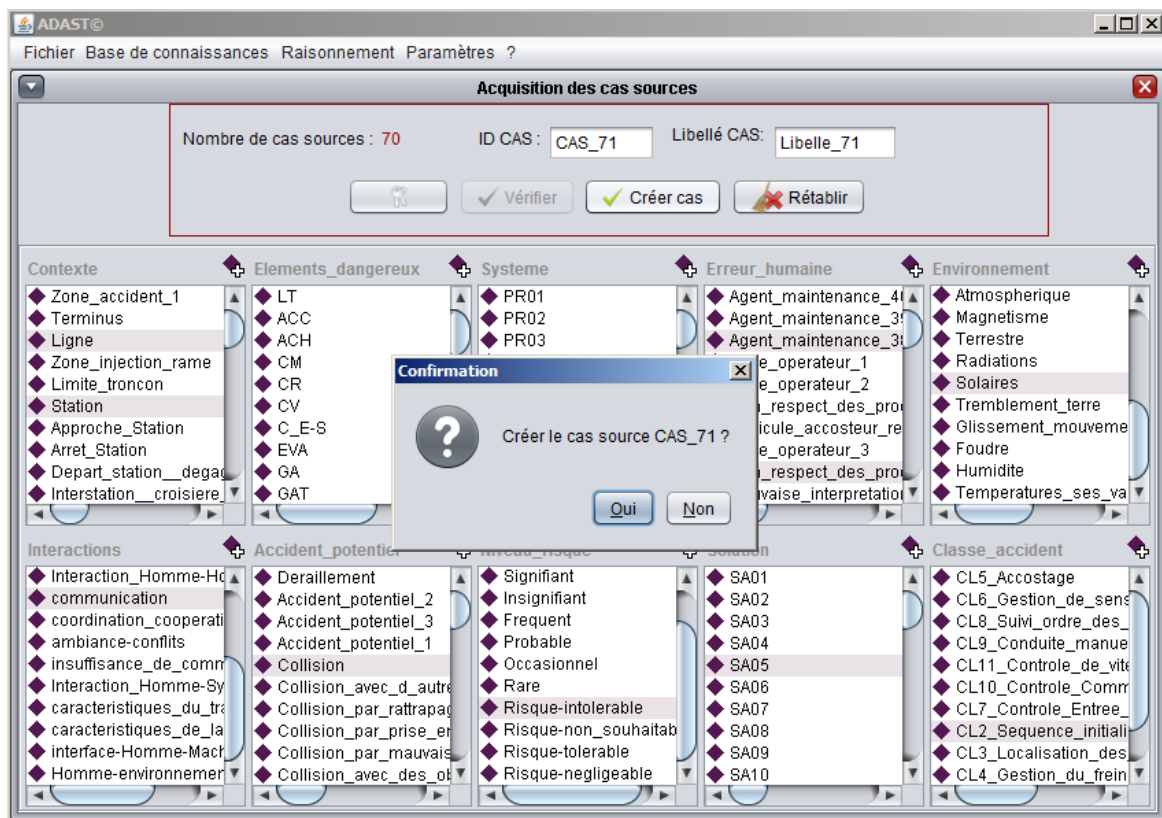


Figure 5.10. Écran d'acquisition de cas sources.

- b. L'utilisateur pourra ainsi faire les différentes opérations relatives à l'acquisition d'un nouveau cas source. Chaque cas source introduit doit être accompagné par un identifiant (ID CAS) et un libellé. La construction d'un cas crée une nouvelle instance du cas dans le modèle de cas, et y relie les

instances qui ont été rajoutées à la description du contexte, l'accident potentiel, le niveau de risque, les causes définies ainsi que la solution et la classe d'appartenance. Dans le cas où il y'aura des instances manquantes, un message d'avertissement sera affiché pour rappeler les éléments nécessaires permettant la construction d'un nouveau cas source (Figure 5.11) :

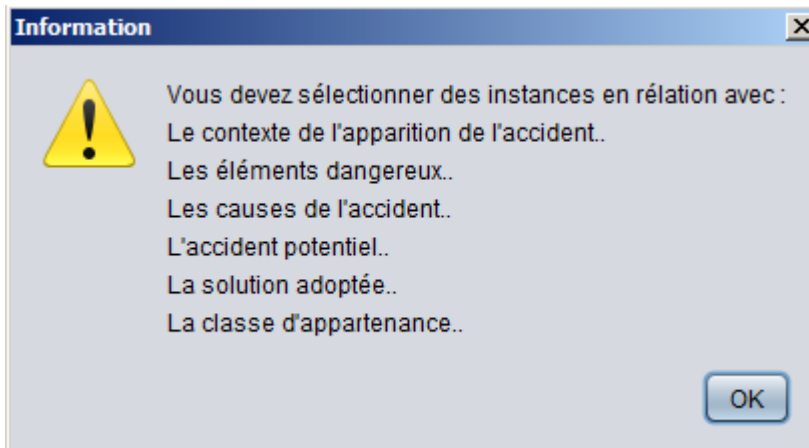


Figure 5.11. Les éléments nécessaires pour acquérir un nouveau cas source.

3. L'utilisateur peut notamment ajouter des nouvelles instances (Figure 5.12). Chaque instance ajoutée possède les éléments de description suivants :
  - a. Un libellé qui indique le concept père de l'instance ;
  - b. Un champ de texte lié à une instance du concept père. L'utilisateur peut saisir une nouvelle instance, qui sera ensuite enregistrée dans l'ontologie.

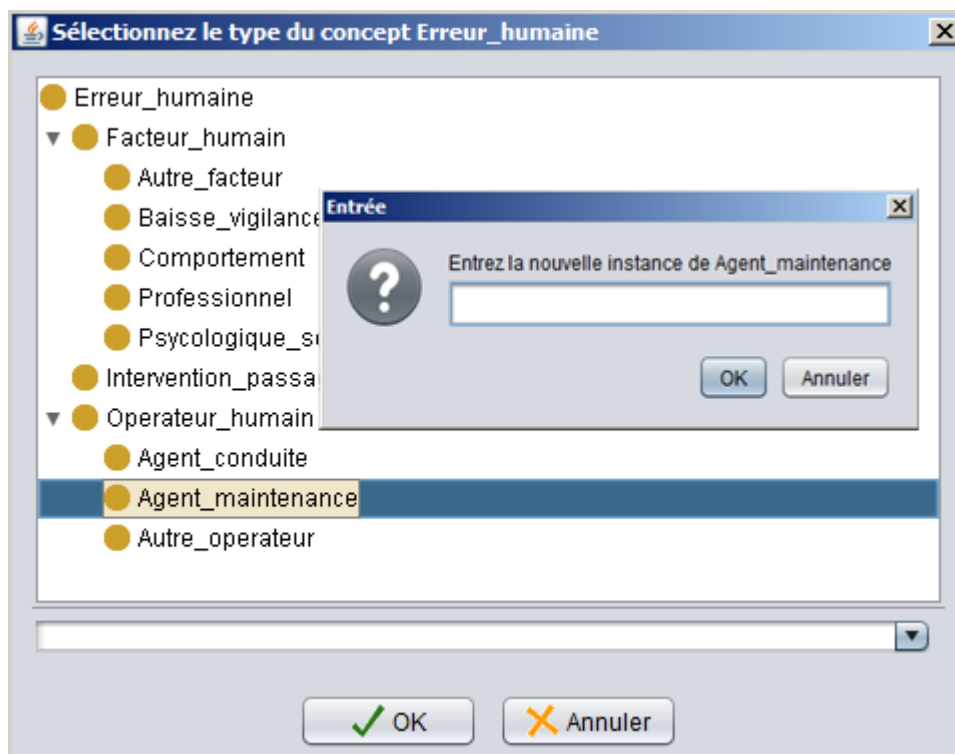


Figure 5.12. Ajout d'une instance à la description d'un cas source.

## - Consultation et suppression des cas sources

La figure 5.13 présente quant à elle l'écran de consultation et notamment la suppression des cas sources. Chaque bouton permet le déclenchement d'une procédure adéquate qui ouvre la connexion avec l'ontologie et effectue l'action demandée. La consultation et la suppression d'un cas passe par les étapes suivantes :

1. Sélectionner la classe correspondante et cliquer sur le bouton « Afficher les cas » pour chercher tous les cas appartenant à cette dernière ;
2. Sélectionner le cas source et cliquer sur le bouton « Afficher les instances ». Une procédure va pointer sur le cas en question dans le fichier Owl grâce à la balise : `<Cas rdf:ID=.. >...</Cas>`. Les instances extraites vont être mises dans un tableau puis affichées dans l'interface ;
3. L'utilisateur pourra supprimer le cas sélectionné à travers le bouton « Supprimer le cas ». Un message de confirmation sera affiché avant de passer à l'action.

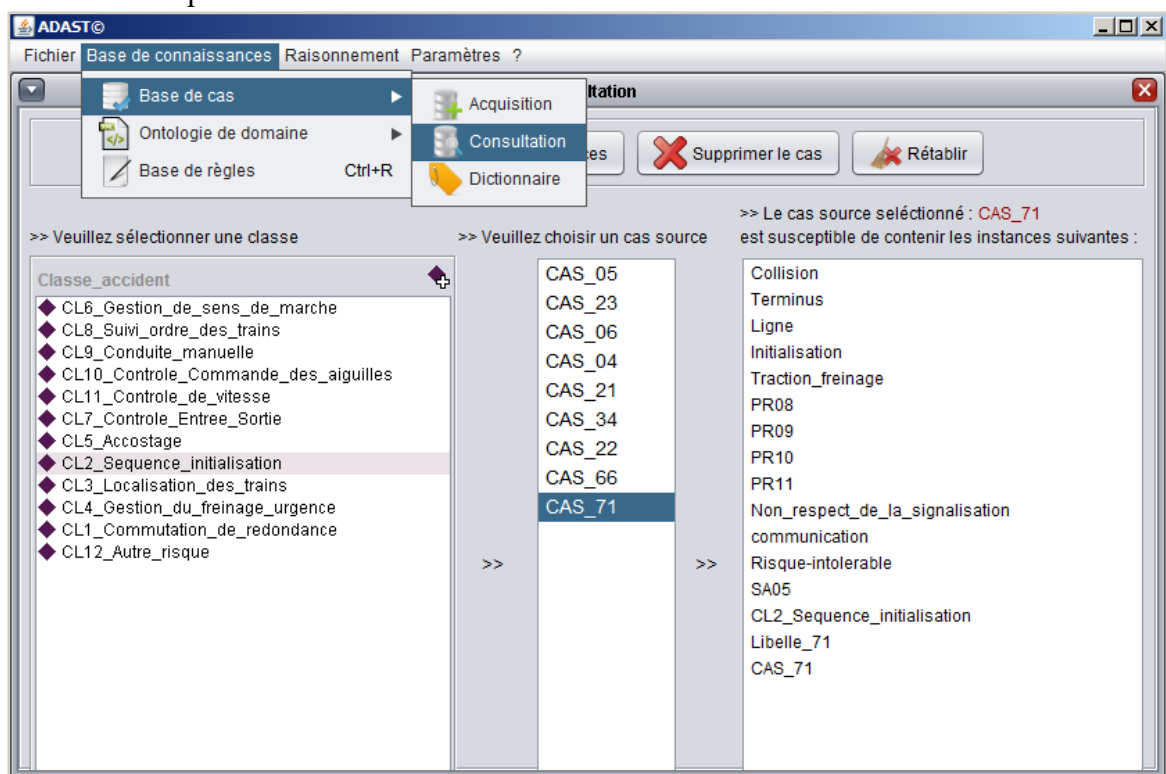


Figure 5.13. Écran de consultation de cas sources.

## 5.4.2. Module d'aide à la décision

Dans ce paragraphe nous allons présenter le module d'aide à la décision à travers les différentes phases de RàPC. Cependant, il convient dans un premier temps de présenter l'écran de préconception relatif aux paramètres de raisonnement.

### 5.4.2.1. Préconception

L'utilisation de la maquette ADAST nécessite en premier lieu une étape de préconception (Figure 5.14) au cours de laquelle nous définissons les différentes valeurs des paramètres et des contraintes utiles au raisonnement.

Dans cette étape, nous pouvons ajuster les paramètres de raisonnement : nous pouvons fixer le seuil de similarité global  $\alpha$  (voir section 4.6.2.2). Nous pouvons notamment changer

le seuil minimal de probabilité pour lequel les règles de classification (à l'issue de l'étape de remémoration) peuvent être prises en considération au moment de l'inférence. Nous pouvons varier finalement les trois paramètres : la confiance, le support et le lift) pour la sélection des règles d'association (voir section 4.6.1).

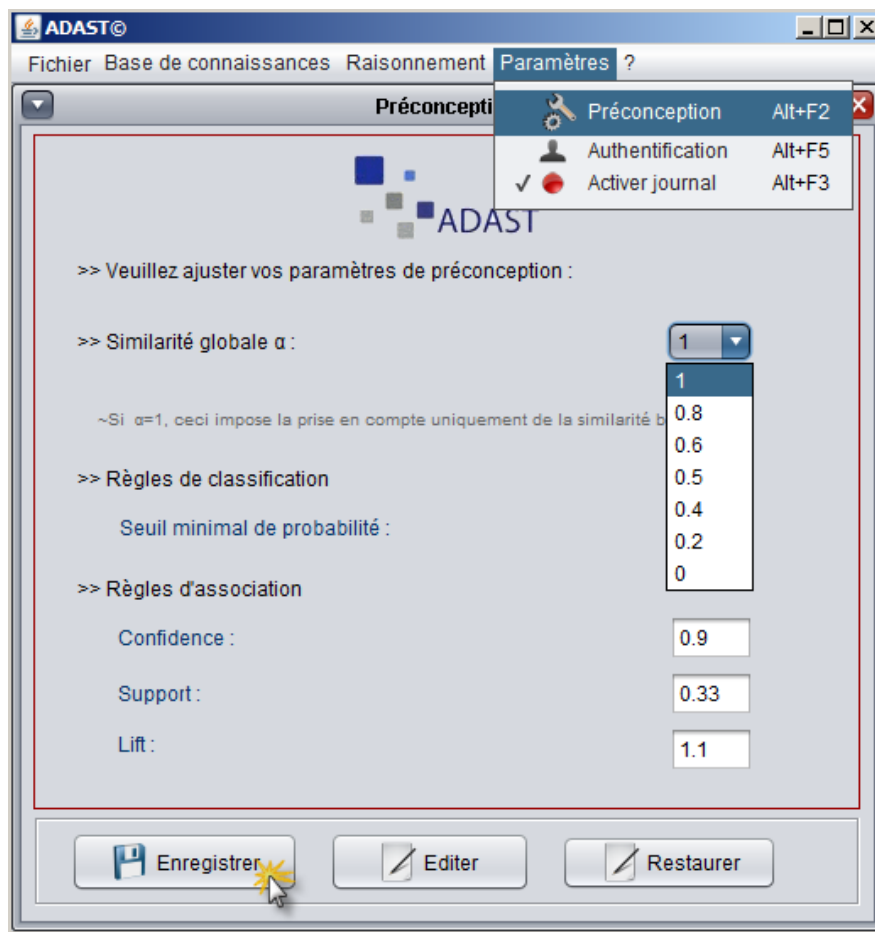


Figure 5.14. Écran de préconception.

Par défaut, les différents paramètres sont pris comme suit :

- $\alpha = 1$  : Le seuil de la similarité globale, il permet de varier l'importance à mettre en considération la similarité basée-slot. Un seuil fixé à 1 permet de prendre en compte seulement ce type de similarité ;
- **Le niveau de probabilité** minimal pour les règles de classification. L'utilisateur peut varier ce niveau de probabilité qui est considéré comme niveau d'acceptabilité des règles. Pour ce paramètre, nous avons considéré dans les tests une valeur très exigeante. En ce sens une valeur de 0,9 donnera une grande confiance dans la classification vue que cette étape a un impact énorme sur la suite des traitements dans le processus RàPC ;
- **La confiance, le support et le lift** : Trois paramètres pour la sélection des règles d'association à partir de la sous-base de règles d'association. Ces paramètres vont être expliqués davantage dans la partie expérimentation au sein de chapitre suivant.

#### 5.4.2.2. Processus de raisonnement

La mise en œuvre du processus de raisonnement, comme nous l'avons déjà vu dans notre approche ADAST, implique le déroulement de cinq étapes consécutives suivantes :

- Élaboration du cas cible ;
- Remémoration des cas similaires ;
- Adaptation ;
- Révision ;
- Et apprentissage.

- **Élaboration d'un cas cible**

Dans cette phase, nous avons suivi deux alternatives pour acquérir un nouveau cas cible. La figure 5.15 présente l'écran dédié à la première méthode de l'élaboration d'un nouveau cas cible tandis que la deuxième est présente dans la figure 5.16.

Dans la première méthode, l'utilisateur sera invité à introduire son cas cible (une description d'un scénario d'accident pour lequel il cherche une solution convenable). Il sélectionne ainsi des instances correspondantes à partir de l'ontologie pour spécifier le cas cible. Il est amené ainsi, à décrire le contexte de l'apparition de l'accident, les différentes causes et aussi l'accident potentiel. En cas de besoin, il pourra aussi introduire de nouvelles instances qui vont être enregistrées dans l'ontologie.

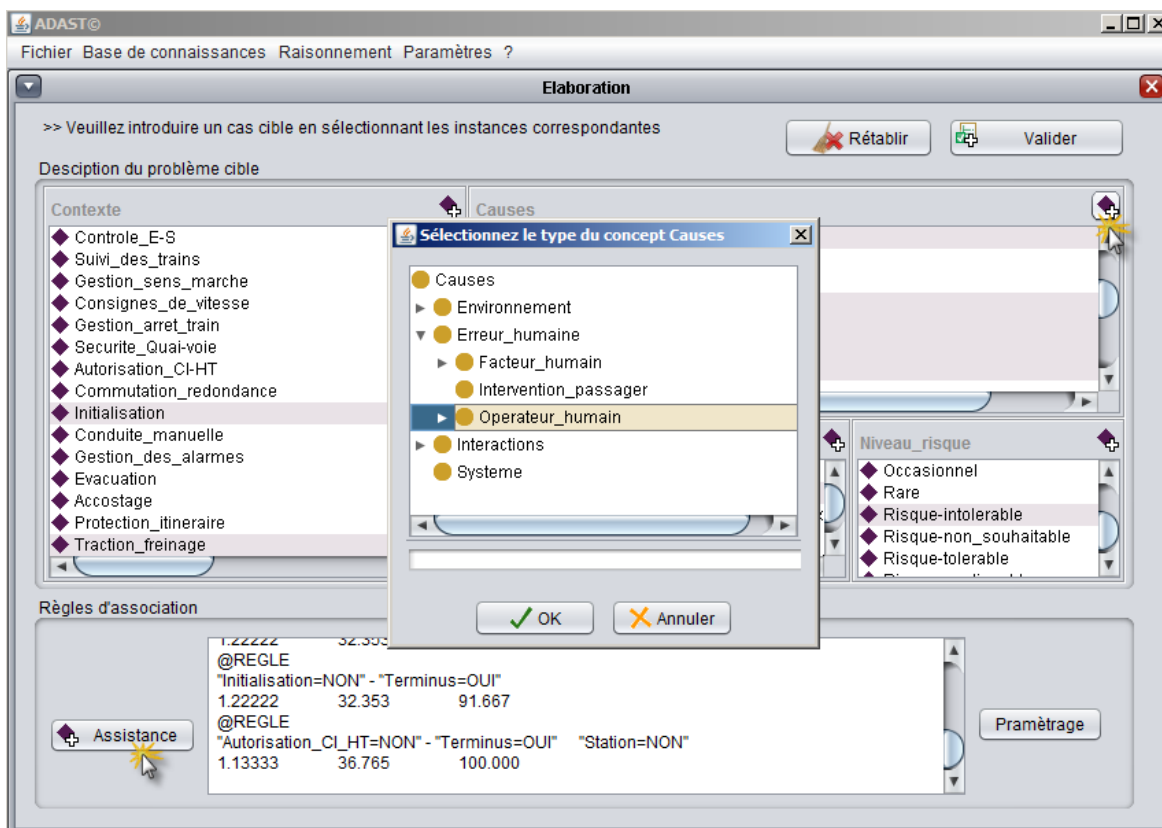


Figure 5.15. Écran dédié à l'élaboration du cas cible (Première méthode).

Le bouton « Assistance » en bas à gauche, permet de déclencher la recherche des règles d'association se rapprochant à la description initiale introduite. Une procédure de recherche va être appliquée selon un mécanisme de chaînage-avant sur un fichier contenant la sous base de règles de classification (section 5.2.1.2).

Pour la deuxième méthode (Figure 5.16), nous avons travaillé plus l'interface à travers les cases à crocher pour que l'utilisateur choisisse préalablement les listes pouvant contenir les instances du nouveau cas cible. L'élaboration du cas cible comporte quatre étapes consécutives :

- L'utilisateur croche les éléments dont il dispose et visualise les listes extraites de l'ontologie à travers le bouton « OK » ;
- Il sélectionne les instances adéquates et les enregistre en cliquant sur le bouton « Enregistrer » ;
- Il déclenche la recherche des règles d'association en relation avec les instances déjà acquises ;
- Il clique sur le bouton « Ajuster » pour spécifier le type et le poids de chaque instance d'attribut (voir section 4.6.2.2). Les poids varient généralement entre 0 (attributs négligeables) et 2 (attributs clés). Les attributs se distinguent par leur type : attributs complexes (C) (Figurant à un niveau non terminal de l'ontologie des attributs et qui préconisent une attention particulière dans le calcul de similarité) et attributs simples (S).

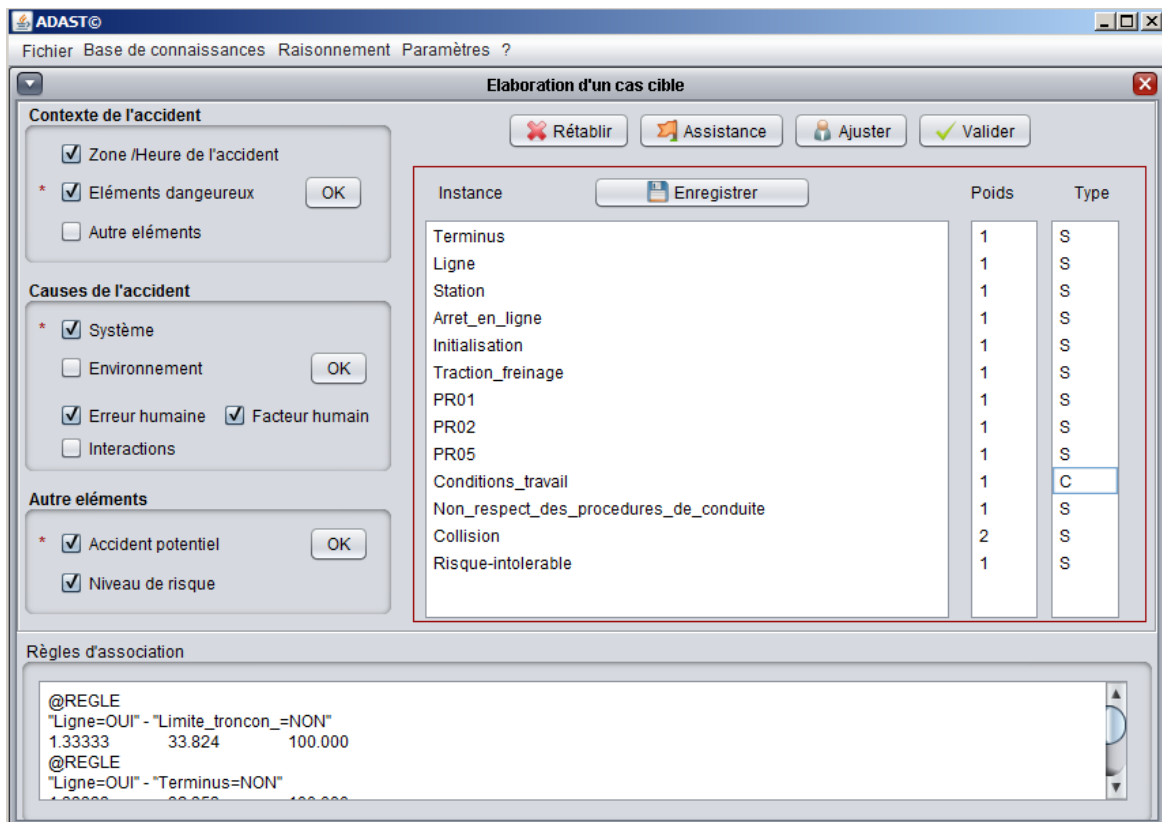


Figure 5.16. Écran dédié à l'élaboration du cas cible (Deuxième méthode).

Une fois la description du cas cible a été bien ajustée par l'utilisateur, il pourra enregistrer le nouveau cas cible en cliquant sur le bouton « Valider ». Un message de confirmation (Figure 5.17) va être affiché à l'écran pour indiquer éventuellement le nombre d'instances. Toutefois, il pourra régler la valeur  $\alpha$  qui représente le seuil de la similarité globale (Figure 5.14) qui, par défaut, est égale à 1.

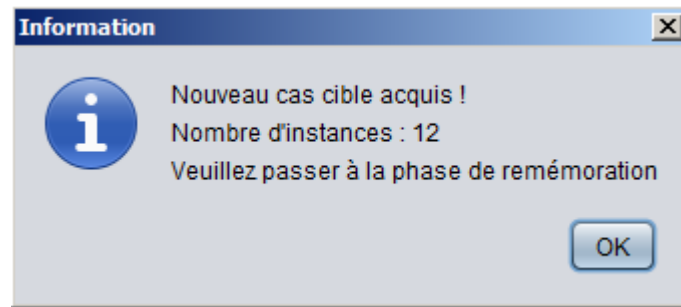


Figure 5.17. Acquisition d'un nouveau cas cible.

### - Remémoration

La figure 5.18 présente l'écran dédié à la deuxième phase de raisonnement, la remémoration des cas similaires. Deux parties se distinguent sur cet écran : la première pour la classification et la seconde pour l'extraction des cas similaires et le calcul de la similarité.

Après la saisie de la description du cas cible, la phase de "classification" peut être lancée. Le résultat de la classification est présenté en figure 5.18. L'utilisateur sera invité dans un premier temps à appuyer sur le bouton « Détecter Classe » pour déclencher la recherche automatique de la classe d'appartenance du cas cible. Un message va être affiché pour indiquer la classe détectée (Figure 5.19).

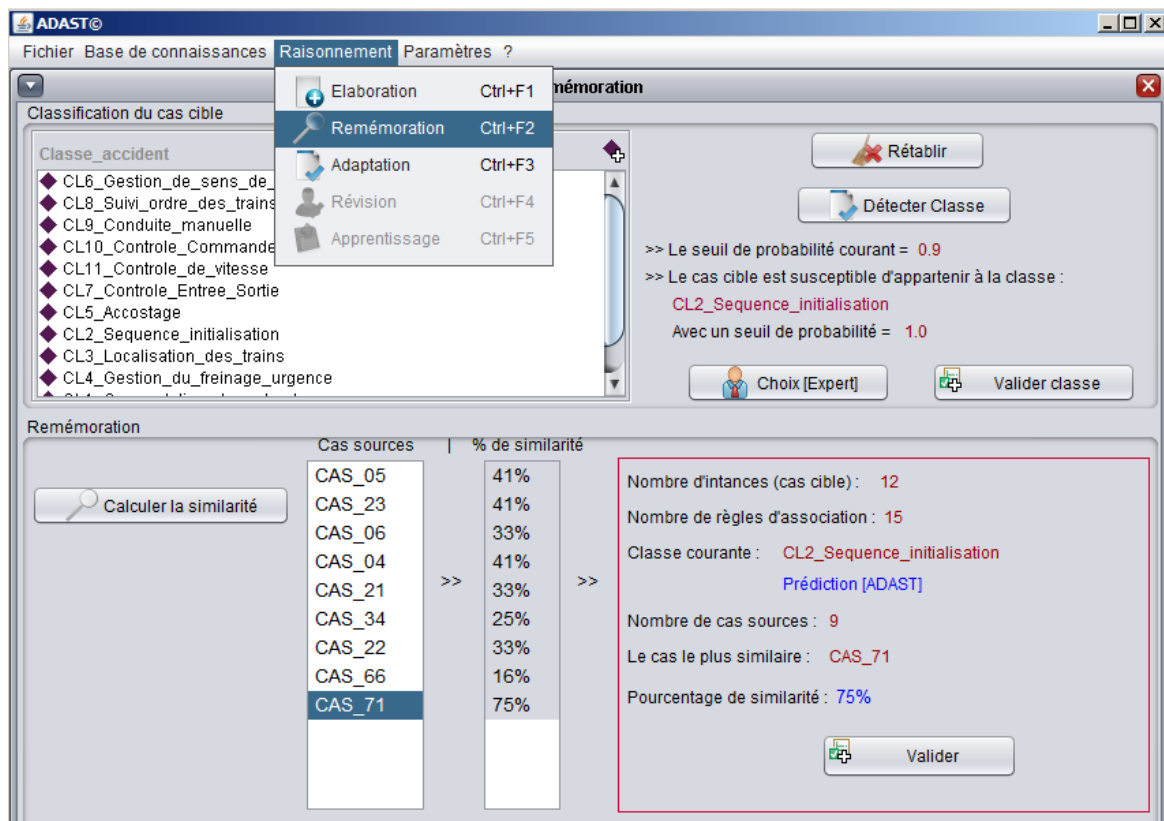


Figure 5.18. Écran dédié à la phase de remémoration.

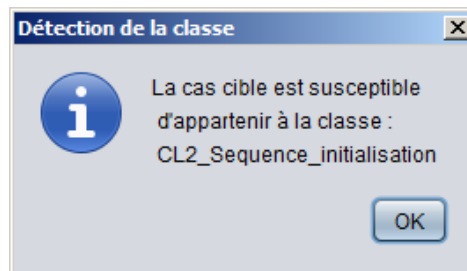


Figure 5.19. Écran dédié à la phase de remémoration.

Toutefois, l'utilisateur aura à choisir d'accepter ou de rejeter la classification proposée. En cas de rejet, il pourra sélectionner une autre classe à partir d'une liste d'instances extraite de l'ontologie (Figure 5.18). Dans la deuxième partie de l'écran, l'utilisateur sera sollicité à déclencher la procédure de recherche du cas le plus similaire au cas cible à travers le bouton « Calculer la similarité ». Le calcul va être opéré dans la classe d'appartenance déjà attribuée au cas cible.

### - Adaptation

La figure 5.20 est dédiée à l'écran d'adaptation. L'utilisateur sera invité à visualiser l'ensemble d'instances correspondantes respectivement au cas cible et au cas le plus similaire. Ensuite, il sera amené à déclencher la procédure d'inférence en activant le bouton « Adapter ». Cette procédure va permettre dans un premier temps, l'application des règles d'adaptation convenables liées au cas source le plus similaire, et ceci, dans le but d'enrichir la description initiale du cas cible, puis la recherche d'une solution adéquate en se basant sur cette nouvelle description pour enfin proposer une solution possible au cas cible.

