HAL
open science

Apport des Systèmes Multi-Agent et de la logique floue pour l'assistance au tuteur dans une communauté d'apprentissage en ligne

Youness Chaabi

To cite this version:

Youness Chaabi. Apport des Systèmes Multi-Agent et de la logique floue pour l'assistance au tuteur dans une communauté d'apprentissage en ligne. Interface homme-machine [cs.HC]. Université de Technologie de Belfort-Montbeliard, 2016. Français. NNT : 2016BELF0293 . tel-01874558

HAL Id: tel-01874558

https://theses.hal.science/tel-01874558v1

Submitted on 14 Sep 2018

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.

SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UNIVERSITÉ DE TECHNOLOGIE BELFORT-MONTBÉLIARD

Apport des Systèmes Multi-Agents et de la logique floue pour l'assistance au tuteur dans une communauté d'apprentissage en ligne

YOUNESS CHAABI

SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UNIVERSITÉ DE TECHNOLOGIE BELFORT-MONTBÉLIARD

N° 2 9 3

THÈSE présentée par

YOUNESS CHAABI

pour obtenir le

Grade de Docteur de
l'Université de Technologie de Belfort-Montbéliard

Spécialité : **Informatique**

Apport des Systèmes Multi-Agents et de la logique floue pour l'assistance au tuteur dans une communauté d'apprentissage en ligne

Unité de Recherche :
Laboratoire Électronique, Informatique et Image

Soutenue publiquement le 11 juillet 2016 devant le Jury composé de :

R. TOUAHNI	Président	Prof de l'enseignement Sup, FSK, Kenitra, Maroc
L. BOULAHIA	Rapporteur	MCF-HDR, Université de Technologie de Troyes, France
M. DAHCHOUR	Rapporteur	Prof de l'enseignement Sup, INPT, Rabat, Maroc
Y. RUICHEK	Rapporteur	Prof des Universités, UTBM, Belfort, France
B. MAZIGH	Examinateur	MCF, FSM,Monastir, Tunisie
R. AJHOUN	Examinateur	Prof de l'enseignement Sup, ENSIAS, Rabat, Maroc
V. HILAIRE	Co-Directeur de thèse	Prof des Universités, UTBM, Belfort, France
R. MESSOUSSI	Co-Directeur de thèse	Prof de l'enseignement Sup, FSK, Kenitra, Maroc

Remerciement

Ce travail de recherche a été réalisé en co-tutelle entre le laboratoire Système et Transport (SET) de l'Université de technologie de Belfort-Montbéliard (UTBM) École d'ingénieurs à Belfort, (France) et le laboratoire Systèmes de télécommunications et Ingénierie de la Décision (LASTID) dans l'Unité de Formation et de Recherche : Sciences et Techniques de Traitement de l'Information de la Faculté des Sciences de Kénitra, (Maroc). Ce travail de recherche a été financé par une allocation dans le cadre du programme volubilis (projet AI n° 33/SI/14).

Je tiens à exprimer ma profonde gratitude à M. Rochdi MESSOUSSI pour sa disponibilité, sa grande rigueur scientifique et pour le soutien qu'il m'a accordé. Je voudrais lui exprimer ma reconnaissance pour l'aide qu'il m'a constamment octroyée tout au long de ce travail, qu'il trouve, en ce mémoire, le témoignage de mes sincères remerciements.

Je tiens à remercier M. Vincent HILAIRE Professeur à L'UTBM pour m'avoir accueilli et dirigé mon doctorat et guidé mes recherches. Ces conseils, critiques, encouragements et sa confiance m'ont permis de mener à bien ce travail de recherche.

J'exprime toute ma gratitude à M. Yassine RUICHEK pour avoir coencadré cette thèse. Ses conseils, critiques, encouragements et sa confiance m'ont permis de mener à bien ce travail de recherche. Il a été pour moi plus qu'un coencadrant de thèse et je tiens à le remercier profondément.

Je tiens à exprimer ma haute considération à Mme. Raja TOUAHNI, Professeur à l'Université Ibn Tofail, pour le soutien qu'elle m'a accordé pendant ces années de thèse et d'avoir présidé ce jury de thèse.

Je remercie tout d'abord Mme. Leïla MERGHEM BOULAHIA et M. Mohamed DAHCHOUR et Raja TOUAHNI pour avoir gentiment accepté d'être rapporteurs de cette thèse et de participer à mon jury.