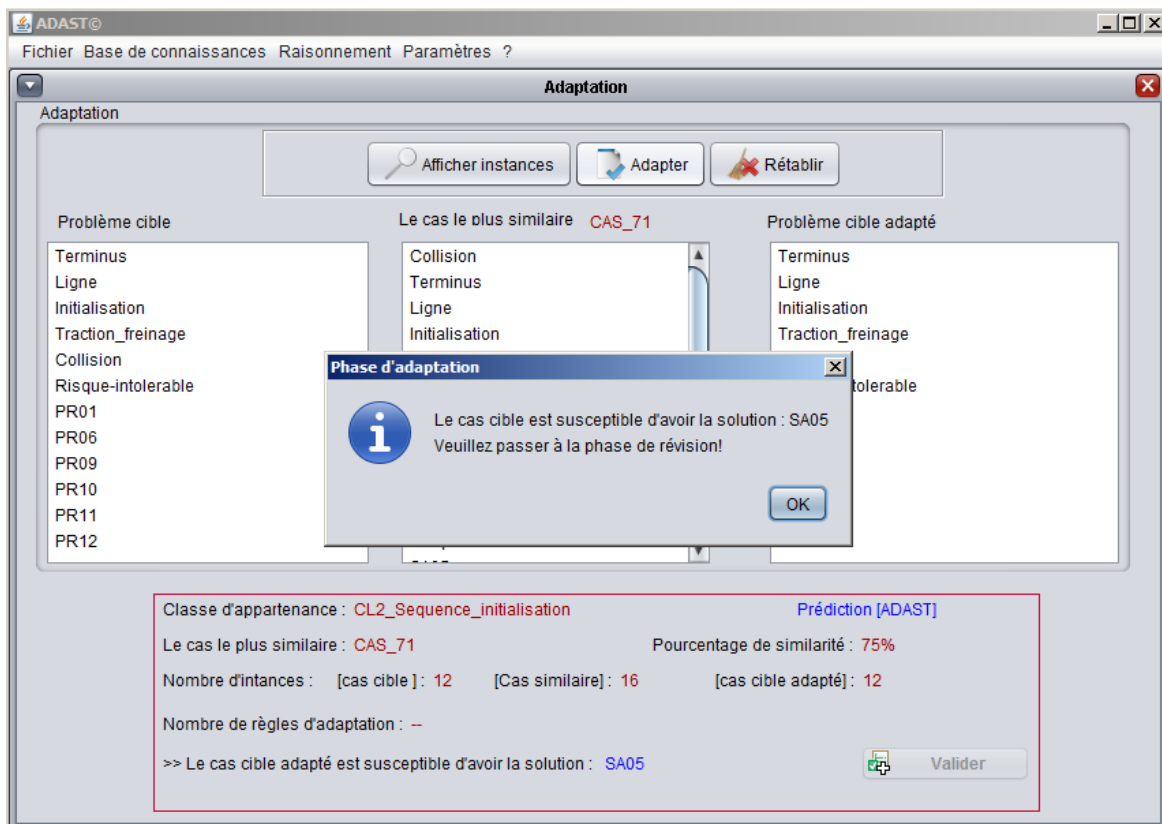


Figure 5.20. Écran dédié à la phase d'adaptation.



## - Révision

La figure 5.21 présente l'écran de révision dans lequel l'utilisateur sera invité conformément à la description que nous avons écrit dans la section 4.6.4, à réviser la solution proposée à l'issue de la phase d'adaptation.

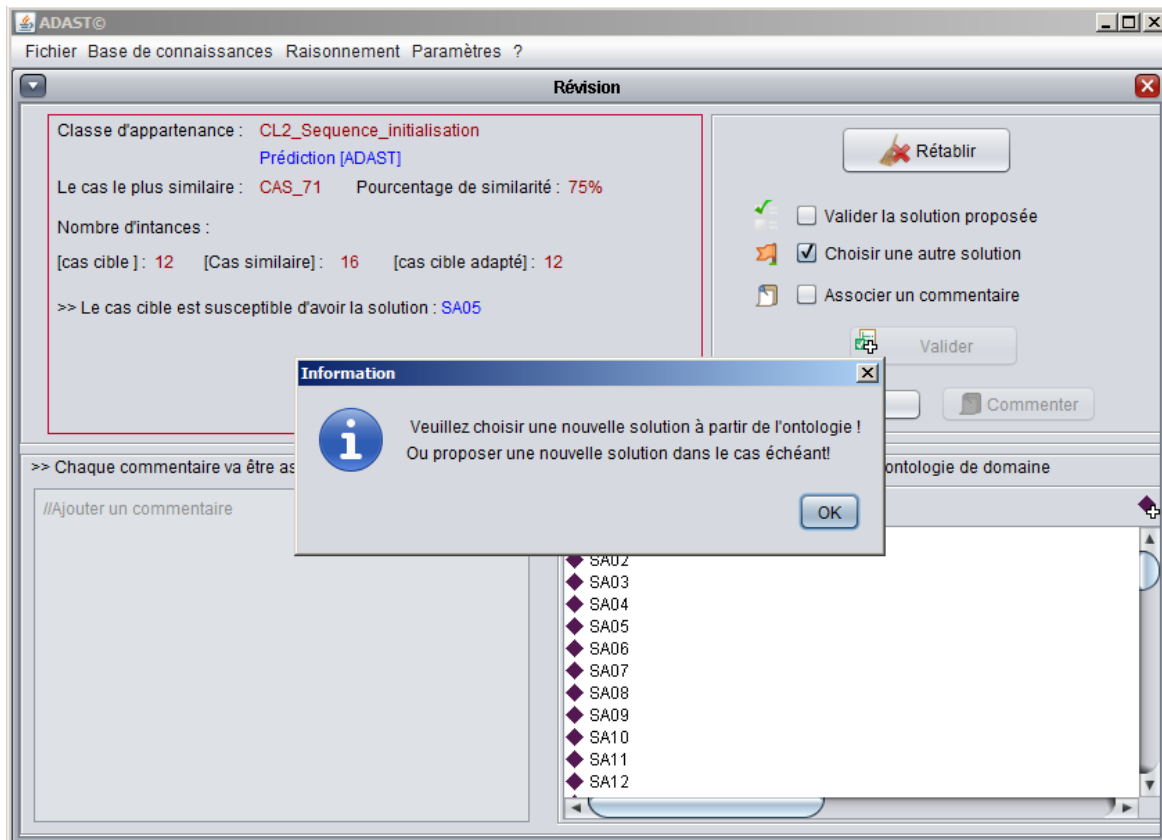


Figure 5.21. Écran dédié à la phase de révision.

Dans la partie droite de l'écran, l'utilisateur aura le choix d'accepter, de commenter ou de rejeter la solution proposée. S'il la rejette, un menu s'affiche pour préciser les suites envisageables à son refus (choix à partir d'une liste d'instances de l'ontologie) (Figure 5.21). En cas de validation, un message de confirmation sera affiché également (Figure 5.22).

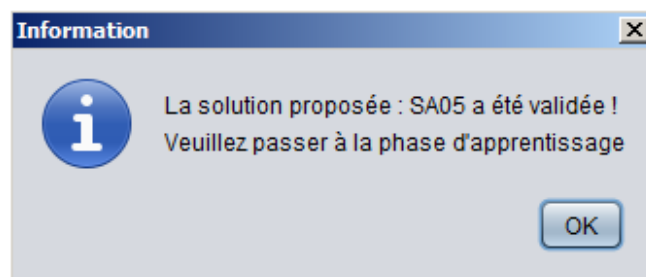


Figure 5.22. Message de validation de la solution proposée

## - Apprentissage

La figure 5.23 présente l'écran d'apprentissage. Conformément à notre démarche d'apprentissage proposée dans ADAST (voir section 4.6.5), deux parties se distinguent sur cet écran : la première pour l'application de l'algorithme d'auto-incrémentation et la seconde pour la prise de décision.

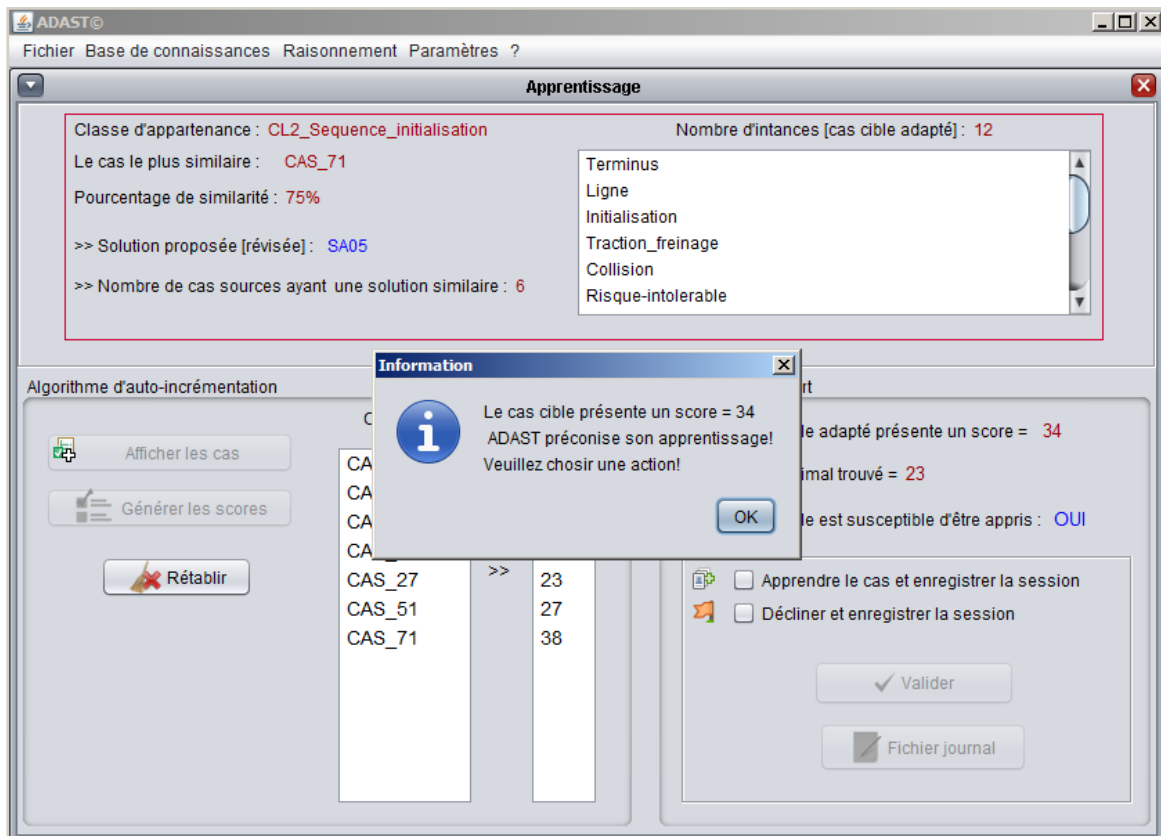


Figure 5.23. Écran dédié à la phase d'apprentissage.

L'utilisateur est amené à visualiser dans un premier temps, les cas sources qui partagent la même solution proposée à travers le bouton « Afficher les cas ». L'activation du deuxième bouton « Générer les scores », va appliquer une procédure qui va calculer dans un premier temps, les scores des cas sources qui partagent la même solution. Deux suites envisageables sont donc présentes :

- En cas d'absence de cas sources, le système va procéder l'apprentissage automatique (Figure 5.24). L'utilisateur est amené alors à saisir un ID-CAS pour pouvoir enregistrer et enrichir la base de cas ;

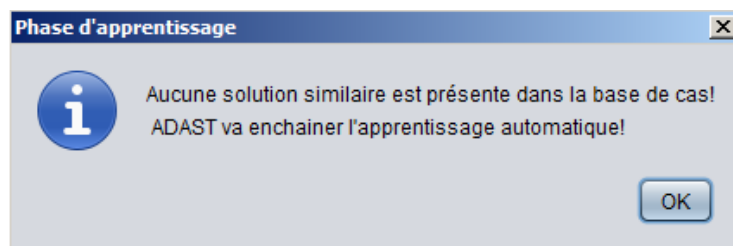


Figure 5.24. Message de confirmation de l'apprentissage automatique.

- S'il y'aura au moins un cas source qui comporte la même solution, le système va calculer les scores correspondants. Une fois l'opération s'est terminée, le système va préconiser l'action convenable (Figure 5.23). L'utilisateur aura ensuite le choix d'appliquer l'action souhaitée. La figure 5.25. présente la suite de messages qui vont être affichés successivement en cas de validation de l'apprentissage.

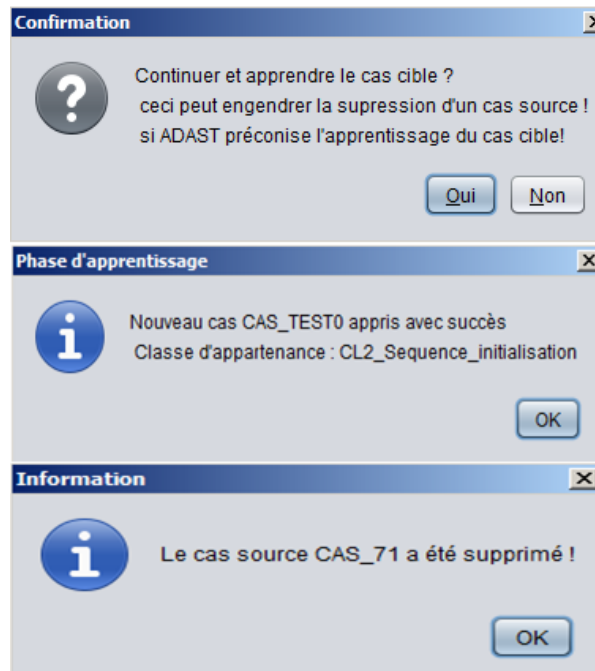


Figure 5.25. Suite de messages relative à l'opération d'apprentissage.

A la fin de chaque session de résolution de problème, le système enregistre dans un fichier journal (Figure 5.26), un résumé signé avec la date et l'heure système qui comprend toutes les informations relatives au raisonnement comme : les instances du cas cible, le résultat de la similarité, la décision d'apprentissage, etc.

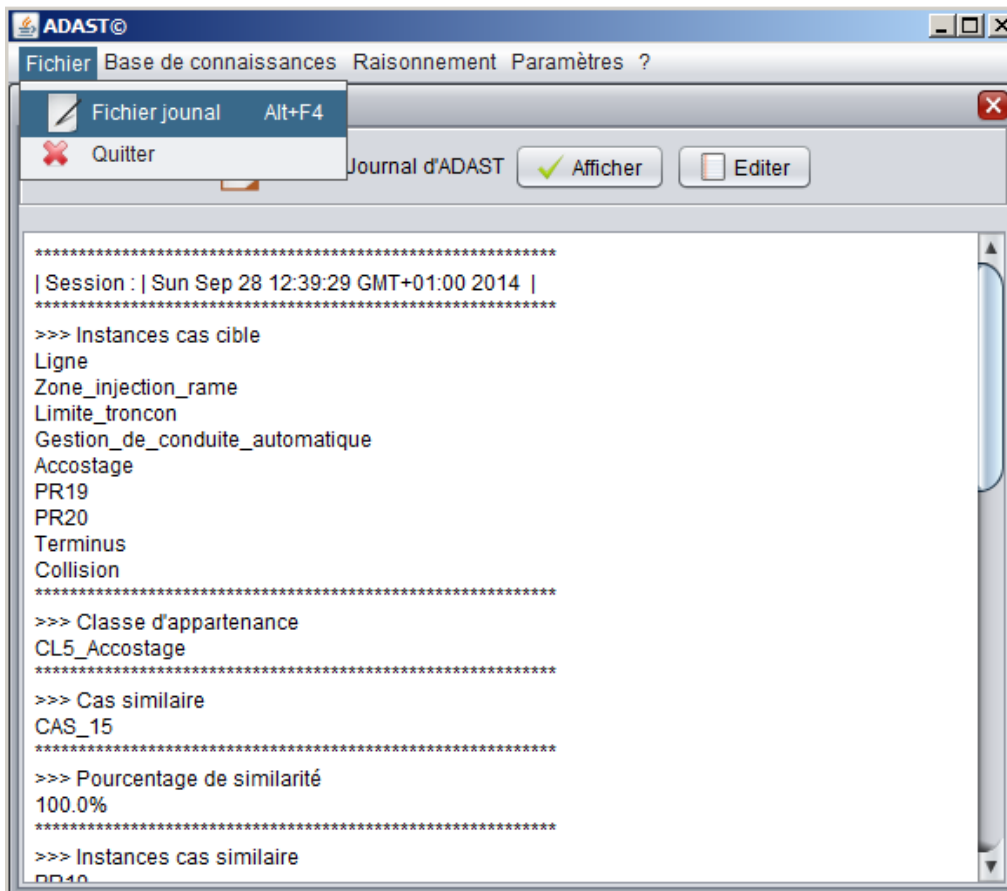


Figure 5.26. Aperçu du fichier journal d'ADAST.

## 5.5. Conclusion

Nous avons étudié dans ce chapitre les spécifications générales de notre système à travers la maquette de faisabilité d'ADAST. Nous avons présenté tout d'abord l'architecture en couches, qui met en évidence trois couches essentielles. La couche de présentation qui établit le lien et l'interfaçage avec les différents types d'utilisateurs, la couche applicative qui regroupe les deux modules principaux de la démarche d'aide à la décision proposée, à savoir, le module d'acquisition des cas et le module d'aide à la décision et enfin la couche de persistance qui fournit les objets nécessaires pour l'interfaçage avec la base de connaissances. Cette dernière est structurée selon trois bases distinctes (une base ontologique, une base de règles et une base de cas qui collecte l'ensemble de cas d'accidents historiques). Nous avons ensuite étalé l'architecture fonctionnelle ainsi que les deux modules principaux consacrés à l'acquisition de connaissances et le processus de raisonnement.

Cette maquette a été testée pour l'analyse de la sécurité en matière d'accident dans le transport ferroviaire. Toutefois, d'un point de vue plus général, notre système est destiné à la résolution de problèmes de représentation, de capitalisation et d'exploitation de connaissances tout en tenant compte des spécificités du domaine étudié. La maquette a été développée dans un environnement Java, profitant ainsi de nombreuses fonctionnalités offertes dans l'environnement Netbeans et notamment par le Framework Jena. Nous avons gardé à l'esprit l'indépendance entre l'interface et les connaissances (la base de cas, la base ontologique et la base de règles) ce qui peut assurer l'adaptabilité à d'autres problèmes similaires et assurer ainsi la généralité d'ADAST.

Nous présentons dans le chapitre suivant les tests et les résultats expérimentaux que nous avons réalisés afin d'évaluer notre approche.

---

# Chapitre 6 : Expérimentation et validation

---

## 6.1. Introduction

Nous avons présenté dans ce rapport notre approche d'aide à la décision ADAST qui s'est basée sur le RàPC et une ontologie de domaine. Nous avons utilisé plusieurs types de connaissances sur lesquelles s'appuie le raisonnement : des connaissances de similarité, des connaissances d'adaptation etc. Comme déjà vu dans les chapitres précédents, nous avons convenu d'utiliser les techniques d'apprentissage automatique ou encore de « fouille de données », par exemple, les règles d'association pour aider à bien caractériser et décrire un cas cible (première phase de RàPC). La technique d'arbre de décision pour déterminer les descripteurs clés ayant le pouvoir de classification des cas cibles saisis. Cette classification nous permet d'opérer la recherche des cas similaires uniquement dans la classe repérée (deuxième phase de RàPC). Nous avons aussi convenu de proposer une approche d'acquisition des connaissances d'adaptation basée sur l'analyse des concepts formels (ACF). Cette approche constitue un apport non négligeable pour remédier aux difficultés liées à l'extraction des connaissances d'adaptation (troisième phase de RàPC). Finalement, nous avons proposé à l'issue de la dernière phase de raisonnement (l'apprentissage), un algorithme qui permet d'assurer un apprentissage auto-incrémental. Les approches que nous avons proposées dans notre travail ont été concrétisées par le développement d'une maquette d'aide à la décision basée sur ADAST (voir chapitre 5). Cette maquette est considérée comme une première étape de validation globale de notre approche.

L'approche ADAST est destinée pour des utilisateurs ayant des connaissances sur le domaine de la sécurité. Actuellement, la maquette de faisabilité basée sur ADAST est utilisée pour pérenniser les connaissances en matière d'accident dans le transport ferroviaire (notre cadre d'application) afin d'y tirer profit en proposant une aide à la décision et des solutions à de nouvelles situations d'insécurité. Les expérimentations que nous avons réalisées avaient ainsi pour but d'évaluer tous les éléments qui constituent notre approche proposée. Pour préparer le terrain des expérimentations, nous avons utilisé l'échantillon d'apprentissage qui comporte une base de scénarios d'accidents qui a été construite et validée par des experts de l'IFSTTAR. La mise en œuvre de notre approche a été réalisée de façon à permettre une évaluation efficace du système.

Ce chapitre présente alors les expérimentations et la validation de notre approche d'aide à la décision.

## 6.2. Cas étudié

Pour pouvoir étudier la faisabilité et le bien fondé de toute la démarche d'aide à la décision proposée à travers l'approche ADAST, nous avons pris les accidents de transport ferroviaire issus du terrain réel en ayant recours à un échantillon d'apprentissage.

Nous allons positionner dans cette section cet échantillon que nous utilisons et notamment l'approche d'expérimentation et de validation que nous poursuivons.

### 6.2.1. Échantillon d'apprentissage

Comme souligné dans la section 5.2.3.1, l'échantillon d'apprentissage dont nous disposons est relatif au dossier technique INRETS-CRESTA CR/A-94-16 (Hadj-Mabrouk H. et Bied-Charreton D., 1993), (Hadj-Mabrouk H., 1994, 1998), (Mejri L., 1995) et (Mejri L. et al., 2009) comprenant soixante-dix scénarios d'accidents issus de terrain réel, dont 64 sont relatifs au risque de collision. A noter que ces scénarios d'accidents sont répertoriés initialement dans 12 classes.

Pour pouvoir utiliser cet échantillon, nous avons convenu de le représenter sous le format CSV<sup>1</sup> (Figure 6.1). Ce format est reconnu par tous les outils de la fouille de données offerts pour la communauté scientifique.

| No. | Gestion_de | Local | contrôle_E/S | Suivi_des_trans | Gestion_sens | Consignes_de | Gestion_arret | Securite_Qua | Autorisatio | Communic | Initialisation | Conduite | Gestion | Evacuation | Accostage | Protect | Tract   | Limits  | Ligne   | Autre_ZA | Classe  |
|-----|------------|-------|--------------|-----------------|--------------|--------------|---------------|--------------|-------------|----------|----------------|----------|---------|------------|-----------|---------|---------|---------|---------|----------|---------|
|     | Nominal    |       | Nominal      | Nominal         | Nominal      | Nominal      | Nominal       | Nominal      | Nom         | Nominal  | Nominal        | Nom      | Nominal | Nominal    | Nominal   | Nominal | Nominal | Nominal | Nominal | Nominal  | Nominal |
| 1   | NON        | OUI   | NON          | NON             | OUI          | NON          | OUI           | NON          | NON         | OUI      | NON            | NON      | OUI     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C1      |
| 2   | NON        | OUI   | NON          | OUI             | NON          | NON          | NON           | NON          | NON         | OUI      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C1      |
| 3   | OUI        | NON   | OUI          | NON             | NON          | NON          | NON           | NON          | NON         | OUI      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C1      |
| 4   | NON        | NON   | OUI          | OUI             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | OUI     | NON     | NON      | C1      |
| 5   | OUI        | NON   | OUI          | OUI             | NON          | NON          | NON           | NON          | NON         | OUI      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C1      |
| 60  | NON        | NON   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | OUI     | NON     | NON     | OUI     | NON      | C10     |
| 61  | NON        | NON   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | OUI     | NON     | NON     | OUI     | NON      | C10     |
| 62  | NON        | NON   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | OUI      | NON     | NON        | NON       | NON     | NON     | NON     | OUI     | NON      | C11     |
| 63  | NON        | NON   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | OUI      | NON     | NON        | NON       | NON     | NON     | NON     | OUI     | NON      | C11     |
| 64  | NON        | NON   | NON          | NON             | NON          | NON          | OUI           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | OUI     | NON      | C11     |
| 6   | NON        | NON   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | OUI            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C2      |
| 7   | NON        | OUI   | NON          | NON             | NON          | NON          | NON           | NON          | OUI         | NON      | OUI            | NON      | OUI     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C2      |
| 8   | NON        | OUI   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | OUI            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C2      |
| 9   | OUI        | NON   | NON          | OUI             | NON          | NON          | NON           | NON          | NON         | NON      | OUI            | NON      | NON     | NON        | NON       | NON     | NON     | OUI     | OUI     | OUI      | C2      |
| 10  | OUI        | OUI   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | OUI            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | OUI     | OUI      | C2      |
| 11  | OUI        | NON   | NON          | NON             | OUI          | NON          | NON           | NON          | NON         | NON      | OUI            | OUI      | NON     | OUI        | NON       | NON     | NON     | NON     | OUI     | NON      | C2      |
| 12  | NON        | OUI   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | OUI            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C2      |
| 13  | NON        | OUI   | OUI          | NON             | NON          | NON          | NON           | NON          | NON         | OUI      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C2      |
| 14  | NON        | OUI   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C3      |
| 15  | NON        | OUI   | NON          | OUI             | NON          | OUI          | OUI           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C3      |
| 16  | OUI        | NON   | NON          | NON             | OUI          | NON          | OUI           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | OUI     | NON     | OUI     | OUI     | OUI      | C3      |
| 17  | NON        | OUI   | NON          | NON             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | OUI     | OUI      | C3      |
| 18  | OUI        | OUI   | NON          | OUI             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C3      |
| 19  | OUI        | OUI   | NON          | OUI             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C3      |
| 20  | OUI        | OUI   | OUI          | OUI             | NON          | NON          | NON           | NON          | NON         | NON      | NON            | NON      | NON     | NON        | NON       | NON     | NON     | NON     | NON     | OUI      | C3      |

Figure 6.1. Représentation de l'échantillon d'apprentissage en format CSV.

## 6.2.2. Approche d'expérimentation

Notre approche d'expérimentation et de validation est scindée en trois volets :

Tout d'abord, nous présentons les étapes de construction et d'acquisition de la base de connaissances à travers les différentes techniques utilisées issues de la fouille de données pour générer les règles d'association (par application de l'algorithme Apriori), les règles de classification par la méthode d'arbre de décision et notamment l'approche d'acquisition de connaissances d'adaptation en ayant recours à quelques sessions de leur acquisition.

En second lieu, nous présentons un scénario d'aide à la décision à travers l'injection d'un cas d'accident réel pour illustrer toutes les étapes de la démarche d'aide à la décision.

Enfin, nous procédons à la validation du système à travers l'évaluation de la **validité** et de l'**utilité** des connaissances produites. A cet effet, nous avons choisi d'utiliser un ensemble de cas de test à partir de l'échantillon d'apprentissage pour les confronter avec les connaissances produites par le système.

<sup>1</sup> Comma-separated values, connu sous le sigle CSV, est un format informatique qui permet la représentation des données sous forme de valeurs séparées par des virgules.

### 6.3. Construction et acquisition de la base de connaissances

#### 6.3.1. Acquisition des règles d'association

Pour générer les règles d'association, nous avons varié les trois paramètres (le lift, le support et la confiance) déjà décrits dans la section 4.6.1 en ayant recours à l'outil de fouille de données Tanagra<sup>1</sup>. Nous nous appuyons à chaque fois sur les variations possibles d'un paramètre et nous imposons à ces variations de se faire en laissant à chaque fois les deux autres paramètres inchangés.

La figure 6.2 présente l'évolution du nombre de règles générées en fonction de la valeur de la confiance sur la base des deux valeurs que nous avons fixées : le support = 0,33 et le lift = 1,1 :

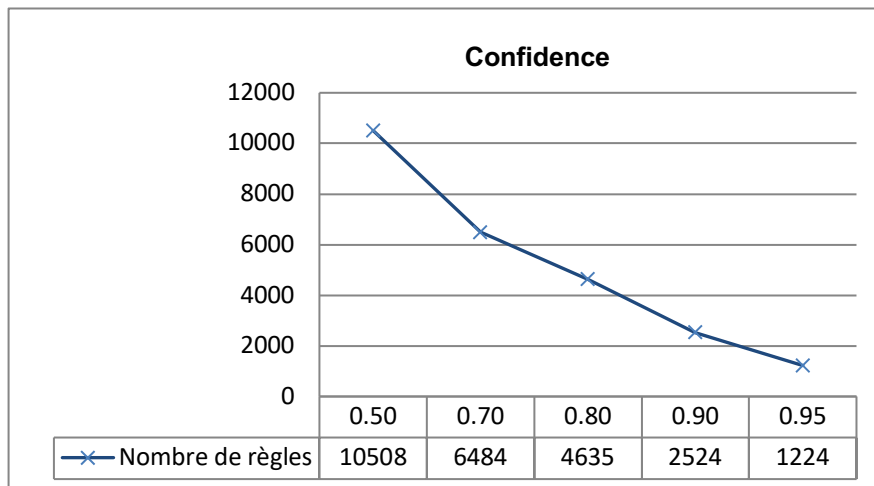


Figure 6.2. Évolution du nombre de règles générées en fonction de la confiance.

Dans la figure 6.3, nous avons varié la valeur de Lift et nous avons laissé les valeurs du support (0,33) et la confiance (0,9) inchangées :

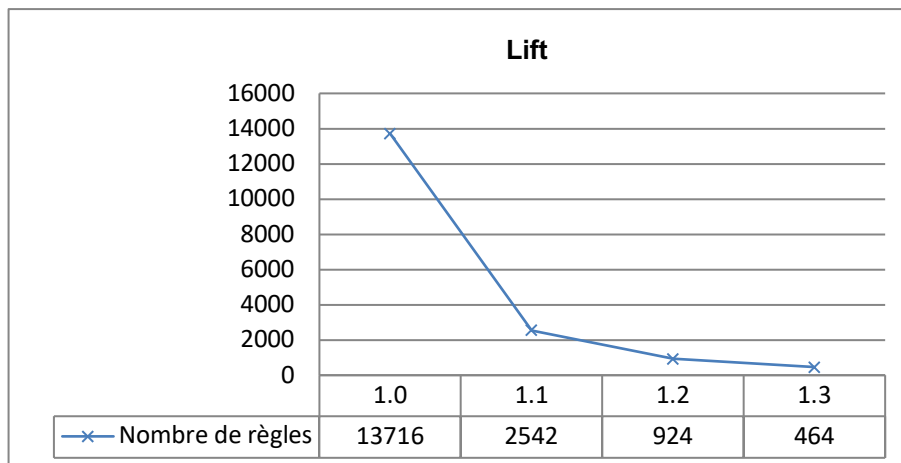


Figure 6.3. Évolution du nombre de règles générées en fonction du lift.

En interprétant les deux figures 6.2 et 6.3, nous pouvons constater que plus la valeur de la confiance augmente, plus le nombre de règles générées diminue et ceci de même pour

<sup>1</sup> TANAGRA est un logiciel gratuit de fouille de données destiné à l'enseignement et à la recherche. <http://eric.univ-lyon2.fr/~ricco/tanagra/fr/tanagra.html>

le Lift. Par contre, le nombre de règles diminue si la valeur de support augmente en laissant toujours le Lift et la confiance inchangés (Figure 6.4) :

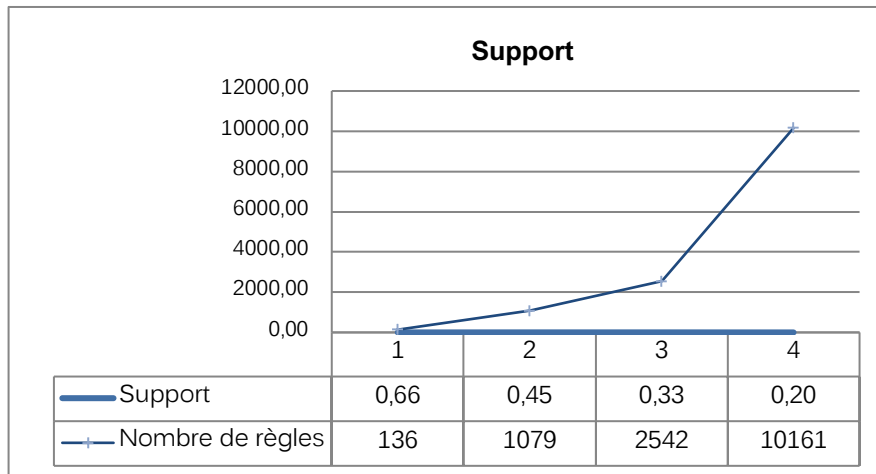


Figure 6.4. Évolution du nombre de règles générées en fonction du paramètre support.

Pour générer le premier ensemble de règles sur lequel nous allons opérer dans la maquette de faisabilité, nous avons convenu de choisir les valeurs des paramètres. Nous avons examiné à cet effet les différents ensembles de règles produits mentionnés dans les figures 6.2, 6.3 et 6.4, et ceci dans le but de contrôler la quantité et la qualité des règles d'association présentes pour une conjonction de descripteurs donnée.

Étant donné que pour chaque attribut d'un cas d'accident, deux valeurs sont possibles pour chaque attribut (OUI ou NON). Nous nous basons actuellement sur l'hypothèse suivante : *Une règle est utile dans notre cas si au moins elle comprend un attribut possédant la valeur OUI, par conséquent existant dans la description du cas cible.*

En suivant cette hypothèse, nous avons convenu de dresser le tableau 6.1 dans lequel nous avons extrait le **taux de règles utiles** pour chaque ensemble de règles testé :

|                  | <i>Paramètres</i>  |                  |             | <i>Nombre de règles</i> | <i>Nombre de règles utiles</i> | <i>Taux de règles utiles (%)</i> |
|------------------|--------------------|------------------|-------------|-------------------------|--------------------------------|----------------------------------|
|                  | <i>Support</i>     | <i>confiance</i> | <i>Lift</i> |                         |                                |                                  |
| E1               | 0,33               | 0,90             | <u>1,0</u>  | 13716                   | 3908                           | 28,00                            |
| <b><u>E2</u></b> | <b><u>0,20</u></b> | <b>0,90</b>      | <b>1,1</b>  | <b>10161</b>            | <b>9212</b>                    | <b>91,00</b>                     |
| E3               | 0,33               | <u>0,80</u>      | 1,1         | 4635                    | 1992                           | 43,00                            |
| E4               | 0,33               | 0,90             | 1,1         | 2524                    | 1140                           | 45,00                            |
| E5               | 0,33               | <u>0,95</u>      | 1,1         | 1224                    | 780                            | 64,00                            |
| E6               | 0,33               | 0,90             | <u>1,3</u>  | 464                     | 84                             | 18,00                            |
| <u>E7</u>        | <u>0,66</u>        | 0,9              | 1,1         | 136                     | 0                              | 00,00                            |

Tableau 6.1. Comparaison des ensembles de règles d'associations générées



Sur la base de ces différents tests, l'ensemble de règles (E2) qui présente le taux le plus supérieur (91%) est choisi. Pour faciliter l'analyse des résultats nous ne considérerons dans ce qui suit que les règles où les deux membres de la prémisse comprennent trois attributs en maximum. Si nous calculons toutes les règles d'association à prémisse composée, nous avons 9212 règles. Le filtrage des résultats nous amène à un nombre de 729 règles.

Cet ensemble sera stocké dans un fichier de données (qui représente la base de règles d'association) selon le format que nous avons présenté dans la section 5.2.1.2 et inclut ensuite dans la maquette.

### 6.3.2. Acquisition des règles de classification

Le rôle du classifieur par arbre de décision est la détection de la classe adéquate pour chaque cas cible saisi. Le processus de classification s'applique sur deux types de données d'apprentissage : données d'apprentissage qui décrivent l'ensemble de cas séparés en classes et les données de test qui représentent les nouveaux cas acquis par les utilisateurs.

Pour effectuer les expérimentations, l'idée est de subdiviser l'échantillon en deux parties : l'échantillon d'apprentissage pour la construction du modèle et l'échantillon de test pour son évaluation (Figure 6.5).

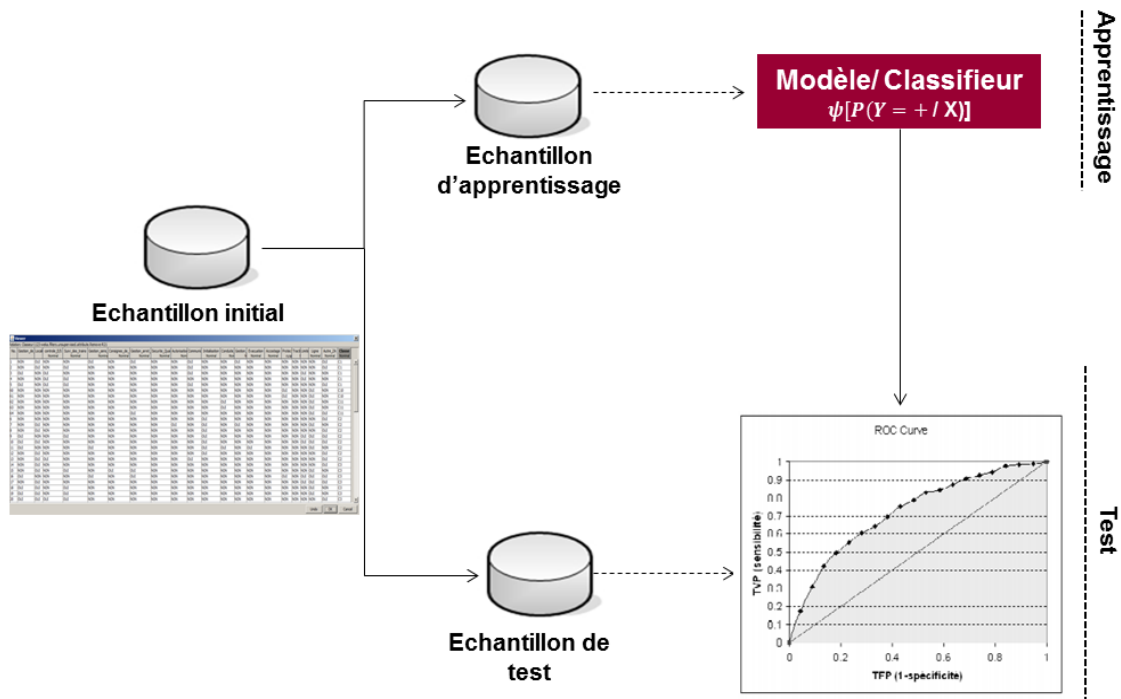


Figure 6.5. Schéma d'évaluation des modèles de classification.

Nous avons convenu de tester différents algorithmes pour consolider le choix adopté dans notre approche. Le test consiste à classer les scénarios d'accidents présents dans l'échantillon d'apprentissage selon 11 classes. Pour réaliser le test, nous profitons des algorithmes intégrés dans les deux outils SIPINA<sup>1</sup> et Weka<sup>2</sup>. Les résultats du test sont présentés dans le tableau 6.2.

<sup>1</sup> SIPINA est un logiciel gratuit de Data Mining spécialisé dans l'induction des arbres de décision

<sup>2</sup> Weka est un logiciel gratuit de Data Mining qui intègre beaucoup d'algorithmes, développé par l'université de Waikato.

Pour chaque algorithme testé, nous avons extrait le taux de classification correct, le taux de fausses classification et la profondeur de l'arbre générée. Le taux d'erreur semble être un indicateur synthétique pertinent, il indique la probabilité de mal classer un individu de la population. Les deux algorithmes C4.5 de Quinlan (Quinlan, 1993) et ASSISTANT 86 de Kass (Kass, 1980) ont donné les meilleurs taux de classifications correctes, respectivement 87% et 85% avec une profondeur de l'arbre égale à 11.

Sur la base de notre expérimentation, nous avons convenu alors de choisir l'algorithme **C4.5**. Ce choix est consolidé en raison de sa bonne performance par rapport aux autres algorithmes d'arbre de décision testés. La méthode de QUINLAN consiste à tester successivement chaque attribut pour connaître celui à utiliser en premier lieu de façon à optimiser le gain d'information, c'est à dire l'attribut qui permet le mieux de distinguer les exemples de différentes classes.

| <i>Algorithme</i>   | <i>Taux correct de classification (%)</i> | <i>Taux de fausse classification (%)</i> | <i>Profondeur de l'arbre</i> |
|---------------------|-------------------------------------------|------------------------------------------|------------------------------|
| ID3                 | 70,00                                     | 30,00                                    | 07                           |
| ASSISTANT 86        | 84,00                                     | 16,00                                    | 11                           |
| CHAID               | 71,00                                     | 29,00                                    | 08                           |
| <b>C4.5</b>         | <b>87,00</b>                              | <b>13,00</b>                             | <b>11</b>                    |
| Improved C4.5       | 77,00                                     | 23,00                                    | 11                           |
| Improved CHAID      | 33,00                                     | 67,00                                    | 03                           |
| Cost sensitive C4.5 | 77,00                                     | 23,00                                    | 09                           |

**Tableau 6.2. Comparaison des performances des algorithmes testés**

Pour consolider le choix de l'algorithme C4.5, nous avons tracé la courbe ROC correspondante. D'après Rakotomalala<sup>1</sup> « *La courbe ROC est un outil d'évaluation et de comparaison des modèles...est Indépendant des matrices de coûts de mauvaise affectation, il faut néanmoins que nous ayons une classe cible (positive) clairement identifiée et que la méthode d'apprentissage puisse fournir un SCORE proportionnel à  $P(Y = +/X)$*  ».

Un exemple est qualifié de positif s'il appartient à une classe particulière ou de négatif s'il n'appartient pas à cette classe. Étant donné que l'outil SIPINA n'offre pas cette possibilité, nous avons utilisé donc l'outil Weka qui permet d'obtenir les différents scores.

Dans la figure 6.6, nous présentons la courbe ROC associée à notre classifieur généré via l'algorithme C4.5. Cette courbe est très supérieure à la diagonale ce qui prouve que notre classifieur peut donner des bons résultats de classifications avec une fraction de faux négatifs très faible. A rappeler que dans un test de classification, un résultat est dit **vrai positif** lorsqu'une classe est correctement détectée par le classifieur. Par opposition aux notions de **faux positif** à travers un cas déclaré positif (bien classé) alors qu'il ne l'était pas

<sup>1</sup> Ricco Rakotomalala : Tutoriels Tanagra - <http://tutoriels-data-mining.blogspot.fr/>

et de **faux négatif** à travers un cas déclaré négatif (mal classé) alors qu'il était en réalité positif et de **vrai négatif** (un résultat correctement déclaré comme négatif).

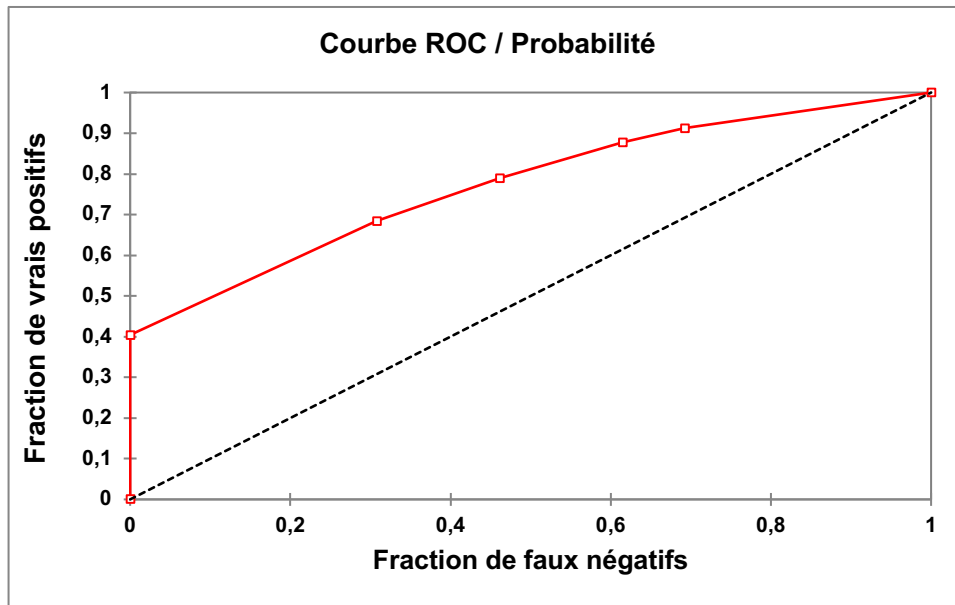


Figure 6.6. Courbe ROC du classifieur généré.

La figure 6.7 montre un aperçu global de l'arbre de décision générée par l'algorithme C4.5.

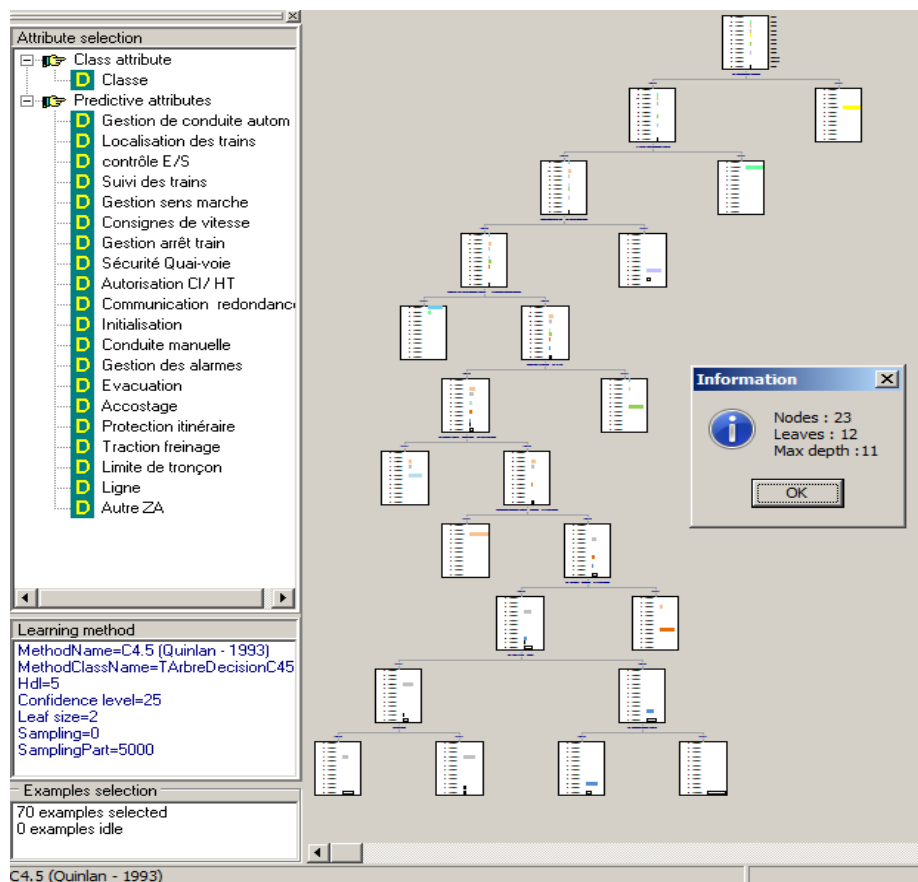


Figure 6.7. Écran visualisation de l'arbre de décision généré via l'outil SIPINA.

L'arbre de décision correspondant est décrit plus en détails ci-après :

- Le premier sommet est appelé la « racine » de l'arbre. Il est situé sur le premier niveau. Nous y observons la distribution de fréquence de la variable à prédire « Classe ». Nous constatons qu'il y a bien 70 observations. Nous citons à titre d'exemple, 5 « CL1 », 8 « CL2 », 10 « CL3 » et 6 « CL ? » (Figure 6.8) ;

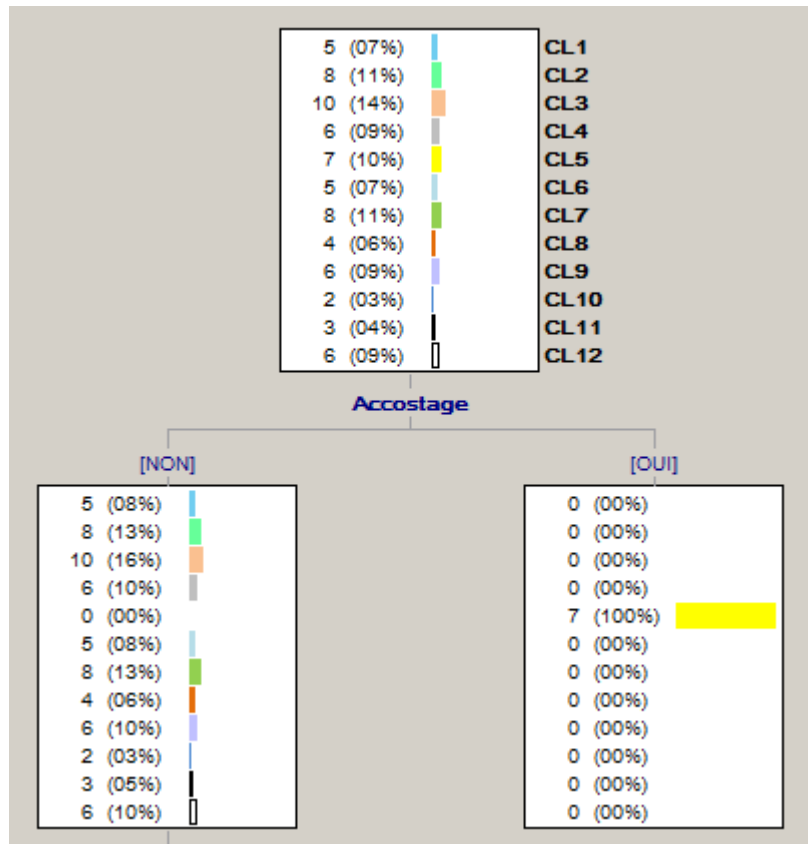


Figure 6.8. Partie de l'arbre de décision générée.

- La variable « Accostage » est la première variable utilisée. Nous parlons de variable de segmentation. Comme elle est composée de 2 modalités {OUI, NON}, elle produit donc 2 sommets enfants ;
- La première arête à droite est produite à partir de la modalité « Accostage ». Le sommet qui en découle couvre 7 observations. La distribution de fréquence nous indique qu'il y'a 100% d'observations qui couvrent la classe CL5.

### 6.3.3. Acquisition de connaissances d'adaptation

Comme nous l'avons déjà vu dans la section 4.6.3.2, notre approche d'acquisition de connaissances d'adaptation est basée sur l'analyse de concepts formels (ACF). L'acquisition se fait en créant des treillis de concepts (treillis de Galois) pour les cas qui appartiennent à la même classe de solution, puis en transformant les concepts formels de ces treillis en conditions d'adaptation associées aux cas.

Nous nous basons sur la répartition que nous avons déjà présentée dans la section 5.2.3.1 et nous présentons la répartition des solutions en fonctions de ces classes dans le tableau 6.3. Un numéro unique *i* est attribué à chaque solution adoptée SA, *i* varie entre 1 et 56 tandis que CL*i* correspond à la classe de solution numéro *i* (Tableau 6.3).

| Classe | Cas |
|--------|-----|
|--------|-----|

|                                       |                                                    |
|---------------------------------------|----------------------------------------------------|
| #CL1 Commutation de redondance        | SA1, SA8, SA23, SA24, SA38                         |
| #CL2 Séquence d'initialisation        | SA4, SA5, SA6, SA19, SA20, SA27, SA19              |
| #CL3 Localisation des trains          | SA3, SA7, SA16, SA17, SA21, SA30, SA45, SA51, SA52 |
| #CL4 Gestion du freinage d'urgence    | SA9, SA18, SA41, SA51                              |
| #CL5 Accostage                        | SA12, SA13, SA14, SA26, SA32                       |
| #CL6 Gestion du sens marche           | SA11, SA15, SA17, SA49                             |
| #CL7 Contrôle d'entrée/sortie         | SA10, SA25, SA28, SA29, SA31, SA36, SA39           |
| #CL8 Suivi de l'ordre                 | SA22, SA33, SA34, SA46                             |
| #CL9 Conduite manuelle                | SA5, SA35, SA40, SA5, SA47                         |
| #CL10 Contrôle/Commande des aiguilles | SA44, SA53                                         |
| #CL11 Contrôle de vitesse             | SA54, SA55, SA56                                   |
| #CL12 Autre classe                    | SA41, SA42, SA43, SA50, SA48                       |

**Tableau 6.3. Répartition des solutions dans les classes de scénarios**

Pour préparer les fichiers qui contiennent les contextes formels des cas, nous utilisons l'outil Contexte Explorer. Cet outil permet de construire un contexte formel ou bien de charger un contexte déjà enregistré, et de lui produire le treillis de concepts correspondant. Pour rappeler, un concept formel est constitué de deux parties : l'**extension** qui contient les individus appartenant au concept et l'**intension** qui contient les propriétés partagées par les individus. Ces concepts construisent un treillis de concepts.

Dans notre approche d'acquisition de connaissances d'adaptation (voir section 4.6.3.2.3), nous utilisons 12 contextes formels associés aux 12 classes de solutions issues de l'échantillon d'apprentissage (Tableau 6.3).

Les connaissances produites par l'approche d'ACF peuvent être filtrées et affinées grâce à l'interaction avec un spécialiste de domaine durant les sessions d'acquisition de ces connaissances. Nous présentons successivement dans les figures 6.9, 6.10, 6.11, 6.12 et 6.13 quelques treillis de concepts parmi ceux que nous avons produits hors-ligne.

La figure 6.9 présente le treillis de concepts de la classe #CL1\_Commuration\_redondance comportant les CAS01, CAS08, CAS29, CAS30 et CAS47.

Nous remarquons l'apparition de plusieurs concepts que nous illustrons dont quelques-uns ci-après :

$$\mathbb{C}_1 = (\{CAS29, CAS47, CAS08, CAS01\}, \{Commutation\_redondance\})$$

$$\mathbb{C}_2 = (\{CAS30, CAS47, CAS29, CAS08\}, \{Limite\_troncon\})$$

$$\mathbb{C}_3 = (\{CAS01, CAS08\}, \{Localisation\_des\_trains\})$$

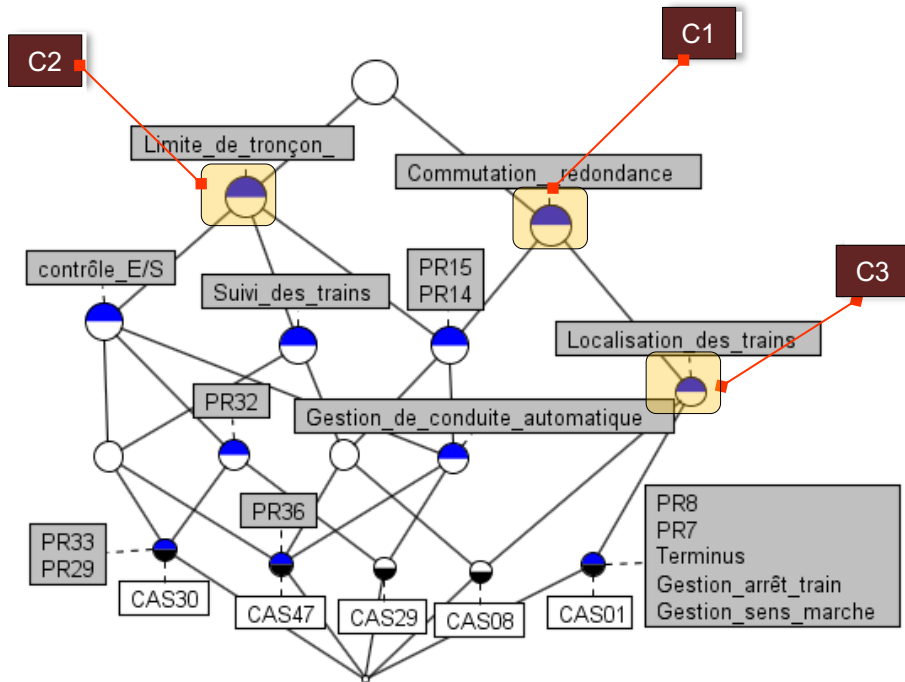


Figure 6.9. Le treillis de concepts de la classe #CL1 Commutation de redondance.

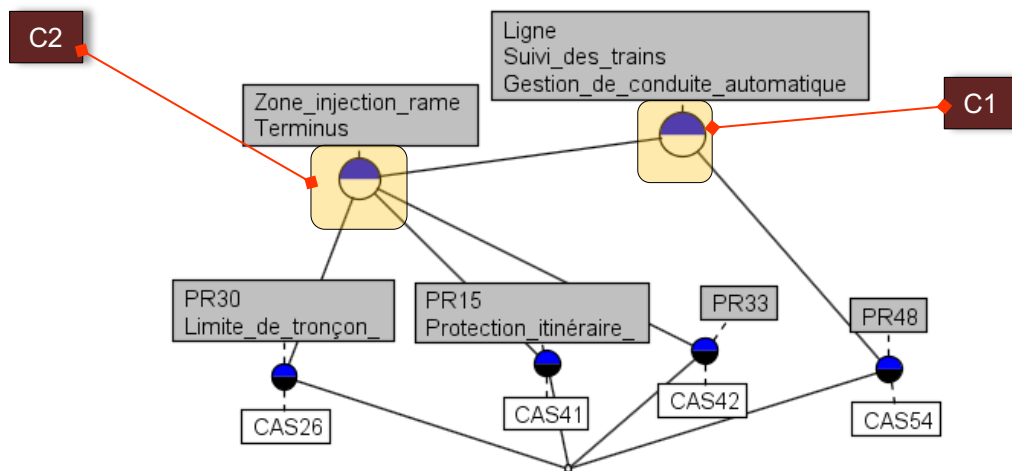


Figure 6.10. Le treillis de concepts de la classe #CL8 Suivi de l'ordre.

La figure 6.11 présente le treillis de concepts de la classe #CL10 Contrôle/Commande des aiguilles. Il comporte 2 cas sources tandis que le treillis de la figure 6.12 correspond à la classe #CL2\_Séquence d'initialisation, il comporte 8 cas sources.

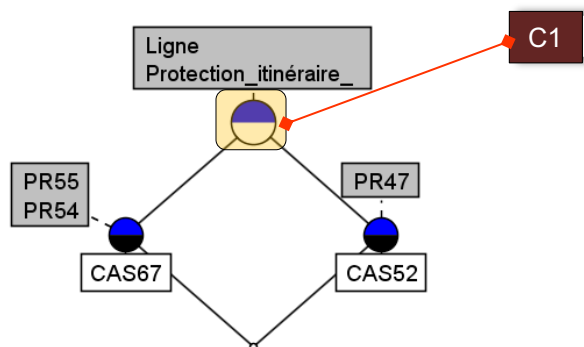


Figure 6.11. Le treillis de concepts de la classe #CL10 Contrôle/Commande des aiguilles.

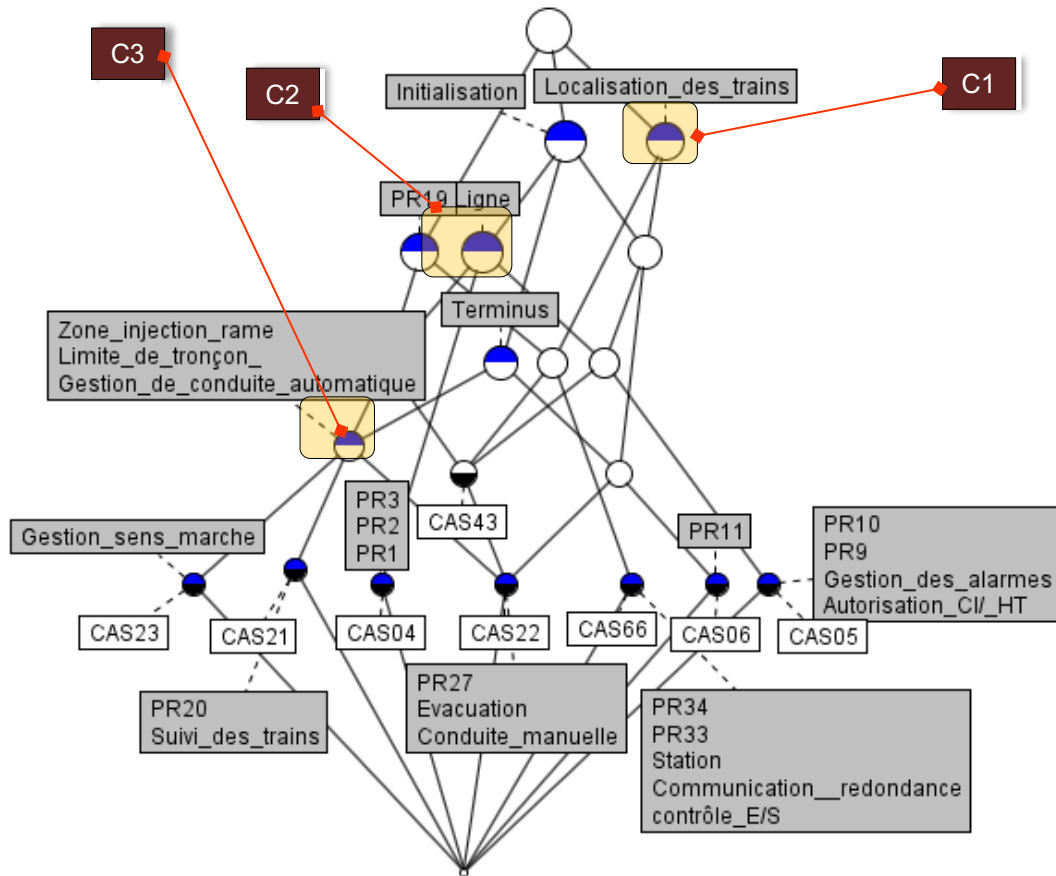


Figure 6.12 .Le treillis de concepts de la classe #CL2 Séquence d'initialisation.