Mes remerciements vont aussi aux Mme. Rachida AJHOUN et M. Bel-Hassen MAZIGH pour avoir gentiment accepté d'être examinateur de cette thèse et de participer à mon jury.

Je tiens à remercier mes proches. Un grand merci à mes chers parents pour m'avoir entouré de leur affection et pour m'avoir offert une éducation qui m'a permis d'aboutir à cette thèse.

Mes remerciements vont aussi aux apprenants de Master qualité logiciel de l'université ibn tofail de kenitra qui ont rendu possible une première expérimentation de mon système en situation réelle. De même, je remercie M. Rochdi MESSOUSSI et Mlle. khadija LEKDIOUI pour l'analyse intuitive des expérimentations.

Pour finir, Je voudrais remercier les membres du LASTID et SET pour m'avoir offert un cadre de travail particulièrement agréable. Un merci aux doctorants, avec qui j'ai eu des discussions fructueuses et amicales tout au long de cette thèse.

Table des matières

1 Introduction Générale 14
1.1 Contexte 15
1.2 Objectif de recherche 16
1.2.1 Critères/Indicateurs 18
1.2.2 SMA 18
1.2.3 Approche méthodologique/Validation 18
1.3 Plan de thèse 19

2 Etat de l'art 21
2.1 Apprentissage collaboratif Assisté par Ordinateur (ACAO) 23
2.1.1 Introduction générale 23
2.1.2 L'apprentissage collaboratif à distance assisté par ordinateur 24
2.1.2.1 Apprentissage collaboratif et approche traditionnelle 24
2.1.2.2 Les Outils du CSCL 25
2.1.3 Les Traces d'interactions dans les formations à distances 25
2.1.3.1 Introduction 25
2.1.3.2 Définition des traces 26
2.1.3.3 Types de traces 27
2.1.3.4 Traitement et l'analyse des traces 30
2.1.4 Les indicateurs 30
2.1.4.1 Définition 30
2.1.4.2 Types d'indicateurs 31
2.1.4.3 Les indicateurs dans les EIAH : quelques exemples dans la littérature EIAH 33
a. La division du travail 33
b. Indicateur de repérage d'interaction 34
c. Indicateur Agent Actif 34
d. Indicateur associé aux actions collaboratives 34
e. Indicateur sur les actions non verbales 35
f. Indicateur sur la contribution d'un acteur 35
g. Indicateur sur le pourcentage de participation 36
h. Un indicateur pour les réseaux sociaux 36
2.1.5 Analyse des forums de discussion 37
2.1.5.1 Introduction 37
2.1.5.2 Une revue de la littérature 38
2.1.5.3 Forum de discussion dans un cadre d'évaluation 39
2.1.5.4 L'analyse automatique des forums 40
2.1.5.5 La théorie des actes de discours et analyse des conversations 41
a. Acte de langage 42
b. Types d'actes de parole 42
2.1.6 Profils Apprenants 44
2.1.6.1 Introduction 44
2.1.6.2 Profils d'apprenants 45
a. Définition de Profil des apprenants. 45
b. Représentation de Profils des apprenants. 45
c. Les dimensions des profils d'apprenants 45
2.2 Systèmes Multi-Agents (SMA) 47
2.2.1 Introduction 47
2.2.2 Concepts de base de la technologie multi-agents 48

2.2.2.1 Définition d'un agent 48
2.2.2.2 Définition de Système Multi-agents 49
2.2.3 Typologie des agents 50
2.2.3.1 Agent cognitif 51
2.2.3.2 Agent réactif 52
2.2.3.3 Agents à réflexes simples 53
2.2.3.4 Agents conservant une trace du monde 53
2.2.3.5 Agents délibératifs 54
2.2.4 Propriétés des systèmes multi-agents 54
2.2.4.1 Interactions entre agents 54
a. Coopération 55
b. Coordination 55
c. La négociation 56
2.2.4.2 Communication 57
2.2.5 Modélisation et Implémentation 58
2.2.5.1 ASPECS : un processus d'ingénierie logicielle pour la modélisation et l'implantation de systèmes complexes [Cossentino et al, 2010] 58
2.2.5.2 CRIO : Un Metamodèle pour l'analyse et la conception de systèmes complexes 60
2.2.5.3 Janus : une plate-forme organisationnelle pour l'implantation et la simulation de systèmes complexes 61
a. Métamodéle de Janus : Domaine de la solution de CRIO 61
b. Architecture et noyau de la plateforme Janus 63
c. Caractéristiques de Janus 65
2.3 Conclusion 66