La figure 6.13 correspond à la classe #CL11 Contrôle de vitesse, il comporte uniquement 3 cas sources.

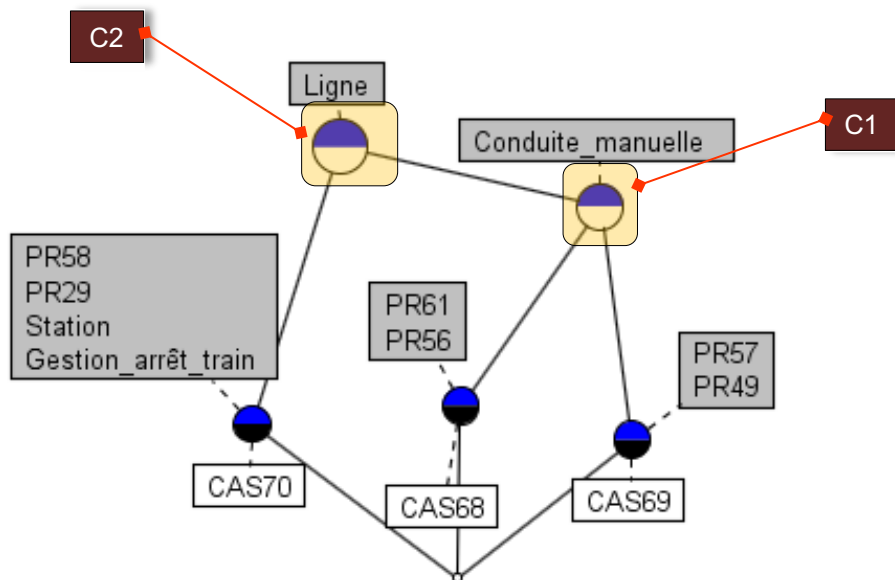


Figure 6.13. Le treillis de concepts de la classe #CL11 Contrôle de vitesse.

Pour chaque treillis, nous écrivons les intensions de chaque cas en forme normale conjonctive (des conjonctions de disjonctions) (voir Annexe 3) pour dégager les différentes conditions d'adaptation.

Après avoir présenté, les différentes étapes que nous avons suivies pour la construction et l'acquisition de la base de connaissances, il convient maintenant d'enchaîner un scénario complet d'aide à la décision.

#### 6.4. Scénario d'aide à la décision

Dans ce paragraphe, nous allons présenter un scénario d'utilisation du système illustrant toute la démarche d'aide à la décision proposée dans ADAST. Le scénario d'aide à la décision consiste à introduire initialement un cas cible (un cas d'accident réel pour lequel nous cherchons une solution convenable) et suivre tout le processus de raisonnement guidé par le système.

##### 6.4.1. Description du scénario d'accident injecté

Le scénario d'accident que nous avons adopté à partir de l'échantillon d'apprentissage (section 6.2.1) et dont la synoptique est décrite dans la figure 6.14, concerne « **la perte d'élément (une rame) suite à une pénétration sur un canton occupé par un autre élément** ».

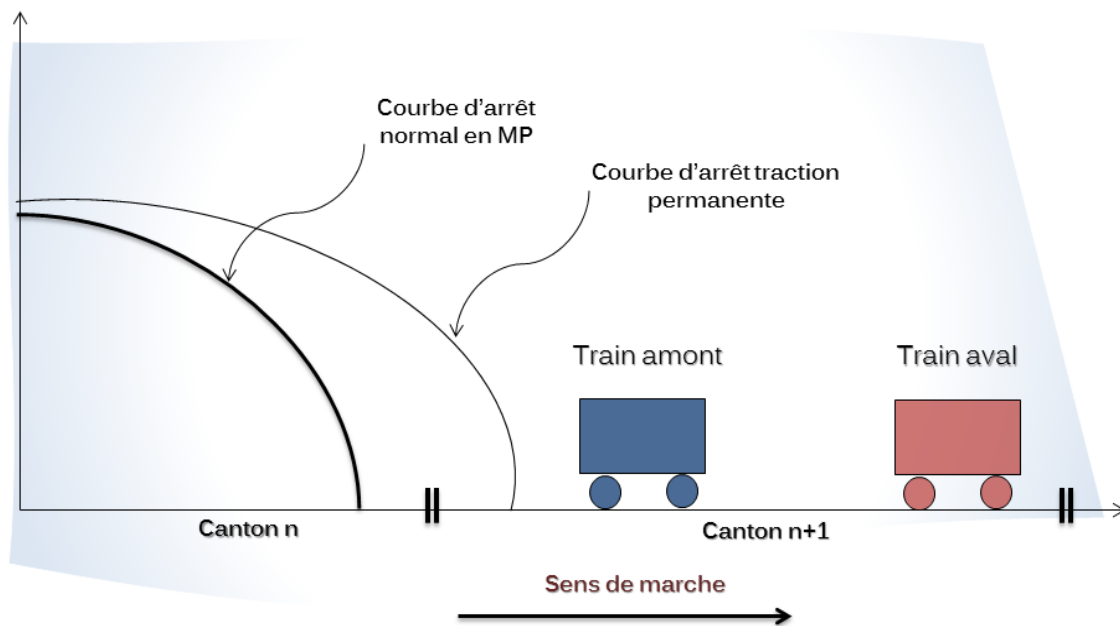


Figure 6.14. Synoptique de l'accident : Perte d'élément - pénétration sur un canton occupé (Dossier technique CR/94-16)

*Soit une rame R1 (train aval) immobilisée en ligne et occupe un canton. Une rame R2 (train amont) pénètre sur ce canton occupé suite à un défaut de son freinage d'urgence (panne de traction permanente). Le pilote automatique (PA) fixe (au sol) coupe la fréquence de sécurité et la haute tension (FS-HT).*

*Le poste central de contrôle (PCC) examine la situation, ne voit pas qu'il y a deux trains sur un même canton (Élément invisible sur la zone de conduite) et rétablit la FS-HT (rétablissement erroné de la FS-HT). Le train aval R1 ne repart pas et reste immobilisé en*



ligne, mais le train amont R2 repart avec un effort de traction maximal, ce qui pourrait engendrer une collision par rattrapage entre trains.

#### 6.4.2. Cycle de raisonnement

Nous allons présenter dans cette section les différents éléments d'aide à la décision fournis par la maquette de faisabilité basée sur l'approche ADAST. Nous allons décrire les cinq principales phases du cycle de raisonnement.

##### - Élaboration d'un cas cible

La première étape du cycle de raisonnement est l'élaboration d'un cas cible. Les descripteurs et les instances du cas cible introduites par l'utilisateur, concernent essentiellement le contexte, la zone de l'apparition de l'accident potentiel, les éléments dangereux et notamment les différentes causes et facteurs favorisant le déclenchement de la situation dangereuse.

Nous avons légèrement modifié pour les besoins de l'évaluation le scénario d'accident déjà présenté en ajoutant d'autres instances non mentionnées lors de la collecte d'informations (Tableau 6.4).

| Problème cible |                              |                                 |                    |                    |
|----------------|------------------------------|---------------------------------|--------------------|--------------------|
|                | Contexte                     | Éléments dangereux              | Causes             | Accident potentiel |
| Instances      | Cantonement fixe             | Suivi trains                    | PA sans redondance | Collision          |
|                | Rame 2                       | Localisation trains             | PCC                | Par rattrapage     |
|                | Ligne                        | Traction freinage               | PR09               |                    |
|                | <del>Limite de tronçon</del> | <del>Consignes de vitesse</del> | PR10               |                    |
|                |                              | <del>Gestion des alarmes</del>  | PR11               |                    |
|                |                              | PR24                            |                    |                    |
|                |                              | PR59                            |                    |                    |

Tableau 6.4. Description initiale du cas cible introduit

A noter qu'une PR représente une « Panne Résumée » résultant du groupement d'un ensemble de pannes élémentaires ayant la même conséquence sur le comportement du système. Avec :

- PR9 : Pénétration d'une rame sur un canton occupé ;
- PR10 : Rétablissement erroné de FS HT ;
- PR11 : Élément invisible sur la zone de conduite automatique intégrale ;
- PR24 : Panne de traction permanente ;
- PR59 : Rame immobilisée en ligne ;

Une première aide à la décision se présente dans la figure 6.15 à travers les règles d'associations générées à l'issue de l'introduction de la description initiale du cas cible. Ces règles vont permettre à l'utilisateur de l'aider à mieux ajuster sa description. Il pourra ainsi, supprimer/ajouter ou même mettre à jour les instances introduites.

Pour le cas cible du tableau 6.4, la figure 6.15 montre que notre système a récupéré 26 règles d'association. L'utilisateur pourra lire ces règles facilement interprétables et ajuster au fur et à mesure sa description.

A partir de cette liste de règles, l'utilisateur est amené à éliminer à partir de la description initiale, les instances suivantes : « Consignes de vitesse », « Gestion des alarmes » et « Limite de tronçon ». La validation de cette étape va nous amener à la phase de remémoration.



Figure 6.15. Règles d'associations générées.

- **Remémoration des cas similaires**

Pour rappeler, nous cherchons tout d'abord dans cette phase à détecter la classe d'appartenance du cas cible pour ensuite opérer la recherche des cas similaires uniquement dans la classe repérée.

Pour le cas cible saisi et affiné du tableau 6.4, la figure 6.15 montre que le système a indiqué que la classe d'appartenance est la classe #CL3 Localisation des trains avec un seuil de probabilité = 1.

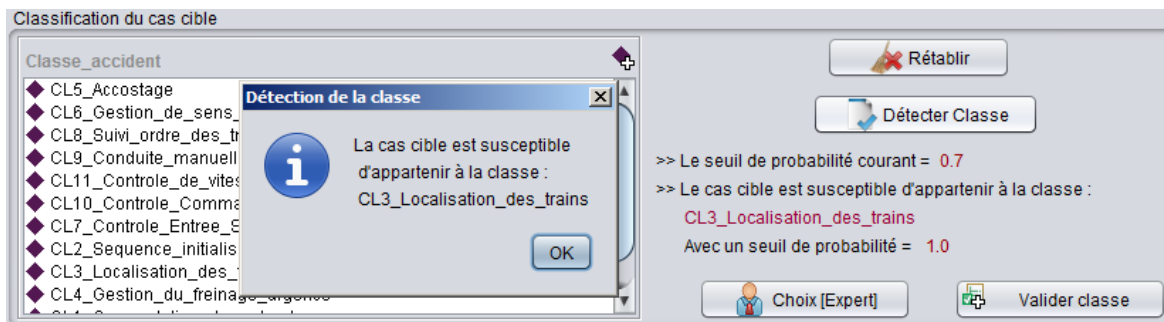


Figure 6.16. Résultat de la classification.

A cet effet, le calcul de similarité sera orienté vers cette classe qui comporte les dix cas sources suivants : CAS\_65, CAS\_38, CAS\_45, CAS\_17, CAS\_03, CAS\_63, CAS\_07, CAS\_25, CAS\_19 et CAS\_53. La figure 6.17 présente les résultats obtenus après le calcul de similarité effectué (avec  $\alpha=1$ ) sur cet ensemble de cas sources :

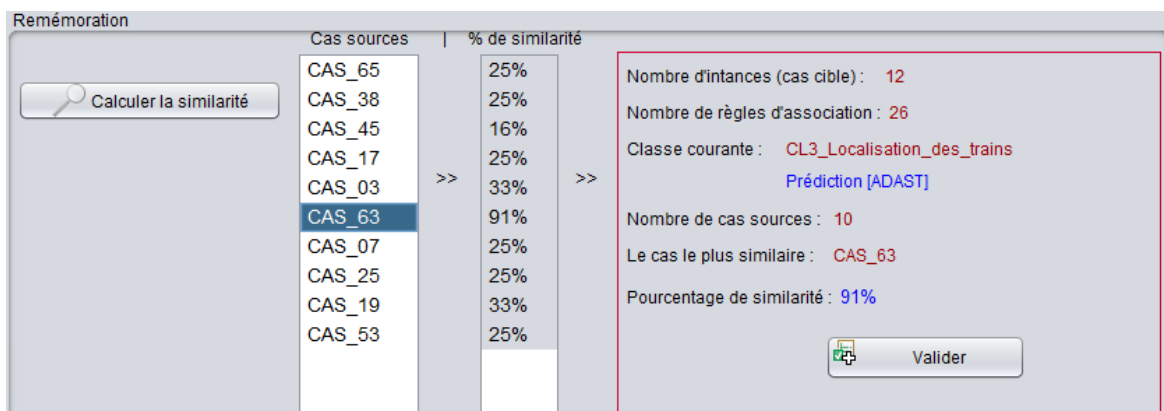


Figure 6.17. Résultat de la remémoration.

Le cas ayant le pourcentage de similarité les plus élevé est le CAS 63 avec un pourcentage de similarité de 90%. Ce cas sera donc retenu pour la phase d'adaptation.

#### - Adaptation

Comme nous l'avons déjà vu (voir sections 4.6.3 et 6.6.3), dans cette étape, les règles d'adaptation construites à l'issue de la phase d'acquisition des connaissances d'adaptation vont être exploitées pour enchaîner une solution possible pour le cas cible. Chaque cas source est muni d'un ensemble de règles d'adaptation  $S$  correspondant à la relation binaire  $r$  entre le problème source et le problème cible.

Sur ce schéma,  $\Delta pb$  et  $\Delta sol$  symbolisent respectivement « ce que cible est à source » et « ce que  $Sol(cible)$  est à  $Sol(srce)$  » (Figure 6.18).

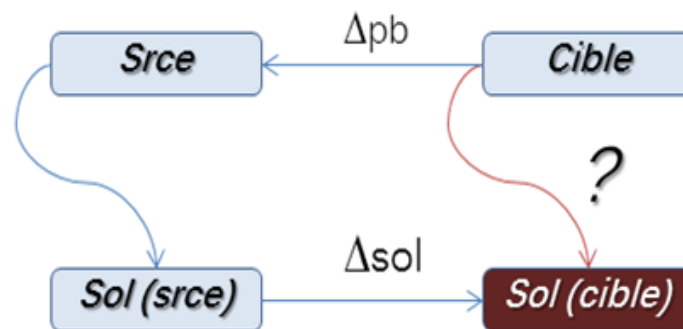


Figure 6.18. Le carré d'analogie pour le problème d'adaptation.

Dans notre cas,  $\Delta pb$  sera résolu à travers l'application des règles d'adaptation accordées au cas source CAS\_63 (retenu comme le cas le plus similaire). Le système va appliquer automatiquement les règles d'adaptations et génère ensuite la solution affectée au cas qui partage le maximum d'attributs en commun (Figure 6.19).

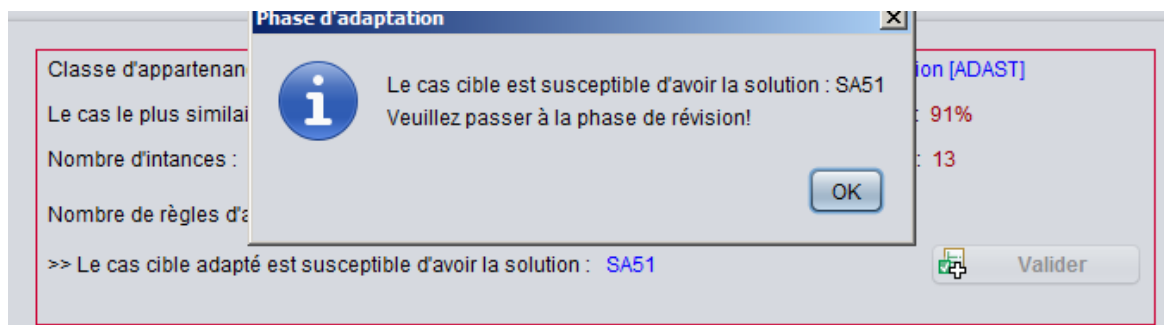


Figure 6.19. Résultat de l'adaptation.

Pour remédier à ce problème (le cas d'accident), la solution adoptée par le système est SA51. Cette solution consiste à *contrôler le courant de traction lors d'un freinage d'urgence (FU) et d'ouvrir les disjoncteurs si nécessaires*.

#### - Révision

Dans la phase de révision, l'utilisateur est amené à valider la solution proposée. Cependant, il a la possibilité de la refuser, dans ce cas, il est invité à sélectionner une autre solution à partir des instances extraites de l'ontologie. Dans le cas échéant, il peut proposer une nouvelle solution qui pourra enrichir l'ensemble de solutions pérennisées.

#### - Apprentissage

Comme convenu, dans cette étape, l'algorithme d'auto-incrémentation de la base de cas proposé dans notre approche ADAST (voir section 4.6.5) sera appliqué. Pour rappeler, cet algorithme va enchaîner le calcul des scores relatifs au degré de représentativité de la partie problème du cas cible et les parties problèmes des cas sources qui partagent la même solution.

Étant donné que seulement les deux cas sources CAS\_63 et CAS\_64 possèdent la solution SA51, le calcul va être opéré sur le cas cible adapté et révisé et ces deux derniers cas.

La figure 6.20 présente le résultat de l'apprentissage. Selon les scores générés par le système (Figure 6.20) : le score du CAS\_cible est supérieur au score du CAS\_64. Un message s'affiche indiquant que : « *le système préconise l'apprentissage du cas cible* ». Cet apprentissage va engendrer la suppression du cas source CAS\_64. Si l'utilisateur décide de valider l'apprentissage, des messages vont être affichés pour présenter les suites envisageables.

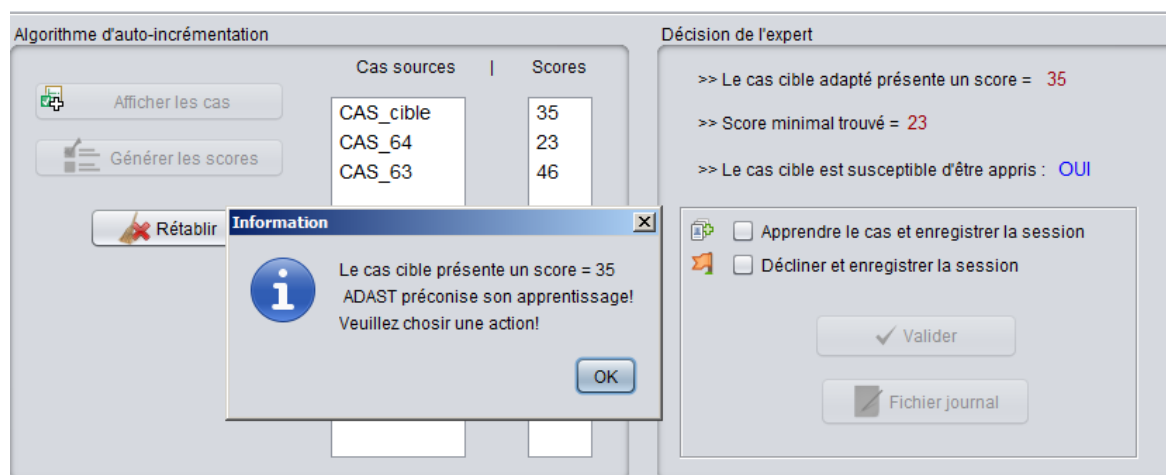


Figure 6.20. Résultat de l'apprentissage.

## 6.5. Évaluation du système

Le présent paragraphe propose une évaluation du système, plus précisément, la démarche d'aide à la décision et le module de résolution de problèmes. Cette évaluation va déboucher sur les limites et notamment les perspectives qui en découlent pour améliorer l'approche proposée.

Très schématiquement, la qualité d'un système d'aide à la décision se mesure en évaluant **la validité et l'utilité** des connaissances produites. Une connaissance est **valide** si elle est adéquate et cohérente par rapport à ce que nous connaissons déjà sur le domaine étudié. Elle est **utile** si elle contribue à réaliser les objectifs définis. Pour tester la représentativité des résultats obtenus par notre système, nous avons choisi d'utiliser des cas de test à partir de l'échantillon d'apprentissage.

Cet échantillon initial a été utilisé à maintes reprises pour élaborer la démarche d'aide à la décision. Nous avons convenu d'utiliser un lot de quatorze scénarios d'accident qui représente le 1/5 de l'échantillon d'apprentissage. Les cas de test choisis sont tirés aléatoirement à partir des 12 classes de scénarios préétablies (voir section 5.2.1.3).

La figure 6.21 représente la répartition des cas de test par rapport aux cas sources.

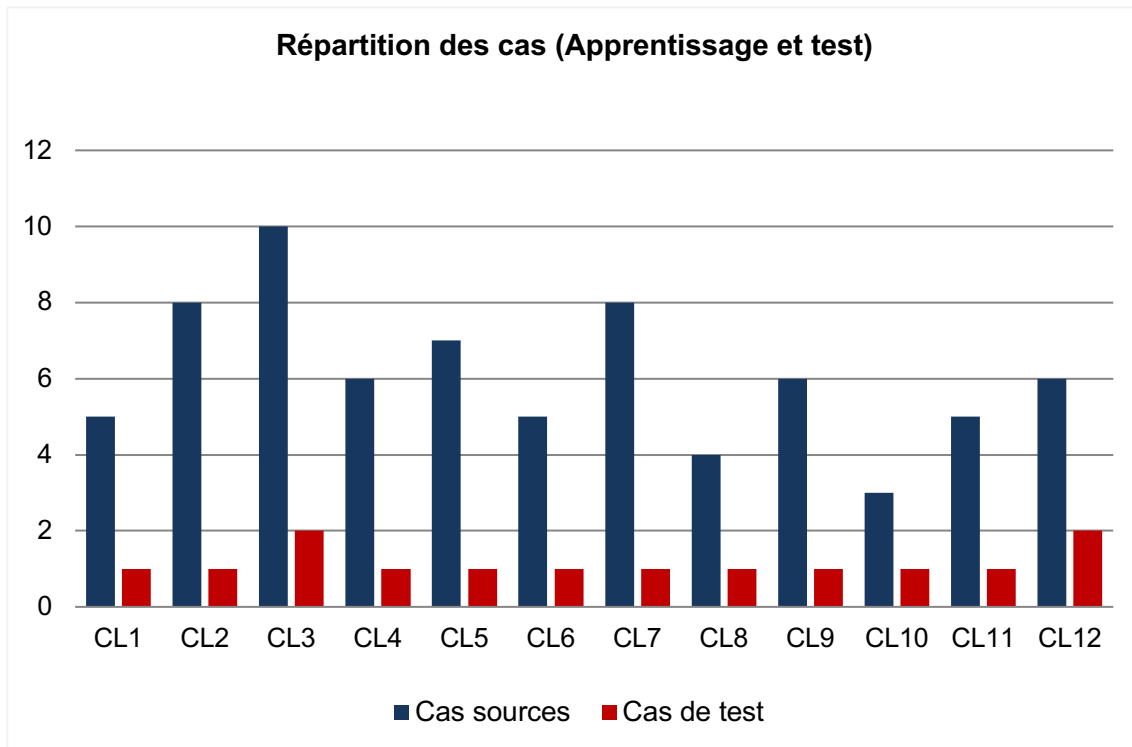


Figure 6.21. Histogramme de la répartition des cas (apprentissage et test).

### 6.5.1. Étude comparative entre l'outil CLASCA et le module de classification d'ADAST

Étant donné que le projet ACASYA<sup>1</sup>, est considéré comme le premier système ayant axé une méthodologie explicite d'analyse d'accidents issue du domaine ferroviaire, il est judicieux voire primordial d'établir une étude comparative entre la maquette de faisabilité basée sur ADAST et en particulier son module de classification d'accidents et l'outil CLASCA (le sous système principal d'ACASYA réservé à la classification). A noter que la classification dans notre approche ADAST est une étape parmi plusieurs dans le processus de raisonnement et la démarche d'aide à la décision. Comme nous l'avons déjà décrit (section 4.6.2, page 90), la classification est une étape assez conséquente, puisqu'elle permet de bien cibler le calcul de la similarité pour extraire les cas les plus similaires au cas cible saisi (la partie problème pour laquelle nous cherchons la solution adéquate).

Nous allons établir une comparaison des résultats de classification obtenus par les deux sous-systèmes (CLASCA et le sous système de classification d'ADAST), pour pouvoir avancer un premier benchmark de classification de scénarios d'accident. Nous procédons à une comparaison des résultats obtenus pour pouvoir situer et éventuellement mettre d'avantage notre contribution dans le cadre de ce travail de thèse.

#### 6.5.1.1. Interprétation des résultats de classification de l'outil CLASCA

L'outil CLASCA (Hadj-Mabrouk H., 1992), (Hadj-Mabrouk H., 1998) et (Mejri L. 1995), (Mejri L. 2009), est présenté comme un système d'apprentissage d'aide à la classification des scénarios d'accidents.

<sup>1</sup> ACASYA : Aide à la Certification par Apprentissage des SYstèmes de transport Automatisés, (développé dans le laboratoire LAMIH de l'université de Valenciennes et l'INRETS devenu IFSTTAR)

Initialement estimé à plusieurs centaines, le nombre d'exemples acquis dans CLASCA est de quarante trois pour la base d'apprentissage et de treize pour les scénarios de test (de validation).

La classification de treize scénarios a été interprétée comme suit (Figure 6.22) :

- 9 scénarios sur 13 sont "bien classés" (un pourcentage de réussite de 69%) : ils sont classés de la même façon par le système et l'expert de certification ;
- 4 scénarios ne sont pas bien classés ("non classés" ou "mal classés") : Selon (Hadj-Mabrouk H., 1992), *d'une part le taux d'adéquation de 2 scénarios avec une classe est inférieur au seuil de similarité courant et le système ne propose aucune classe d'appartenance. D'autre part, la classification du système pour les deux autres scénarios n'agrée pas l'expert certifieur.* La principale cause de cet échec provient du non représentativité des descriptions des classes apprises sur la base d'un petit nombre de scénarios.

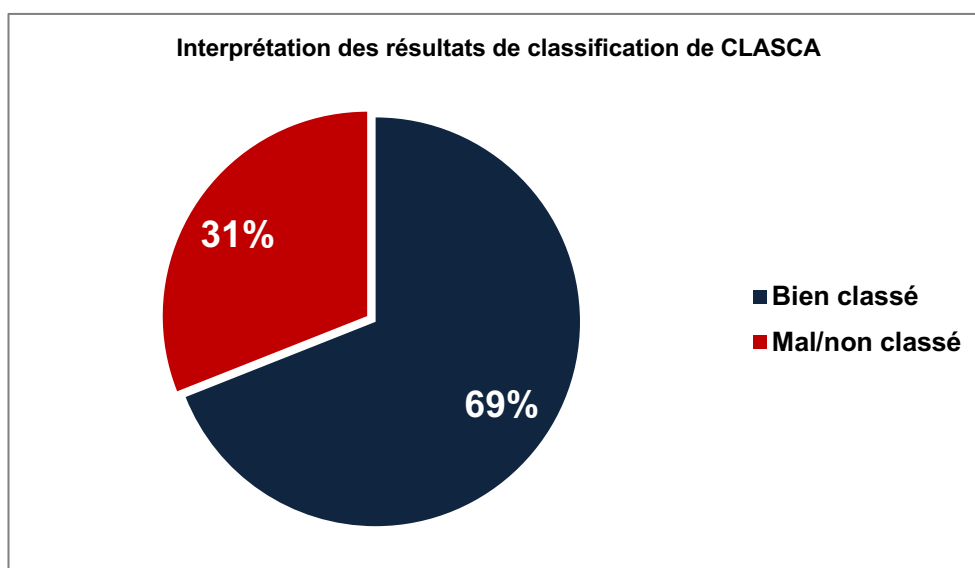


Figure 6.22. Interprétation des résultats de classification obtenus par CLASCA.

#### 6.5.1.2. Interprétation des résultats du module de classification d'ADAST

Pour l'évaluation de la maquette basée sur ADAST, nous rappelons que nous avons convenu d'utiliser un lot de quatorze scénarios d'accident qui représente le 1/5 de l'échantillon d'apprentissage. Les cas de test choisis sont tirés aléatoirement à partir des 12 classes de scénarios préétablies (voir section 5.2.1.3).

Dans le tableau 6.5, nous présentons également les résultats de classification obtenus par le sous module de classification de la maquette basée sur ADAST.

Nous recensons quatre principales rubriques :

- Le nombre de règles d'association que le système avait récupéré pour chaque cas de test introduit ;
- La prédiction de sa classe d'appartenance ;
- La classe originale attribuée par les experts (l'échantillon d'apprentissage) ;
- Le résultat de la validité de la classe prédite par rapport à la classe censée être l'originale.

| Cas de test<br>[Cas cible] |                                                                    | Élaboration<br>d'un cas<br>cible | Résultat de<br>classification        |                     |                         | Validité                 |
|----------------------------|--------------------------------------------------------------------|----------------------------------|--------------------------------------|---------------------|-------------------------|--------------------------|
| N°                         | Libellé cas                                                        |                                  | Nombre de<br>règles<br>d'association | Prédiction<br>ADAST | Seuil de<br>probabilité |                          |
| CAST_47                    | Défaut de transmission entre deux PA adjacents                     | 15                               | CL01                                 | 0,8                 | CL01                    | Valide                   |
| CAST_04                    | Recul d'une Rame sur un tronçon en initialisation                  | 26                               | CL02                                 | 1,0                 | CL02                    | Valide                   |
| CAST_17                    | Recul de l'élément cible au changement d'itinéraire                | 15                               | CL6                                  | 1,0                 | CL3                     | <b><u>Non valide</u></b> |
| CAST_19                    | Cible délocalisée                                                  | 15                               | CL3                                  | 1,0                 | CL3                     | Valide                   |
| CAST_24                    | Recul d'élément repéré                                             | 32                               | CL4                                  | 1,0                 | CL4                     | Valide                   |
| CAST_13                    | Perte d'élément en phase d'accostage                               | 15                               | CL5                                  | 1,0                 | CL5                     | Valide                   |
| CAST_55                    | Cible à contre-sens                                                | 15                               | CL6                                  | 0,83                | CL6                     | Valide                   |
| CAST_11                    | Perte d'un élément suite à une panne d'émetteur d'anticollision    | 15                               | CL7                                  | 0,8                 | CL7                     | Valide                   |
| CAST_26                    | Effacement trop tôt d'un élément                                   | 15                               | CL8                                  | 0,7                 | CL8                     | Valide                   |
| CAST_27                    | Franchissement d'un point d'arrêt en conduite manuelle             | 15                               | CL9                                  | 0,75                | CL9                     | Valide                   |
| CAST_67                    | Déraillement ou collision suite à un défaut de contrôle d'aiguille | 26                               | CL10                                 | 1,0                 | CL10                    | Valide                   |
| CAST_70                    | Captage par le train des informations émises sur le canton aval    | 26                               | ?                                    | ?                   | CL11                    | <b><u>Non valide</u></b> |
| CAST_56                    | Demande d'évacuation en démarrage de station                       | 0                                | CL12<br>Autre                        | 0,8                 | CL12<br>Autre           | Valide                   |
| CAST_62                    | Train avec voyageurs en garage                                     | 15                               | CL12<br>Autre                        | 0,8                 | CL12<br>Autre           | Valide                   |

**Tableau 6.5. Interprétation des résultats de classification des cas cible**

Les résultats de la classification peuvent être interprétés comme suit (Figure 6.23) :

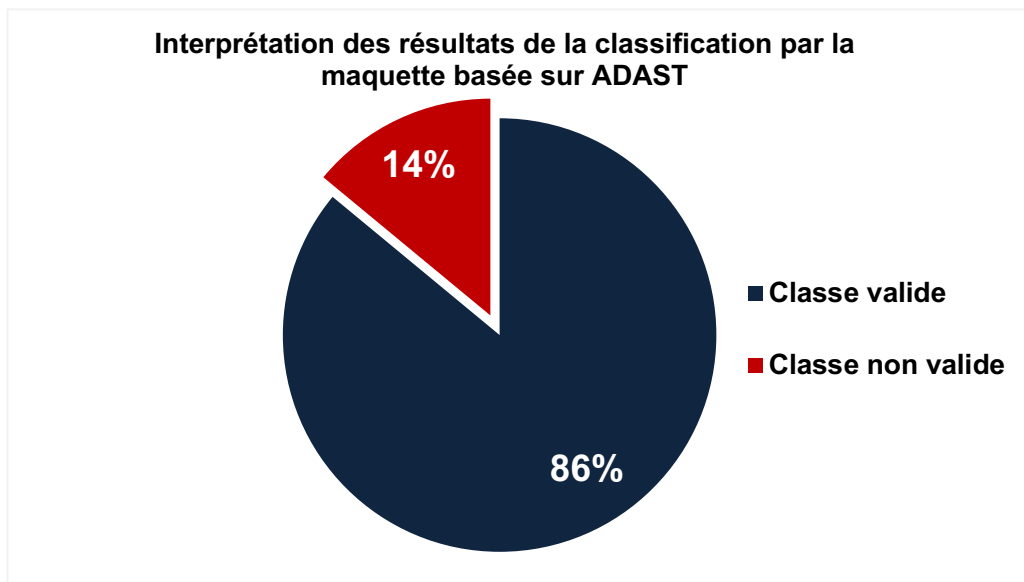


Figure 6.23. Interprétation des résultats de classification obtenus par le système.

- 12 classes proposées sont "conformes" sur 14 (un pourcentage à peu près égal à 86% de réussite) : ces classes sont correctes comme préconisé par la classification dont nous disposons initialement déjà validée par l'expert de domaine (elles représentent les classes originales attribuées à ces cas sources) ;
- 2 classes sont "non conformes" (un pourcentage de 14%) : le système propose une classe qui n'est pas conforme à la classe originale attribuée au cas de test. La principale cause de cet échec pour ces deux cas (CAST17 et CAST70) provient du classifieur que nous avons adopté (voir section 6.3.2), qui donne un taux correct de classification égale à 87%.

### 6.5.2. Interprétation des résultats de la démarche d'aide à la décision

Nous recensons également, quatre champs pour le tableau 6.6 ; les classes prédites, les pourcentages de similarité relatives aux trois cas sources remémorés les plus similaires et notamment la solution proposée par le système pour résoudre le problème. Nous donnons :

- SP : Isoler un PA ;
- SA4 : Faire le bilan des rames présentées en ligne ;
- SA5 : Renforcement des consignes ;
- SA10 : Le canton amont ne doit être libéré que si l'aval est occupé ;
- SA12 : Vérifier l'accostage effectif ;
- SA16 : Tenir compte de la distance de recul ;
- SA17 : Élément prend comme cible un élément inconnu FU ;
- SA18 : Compter le nombre de FU ;
- SA38 : Un PA considère le PA aval en mode élément sortant s'il est en défaut de réception avec lui ;
- SA41 : Immobiliser le véhicule tant que l'alarme d'évacuation n'est pas acquittée ;
- SA49 : Vérifier le sens de la cible ;
- SA53 : Tester périodiquement les capteurs de position d'aiguille ;
- SA56 : Éviter par conception tout couplage possible entre 2 cantons successifs.



| Cas de test | Résultat de remémoration |                             |                     | Résultat d'adaptation       |                   |                        | Utilité          | Résultat d'apprentissage           |                |                        |
|-------------|--------------------------|-----------------------------|---------------------|-----------------------------|-------------------|------------------------|------------------|------------------------------------|----------------|------------------------|
|             | Prédiction ADAST         | Les cas les plus similaires | Taux de Similarité  | Nbre de règles d'adaptation | Proposition ADAST | Originale [Par Expert] |                  | Cas sources ayant la même solution | Score          | Décision apprentissage |
| CAST_47     | CL01                     | CAS_47<br>CAS_08<br>CAS_29  | 100%<br>66%<br>33%  | 4                           | SA38              | SA38                   | Utile            | CAS_47                             | 30             | NON                    |
| CAST_04     | CL02                     | CAS_04<br>CAS_23<br>CAS_05  | 100%<br>60%<br>40%  | 5                           | SA4               | SA4                    | Utile            | CAS_04                             | 24             | NON                    |
| CAST_17     | CL3                      | CAS_16<br>CAS_55<br>CAS_18  | 88%<br>66%<br>55%   | 4                           | SA15              | SP<br>SA16             | <b>Non utile</b> | CAS_16                             | 33             | NON                    |
| CAST_19     | CL3                      | CAS_19<br>CAS_17<br>CAS_65  | 100%<br>71%<br>57%  | 4                           | SP<br>SA17        | SP<br>SA17             | Utile            | CAS_18<br>CAS_19                   | 27<br>24       | NON                    |
| CAST_24     | CL4                      | CAS_24<br>CAS_20<br>CAS_50  | 100%<br>80%<br>80%  | 5                           | S18               | SA18                   | Utile            | CAS_10<br>CAS_20<br>CAS_24         | 35<br>32<br>21 | NON                    |
| CAST_13     | CL5                      | CAS_13<br>CAS_14<br>CAS_15  | 100%<br>88%<br>55%  | 4                           | SA12<br>SA13      | SA12                   | Utile            | CAS_13                             | 35             | NON                    |
| CAST_55     | CL6                      | CAS_55<br>CAS_16<br>CAS_02  | 100%<br>88%<br>71%  | 5                           | SA49<br>SA17      | SA49                   | Utile            | CAS_55                             | 23             | NON                    |
| CAST_11     | CL7                      | CAS_11<br>CAS_37<br>CAS_44  | 100%<br>55%<br>55%  | 6                           | SA10              | SA10                   | Utile            | CAS_11                             | 29             | NON                    |
| CAST_26     | CL8                      | CAS_26<br>CAS_42<br>CAS_41  | 100%<br>85%<br>71%  | 5                           | SA22              | SA22                   | Utile            | CAS_26                             | 30             | NON                    |
| CAST_27     | CL9                      | CAS_27<br>CAS_28<br>CAS_43  | 100%<br>100%<br>80% | 4                           | SA5               | SA5                    | Utile            | CAS_05<br>CAS_23<br>CAS_27         | 35<br>44<br>20 | NON                    |
| CAST_67     | CL10                     | CAS_67<br>CAS_52            | 100%<br>80%         | 5                           | SA53              | S53                    | Utile            | CAS_67                             | 18             | NON                    |
| CAST_70     | ?                        | ?                           | -                   | -                           | ?                 | SA56                   | <b>Non utile</b> | -                                  | -              | NON                    |
| CAST_56     | CL12<br>Autre            | CAS_56<br>CAS_59            | 100%<br>50%         | 2                           | SA41              | SA41                   | Utile            | CAS_50<br>CAS_56                   | 36<br>24       | NON                    |
| CAST_62     | CL12<br>Autre            | CAS_62<br>CAS_59<br>CAS_58  | 100%<br>71%<br>57%  | 4                           | SA50              | SA50                   | Utile            | CAS_62<br>CAS_59                   | 27<br>20       | <b>OUI</b>             |

Tableau 6.6. Interprétation des résultats de la démarche d'aide à la décision

D'une manière générale, nous pouvons interpréter les résultats de l'injection de ces quatorze cas de test à travers la figure 6.24 :

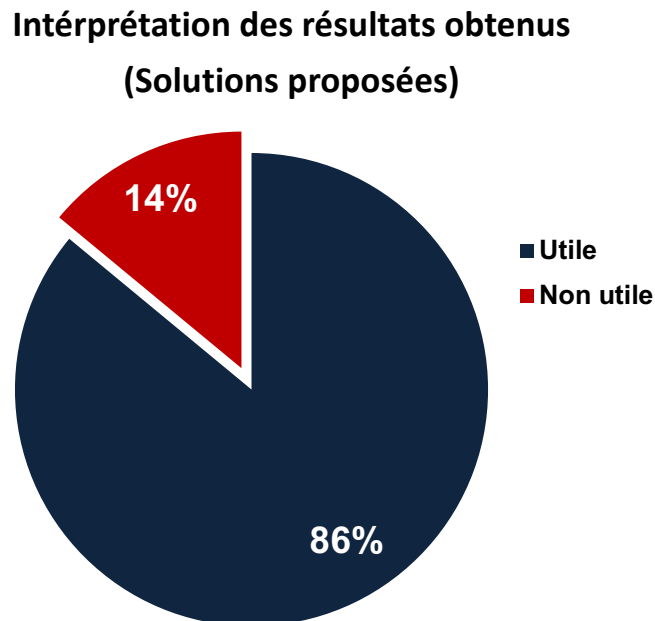


Figure 6.24. Interprétation des résultats obtenus par le système.

Nous pouvons interpréter les résultats obtenus comme suit :

- 12 solutions proposées sont "utiles" sur 14 (un pourcentage à peu près égal à 86% de réussite) : ces solutions sont correctes et peuvent résoudre le problème (le scénario d'accident) comme préconisé par l'expert de domaine (elles représentent les solutions originales attribuées à ces cas sources) ;
- 2 solutions sont "non utiles" (un pourcentage de 14%) : le système propose une solution qui n'est pas conforme à la solution originale attribuée au cas de test. La principale cause de cet échec provient de la non-conformité de la classe prédite par le système (CAST17 et CAST70) qui a conduit vers un calcul de similarité dans une classe différente à celle censée être la classe d'appartenance adéquate. Ceci peut entraîner le repérage des cas sources avec un pourcentage de similarité faible d'où des règles d'adaptations non applicables pour résoudre les problèmes « cible » ;
- Pour le résultat d'apprentissage, sur 14 cas de test, le système a préconisé l'apprentissage seulement du dernier cas CAST\_62. En effet, selon la démarche d'auto-incrémentation que nous avons proposée, ce dernier présente un score =22 qui est supérieur au score associé au cas source CAS\_59 qui présente la même solution SA50.

Bien que cette première évaluation de tout le système ait montré le bien-fondé de la démarche d'aide à la décision adoptée dans l'approche ADAST, une évaluation plus fine est nécessaire pour mesurer et à terme optimiser la performance du système.

Cette nécessité est due essentiellement à la pauvreté de la base d'apprentissage et à la qualité des scénarios d'accident. Pour cela, il est impératif d'enrichir la base d'exemples d'apprentissage par des cas supplémentaires relatifs au risque de collision ainsi que tous les autres risques d'accidents.

Au cours du paragraphe suivant, nous allons positionner les différentes limites et les problèmes rencontrés dans notre démarche d'aide à la décision proposée pour déboucher éventuellement sur des perspectives de recherche à court et à moyen terme.

## **6.6. Limites de la démarche d'aide à la décision**

Bien que l'approche ADAST que nous avons proposé ait montré un intérêt en ce qui concerne la démarche d'aide à la décision et la résolution de problème, elle a aussi révélé des carences et des limites. Dans ce paragraphe, nous discutons ces limites pour déboucher en final sur des perspectives envisageables pouvant améliorer sensiblement toute la démarche d'aide à la décision.

Notre approche a révélé quelques problèmes. Ceux-ci relèvent de plusieurs aspects de la démarche suivie. Ces problèmes peuvent se résumer dans les quatre points suivants : la qualité et la quantité des cas d'accidents, la diversité des techniques et des méthodes utilisées, la validation globale et la complexité des algorithmes proposés.

### **6.6.1. Problèmes de la qualité et de la quantité des cas d'accidents**

Un inconvénient que nous pouvons mentionner est en relation avec l'échantillon d'apprentissage. En effet, les résultats expérimentaux ont montré que le déploiement et l'utilisation actuelle des exemples d'apprentissage n'ont pas atteint la réussite espérée.

Bien que les données collectées et acquises dans l'échantillon initial sont relatives à un seul type de risque jugé représentatif des autres risques, cet échantillon reste très limité en terme de cardinalité et surtout que les scénarios du risque de collision sont répartis sur 11 classes dont la taille moyenne est de 6 scénarios. Ce problème de quantité insuffisante de scénarios est couplé à un autre problème de qualité de ces scénarios. Certains scénarios sont mal décrits et comportent parfois des erreurs de modélisation. Même si nous avons enrichi cet échantillon par d'autres scénarios d'insécurité, ça reste toujours limité et ne résout pas le problème en profondeur. Cette lacune peut être améliorée si nous collectons et construisons une nouvelle base d'exemples d'apprentissage à partir des enquêtes techniques des accidents récemment survenus.

D'une manière générale, l'enrichissement de l'échantillon d'apprentissage par des exemples issus du terrain réel ne peut être que bénéfique pour toute la démarche d'aide à la décision préconisée.

### **6.6.2. La diversité des techniques et des méthodes appliquées**

Bien que les techniques et les méthodes utilisées dans notre approche aient fourni des résultats intéressants, néanmoins, des limites peuvent être dégagées. La première méthode utilisant des règles d'association apparaît intéressante dans le sens où l'application de ces règles est simple et n'est pas assujettie à la problématique des événements d'insécurité rares. La deuxième méthode se basant sur l'arbre de décision semble facile à mettre en œuvre malgré les difficultés de répertorier initialement les classes et le nombre limité de scénarios d'accidents.

La méthode d'acquisition des connaissances d'adaptation est plus complexe à mettre en œuvre et nécessite un effort non négligeable pour valider les règles d'adaptation issues de chaque treillis de concept. Le travail effectué hors-ligne est manuel, il nécessite beaucoup de concentration et une bonne analyse pour tirer les règles de substitution et d'enrichissement convenables et applicables. Cependant, d'autres techniques peuvent être envisagées pour automatiser la tâche d'analyse des treillis de concepts et générer les règles d'adaptation.

### 6.6.3. La validation globale

Étant donné qu'il y'a eu une séparation effective entre la base de connaissances et le module de raisonnement, et pour couvrir d'autres domaines similaires, une phase d'acquisition de connaissances est indispensable pour préparer les ressources nécessaires au raisonnement en particulier, les règles d'association et les règles de classification. De même pour la phase la plus délicate, l'acquisition de connaissances d'adaptation peut ne pas être adéquate si un changement majeur dans la description des exemples d'apprentissage survient.

Au second niveau, la validation globale de la démarche d'aide à la décision implique l'expert du domaine. En pratique, cette validation s'avère très délicate dans la mesure où l'expert rencontre d'énormes difficultés à apprécier la sémantique d'un ensemble de règles vu le nombre élevé de ces règles qui constituent le modèle et le niveau de granularité qui ne permet pas de rendre compte de l'enchaînement de ces règles par le mécanisme d'inférence. Il pourrait être judicieux d'intégrer une phase d'argumentation des règles générées et des résultats trouvés par l'approche ADAST.

### 6.6.4. Complexité des algorithmes proposés

Dans le but de conforter la démarche d'aide à la décision proposée, plus particulièrement le processus de raisonnement à partir de cas, plusieurs algorithmes ont été implémentés et d'autres ont été conçus et testés manuellement. Bien que ces algorithmes permettent d'aboutir aux résultats escomptés, ils n'ont pas fait l'objet d'études spécifiques pour optimiser leur complexité et minimiser la lourdeur de leur utilisation. Ces algorithmes portent notamment sur :

- La recherche des règles d'association valables pour le cas cible ;
- La recherche des règles de classification applicables ;
- La recherche des cas similaires ;
- L'algorithme d'auto-incrémentation pour la phase d'apprentissage.

Ces problèmes d'efficacité et de lourdeur d'algorithmes sont "tolérables" dans des cas de l'étude de faisabilité de l'ensemble de la démarche proposée mais deviendront vite pénalisants lors d'une exploitation réelle sur des classes de cas d'accidents volumineuses. Dans cette perspective, une étude particulière devra leur être consacrée.

## 6.7. Conclusion

Ce dernier chapitre a présenté les expérimentations et la validation de notre approche ADAST ainsi que toute la démarche d'aide à la décision retenue. Cette démarche basée sur l'utilisation conjointe du raisonnement à partir de cas et d'une ontologie de domaine semble être intéressante dans le but qu'elle permet de remédier aux lacunes et limites rencontrées en utilisant les outils et les approches conventionnelles de résolution de problème.

L'approche d'expérimentation et de validation que nous avons poursuivie est scindée en trois volets: nous avons présenté premièrement les étapes de construction et d'acquisition de la base de connaissances à travers les différentes techniques utilisées issues de la fouille de données pour générer les règles d'association, les règles de classification et notamment l'approche d'acquisition de connaissances d'adaptation. En second lieu, nous avons illustré un scénario d'aide à la décision à travers l'injection d'un cas d'accident réel pour étudier toutes les étapes de raisonnement. Enfin, nous avons procédé à la validation du système à travers l'évaluation de la **validité** et de l'**utilité** des connaissances produites. A défaut de cas

témoins, l'évaluation s'est appuyée sur un lot de scénarios de tests pris à partir de la base des scénarios historiques.

Cette évaluation a montré l'intérêt majeur des techniques d'apprentissage et de fouilles de données utilisées pour tirer profit de l'échantillon d'apprentissage. Elle a aussi montré la faisabilité de l'ensemble de la démarche d'aide à la décision proposée. La maquette basée sur notre approche ADAST est considérée comme un système qui favorise la recherche de solutions convenables pour des situations d'insécurité susceptibles de rendre exhaustive l'analyse des risques d'accidents, par conséquent d'améliorer la sécurité.

---

# Conclusion générale

---

Ce mémoire a présenté notre contribution à l'amélioration des méthodes usuelles d'analyse et d'évaluation de la sécurité. Cette contribution, basée sur une approche de raisonnement à partir de cas (RàPC), les ontologies ainsi qu'un panorama d'outils et d'approches, s'est concrétisée par l'élaboration d'ADAST, (*une nouvelle approche générique d'aide à la décision pour l'amélioration de la sécurité*), dans un domaine où la connaissance experte est dispersée, évolutive et intuitive. Notre approche est organisée de façon à reproduire, en grande partie, la stratégie utilisée par les experts de domaine. Le foisonnement des méthodes, techniques ainsi que l'ambiguïté du vocabulaire employé compliquent le choix de l'utilisateur d'un système adapté à son domaine d'étude. Pour simplifier ce choix, nous avons proposé une vision synthétique du domaine étudié. Cette approche, qui reste encore à améliorer, nous a cependant permis d'imposer plusieurs outils et techniques issues de la fouille de données et notamment l'analyse des concepts formels (ACF) appliquée à l'acquisition des connaissances d'adaptation.

Après **l'introduction générale**, dans laquelle, nous avons présenté le contexte, la problématique et défini les objectifs, dans **le premier chapitre** de ce mémoire, nous avons introduit le domaine de notre recherche, la sécurité et le retour d'expérience (Rex). Ce chapitre a débouché sur le choix d'une part de l'ontologie pour l'acquisition de connaissances et la structuration des cas qui, appliquée au domaine de la sécurité, s'est révélée efficace pour l'extraction et la formalisation des scénarios d'accidents. D'autre part sur le RàPC, comme méthode parente de Rex permettant de combler les problèmes liés à la résolution de problèmes et l'exploitation des connaissances issues de Rex.

Au sein du **deuxième chapitre**, nous avons dressé une étude bibliographique exhaustive sur les concepts de base adoptés, à savoir, les ontologies et le RàPC à travers leur aspects importants et plus particulièrement celles que nous abordons dans notre travail de thèse. Notre étude s'est focalisée ensuite sur la fouille de données et les méthodes d'extraction de connaissances à partir de données, plus précisément ceux adoptées dans le cadre de notre approche, les règles d'association, la méthode d'arbre de décision et la technique d'Analyse des Concepts Formels (ACF). Cette étude théorique au sein de ce deuxième chapitre a débouché enfin sur une synthèse et un bilan des choix méthodologiques et des techniques effectués.

Dans le **troisième chapitre**, nous avons étudié les différentes approches de RàPC qui se sont basées d'une façon ou d'une autre sur les ontologies et celles liées à notre domaine d'application, la sécurité des transports. Nous avons synthétisé au final ces travaux pour montrer les limites et les carences dans le but de bien situer notre approche.

Dans le **quatrième chapitre**, nous avons présenté notre approche d'aide à la décision baptisée, ADAST (**A**ide à la **D**écision pour **A**méliorer la **S**écurité des **T**ransports). Nous avons rappelé tout d'abord les motivations, les objectifs attendus et nos choix méthodologiques adoptés. Nous avons ensuite présenté l'architecture fonctionnelle, les modèles de connaissances élaborés et les articulations entre les phases de raisonnement et la démarche d'aide à la décision retenue.

Au sein du **cinquième et du sixième chapitre**, nous avons présenté les premiers pas de validation de notre démarche d'aide à la décision à travers la maquette de faisabilité basée sur ADAST et aussi sur les tests expérimentaux et l'évaluation que nous avons menée pour

---

soutenir et consolider tout le travail proposé. Nous avons enfin dégagé les limites liées à notre approche ADAST dans le but de bien positionner les perspectives éventuelles pouvant évoluées tout le travail proposé.

## A. Contributions

Dans ce paragraphe, nous résumons les résultats achevés ainsi que les différentes contributions autour de processus de raisonnement à partir de cas (RàPC) et notamment toute la démarche d'aide à la décision.

Dans la phase d'élaboration, nous avons convenu de solliciter l'utilisateur à décrire le cas cible à travers l'instanciation de l'ontologie de domaine. Cette ontologie peut être considérée comme un bon point d'accès pour représenter des connaissances de domaine. Nous avons convenu également d'assister l'utilisateur moyennant les règles d'association dans le cas où sa description initiale est bruitée ou insuffisante à élaborer un cas cible. A cet effet, l'application de l'algorithme Apriori (Agrawal R. et Srikant R., 1993) a pour but de produire à partir de l'échantillon d'apprentissage une liste des instances pertinentes à utiliser sous forme de règles pour enrichir d'avantage la description initiale. Pour mener efficacement l'extraction des règles d'association, nous avons utilisé trois paramètres complémentaire ; la confiance, le support et le lift. Ces trois paramètres favorisent une bonne précision des règles produites. Cette extension permet non seulement d'optimiser la phase d'élaboration (le traitement des données manquantes ou entachées d'erreurs) mais aussi d'améliorer la qualité des connaissances en entrée pour la prochaine phase, la remémoration.

Dans notre approche, la phase de remémoration (deuxième phase du RàPC) est plus ciblée, elle opère la recherche des cas similaires uniquement dans la classe d'appartenance du cas cible. La technique d'arbre de décision (issue de la fouille de données) est utilisée pour générer les règles de décision pouvant opérer la détection de la classe adéquate du cas cible. L'algorithme C4.5 (Quinlan R., 1993) est alors appliqué pour générer les règles de décisions convenables. Une fois le repérage de la classe d'appartenance est effectué, nous opérons une recherche des cas jugés les plus similaires à partir de la classe déduite. Un algorithme de calcul de similarité a été proposé à cet effet. L'approche de calcul de similarité retenue utilise deux mesures complémentaires : une mesure basé concept et une mesure basé slot. L'apport de notre approche pour la phase de remémoration est double ; d'une part, nous limitons l'utilisation des ressources et en d'autre part, nous gagnons en termes de sémantique puisque les cas appartenant à une même classe partagent généralement les mêmes caractéristiques.

En ce qui concerne notre approche d'adaptation que nous avons suivie, elle est scindée en deux étapes : une première étape, dans laquelle nous préparons les différentes règles d'adaptation et une deuxième, dans laquelle, le système infère en-ligne sur les règles d'adaptation générées pour solliciter une solution possible au cas cible. Pour acquérir les connaissances d'adaptation, nous utilisons l'approche d'analyse des concepts formels (ACF). L'ACF permet la découverte automatique des attributs partagés entre des cas sources appartenant à une même classe de solution et des régularités entre ces cas à travers des règles de substitutions et/ou des règles d'enrichissement possibles. Notre travail sur l'acquisition des connaissances d'adaptation est proche du travail réalisé dans Cabamaka (d'Aquin M. et al. 2007, Badra F. et al. 2007) et inspiré également des travaux de COBRA (Abou-Assali A., 2010) et (Abou-Assali A. et al, 2013) dans le sens où tous ces travaux s'appuient sur les techniques de l'analyse de concepts formels pour extraire les connaissances d'adaptation à partir de la base de cas. Cabamaka vise, dans un premier temps, à extraire les motifs fermés fréquents de l'ensemble des couples de cas différents

constitués à partir de la base de cas. Ces motifs expriment les similarités et les dissimilarités obtenues en examinant chaque couple de cas. Ils sont ensuite interprétés et validés par un analyste pour former des règles d'adaptation. COBRA utilise les connaissances d'adaptation au sein même de la phase de remémoration pour chercher les cas similaires considérés comme les plus adaptables. Dans notre approche, nous jugeons que la phase de remémoration est optimisée de telle sorte qu'elle nous permet de repérer des cas très similaires au cas cible d'où une probabilité d'adaptation élevée. De ce fait, notre approche nous permet d'éviter ce coût élevé en ne tenant compte que des similarités entre les cas. Le fait d'ignorer les dissimilarités est raisonnable parce que les cas sont séparés initialement selon leur classe d'appartenance.

Dans la phase de révision, la solution possible au cas cible proposée à l'issue de la phase d'adaptation sera évaluée manuellement par l'utilisateur. Il se peut qu'elle ne convienne pas à l'utilisateur ou bien qu'une fois mise en application, elle s'avère incapable de résoudre le problème posé. L'utilisateur a donc, pendant la phase de révision, l'opportunité de modifier, affiner ou même refuser la solution proposée.

Pour bien mener la phase d'apprentissage, nous avons proposé un algorithme d'auto-incrémentation de la base de cas. Cet algorithme s'inspire des travaux réalisés sur la maintenance de la base par (Haouchine M.K. et al., 2009). Considérant que notre base de cas regroupe un ensemble de classes de cas, d'où une structuration initiale suffisamment préinstallée, nous avons convenu d'éliminer la méthode de structuration des cas proposée par cet auteur en mettant en place une méthode d'auto-incrémentation. Cette méthode permet de donner une autre optique qui prend en considération plus l'aspect sémantique des connaissances. Chaque décision d'apprentissage (enrichissement de la base de cas) se fait en se basant sur un score dédié au calcul de la représentativité de la partie problème du cas cible par rapport à la partie problème du cas source partageant la même solution.

Pour consolider et soutenir toute l'approche proposée, nous avons proposé un premier prototype d'une maquette de faisabilité basée sur ADAST. Cette maquette constitue à ce jour, les premiers pas de validation globale de l'approche. Nous avons mis en œuvre cette maquette en séparant le module de raisonnement de la base de connaissances. Ceci est dans le but de fournir un outil qui peut couvrir d'autres domaines d'études similaires.

L'expérimentation que nous avons menée quant à elle, a permis de consolider la première validation effectuée à travers la maquette de faisabilité basée sur ADAST. Elle nous a permis également d'évaluer toute la démarche d'aide à la décision à travers la mesure de **la validité et d'utilité** des connaissances produites.

## B. Perspectives

Afin que notre démarche d'aide à la décision puisse être évaluée en milieu industriel, il est nécessaire de prévoir certaines améliorations pour le module d'aide à la décision d'une part et d'autre part de consolider la quantité et la qualité des connaissances détenues par le système. Les points que nous venons d'évoquer permettent de déclencher une réflexion en termes de perspectives sur l'ensemble des problèmes posés par l'évaluation.

Le travail présenté dans cette thèse ouvre plusieurs perspectives que nous essaierons d'explorer dans le futur. Ces perspectives peuvent être regroupées autour de cinq points :

### a. Enrichir et consolider la base historique de scénarios

Initialement estimé à plusieurs centaines, le nombre de scénarios (cas d'accidents) acquis à ce jour s'élève à 70. Cette révision à la baisse influe évidemment sur la qualité des résultats.



En termes de perspective, il faut que les travaux dans ce sens se focalisent sur une action visant l'amélioration de la quantité des cas d'accidents fournis. Il faut définir des critères minimaux de complétude et d'exhaustivité de la base de scénarios et des classes qui la composent. Les critères **minimaux de complétude** fixent le nombre minimum de scénarios que la base doit contenir et les critères **minimaux d'exhaustivité** renseignant sur les seuils à partir desquels, la base de scénarios peut être jugée suffisamment représentative de l'ensemble d'accidents (collision, déraillement, chute, évacuation d'urgence, etc.) et de l'ensemble des problèmes de sécurité (protection des itinéraires, accostage, initialisation, aiguillage, etc.).

#### b. Intégrer et optimiser les algorithmes utilisés

Il est très intéressant de faire des études spécifiques pour optimiser la complexité et minimiser la lourdeur de l'utilisation des différents algorithmes utilisés dans toute la démarche d'aide à la décision. Dans un autre volet, une démarche d'intégration des algorithmes de fouilles de données dans la maquette ADAST pourra être utile en termes de portabilité. En effet, plusieurs outils de fouilles de données telle que Weka, Research Sipina ou encore Tanagra offrent des codes source à des fins de réutilisation. Ceci peut nous être un bon point d'accès pour atteindre les objectifs envisagés dans cette perspective.

#### c. Apprendre par le succès et l'échec pour améliorer la démarche d'aide à la décision

Cette approche d'apprentissage par le succès et l'échec est initialement introduite par (Kodratoff Y., 1986) adoptée ensuite dans les travaux de (Mejri L., 2009). C'est une méthode basée sur les techniques de "récompense" et de "sanction". La technique de récompense consiste à donner plus de poids aux connaissances sollicitées dans le processus de résolution du problème. La technique de sanction pénalise les connaissances menant à un échec dans la résolution du problème. Apprendre par le succès et l'échec, c'est apprendre à partir de ses propres réussites et tirer profit de ses erreurs, bref de son expérience.

Cette forme d'apprentissage permettrait au système de prendre en compte ses réussites, c'est-à-dire les solutions générées conformes qui sont jugées valables, en utilisant la technique de la récompense. Celle-ci aurait pour effet de donner plus de poids aux règles d'adaptation qui ont été sollicitées dans la phase d'adaptation. Ce type d'apprentissage permettrait aussi au système d'aide de tirer profit de ses échecs, c'est à dire les solutions non conformes et qui sont rejetées par les experts, moyennant la technique de punition. Une sanction, dans ce cas, permet d'écarter de plus en plus les connaissances ou les règles qui ont contribué à la considération d'un chemin de raisonnement menant à un échec. La mise en œuvre de cette forme d'apprentissage pourrait être envisagée à travers un algorithme conçu pour conclure toute la session de résolution du problème.

#### d. Prise en compte de la modélisation dynamique des cas d'accidents

A l'état actuel de nos travaux de thèse, uniquement l'aspect statique des cas d'accidents est pris en considération dans la modélisation. Une perspective visant à joindre la modélisation dynamique au formalisme actuel moyennant l'ontologie de domaine pourrait être d'une grande utilité pour bien comprendre, analyser le processus d'aboutissement d'un accident. Dans le projet « ACASYA » d'Aide à la Certification par Apprentissage des SYstèmes de transport Automatisés (Hadj-Mabrouk H., 1992), (Mejri L., 1995), (Hadj-Mabrouk H., 2007) et (Mejri L., 2009), une ébauche d'une modélisation dynamique via les réseaux de pétri fût utilisée conjointement avec une fiche de description statique (Annexe

1). Dans cette optique, nous pouvons enrichir notre modèle de connaissances en développant une deuxième ontologie qui servira pour la description dynamique de l'aboutissement d'un cas d'accident. Les deux ontologies peuvent être impliquées donc dans toute la démarche d'aide à la décision proposée.

#### e. Ontology Web Language- Services (OWL-S)

Il s'agit d'une perspective de recherche qui permettrait une bonne extension de notre travail. Cette perspective semble être intéressante dans le cadre d'une démarche d'aide à la décision dans le domaine de la sécurité des transports faisant recours aux services web.

Rappelons que pour notre contexte, plusieurs acteurs (décideurs, expert de l'analyse de la sécurité, constructeur, responsable de collectivité publique, etc.), participent aux différentes activités d'analyse de la sécurité. Il serait donc opportun, de positionner une approche d'aide à la décision, ADAST en l'occurrence, dans le cadre d'un usage distribué dans les services dans le web. Les acteurs dans notre cas et les agents logiciels devraient être en mesure de découvrir, invoquer et surveiller en temps réel les ressources Web offrant des services en relation avec le retour d'expérience ferroviaire (les accidents et les situations d'insécurité) et devraient être en mesure de les exploiter en vue toujours de proposer l'aide à la décision adéquate, permettant aux différents acteurs d'améliorer la sécurité.

De puissants outils doivent être activés par des descriptions de services, à travers le cycle de vie du service Web. La mise en œuvre d'une ontologie de services OWL-S, permettra non seulement d'aider les agents logiciels à découvrir le service Web qui répondrait à un besoin spécifique de l'analyse de la sécurité, sans avoir besoin d'intervention humaine, mais aussi aider dans la composition et l'interopérabilité des services d'une manière qui permettra l'exécution automatique de ces tâches. L'ontologie OWL-S comporte trois parties principales : le profil du service (utilisé pour décrire ce que le service fait), le modèle de processus (décrit le fonctionnement du service) et le mode d'accès (les informations dont l'acteur a besoin pour invoquer le service).

---

# Bibliographie

---

- Aamodt, A. & Plaza, E. (1994). "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches". *AI Communications*, 7(1): pp.39–59.
- Aamodt, A. (1991). "A Knowledge-Intensive, Integrated Approach to Problem Solving and Sustained Learning". PhD thesis, University of Trondheim, Norway.
- Aamodt, A. (2004). "Knowledge-Intensive Case-Based Reasoning in CREEK". *Advances in Case-Based Reasoning*, volume 3155/2004: pp1–15.
- Abd-Elrahman, E., Samhaa, R., Mahmoud, R., & Osman, H. (2007). "Applying data mining for ontology building". The 42nd Annual Conference On Statistics, Computer Science, and Operations Research.
- Abou-Assali, A. (2010). "Acquisition des connaissances d'adaptation et Traitement de l'hétérogénéité dans un système de RàPC basé sur une Ontologie, Application au diagnostic de la défaillance de détecteurs de gaz". Thèse de doctorat à l'INRIA 2010.
- Abou-Assali, A., Lenne, D. & Debray, B. (2008). "Ontology Development for Industrial Risk Analysis". IEEE International Conference on Information & Communication Technologies: from Theory to Applications (ICTTA'08), Damascus, Syria, April.
- Abou-Assali, A., Lenne, D. & Debray, B. (2013), "Adaptation Knowledge Acquisition in a CBR System". *International Journal on Artificial Intelligence Tools*, Vol. 22, No. 1, 1250041.
- Abou-Assali, A., Lenne, D., Debray, B. & Sébastien B. (2009). "COBRA : Une plate-forme de RàPC basée sur des ontologies". IC 2009, Hammamet, Tunisie.
- Agrawal, R. & Srikant, R. (1994). "Fast Algorithms for Mining Association Rules". Proceedings of the 20th VLDB Conference, Santiago, Chile.
- Aime, X. (2011). "Gradients de prototypicalité, mesures de similarité et de proximité sémantique : une contribution à l'Ingénierie des Ontologies". Thèse de doctorat en Informatique, École polytechnique de l'Université de Nantes, Le 8 avril 2011.
- Althoff, K.D., Auriol, E., Barletta, R. & Manago, M. (1995). "A Review of Industrial Case-Based Reasoning Tools". *AI Intelligence*, Oxford, UK.
- Amalberti, R. & Barriquault, C. (1999). "Fondements et limites du retour d'expérience". *Annales des Ponts et Chaussées* n° 91, pp.67-75.
- Amancio, B., Gerald, R., Abraham, B. & Harald, G. (2008). "SemTree: Ontology-Based Decision Tree Algorithm for Recommender Systems". Proceedings of International Semantic Web Conference (Posters & Demos), volume 401 of CEUR Workshop Proceedings, CEUR-WS.org,
- Arnaud, V. & Napoli, A. (2013). "Modélisation ontologique pour l'analyse de comportements de navires à risques". Colloque SAGEO 2013 - Spatial Analysis and GEOMatics, Brest, France.
- Arpirez, J., Corcho, O., Fernandez-Lopez, M. & Gomez-Perez, A. (2001). "WebODE: a Workbench for Ontological Engineering". In First international Conference on Knowledge Capture, Victoria, Canada.
- Bachimont, B. (2000). "Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances". *Évolution Récentes et nouveaux défis Paris* : Eyrolles, France, pp.305-323.
- Badra F., Cordier, A. et Lieber, J. (2009). "Découverte opportuniste de connaissances

- d'adaptation*". 17ème atelier de raisonnement à partir de cas, Paris, juin, pp.23–34.
- Badra, F. & Lieber, J. (2008). "Representing Case Variations for Learning General and Specific Adaptation Rules", Fourth Starting AI Researcher's Symposium (STAIRS 2008), IOS Press, Patras Grèce, pp.1–11.
- Badra, F. (2009). "Extraction de connaissances d'adaptation en raisonnement à partir de cas". Thèse de doctorat, 20 novembre 2009, université Henri Poincaré, Nancy 1.
- Badra, F., Bendaoud, R., Bentebitel, R., Champin, P.A., Cojan, J., Cordier, A., Després, S., Jean-Daubias, S., Lieber, J., Meilender, T. (2008). "Taaable : Text Mining, Ontology Engineering, and Hierarchical Classification for Textual Case-Based Cooking". ECCBR Workshops, Workshop of the First Computer Cooking Contest, pp.219–228.
- Badra, F., Cojan, J., Cordier, A., Lieber, J., Meilender, T., Mille, A., Molli, P., Nauer, E., Napoli, A., Skaf-Molli, H. & Toussaint, Y. (2009). "Knowledge acquisition and discovery for the textual case-based cooking system WikiTaaable". 8th International Conference on Case-Based Reasoning - ICCBR 2009, Workshop Proceedings.
- Badra, F., Lieber, J. & Napoli, A. (2007). "Extraction de connaissances d'adaptation par analyse de la base de cas". Actes des cinquièmes journées Extraction et Gestion des Connaissances (EGC'07), pp.751–760.
- Bahloul, D. (2006). « Une approche hybride de gestion des connaissances basée sur les ontologies: application aux incidents informatiques ». Thèse en informatique. Institut National des Sciences Appliquées de Lyon.
- Bahri, S. & Hadj-Mabrouk, H. (2009). "L'intégration de la Sécurité dans les Systèmes Technologiques de l'Information et de la Communication". SETIT2009, Hammamet, Tunisie.
- Barbut, M. & Monjard, B. (1970). "Ordre et classication : Algèbre et combinatoire". Editeur Hachette, 176 p.
- Bart, G. & Alireza, S. (2009). "Ontology Granulation Through Inductive Decision Trees I". 5th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW2009).
- Bayardo, R. J. (1998). "Efficiently mining long patterns from databases". in proceedings of the 1998 ACM SIGMOD international conference on Management of Data, ACM Presss, June, pp.85-93.
- Belacel N. (2000). "Méthodes de Classification Multicritère Méthodologie et Applications à l'Aide au Diagnostic Médical". Thèse présentée en vue de l'obtention du grade de Docteur en Sciences orientation Recherche Opérationnelle. Université libre de Bruxelles.
- Beler, C. (2008). "Modélisation générique d'un retour d'expérience cognitif. Application à la prévention des risques". Thèse de doctorat de l'université de Toulouse.
- Ben-yahya, S. & Engelbert, M. N. (2004). "Approches d'extraction de règles d'association basées sur la correspondance de Galois". Motifs dans les bases de données, pp.23-55.
- Bergenti, F. & Poggi, A. (2000), "UML in the design of multi-agent systems". In Engineering Societies in the Agents world, Lecture Notes in Computer Science 1972, Springer, 2000, pp.106-113.
- Bergmann, R. & Stahl, A. (1998). "Similarity measures for object-oriented case representations". Dans Proceedings of the European Workshop on Case-Based Reasoning, EWCBR' 98.
- Bergmann, R. (2002). "Experience Management: Foundations, Development Methodology, and Internet based applications". LNCS, 2432, Springer-Verlag Berlin Heidelberg.
- Bergmann, R., Althoff, K.D., Breen, S., Göker, M., Manago, M., Traphöner, R. & Weiss, S. (2003). "Developing Industrial Case-Based Reasoning Applications: The INRECA Methodology". Lecture Notes in Artificial Intelligence, LNAI 1612, Springer Verlag, Berlin.

- Birkhoff, G. (1950). "*Lattice theory*". American Mathematical, Society Colloquium Publications 25 no. 2, pp.204-206.
- Blazquez, M., Fernández, M., Garcia-Pinar, J.M. & Gomez-Perez, A. (1998). "*Building Ontologies at the Knowledge Level using the Ontology Design Environment*". In Proceedings of the Banff Workshop on Knowledge Acquisition for Knowledge-based Systems, Canada.
- Borst, W. (1997). "*Construction of Engineering Ontologies for Knowledge Sharing and Reuse*". Ph.D. Dissertation, University of Twente.
- Boulanger, J-L. (2006). "*Expression et validation des propriétés de sécurité logique et physique pour les systèmes informatiques critiques*". Thèse de Doctorat de l'Université de Technologie Compiègne.
- Boussaid, I. (2005). « *Personnalisation de l'information et gestion des profils utilisateurs : une approche fondée sur les ontologies* ». Université de Mentouri Constantine Master en informatique, Option Computation & Information.
- Brachman, R., Schmolze, J.G. (1985). "*An overview of the KI-One knowledge representation system*". Cognitive Science, vol. 9, n°2, pp.171-216.
- Brickley, B. (2004). "*RDF Vocabulary Description Language 1.0: RDF Schema*". W3C Recommendation 10 February.
- Brickley, D. (2002). "*RDF vocabulary description language 1.0: RDF schema*". [URL : <http://www.w3.org/tr/rdf-schema/>].
- Brin, S., Motwani, R. & Silverstein, C. (1997). "*Beyond market baskets: generalized associations rules to correlations*". In Proceedings of ACM SIGMOD'97.
- Brin, S., Motwani, R., Ullma, J.D. & Tsur, S. (1997). "*Dynamic itemset counting and implication rules for market basket data*". In Proceedings of the 1997 ACM SIGMOD international conference on Management of data, pp.255-264.
- Bringaud, V. (2012). "*20 propositions pour organiser un dispositif de retour d'expérience*". Communication au 18ème congrès de maîtrise des risques et de sûreté de fonctionnement – Lambda mu, Tours, octobre.
- Cabeau, E. (1999). "*Introduction à la conception de la sûreté*". Cahier technique Schneider Electric n°144.
- Cain, T., Pazzani, M. J., Silverstein, G. (1991) "Using domain knowledge to influence similarity judgement". Proceedings of the Case-Based Reasoning Workshop. Morgan Kaufmann, Washington, DC, pp.191-202
- Capsus, L. & Tourigny, N. (2000). "*Le raisonnement à partir de cas : une aide à la formation en analyse de sécurité routière*". Technologies de l'information et de la communication dans les enseignements d'ingénieurs et dans l'industrie. Colloque international No2, Troyes, France, pp.227-235.
- Carbonell, J.G. (1983). "*Derivational Analogy and its Role in Problem Solving*". Proceedings of the third Annual National Conference on Artificial Intelligence, Washington, D.C., USA, August, AAAI-83, pp.64–69,
- Cauffriez, L., Ciccotelli, J., Conrard, B. & Bayart, M. (2004). "*Design of intelligent distributed systems: a dependability point of view*". Reliability Engineering and System Safety, Eselvier Editions, Vol. 84: pp.19–32.
- Ceausu, V. & Després, S. (2004). "*Une approche Mixte pour la construction d'une ressource terminologique Domaine d'application : Accidentologie routière*". IC2004. [URL : <http://liris.cnrs.fr/~ic04/programme/articles/Ceausu-IC2004.pdf>].
- Cestnik, B., Kononenko, I. & Bratko, I. (1987). "*ASSISTANT 86: A Knowledge Elicitation Tool for Sophistical Users*". Proceedings of the 2nd European Working Session on Learning, pp.31-

45.

- Chauchat, J.H., Rakotomalala, R., Carloz, M. & Pelletier, C. (2001). "*Targeting customer groups using gain and cost matrix: a marketing application*". Proceedings of Data Mining for Marketing Applications Workshop, PKDD'2001, pp.1-13.
- Chaudet, F. (2002). "*Apport du retour d'expérience à la maîtrise des risques relatifs l'hygiène, la sécurité et l'environnement, dans le petit établissement industriels. Application à l'industrie du traitement thermique*". Thèse de doctorat l'institut national des sciences appliquées de Lyon et l'école nationale supérieure des mines de Saint-Etienne
- Chein, M. & Mugnier, M.L. (1992). «*Conceptual graphs : fundamental notions* ». Revue d'Intelligence Artificielle 1992, vol. 6 (4), pp.365-406.
- Chen, K. & Barthès, J.B. (2007). "*MemoPA: Intelligent Personal Assistant Agents with Case Memory Mechanism*". Lecture Notes in Computer Science, 4682: pp.1357–1367.
- Chen, K. (2008). "*Contribution à la Conception de la Mémoire d'un Agent Assistant Personnel*". PhD thesis, Université de Technologie de Compiègne, France.
- Chevalier, M., Garnier, R., Chang, P., Lusson, B. (2004). "*La Sûreté de Fonctionnement (SdF)*". Le magazine Schneider Electric de l'enseignement technologique et professionnel. [URL : [http://www.intersections.schneider-electric.fr/stock\\_images/elec/1/n3/GT12.pdf](http://www.intersections.schneider-electric.fr/stock_images/elec/1/n3/GT12.pdf)]
- Chopard-Guillaumot, G., Hadj-Mabrouk H. (1996). "*Définition des principaux concepts relatifs à la notion de sécurité dans les transports guidés*". Revue Générale des Chemins de Fer, Paris, n° 2, Éditions Dunod, Février 1996, pp.23-36.
- Clay, S. (2006). "L'ontologie est surfaite". Traduction Christophe Ducamp. [URL : <http://www.elanceur.org/Articles/OntologieSurfaite.html>].
- Cordier, A. (2008). "*Interactive Knowledge Acquisition in Case Based Reasoning*". Thèse de doctorat de l'Université Claude Bernard Lyon 1.
- Cordier, A., Mascaret, B. & Mille, A. (2009). "*Étendre les possibilités du raisonnement à partir de cas grâce aux traces*". 17ème Séminaire Raisonnement à Partir de Cas.
- Cordier, A., Fuchs, B. & Mille, A. (2006). "*Engineering and learning of adaptation knowledge in Case-Based Reasoning, Managing Knowledge in a World of Networks*". Lecture Notes in Computer Science, pp.303–317.
- Cordier, A., Fuchs, B., Lieber, J. & Mille, A. (2007). "*Acquisition interactive des connaissances d'adaptation intégrée aux sessions de raisonnement à partir de cas - Principes, architecture IAKA et prototype KAYAK*". Actes du 15ème atelier de raisonnement à partir de cas (RàPC'07), Grenoble, France, pp.71–84,
- Cordier, A., Fuchs, B., Lieber, J. & Mille, A. (2007). "*Failure analysis for domain knowledge acquisition in a knowledge-intensive CBR system*". Proceedings of the 7th International Conference on Case-Based Reasoning, (ICCB'07), pp.463–477.
- Cover, T.M. & Joy A.T., (2006). "*Elements of Information Theory, 2nd Edition*". Wiley-Interscience, 776 pages.
- Cranefield, S., Purvis, M. (1999). "UML as an Ontology Modelling Language". Department of Information Science, University of Otago, New Zealand.
- Craw, S., Wiratunga, N. & Rowe, R. (2006). "*Learning adaptation knowledge to improve case-based reasoning*". Artificial Intelligence, 170(16-17): pp.1175–1192.
- Craw, S., Wiratunga, N. & Rowe, R. (2006). "*Learning adaptation knowledge to improve case-based reasoning*". Artificial Intelligence, 170(16-17):pp.1175–1192.
- d'Aquin M., Badra F., Lafrogne S., Lieber J., Napoli A., & Szathmary, L. (2007). "*Case base mining for adaptation knowledge acquisition*". Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI 2007), pp.750–755.

- d'Aquin, M., Badra, F., Lafrogne, S., Lieber, J., Napoli, A. & Szathmary, L. (2007). "Case base mining for adaptation knowledge acquisition". Proceedings of the International Conference on Artificial Intelligence, IJCAI'07, pp.750–756.
- d'Aquin, M., Lieber, J. & Napoli, A. (2005). "Decentralized Case-Based Reasoning for the Semantic Web". Proceedings of International Semantic Web Conference, ISWC 2005, Lecture Notes in Computer Science, pp.142–155.
- d'Amato, C., Staab, S. & Fanizzi, N. (2008). "On the Influence of Description Logics Ontologies on Conceptual Similarity". In EKAW 2008, International Conference on Knowledge Engineering and Knowledge Management Knowledge Patterns, pp.48-63.
- Dameron, O. (2003). "Modélisation, représentation et partage de connaissances anatomiques sur le cortex cérébral". Thèse de doctorat de l'université de Rennes.
- Després, S. & Ceausu, V. (2004). "Raisonnement à partir de cas pour contribuer à améliorer l'aménagement du réseau urbain en prenant en compte la sécurité". [URL :<http://old-lipn.univ-paris13.fr/seminaires/AtelierRaPC/Articles/despres.pdf>].
- Desroches, A., Leroy, A. & Vallé, F. (2003). "La gestion des risques ; principes et pratiques", Edition Lavoisier, pp.29–44,
- Díaz-Agudo, B. & González-Calero, P.A. (2000). "An architecture for knowledge intensive CBR systems". Advances in Case-Based Reasoning (EWCBR'00), Springer-Verlag, Berlin Heidelberg New York.
- Díaz-Agudo, B. & González-Calero, P.A. (2002). "CBROnto : a task/method ontology for CBR". Proceedings of the 15th International FLAIRS, 2 : pp.101–106.
- Díaz-Agudo, B., González-Calero, P.A., Recio-García, J.A. & Sánchez-Ruiz, A.A. (2007). "Building CBR systems with jcolibri. Science of Computer Programming". Special issue on Experimental Software and Toolkits, 69(1-3): pp.68–75.
- Dominati, A., Bonneau, A. & Lewkowitch-Orlandi A. (1996). "SACRE : une base de données sur les incidents du parc nucléaire d'EDF au service du retour d'expérience facteur humain". Colloque national de fiabilité et maintenabilité, n° 10, octobre 1996.
- Domingue, J. & Tadzebao. (1998). "WebOnto: Discussing, Browsing, and Editing Ontologies on the Web". the 11th Workshop on Knowledge Acquisition, Modeling and Management, KAW'98, Banff, Canada.
- Ehrig, M., Haase, P., Hefke, M. & Stojanovic, N. (2004). "Similarity for ontology-a comprehensive framework". In Workshop Enterprise Modelling and Ontology: Ingredients for Interoperability.
- Essam, A., LotfyAbdrabou, M. & AbdEl-Badeeh, M.S. (2010). "A Breast Cancer Classifier based on a Combination of Case-Based Reasoning and Ontology Approach". Proceedings of the International Multiconference on Computer Science and Information Technology, pp. 3-10
- Euzenat, J. (2001). "Towards a principled approach to semantic interoperability". In Proceedings of the IJCAI Workshop on Ontology and Information Sharing, pages 1925, Seattle, US.
- Euzenat, J. (2007). "Semantic precision and recall for ontology alignment evaluation", In Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007), Hyderabad, India, pp. 348-353.
- Farquhar, A., Fikes, R., Pratt, W. & Rice, J. (1995). "Collaborative Ontology Construction for Information Integration Technical Report KSL-95-10". Knowledge Systems Laboratory, Stanford University. CA.
- Fensel, D., Harrick's, I., Van Harmelen, F., Decker, S., Erdmann, M. & Klein, M. (2000). "OIL in a nutshell". In Proceedings of European Knowledge Acquisition Workshop (EKAW'2000), Juan-les-Pins, France, Springer-Verlag LNAI 1937, pp.1-16.

- Fernandez, M. (1999). "Overview of Methodologies for Building Ontologies". Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends, (IJCAI99), August.
- Fernandez, M., Gomez-Perez, A. & Juristo, N. (1997). "Methontology: From Ontological Art Towards Ontological Engineering". Proceedings of the AAAI97, Spring Symposium Series on Ontological Engineering, pp.33–40.
- Ferrandez, F. (1999). "L'apport des études détaillées d'accidents aux retours d'expérience en sécurité routière". Annales des Ponts et Chaussées n° 91, 1999, pp.36-42.
- Ferré, S. & Ridoux, O. (1999). "Une généralisation logique de l'analyse de concepts formels". Rapport de recherche - INRIA, ISSN 0249-6399.
- Fournier, P.V. & Vincent, S.T. (2012). "Mining Top-K Non-Redundant Association Rules". Proceedings of 20th International Symposium, ISMIS 2012, Macau, China, December, pp.31-40.
- Fuchs, B. & Mille, A. (1999). "Une modélisation au niveau connaissance de la tâche d'adaptation en raisonnement à partir de cas". Dans Actes du 7ème atelier de raisonnement à partir de cas (RàPC'99), pp.27–36.
- Fuchs, B., Lieber, J., Mille A. & Napoli, A. (2006). "Une première formalisation de la phase d'élaboration du raisonnement à partir de cas". Actes du 14ième atelier du raisonnement à partir de cas, Besançon.
- Fuchs, B., Lieber, J., Mille, A. & Napoli, A. (2006). "A general strategy for adaptation in case-based reasoning". Rapport technique RR-LIRIS-2006-016, LIRIS UMR 5205 CNRS/INSA de Lyon/Université Claude Bernard Lyon 1/Université Lumière Lyon 2/Ecole Centrale de Lyon, 2006.
- Fürst, F. & Trichet, F. (2006). "Raisonnement sur des ontologies lourdes à l'aide de Graphes Conceptuels, Une application à l'alignement d'ontologies de domaine", INFORSID, pp.879-894.
- Ganascia, J.G. (1991). "L'apprentissage symbolique". Encyclopédie de la communication - PUF, sous la direction de Lucien SFEZ.
- Ganascia, J.G., Puget, J.F. & Helft, N. (1990). "Comportement des systèmes d'apprentissage : vers une modélisation générale". Actes des 3èmes Journées Nationales PRC-GDR Intelligence Artificielle, 5-7 Mars Edition HERMES, Chapitre 5, pp.237-269.
- Ganter, B. & Wille, R. (1997). "Applied Lattice Theory: Formal Concept Analysis". General Lattice Theory, G. Grätzer editor, Birkhäuser. Preprints.
- Gasmi, G., Ben-Yahia, S., Mephu, E. & Slimani, Y. (2005). "IGB: une nouvelle base générique informative des règles d'association". Springer, pp.81-90.
- Gayon A. (2001). "Importance de la sécurité dans les entreprises". Référence AG4600, Date de publication : 10 juil. 2001. [URL : <http://www.techniques-ingenieur.fr/base-documentaire/environnement-securite-th5/management-de-la-securite-42154210/importance-de-la-securite-dans-les-entreprises-ag4600/>].
- Gilles C. (2009). « Cours XML ». [<ftp://ftp-developpez.com/gilles-chagnon/cours/xml/dtd-et-schemas/format-document-xml-dtd-et-schemas.pdf>].
- Goel V., Buchel C., Frith C. & Dolan RJ. (2000). "Dissociation of mechanisms underlying syllogistic reasoning". Neuroimage 2000; 12: 504–14
- Goh, Y. & Chua, D. (2010). "Case-Based Reasoning Approach to Construction Safety Hazard Identification: Adaptation and Utilization". Journal of Construction Engineering and Management, Vol. 136, No. 2, pp.170-178,
- Gómez-Gauchía, H., Díaz-Agudo, B. & González-Calero, P. (2006). "Ontology-Driven Development, of Conversational CBR Systems". Advances in Case-Based Reasoning,



- Proceedings of the 8th European Conference (ECCBR'06), Turkey, Springer, volume 4106/2006, pp.309–324,
- Gruber, T.R. (1993). "*A translation approach to portable ontology specifications*". Knowledge Acquisition, pp.199–220.
- Hadj-M'tir R. (2010). "*Ingénierie Ontologique pour la Réutilisation à Base de Cas dans les Systèmes de E-Learning*". Thèse de doctorat en Informatique, LIRIS, RIADI-ENSI, Tunisie.
- Hadj-Mabrouk, A. & Hadj-Mabrouk, H. (2003). "*Approche d'intégration de l'erreur humaine dans le retour d'expérience, Application au domaine de la sécurité des transports ferroviaires*". Synthèse n° 43, INRETS-IFSTTAR.
- Hadj-Mabrouk, H. & Bied-Charreton, D. (1993). "*Mise à jour de la BCHS : Base de Connaissances Historiques des Scénarios d'accidents potentiels*". Convention Région Nord-Pas de Calais/INRETS-CRESTA/LAIH-UVHC. Dossier technique INRETS-CRESTA, Indice C, CR/A-93-36, Arcueil Mai 1993.
- Hadj-Mabrouk, H. & Darricau, M. (1996). "*SAUTREL : outil d'aide aux analyses des effets des erreurs de logiciels de sécurité dans les transports guidés*". LM 10, 10ième Colloque national de fiabilité et maintenabilité, France, Saint-Malo, tome 2, pp.790-797.
- Hadj-Mabrouk, H. & Triki, I. (2003). "*La réglementation européenne en matière de sécurité ferroviaire*". Revue Générale des Chemins de Fer, Éditions DELVILLE, avril 2003, pp.5-20.
- Hadj-Mabrouk, H. (1992). "*Apprentissage automatique et acquisition des connaissances : deux approches complémentaires pour les systèmes à base de connaissances*". Thèse de doctorat en Automatique Industrielle et Humaine, Université de Valenciennes.
- Hadj-Mabrouk, H. (1994). "*Bilan sur les connaissances acquises pour le développement d'un système d'aide à l'examen de la sécurité des transports guidés*". Résultat des sessions de recueil de connaissances du 15/02/90 au 22/02/94. Rapport de convention Région Nord-Pas de Calais/INRETS-CRESTA/LAIH-UVHC. CR/A-94-28, Arcueil 22 Avril 1994, 34p.
- Hadj-Mabrouk, H. (1995). "*La maîtrise des risques dans le domaine des automatismes des systèmes de transport guidés*". Revue RTS, numéro 49, pp.101-112.
- Hadj-Mabrouk, H. (1998). « *Acquisition et évaluation des connaissances de sécurité des systèmes industriels. Application au domaine de la certification des systèmes de transport guidés* ». Thèse d'Habilitation à Diriger des Recherches, Université de Technologie de Compiègne.
- Hadj-Mabrouk, H. (2007). "*ACASYA et SAUTREL : Deux mécanismes d'apprentissage d'aide au retour d'expérience*". SETIT 2007: 4rth International Conference: Sciences of Electronic Technologies of Information and Telecommunications, March 25-29, TUNISIA.
- Hadj-Mabrouk, H. (2007). "*Ouvrage collectif : Chapitre 4 : Contribution du raisonnement à partir de cas à l'analyse des effets des erreurs du logiciel. Application à la sécurité des transports ferroviaires*". Éditions Hermes - Lavoisier, pp.123-148.
- Hadj-Mabrouk, H. (2010). "*Introduction à la sécurité et à l'analyse des risques technologiques et humains Application à la certification des systèmes de transports ferroviaire*". 3ème Symposium International sur la Maintenance et la Maîtrise des Risques, Rabat, Maroc
- Hadj-Mabrouk, H., Maalel, A. & Dogui, M. (2012). "*Contribution de la chronobiologie de la vigilance à la sécurité des transports*". Proceedings of the Sixth International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, SETIT'12, Sousse, Tunisia
- Hammond, K.J. (1986). "*CHEF: A model of case-based planning*". Proceedings of the 4th European Workshop on Advances in Case-Based Reasoning, Springer-Verlag, London, UK, 1998, pp. 184-195.
- Hanney, K. & Keane, M.T. (1996). « *Learning Adaptation Rules from Cases* », Proceedings of the

- 3rd European Workshop on Case- Based Reasoning (EWCBR'96), tome 1168 de LNAI. Springer.
- Hanney, K. (1997). « *Learning Adaptation Rules from Cases* ». Thèse de maître, Trinity College, Dublin.
- Haouchine, M.K. (2009). « *Remémoration guidée par l'adaptation et maintenance des systèmes de diagnostic industriel par l'approche du raisonnement à partir de cas* ». Thèse en Automatique à l'UFR des Sciences et Techniques de l'Université de Franche-Comté. [URL : <http://www.elanceur.org/Articles/OntologieSurfaite.html>].
- Haouchine, M.K., Chebel-Morello, B. & Zerhouni, N. (2009). "Autoincrement of expertise for failure diagnostic". 13th IFAC Symposium on Information Control Problems in Manufacturing INCOM'09, Moscow, Russie, Juin.
- Hendler, J. & McGuinness, D.L. (2000). "The DARPA Agent Markup Language". IEEE Intelligent Systems- Trends and Controversies", 15(6): pp.67-73.
- Holitiana, R. (2004). "Intégration du Retour d'Expérience dans les processus industriels, Application à Alstom Transport". Thèse réalisée en Convention Industrielle de Formation par la Recherche, Département Qualité, Alstom Transport SA.
- Hollnagel, E. (1999). "Accident and barriers". In 7th European Conference on Cognitive Science Approaches to Process Control, Villeneuve d'Ascq, France, pp.175–180,
- Horrocks, I. & Patel-Schneider, P.F. (2003). "Reducing OWL Entailment to Description Logic Satisfiability". In D. Fensel, K. Sycara et J. Mylopoulos, editors, 2nd International Semantic Web Conference (ISWC'03), volume (2870) of Lecture Notes in Computer Science, pp.17-29. [URL: [http://www.irmagrenoble.com/05documentation/06glossaire\\_entier.php](http://www.irmagrenoble.com/05documentation/06glossaire_entier.php)].
- Iyan, J., Joel, A., Brigitte, C., Sebastien, D. & Rallou, T. (2010). "Making Ontology-Based Knowledge and Decision Trees interact: an Approach to enrich knowledge and increase expert confidence in data-driven models". In proceedings of KSEM Conference.
- Jabrouni, H. (2012). "Exploitation des connaissances issues des processus de retour d'expérience industriels". Thèse de doctorat de l'université de Toulouse, discipline : Systèmes industriels.
- Jiang, J. & Conrath, D. (1997). "Semantic similarity based on corpus statistics and lexical taxonomy". In Proceedings of International Conference on Research in Computational Linguistics, Taiwan.
- Joffrey, C. (2009). "Prise en compte des séquences de défaillances pour la conception de systèmes d'automatisation. Application au ferroutage". Thèse de doctorat en automatique, génie informatique et traitement de signal et images, de l'université des Sciences et des Technologies de Lille.
- Joing M. (1991), "Le retour d'expérience à la SNCF", Colloque : la sécurité des transports collectifs, décembre 1991, Paris, 178-180.
- Joing, M. & Keravel, F. (1993). "Retour d'expérience et analyse du facteur humain". Revue Générale des Chemins de Fer, juin 1993, pp.5-8.
- Kass, G. (1980). "An exploratory technique for investigating large quantities of categorical data". Applied Statistics, 29(2), pp.119-127.
- Kassel, G. (2002). "OntoSpec : une méthode de spécification semi-informelle d'ontologies". Dans Actes des journées francophones d'Ingénierie des Connaissances (IC'02), pp.75–87.
- Kassel, G., & Perpette, S. (1999). "Co-operative ontology construction needs to carefully articulate terms, notions and objects". In Proceedings of the International Workshop on Ontological Engineering on the Global Information Infrastructure, Dagstuhl, Castle, Germany, pp.57–70.
- Khatraoui, M., Bousbia, N. & Balla, A. (2008). "Détection de la similarité sémantique entre pages

- visitées durant une session d'apprentissage". 8èmes Journées Francophones Extraction et Gestion des Connaissances, Sophia Antipolis.
- Kifer, M., Lausen, G. & Wu, J. (1995). "Logical Foundations of Object-Oriented and Frame-Based Languages". Journal of the Association for Computing Machinery.
- Kinley, A. (2001). « *Learning to improve case adaptation* ». Thèse de doctorat, Computer Science Department - Indiana university.
- Kodratoff, Y. (1986). "Leçons d'apprentissage symbolique automatique", Cepadues – Édition. Collection techniques avancées de l'informatique, 191p.
- Kolli, M. (2011). « *Proposition d'un système de détection des relations sémantiques dans les ontologies basée sur le Raisonnement à partir de cas* ». Thèse de doctorat En Sciences En Informatique à l'Université Mentouri, Constantine.
- Kolodner, J. (1992). "An Introduction to Case-Based Reasoning". Artificial Intelligence Review, 6(1): pp.3–34.
- Kowalski, M. (2012). "Applying of an Ontology-driven Case-based Reasoning System in Logistics". International Journal of Computers & Technology, Volume 3 No. 2, Oct.
- Krishnakumar, R., Prezelin, N. & Razafindrakoto, H. (2005). "Les outils de sûreté de fonctionnement". DESS Sécurité des transports, Université de Versailles Saint-Quentin en Yvelines.
- Lallich, S. & Teytaud, O. (2004). « *Évaluation et validation de l'intérêt des règles d'association* », revue RNTI. [URL :<http://www.techno-science.net/?onglet=ouvrages&ID=2854286464>, 2004.<http://perso.wanadoo.fr/olivier.teytaud/publis/evaluationEtValidationDeLInteretDesReglesDAssociation.pdf>]
- Lamontagne, L. (2001). "Raisonnement à base de cas textuel pour la réponse automatique au courrier électronique", Proposition de projet doctoral, pp.13-14. [URL : <http://rali.iro.umontreal.ca/rali/sites/default/files/publis/0predocLamontagne.pdf>].
- Lamontagne, L. (2004). "Une approche CBR textuel de réponse au courrier électronique". Thèse de doctorat, Département d'informatique et de recherche opérationnelle, Université de Montréal.
- Laprie J.C. (1995b). "LIS, Guide de la sûreté de fonctionnement". Toulouse, mai 1995, 2e éd., 369 p. (ISBN 978-2-85428-382-2).
- Laprie, J.C., J., Arlat, J.P. Blanquart, A. Costes, Y. Crouzet, Deswarte Y., J.C. Fabre, H. Guillermain, K. Kanoun, C. Mazet, D. Powell, C. Rabéj Et P, (1995), "Guide de la sûreté de fonctionnement", Laboratoire d'Ingénierie de la Sûreté de fonctionnement, Toulouse, France, Cépaduès-Editions, ISBN 2-85428- 382-1.
- Leacock, C. & Chodorow, M. (1998). "Combining Local Context and WordNet Similarity for Word Sense Identification". In WordNet: An Electronic Lexical Database, C. Fellbaum, MIT Press.
- Leake, D.B. & Powell, J.H. (2007). "Mining large-scale knowledge sources for case adaptation knowledge". Case-Based Reasoning Research and Development, 7th International Conference on Case-Based Reasoning, (ICCBR'07), pp.209–223.
- Leake, D.B. & Powell, J.H. (2008). "Knowledge planning and learned personalization for web-based case adaptation". Advances in Case-Based Reasoning, 9th European Conference, ECCBR 2008, Trier, Germany, September 1-4, 2008. Proceedings, pp.284–298.
- Leake, D.B. & Powell, J.H. (2008). "Knowledge planning and learned personalization for web-based case adaptation". Proceedings of Advances in Case-Based Reasoning, 9th European Conference, ECCBR 2008, Trier, Germany, September, pp.1-4.
- Leake, D.B., Kinley, A. & Wilson, D. (1996). "Acquiring Case Adaptation Knowledge: A Hybrid

- Approach*”. in AAAI/IAAI, tome 1.
- Lenz, M. (1999). “*Case Retrieval Nets as a Model for Building Flexible Information Systems*”. Thèse de doctorat de l’université de Humboldt.
- Leonid, L., Chepelev, Dana, K., & Michel, D. (2011). “*Hazard Estimation and Method Comparison with OWL-Encoded Toxicity Decision Trees*”. Eighth International Workshop San Francisco, California, USA June 5-6.
- Levenshtein, V. (1966). « *Binary codes capable of correcting deletion, insertions, and reversals* ». In Soviet Physics Doklady, 1966, vol. 10, n° 8, pp.707-710.
- Lieber, J. (1999). “*Reformulations and Adaptation Decomposition: Formalisation of Adaptation in Case-Based Reasoning*”. Third International Conference on Case-Based Reasoning Workshop, ICCBR-99 Workshop number 3.
- Lieber, J. M., d’Aquin, S., Brachais & Napoli, A. (2004). « *Une étude comparative de quelques travaux sur l’acquisition de connaissances d’adaptation en raisonnement à partir de cas* ». Dans Actes du 12me atelier de raisonnement à partir de cas (RàPC’04).
- Lieber, J.M., d’Aquin, S., Badra, F. & Napoli, A. (2008). “*Modeling adaptation of breast cancer treatment decision protocols in the kasimir project*”. Applied Intelligence, 28(3) :pp.261–274.
- Lin, D. & Kedem, Z. (1998). “*Princer Search: A new algorithm for discovering the maximum frequent set*”. In proceedings of the 6th Biennial international conference on Extending Database Technology, Springer Verlag, March. LNCS, vol. 1377, pp.105-119.
- Lin, D. (1998). “*An Information-Theoretic Definition of Similarity*”. Dans Proceedings of the 15th International Conference on Machine Learning, Morgan Kaufman Publishers Inc, pp. 296–304.
- Lin, Y.C. (2010). “*Design and Implementation of an Ontology-Based Psychiatric Disorder Detection System*”. Wseas Transactions on Information Science And Applications.
- Maalel, A. & Hadj-Mabrouk, H. (2010). “*Contribution of Case Based Reasoning (CBR) in the Exploitation of Return of Experience. Application to Accident Scenarii in Rail Transport*”. Proceedings of the 3rd International Conference on Information Systems and Economic Intelligence, SIIE’2010, Sousse, Tunisia, pp.99-109.
- Maalel, A. (2009). “*Apport de l’intelligence artificielle au retour d’expérience. Application à la sécurité des transports ferroviaires*”. Mémoire de mastère de recherche en sciences de transports et de la logistique, ISTLS, université de Sousse.
- Maalel, A. (2010). “*Étude du projet ACASYA*”. Rapport interne RII-2010, laboratoire RIADI, ENSI.
- Maalel, A., Hadj-Mabrouk, H., Mejri, L. & Ben Ghézala, H. (2012a). “*Toward a Knowledge Management Approach Based on an Ontology and Case-based Reasoning (CBR)*”. In the proceedings of Sixth International Conference on Research Challenges in Information Science, May 16-18, Valencia, Spain.
- Maalel, A., Hadj-Mabrouk, H., Mejri, L. & Ben Ghézala, H. (2011). “*Development of Ontology to Assist the Modeling of an Accident Scenario: Application on Railroad Transport*”. Journal of Computing, Volume 3, Issue 7, July.
- Maalel, A., Hadj-Mabrouk, H., Mejri, L. & Ben Ghézala, H. (2012b). “*Towards an Ontology of Help to the Modeling of Accident Scenarii: Application on Railroad Transport*”. Proceedings of the Sixth International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, SETIT’12, Sousse, Tunisia.
- Maalel, A., Hadj-Mabrouk, H., Mejri, L. & Ben Ghézala, H. (2012c). “*Towards a Case-Based Reasoning Approach Based On Ontologies. Application to Railroad Accidents*”. Proceedings of the International Conference on Data and Knowledge Engineering, ICDKE12, Wu Yi Shan, Fujian, China, Springer, Lecture Notes in Computer

- Science, pp.48-55.
- Maalel, A., Mhiri, W. & Hadj-Mabrouk, H. (2011). "*The Contribution of Machine Learning to Analyze and Evaluate the Safety of Automated Transport Systems*". Proceedings of The IEEE International Conference on Communications, Computing and Control Applications, IEEE CCCA'2011, Hamammet.
- Madhavan, J., Berntein, P. & Rahm, E. (2001). "*Generic schema matching with Cupid*". In Proceedings of the 27th International Conference on Very Large Data Base (VLDB)), Roma, Italy, pp.48-58.
- Maedche, A., Motik, B., Stojanovic, L., Studer, R. & Volz, R. (2003). "*Ontologies for Enterprise Knowledge Management*". IEEE Intelligent Systems, 18(2): pp.26-33.
- Malek, M. (2008). "*Chapitre 1.Fouille de données : Approche de règles d'association*". EISTI, Deuxième année.
- Manola, F. & Miller, E. (2004). "*RDF Primer*". W3C Recommendation 10 February.
- Mazouni, M.H. (2008). "*Pour une Meilleure Approche du Management des risques : de la Modélisation Ontologique du Processus Accidentel Au système Interactif d'Aide à la Décision*". Thèse de Doctorat à l'Institut National Polytechnique de Lorraine, Novembre 2008.
- Mbaye, S. (2010). "*Le retour d'expérience, analyse comparée des pratiques de REX dans les industries chimiques et nucléaires*". Les cahiers de la sécurité industrielle, 5/2010.
- Mejri, L. (1995). « *Une démarche basée sur l'apprentissage automatique pour l'aide à l'évaluation et à la génération de scénarios d'accidents. Application à l'analyse de sécurité des systèmes de transport automatisés* ». Thèse de doctorat de l'université de valenciennes.
- Mejri, L., Hadj-Mabrouk, H. & Caulier, P. (2009). "*Un modèle générique unifié de représentation et de résolution de problèmes pour la réutilisation des connaissances de sécurité*". Revue Recherche Transports Sécurité, n°103, Éditions LAVOISIER, 2009, pp.131-148.
- Mendes, O. (2003). "*État de l'art sur les méthodologies d'ingénierie ontologique*". Montréal, Québec, Canada: Centre de recherche LICEF.
- Messai, N. (2009) ; "*Analyse de concepts formels guidée par des connaissances de domaine : Application µa la découverte de ressources génomiques sur le Web*". Thèse de doctorat de l'université Henri Poincarée, Nancy 1.
- Mille, A. (1999). "*Tutorial CBR : Etat de l'art de raisonnement à partir de cas*". Plateforme AFIA'99, Palaiseau.
- Mille, A. (2006). "*Traces based reasoning (TBR) definition, illustration and echoes with storytelling*". Rapport Technique RR-LIRIS-2006-002, LIRIS UMR 5205 CNRS/INSA de Lyon/Université Claude Bernard Lyon 1/Université Lumière Lyon 2/Ecole Centrale de Lyon, January.
- Miller, G.A. (1990). "*Nouns in WordNet: a lexical inheritance system*". In International Journal of Lexicography, Vol. 3, N°4 pp.245–264.
- Minsky, M. (1975). "*A Framework for Representing Knowledge*". In Edited by P. Winston. The Psychology of Computer Vision. New-York: Mc-Graw-Hill, pp.211-281.
- Monge, A., Elkan, C. (1996). « *The field-matching problem: algorithm and applications* ». In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, pp.267-270.
- Nagypál, G. (2005). "*Methodology for building SWS ontologies in DIP*". Data, Information and Process Integration with Semantic Web Services, FP6 – 507483.
- Neeches, R., Fikes, R.E., Finin, T., Gruber, T.R., Senator, T. & Swartout, W.R. (1991). "Enabling technology for knowledge sharing". AI Magazine. 12: pp.36- 56.

- Noy, N.F. & McGuinness, D.L. (2001). "*Ontology Development 101: A Guide to Creating Your First Ontology*". Technical Report KSL-01-05 and SMI-2001-0880, Stanford Knowledge Systems Laboratory and Stanford Medical Informatics.
- Oberle D., (2006) "Ontologies. Dans Semantic Management of Middleware", volume 1 de Semantic Web and Beyond, Springer US, pp.33–53.
- Pasquier N, (2000), "Data Mining. Algorithmes d'extraction et de reduction des règles d'association dans les bases de données ", Thèse de doctorat en informatique.
- Pasquier, N., Bastide, Y., Taouil, R. & Lakhal, L. (1999). "*Discovering frequent closed itemsets for association rules*". In Proceedings of the 7th biennial international conference on database theory, Springer-verlag, January, pp.398-416.
- Patwardham, S. (2003). "*Incorporating Dictionary and Corpus Information in a Measure of Semantic Relatedness*". M.S. Thesis, August.
- Perez, A. G. & Benjamins V. R. (1999). "*Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods*". In Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (Stockholm, Sweden, 1999).
- Perez, A. G., Fernandez-Lopez, M. & Corcho, O. (2003). "*Ontological Engineering*". Springer, Advanced Information and Knowledge Processing.
- Psyché, V. (2007). « *Rôle des ontologies en ingénierie des EIAH : Cas d'un système d'assistance au design pédagogique* ». Thèse de doctorat, Université du Québec, Canada.
- Psyché, V., Mendes, O. & Bourdeau, J. (2003). "*Apport de l'ingénierie ontologique aux environnements de formation à distance STICEF/Hors-série*". Technologies et formation à distance.
- Purvis, L. & Pu, P. (1998). « *Composer: A case-based reasoning system for engineering design. Robotica* ». 16(3): pp.285–295. [URL:<http://dx.doi.org/10.1017/S0263574798000368>].
- Quinlan, R. (1968). « *Semantic Memory* ». In Edited by M. Minsky. Semantic Information Processing, Cambridge: MIT Press, pp.227-270.
- Quinlan, R. (1979). "*Discovering rules by induction from large collections of examples*". D. Michie ed., Expert Systems in the Microelectronic age, pp.168-201.
- Quinlan, R. (1993). "*C4.5: Programs for Machine Learning*", Morgan Kaufman Publishers.
- Rada, R., Mili, H., Bicknell, E. & Blettner, M. (1989). "*Development and application of a metric on semantic nets*". IEEE Transactions on Systems, Man and Cybernetics, 19(1): pp.17–30.
- Rakotomalala, R. (2005). "*Arbres de décision* ", in Revue Modulad, n°33.
- Rakotomalala, R., Lallich, S. & Palma, D. (1999). "*Studying the behavior of generalized entropy in induction trees using a m-of-n concept*". Proceedings of 3rd European Conf. on KDD, pp. 510-517.
- Recio-García, J.A., Díaz-Agudo, B., González-Calero, P. & Sanchez, A. (2006). "*Ontology based CBR with jCOLIBRI*". Applications and Innovations in Intelligent Systems, 14: pp.149–162.
- Renaud, J., Morello, B.C., Fuchs, B. & Lieber, J. (2007). "*Raisonnement à Partir de Cas I : conception et configuration de produits*". Hermes-Lavoisier, February.
- Research teams from the FonCSI programme Operational Experience Feedback. (2008). "*Sociocultural Success Factors for Operational Experience Feedback, Seven Field Studies Keywords industrial safety, operational experience feedback, learning from experience, lessons learned*". September.
- Resnik, P. (1999). "*Semantic similarity in a taxonomy: an information based measure and its application to problems of ambiguity in natural language*". Journal of Artificial Intelligence

- Research 11, pp.95–130.
- Ricco, R. (2005). *"Arbres de Décision"*. Revue MODULAD.
- Riesbeck, C.K. & Schank, R.C. (1989). "Inside Case-Based Reasoning", Lawrence Erlbaum Associates.
- Rifqi, M. (1996). *"Mesures de comparaison, typicalité et classification d'Objets flous : théorie et pratique"*. Thèse d'université, Université Pierre et Marie Curie, Paris VI, Paris.
- Sabou, M., Lopez, V. & Motta, E. (2006). « *Ontology Selection on the Real Semantic Web: How to Cover the Queens Birthday Dinner?* ». In Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management.
- Safar, S., Reynaud, C. & Calvier, F. (2007). « *Techniques d'alignement d'ontologies basées sur la structure d'une ressource complémentaire* ». Actes des 1ères Journées Francophones sur les Ontologies, 2007, Sousse, Tunisie, pp.21-35.
- Savasere, A., Omiecinski, E. & Navathe, S. (1995). *"An efficient algorithm for mining association rules in large databases"*. in Proceedings of the 21st international conference on Very Large Data Bases (VLB'95), 4326444, Morgan Kaufmann.
- Schank, R.C. & Abelson, R.P. (1977). *"Scripts, plans, goals and understanding : An inquiry into human knowledge structures"*. Lawrence Erlbaum Associates Hillsdale, NJ, Dynamic memory. Cambridge University.
- Schank, R.C. (1986). *"Explanation patterns : Understanding mechanically and creatively"*. Lawrence Erlbaum Associates.
- Scharff, C. (2004). *"Introduction à la fouille de données"*. IFI. Juin [URL :<http://www.csis.pace.edu/~scharff/DMIFI/fouille2.ppt>]
- Schmidt, R. & Vorobieva, O. (2005). *"Adaptation and medical case-based reasoning focusing on endocrine therapy support"*. Artificial Intelligence in Medicine, 10th Conference on Artificial Intelligence in Medicine (AIME'05).
- Sebag, M., Schoenauer, M. (1990). *"Apprentissage de règles par découverte"*. 5èmes Journées Françaises de l'Apprentissage : JFA 90, Lannion, pp.213-231
- Smyth, B. & Keane, M.T. (1993). *"Retrieving adaptable cases: The role of adaptation knowledge in case retrieval"*. In EWCBR, tome 837 de Lecture Notes in Computer Science, Springer, 1993. ISBN 3-540-58330-0, pp.209–220.
- Smyth, B. & McKenna, E. (2001). *"Competence guided incremental footprint-based retrieval. Knowledge-Based Systems"*, 14(3-4), pp.155-161.
- Smyth, F.C., Lottaz, L. & Faltings, B. (1995). *"Spatial composition using cases: Idiom. Dans Case-Based Reasoning Research and Development"*. Proceedings of First International Conference, ICCBR-95, Sesimbra, Portugal, October 23-26.
- Sowa, J. (1984). *"Conceptual Structures: information processing in mind and machine"*. Reading: Addison Wesley publishing Company, 481p.
- Sowa, J.F. (1999). *"Knowledge Representation: Logical, Philosophical, and Computational Foundations"*. Brooks Cole Publishing Co., Pacific Grove, CA.
- Stahl, A. & Bergmann, R. (2000). *"Applying recursive cbr for the customization of structured products in an electronic shop"*. In proceedings of Advances in Case-Based Reasoning, 5th European Workshop, EWCBR 2000, Trento, Italy, September 6-9.
- Studer, R., Benjamins, V.R. & Fensel, D. (1998). *"Knowledge engineering: Principles and methods"*. Data Knowledge Engineering. vol. 25, no 1-2, pp.161-197.
- Swartout, B., Patil, R., Knight, K. & Russ, T. (1997). *"Towards Distributed Use of Large Scale Ontologies"*. Spring Symposium Series on Ontological Engineering, Stanford University,

- CA. pp.138-138.
- Sylvie, J., Tao-Yan, J., Dominique, L., George, L. & Oumar, S. (2005). "*Extraction de règles d'association pour la prédiction de valeurs manquantes*". Novembre 2005, Numéro spécial CARI'04 Revue ARIMA, pp.103-124.
- Taoivonen, H. (1996). "*Discovery of frequent patterns in large data collection*". PhD thesis, university of Helsinki.
- Tea, C. (2009). "*Retour d'expérience et données subjectives : quel système d'information pour la gestion des risques ? Le cas de la sécurité dans le transport ferroviaire*". Thèse de doctorat de l'École Nationale Supérieure d'Arts et Métiers Spécialité "Sciences de gestion".
- Tittel, L. (1992). "*LZ 129 Hindenburg. Schriften zur Geschichte der Zeppelin Luftschiffahrt*". N5.
- Uschold, M. & Grüninger, M. (1996), "*Ontologies: principles, methods, and applications*", *Knowledge Engineering Review*, 11(2): pp.93–155.
- Valmi, D.L. (2010). "*Améliorer la mémorisation par enrichissement de l'ontologie du domaine*". 18ème atelier de Raisonement à Partir de Cas - RàPC2010, Strasbourg : France.
- W3C OWL Working Group. (2012). "*OWL 2 Web Ontology Language*". Document Overview (Second Edition) W3C Recommendation 11 December.
- Wanner, J.C. (2000). "*La recherche désespérée du risque nul*". Annales des mines, juillet 2000, pp.49-58.
- Weiss, S.M. & Kulikowski, C.A. (1991). "*Computer systems that learn. Artificial intelligence*". Morgan-Kaufmann, vol. 62, no2, pp.363-378.
- Wilke, W. & Bergmann, R. (1998). "*Techniques and knowledge used for adaptation during case-based problem solving*". Lecture notes in computer science, pp.497–506.
- Wille, R. (1982). "*Restructuring lattice theory: an approach based on hierarchies of concepts*", Rival, i. (ed.) : Ordered sets, pp.445-470.
- Wiratunga, N. & Craw. (2002). "*Rowe. Learning to adapt for case-based design*". 6th European Conference on Advances in Case-Based Reasoning, ECCBR 2002 Aberdeen, Scotland, pp.421–435.
- Wu, Z. & Palmer, M. (1994). "*Verbs Semantics and Lexical Selection*". In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, pp.133–138, New Mexico, USA.
- Wybo, J.L. (2012). "*Le retour d'expérience : un processus d'acquisition de connaissances et d'apprentissage*". HAL Id : hal-00614238 [URL : <https://hal-mines-paristech.archives-ouvertes.fr/hal-00614238>].
- Yanping, C., Zhenmin, T. & Haibing D. (2005). "*Case-Based Reasoning and Rule-Based Reasoning for Railway Incidents Prevention*". Proceedings OF Services Systems and Services Management.
- Zaki, M.J., Parthasarath, H., Ogihara, M. & Li, W. (1997). "*New algorithms for discovery of association rules*". In proceedings of the the 3d international conference on knowledge discovery and data mining, pp.283-286.
- Zargayouna, H. & Salotti, S. (2004). "*Mesure de similarité dans une ontologie pour l'indexation sémantique de documents XML*". Manuscrit auteur, publié dans "Ingénierie des Connaissances, Lyon : France.
- Zghal, S. (2010). "*Contributions à l'alignement d'ontologies OWL par agrégation de similarités*". Thèse de doctorat en informatique, Faculté des Sciences de Tunis & CRIL-CNRS, Lens, 21 décembre 2010.
- Zhang, J. & Vasant, H. (2003). "*Learning Decision Tree Classifiers from Attribute Value*



- Taxonomies and Partially Specified Data*". Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC.
- Zhang, J., Adrian, S. & Vasant, H. (2002). "*Ontology-Driven Induction of Decision Trees at Multiple Levels of Abstraction*". Springer-Verlag Berlin Heidelberg, LNAI 2371, pp.316-323.
- Zhang, K., Tang, J., Hong, M., Li, J. & Wei, W. (2006), "*Weighted Ontology-Based Search Exploiting Semantic Similarity*". Lecture Notes In Computer Science.
- Ziani, M., Boulanger, D. & Talens, G. (2010). "*Système d'aide à l'alignement d'ontologies métier. Application au domaine géotechnique*". Actes du XXVIII<sup>e</sup> congrès INFORSID, Marseille, Mai 2010.
- Zighed, D. & Ricco R. (1996). "*Sipina-W for Windows: User's Guide*". Laboratory ERIC, University of Lyon 2.
- Zubair, M. & Khan, M.J. (2012). "*Prediction and Analysis of Air Incidents and Accidents Using Case-Based Reasoning*". (GCIS), Third Global Congress on, pp.315 – 318.

# Publications

Les publications qui ont trait à notre travail de thèse jusqu'à aujourd'hui, sont présentées dans la liste ci-dessous. Une mise à jour complète de cette liste est également disponible à l'adresse suivante :

<http://sites.google.com/site/maalelahmed/publications>

## Article revue

1. Maalel A., Mejri, L. & Ben Ghézala H. (in press). "Adast: A Decision Support Approach Based on an Ontology and Case-based Reasoning (CBR). Application to Railroad Accidents". **International Journal of Information and Decision Sciences, IJIDS, Inderscience publishers**. *Scopus, DBLP*

## Articles dans des conférences (indexées et/ou classées)

1. Maalel A., Mejri, L., Hadj-Mabrouk H. & Ben Ghézala H. (2014). "Towards a Decision Support System for Security Analysis Application to Railroad Accidents", **In Information Systems for Crisis Response and Management in Mediterranean Countries**, Springer International Publishing, Lecture Notes in Business Information Processing, LNBIP, Volume 196, pp. 46–56. *Springer, DBLP*
2. Maalel A., Mejri, L., Hadj-Mabrouk H. & Ben Ghézala H. (2012). "Towards a Case-Based Reasoning Approach Based On Ontologies. Application to Railroad Accidents". **Proceedings of the International Conference on Data and Knowledge Engineering, ICDKE12, Springer, LNCS 7696**, pp. 48–55. *Springer, DBLP*
3. Maalel A., Mejri, L., Hadj-Mabrouk H. & Ben Ghézala H. (2012). "Toward a Knowledge Management Approach Based on an Ontology and Case-based Reasoning (CBR)". **Proceedings of the Sixth IEEE International Conference on Research Challenges in Information Science, IEEE RCIS12, Valencia, Spain**. *Classe B, IEEE Xplore. DBLP*
4. Maalel A., Mejri, L., Hadj-Mabrouk H. & Ben Ghézala H. (2012). "Towards an Ontology of Help to the Modeling of Accident Scenarii: Application on Railroad Transport". **Proceedings of the Sixth International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, SETIT'12, Sousse, Tunisia**. *IEEE Xplore*.
5. Mejri, L., Maalel A. & Ben Ghézala H. (2012). "Simulation de cas d'accidents pour l'aide à la décision. Application à la sécurité des systèmes de transport ferroviaires automatisés". Atelier aide à la Décision à tous les Etages, **AIDE@EGC'2012, Acte de la 12ème Conférence internationale francophone sur l'extraction et la gestion des connaissances**, Bordeaux.
6. Mejri, L., Maalel A., Hadj-Mabrouk H. & Ben Ghézala H. (2011). "Construction d'une ontologie d'aide au renforcement de la sécurité des systèmes de transport automatisés". **Acte de la 11ème Conférence internationale francophone sur l'extraction et la gestion des connaissances, EGC'2011**. *DBLP*

# Annexe1

## Projet ACASYA

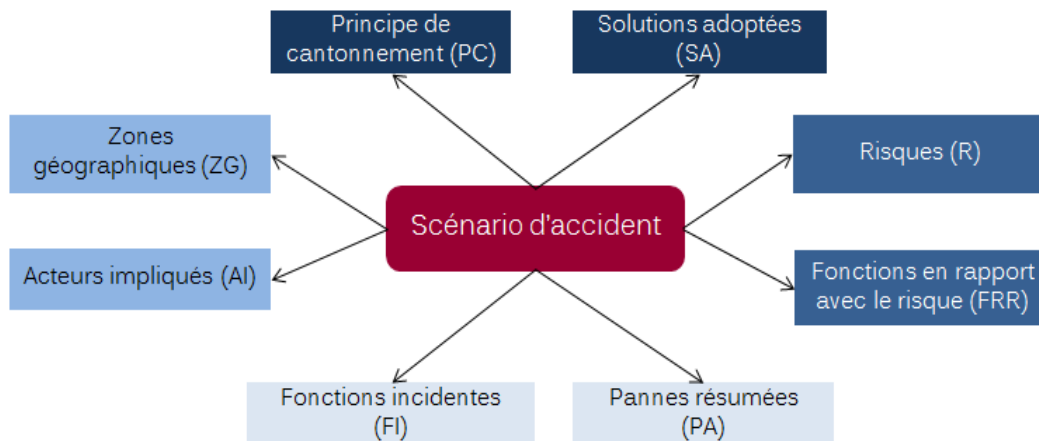
Le projet ACASYA d'Aide à la Certification par Apprentissage des SYstèmes de transport Automatisés (Hadj-Mabrouk H, 1992), (Mejri, 1995), (Hadj-Mabrouk H, 1997), (Hadj-Mabrouk H, 2007) et (Mejri, 2009) est considéré comme projet initiateur reposant sur l'utilisation conjointe des techniques d'acquisition de connaissances et d'apprentissage automatique dans le domaine de la sécurité des systèmes de transports ferroviaires. ACASYA est un système destiné à coopérer avec un expert pour l'assister dans sa décision.

### Modélisation des connaissances dans ACASYA

ACASYA repose sur l'utilisation d'une modélisation statique et dynamique d'un scénario d'accident.

#### Modélisation statique

Selon la description statique, un scénario d'accident est caractérisé par un contexte et un ensemble d'événements et de paramètres. La phase de Micro-extraction des connaissances a débouché sur la mise en forme d'un modèle basé notamment sur l'identification des huit paramètres décrivant un scénario d'accident (figure A).



**Figure A.** Paramètres descriptifs d'un scénario d'accident

La description statique recense les principaux paramètres d'un scénario d'accident, dans une "fiche descriptive" élaborée sous forme d'un tableau de données (Attributs/Valeurs) (figure B). Les attributs correspondent aux huit paramètres et à chaque attribut est associée une liste des valeurs possibles. Cette fiche a par la suite été utilisée comme formulaire de base pour l'acquisition des scénarios ; elle se veut exhaustive et permet d'aborder l'étude de tout type de système de transport automatisé (VAL, MAGGALY, POMA) (Hadj-Mabrouk H, 1992).

|                   | LISTE DES ATTRIBUTS                         | LISTE DES VALEURS POSSIBLES        |                                                      | Valeurs choisies (X) | concepts(*) |
|-------------------|---------------------------------------------|------------------------------------|------------------------------------------------------|----------------------|-------------|
| 1                 | PRINCIPE DE CANTONNEMENT                    | Canton fixe                        |                                                      | X                    |             |
|                   |                                             | Canton mobile                      |                                                      |                      |             |
| 2                 | RISQUES                                     | Collision                          |                                                      | X                    | *           |
|                   |                                             | Déraillement                       |                                                      |                      |             |
|                   |                                             | Evacuation d'urgence mal contrôlée |                                                      |                      |             |
|                   |                                             | Chute dans le véhicule             |                                                      |                      |             |
|                   |                                             | Chute sur la voie                  |                                                      |                      |             |
|                   |                                             | Entraînement de l'individu         |                                                      |                      |             |
|                   |                                             | Electrocution                      |                                                      |                      |             |
|                   |                                             | Heurt à la fermeture des portes    |                                                      |                      |             |
| 3                 | FONCTIONS EN RAPPORT A VEC LE RISQUE        | Gestion de conduite automatique    |                                                      |                      |             |
|                   |                                             | Localisation des trains            |                                                      | X                    |             |
|                   |                                             | Contrôle d'entrée/sortie           |                                                      |                      |             |
|                   |                                             | Suivi des trains                   |                                                      |                      |             |
|                   |                                             | Gestion de sens de marche          |                                                      | X                    |             |
|                   |                                             | Consigne de vitesse                |                                                      |                      |             |
|                   |                                             | Gestion de l'arrêt des trains      |                                                      | X                    |             |
|                   |                                             | Sécurité Quai-Voie                 |                                                      |                      |             |
|                   |                                             | Autorisation CI/HT                 |                                                      |                      |             |
|                   |                                             | Commutation de redondance          |                                                      | X                    | *           |
|                   |                                             | Initialisation                     |                                                      |                      |             |
|                   |                                             | Conduite manuelle                  |                                                      |                      |             |
|                   |                                             | Gestion des alarmes                |                                                      | X                    |             |
|                   |                                             | Evacuation                         |                                                      |                      |             |
|                   |                                             | Accostage                          |                                                      |                      |             |
|                   |                                             | Protection des itinéraires         |                                                      |                      |             |
| Traction/Freinage |                                             |                                    |                                                      |                      |             |
| 4                 | ZONES GEOGRAPHIQUES SPECIFIQUES AU SCENARIO | Terminus                           |                                                      | X                    | *           |
|                   |                                             | Station                            |                                                      |                      |             |
|                   |                                             | Ligne                              |                                                      |                      |             |
|                   |                                             | Zone d'injection de rame           |                                                      |                      |             |
|                   |                                             | Limite de tronçon                  |                                                      |                      |             |
| 5                 | ACTEURS IMPLIQUES DANS LE SCENARIO          | Nombre de rames                    |                                                      | 2                    | *           |
|                   |                                             | Opérateur au PCC                   |                                                      | X                    | *           |
|                   |                                             | Opérateur itinérant                |                                                      |                      |             |
|                   |                                             | PA avec redondance                 |                                                      | X                    | *           |
|                   |                                             | PA sans redondance                 |                                                      |                      |             |
| 6                 | FONCTIONS INCIDENTES                        | Gestion des itinéraires            |                                                      | X                    | *           |
|                   |                                             | Régulation de trafic               |                                                      |                      |             |
|                   |                                             | Consignes (consistance, vigilance) |                                                      | X                    | *           |
|                   |                                             | Communication (transmission)       |                                                      |                      |             |
| 7                 | PANNES RESUMÉES ENGENDREES PAR LE SCENARIO  | PR7                                | Maintien d'un Itinéraire à tort par le PA en attente | X                    | *           |
|                   |                                             | PR8                                | Commutation à tort du PA fixe par le PCC             | X                    | *           |
|                   |                                             |                                    |                                                      |                      |             |
| 8                 | SOLUTIONS PROVISOIRES                       | SP                                 | Isoler un PA                                         | X                    |             |
|                   | SOLUTIONS OPTIMALES                         | SO                                 | Faire intervenir la position d'aiguillage            | X                    |             |

Figure B. Fiche descriptive des paramètres d'un scénario d'accident (Hadj-Mabrouk A., 1992)

## Annexe 2

### Génération des règles d'adaptation

A partir du treillis de concept généré (figure F), nous pouvons illustrer donc les concepts suivants :

$$\mathbb{C}_1 = (\{cas_1, cas_{47}\}, \{Terminus\})$$

$$\mathbb{C}_2 = (\{cas_{46}, cas_{47}\}, \{PR15, PR36, ST, C.ES, GCA\})$$

$$\mathbb{C}_3 = (\{cas_{01}, cas_{46}\}, \{CR\})$$

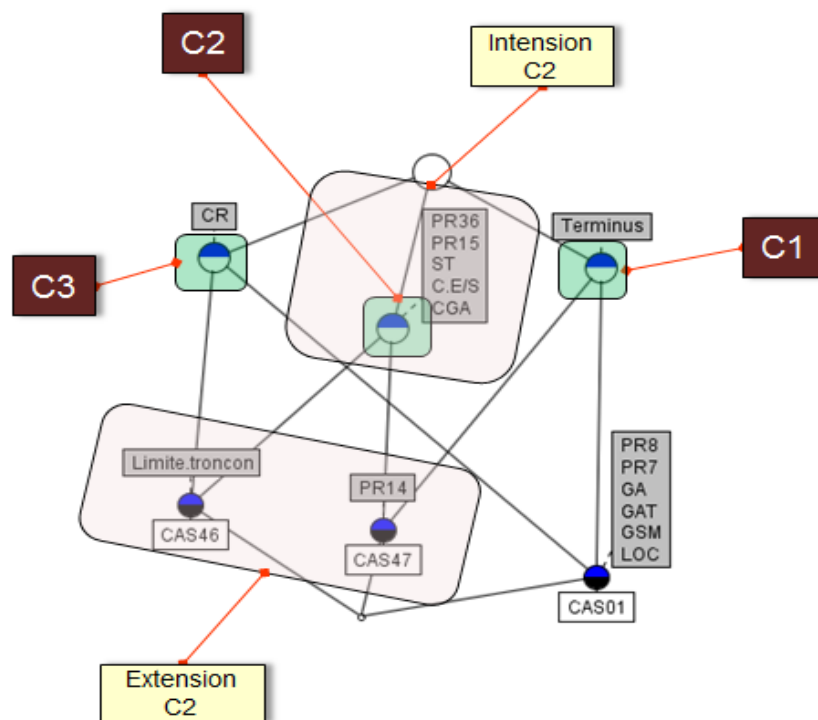


Figure C. Identification de l'extension et l'intension

Selon l'approche de de l'analyse de concept logique, l'**intension** du concept  $int(\mathbb{C})$  est une formule propositionnelle qui correspond à la disjonction des parties problème des cas.

Nous écrivons les **intensions** du cas 47 ( $\mathbb{C}_1$  et  $\mathbb{C}_2$ ) en forme normale conjonctive (des conjonctions de disjonctions) :

Nous avons :

$$int(\mathbb{C}_1) = cas_{01} \vee cas_{47}$$

$$= (LOC_{01} \wedge GSM_{01} \wedge GAT_{01} \wedge GA_{01} \wedge PR7_{01} \wedge PR8_{01} \wedge Terminus_{01}) \vee (PR14_{47} \wedge Terminus_{47})$$

En se basant sur distributivité de « ou » par rapport à « et » :

$$(P \vee (Q \wedge R)) \Leftrightarrow ((P \vee Q) \wedge (P \vee R))$$

Nous avons :

$$\mathit{int}(C1) = \mathit{cas\_01} \vee \mathit{cas\_47}$$

$$= ( \text{LOC\_01} \vee \text{PR14\_47} ) \wedge ( \text{LOC\_01} \vee \text{Terminus\_47} ) \wedge ( \text{GSM\_01} \vee \text{PR14\_47} ) \wedge ( \text{GSM\_01} \vee \text{Terminus\_47} ) \wedge ( \text{GAT\_01} \vee \text{PR14\_47} ) \wedge ( \text{GAT\_01} \vee \text{Terminus\_47} ) \wedge ( \text{GA\_01} \vee \text{PR14\_47} ) \wedge ( \text{GA\_01} \vee \text{Terminus\_47} ) \wedge ( \text{PR7\_01} \vee \text{PR14\_47} ) \wedge ( \text{PR7\_01} \vee \text{Terminus\_47} ) \wedge ( \text{PR8\_01} \vee \text{PR14\_47} ) \wedge ( \text{PR8\_01} \vee \text{Terminus\_47} ) \wedge ( \text{Terminus\_01} \vee \text{PR14\_47} ) \wedge ( \text{Terminus\_01} \vee \text{Terminus\_47} )$$

Nous éliminons toutes les valeurs qui ne représentent pas le même attribut/concept selon la première règle de l'ACL (voir section 4.5.3.2) :

$$= ( \text{LOC\_01} \vee \text{PR14\_47} ) \wedge ( \text{LOC\_01} \vee \text{Terminus\_47} ) \wedge ( \text{GSM\_01} \vee \text{PR14\_47} ) \wedge ( \text{GSM\_01} \vee \text{Terminus\_47} ) \wedge ( \text{GAT\_01} \vee \text{PR14\_47} ) \wedge ( \text{GAT\_01} \vee \text{Terminus\_47} ) \wedge ( \text{GA\_01} \vee \text{PR14\_47} ) \wedge ( \text{GA\_01} \vee \text{Terminus\_47} ) \wedge ( \text{PR7\_01} \vee \text{PR14\_47} ) \wedge ( \text{PR7\_01} \vee \text{Terminus\_47} ) \wedge ( \text{PR8\_01} \vee \text{PR14\_47} ) \wedge ( \text{PR8\_01} \vee \text{Terminus\_47} ) \wedge ( \text{Terminus\_01} \vee \text{PR14\_47} ) \wedge ( \text{Terminus\_01} \vee \text{Terminus\_47} )$$

De même

$$\mathit{int}(C2) = \mathit{cas\_46} \vee \mathit{cas\_47}$$

$$= ( \text{limite.troncon\_46} \wedge \text{GCA\_46} \wedge \text{C.ES\_46} \wedge \text{ST\_46} \wedge \text{PR15\_46} \wedge \text{PR36\_46} ) \vee ( \text{GCA\_47} \wedge \text{C.ES\_47} \wedge \text{ST\_47} \wedge \text{PR15\_47} \wedge \text{PR36\_47} \wedge \text{PR14\_47} )$$

$$= ( \text{limite.troncon\_46} \vee \text{GCA\_47} ) \wedge ( \text{limite.troncon\_46} \vee \text{C.ES\_47} ) \wedge ( \text{limite.troncon\_46} \vee \text{ST\_47} ) \wedge ( \text{limite.troncon\_46} \vee \text{PR15\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR36\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR14\_47} ) \wedge ( \text{GCA\_46} \vee \text{GCA\_47} ) \wedge ( \text{GCA\_46} \vee \text{C.ES\_47} ) \wedge ( \text{GCA\_46} \vee \text{ST\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR15\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR36\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR14\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR36\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR14\_47} ) \wedge ( \text{C.ES\_46} \vee \text{GCA\_47} ) \wedge ( \text{C.ES\_46} \vee \text{C.ES\_47} ) \wedge ( \text{C.ES\_46} \vee \text{ST\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR15\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR36\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR14\_47} ) \wedge ( \text{ST\_46} \vee \text{PR36\_47} ) \wedge ( \text{ST\_46} \vee \text{PR14\_47} ) \wedge ( \text{ST\_46} \vee \text{GCA\_47} ) \wedge ( \text{ST\_46} \vee \text{C.ES\_47} ) \wedge ( \text{ST\_46} \vee \text{ST\_47} ) \wedge ( \text{ST\_46} \vee \text{PR15\_47} ) \wedge ( \text{ST\_46} \vee \text{PR36\_47} ) \wedge ( \text{ST\_46} \vee \text{PR14\_47} ) \wedge ( \text{PR15\_46} \vee \text{PR36\_47} ) \wedge ( \text{PR15\_46} \vee \text{PR14\_47} ) \wedge ( \text{PR15\_46} \vee \text{GCA\_47} ) \wedge ( \text{PR15\_46} \vee \text{C.ES\_47} ) \wedge ( \text{PR15\_46} \vee \text{ST\_47} ) \wedge ( \text{PR15\_46} \vee \text{PR15\_47} ) \wedge ( \text{PR15\_46} \vee \text{PR36\_47} ) \wedge ( \text{PR15\_46} \vee \text{PR14\_47} ) \wedge ( \text{PR36\_46} \vee \text{PR36\_47} ) \wedge ( \text{PR36\_46} \vee \text{PR14\_47} ) \wedge ( \text{PR36\_46} \vee \text{GCA\_47} ) \wedge ( \text{PR36\_46} \vee \text{C.ES\_47} ) \wedge ( \text{PR36\_46} \vee \text{ST\_47} ) \wedge ( \text{PR36\_46} \vee \text{PR15\_47} ) \wedge ( \text{PR36\_46} \vee \text{PR36\_47} ) \wedge ( \text{PR36\_46} \vee \text{PR14\_47} )$$

$$= ( \text{limite.troncon\_46} \vee \text{GCA\_47} ) \wedge ( \text{limite.troncon\_46} \vee \text{C.ES\_47} ) \wedge ( \text{limite.troncon\_46} \vee \text{ST\_47} ) \wedge ( \text{limite.troncon\_46} \vee \text{PR15\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR36\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR14\_47} ) \wedge ( \text{GCA\_46} \vee \text{GCA\_47} ) \wedge ( \text{GCA\_46} \vee \text{C.ES\_47} ) \wedge ( \text{GCA\_46} \vee \text{ST\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR15\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR36\_47} ) \wedge ( \text{GCA\_46} \vee \text{PR14\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR36\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR14\_47} ) \wedge ( \text{C.ES\_46} \vee \text{GCA\_47} ) \wedge ( \text{C.ES\_46} \vee \text{C.ES\_47} ) \wedge ( \text{C.ES\_46} \vee \text{ST\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR15\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR36\_47} ) \wedge ( \text{C.ES\_46} \vee \text{PR14\_47} )$$

$$\begin{aligned}
& (\overline{C.ES_{46}} \vee \overline{PR36_{47}}) \wedge (\overline{C.ES_{46}} \vee \overline{PR14_{47}}) \wedge (\overline{ST_{46}} \vee \overline{PR36_{47}}) \wedge (\overline{ST_{46}} \vee \overline{PR14_{47}}) \wedge \\
& (ST_{46} \vee GCA_{47}) \wedge (ST_{46} \vee C.ES_{47}) \wedge (ST_{46} \vee ST_{47}) \wedge (\overline{ST_{46}} \vee \overline{PR15_{47}}) \wedge (\overline{ST_{46}} \vee \\
& \overline{PR36_{47}}) \wedge (\overline{ST_{46}} \vee \overline{PR14_{47}}) \wedge (PR15_{46} \vee PR36_{47}) \wedge (PR15_{46} \vee PR14_{47}) \wedge (\overline{PR15_{46}} \\
& \vee \overline{GCA_{47}}) \wedge (\overline{PR15_{46}} \vee \overline{C.ES_{47}}) \wedge (\overline{PR15_{46}} \vee \overline{ST_{47}}) \wedge (PR15_{46} \vee PR15_{47}) \wedge (PR15_{46} \\
& \vee PR36_{47}) \wedge (PR15_{46} \vee PR14_{47}) \wedge (PR36_{46} \vee PR36_{47}) \wedge (PR36_{46} \vee PR14_{47}) \wedge \\
& (\overline{PR36_{46}} \vee \overline{GCA_{47}}) \wedge (\overline{PR36_{46}} \vee \overline{C.ES_{47}}) \wedge (\overline{PR36_{46}} \vee \overline{ST_{47}}) \wedge (PR36_{46} \vee PR15_{47}) \wedge \\
& (PR36_{46} \vee PR36_{47}) \wedge (PR36_{46} \vee PR14_{47})
\end{aligned}$$

Le  $cas_{47}$  possède cet ensemble de conditions. Nous montrons les conditions qui correspondent à  $\mathbb{C}_1$  et  $\mathbb{C}_2$  dans le tableau suivant (Tableau B) :

|                    | <i>Condition de substitution</i>                                                                                                                                                                                                                                      | <i>Substitution (<math>\alpha \in \mathcal{V}</math>)</i>                                                                                         |
|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| $\mathbb{C}_{A_1}$ | <b>Condition d'égalité</b><br>• Zone d'accident                                                                                                                                                                                                                       | $Terminus \rightsquigarrow \alpha$                                                                                                                |
|                    | <b>Appartenance au concept</b><br>• Causes liées au système <ul style="list-style-type: none"> <li>▪ Pilotage automatique</li> </ul>                                                                                                                                  | $PR7 \rightsquigarrow \alpha$<br>$PR8 \rightsquigarrow \alpha$                                                                                    |
| $\mathbb{C}_{A_2}$ | <b>Appartenance au concept</b><br>• Éléments dangereux <ul style="list-style-type: none"> <li>▪ Conduite automatisée :</li> <li>▪ Conduite manuelle :</li> </ul> • Causes liées au système <ul style="list-style-type: none"> <li>▪ Pilotage automatique :</li> </ul> | $GCA \rightsquigarrow \alpha$<br><br>$C.ES \rightsquigarrow \alpha$<br>$ST \rightsquigarrow \alpha$<br>$PR14, PR15, PR36 \rightsquigarrow \alpha$ |

**Tableau A.** Conditions d'adaptations trouvées

# Annexe 3

## Captures écrans (éditeur protégé)

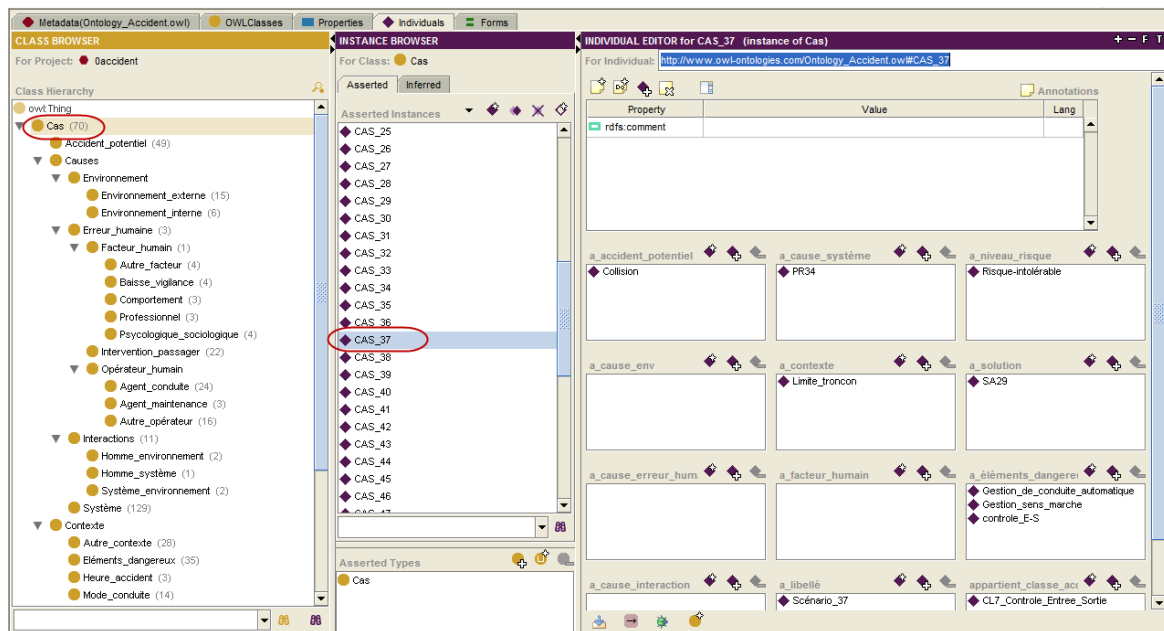


Figure D. Représentation d'un cas source au sein de l'ontologie.

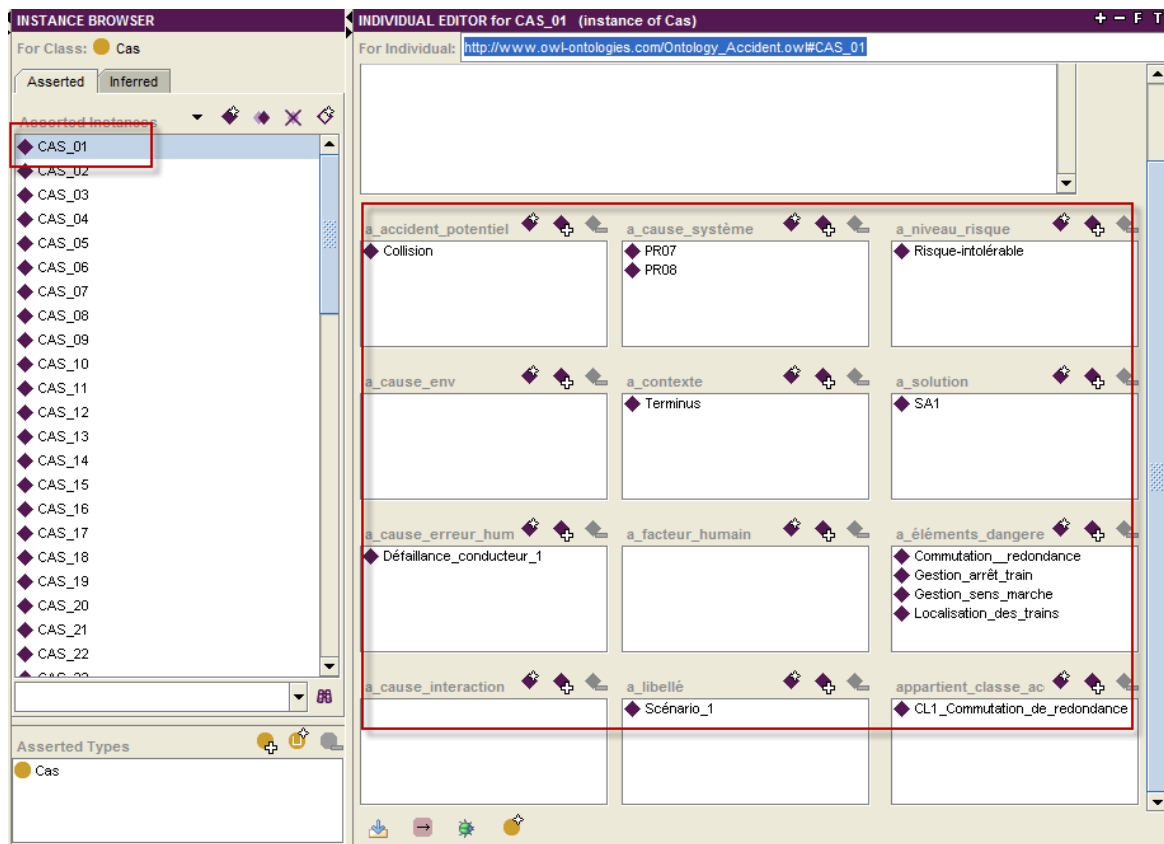


Figure E. Instances attribuées au cas source cas\_01 au sein de l'ontologie



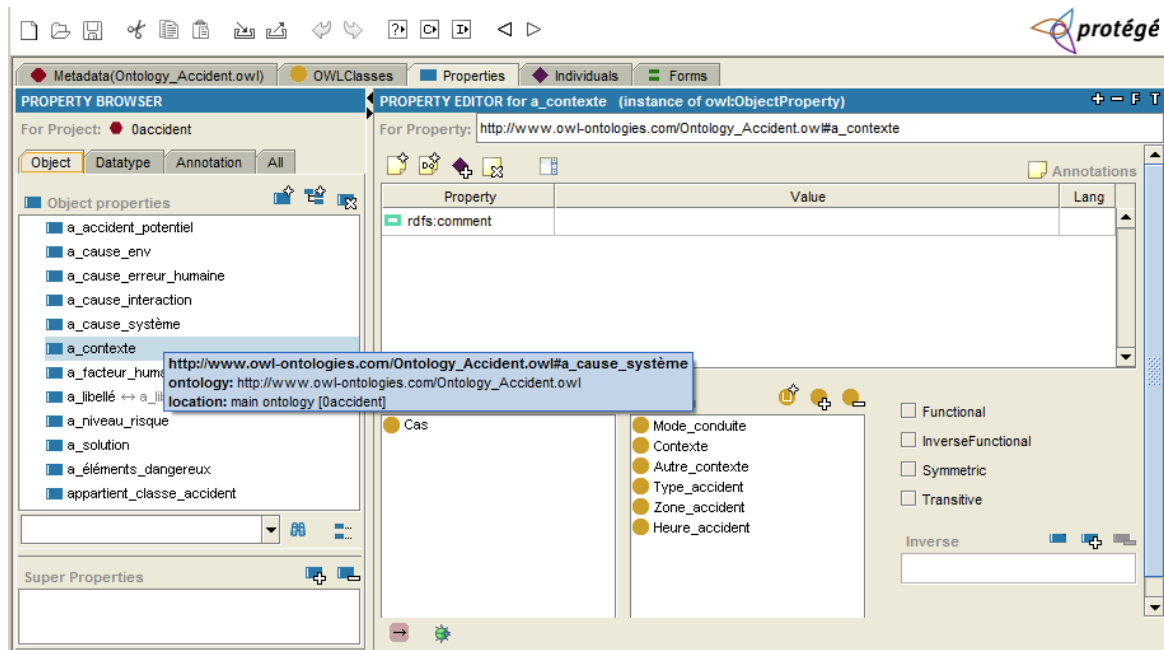


Figure F. Propriétés définies de l'ontologie dans protégé

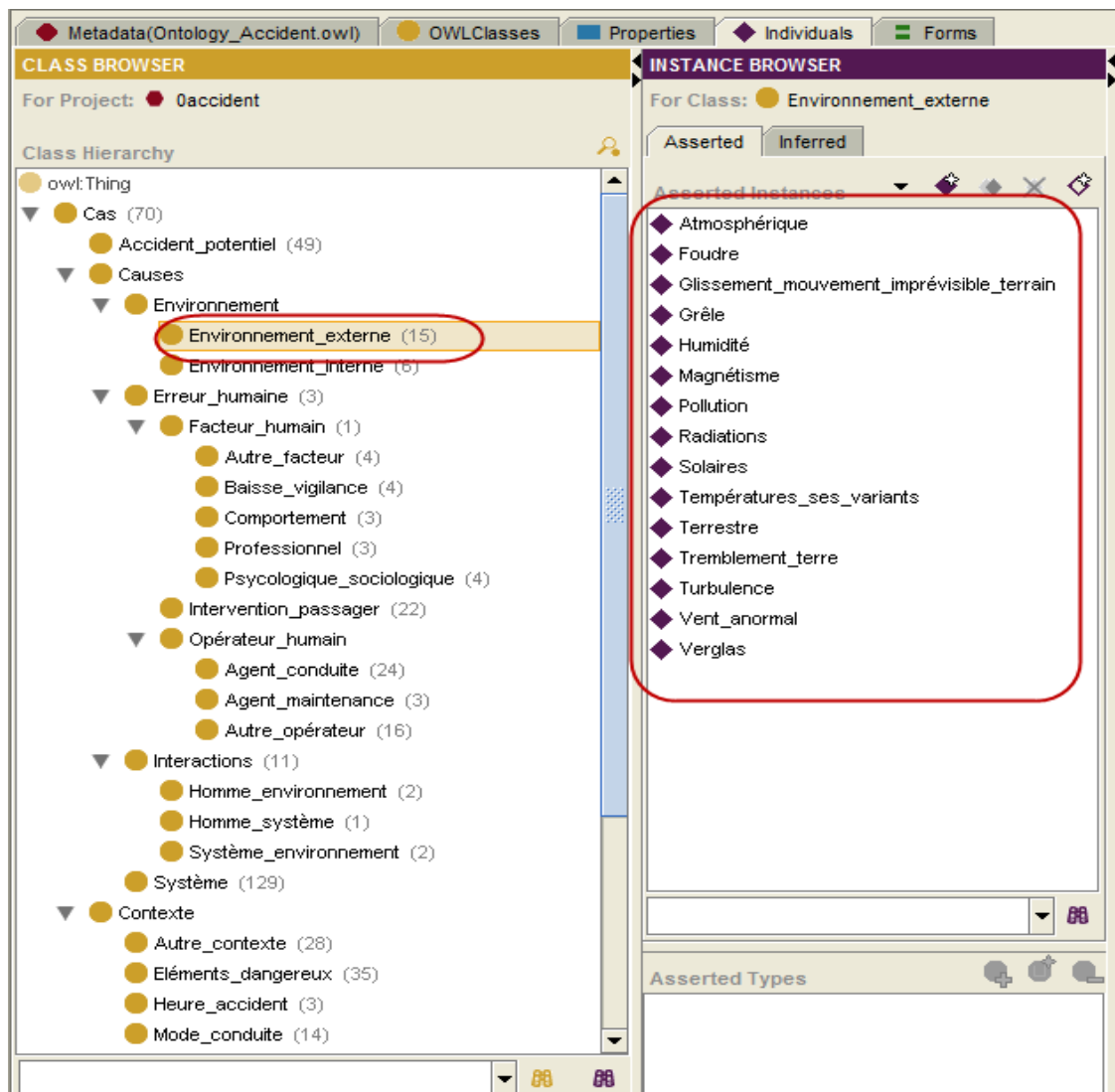


Figure G. Liste d'instances du concept « Environnement externe »