3 Classification automatique de texte et logique floue 68
3.1 Classification automatique de texte 69
3.1.1 Introduction 69
3.1.2 L'apprentissage automatique 69
3.1.3 L'apprentissage supervisé 70
3.1.4 La catégorisation est un problème de classification supervisée 70
3.1.5 Comment classer ? 70
3.1.6 Codage des textes 71
3.1.6.1 Introduction 71
3.1.6.2 Le texte 71
3.1.6.3 Prétraitements 72
3.1.6.4 Définition de descripteurs 73
a. Représentation en « Sac de mots » « bag of words » 73
b. Représentation des textes par des phrases 74
c. Représentation des textes avec des racines lexicales (stemming) 74
d. Représentation des textes avec des lemmes (lemmatisation) 75
e. Représentation des textes avec la méthode des n-grammes 75
3.1.6.5 Codage des termes (Calcul du poids) 76
a. Le codage TFxIDF 76
b. Codage TFC 77
c. L'entropie 77
3.1.6.6 Réduction de la dimension 77
3.1.7 Différents Modèles de classifieurs 78
3.1.7.1 Réseaux bayésiens classifieurs 79
3.1.7.2 Machines à Vecteurs Support – SVM 79
3.1.7.3 Rocchio 80
3.1.7.4 K plus proches voisins - kPPV 80
3.1.7.5 Arbres de décision 81
3.1.7.6 Les approches neuronales 81
3.1.8 Avantage de la méthode adoptée (Naïve Bayes) 82

3.2 Logique floue 83
3.2.1 Introduction 83
3.2.2 Logique classique et logique floue 84
3.2.3 Concepts et définitions 85
3.2.3.1 Sous-ensembles flous ou Fuzzification 86
3.2.3.2 Variables linguistiques 87
3.2.3.3 Le raisonnement en logique floue 88
a. Modus ponens généralisé 89
b. Méthodes d'inférence floue 89
3.2.3.4 La défuzzification 90
3.3 Conclusion 91

4 Modélisation du système d'aide au tuteur 94
4.1 Introduction 95
4.2 La problématique du suivi des apprenants 95
4.3 Analyse des conversations textuelles asynchrones d'apprenants pour déterminer leur comportement social 98
4.3.1 Les activités de collaboration 98
4.3.2 Objectifs recherchés dans la détermination de profils de comportements sociaux 100
4.3.3 L'analyse qualitative et quantitative des conversations textuelles asynchrones 101
4.3.4 Choix des indicateurs 103
4.3.5 Classification par approche Bayésienne 104
4.3.6 Niveau logique floue 106
4.4 Approche méthodologique 108
4.5 Conclusion 109

5 Définition des mécanismes de base des différents agents 111
5.1 Introduction 112
5.2 Architecture multi-agents pour l'analyse et l'assistance 112
5.3 Les agents du système 114
5.3.1 Agent Récupération 114
5.3.2 Agent Filtrage 116
5.3.3 Agent Lemmatisation 119
5.3.4 Agent Classification de messages 123
5.3.4.1 Classification par actes de langage 123
5.3.4.2 Classification Bayésienne 125
5.3.5 Agent indicateur 135
5.3.5.1 Volume d'interventions : 135
5.3.5.2 Type d'interventions 136
5.3.5.3 Réactions Entrainées 137
5.3.6 Agent Logique floue 139
5.3.6.1 Première étape : Fuzzification 140
5.3.6.2 Deuxième étape : Le moteur d'inférence 143
5.3.6.3 Troisième étape : défuzzification 144
5.4 Conclusion 144

6 Expériences & résultats 146
6.1 Introduction 147
6.2 Cadre théorique 147
6.2.1 Les interactions dans un forum de discussion éducatif 148
6.2.2 Problématique 149
6.3 Méthodologie 150
6.3.1 Contextes d'expérimentation 151
6.3.2 Organisation humaine dans le projet 151
6.3.3 Structuration du projet 152

6.3.4 Les outils mis à la disposition des apprenants 153
6.3.5 Le protocole de recueil de données 153
6.3.6 Description du corpus 154
6.4 Analyse et résultats de l'expérimentation 154
6.4.1 Analyse intuitive des tuteurs 154
6.4.2 Analyse de système 156
6.4.2.1 Analyse par actes de langage. 158
6.4.2.2 Analyse par naïve bayes 159
6.4.3 Comparaissant des deux méthodes de classification 159
6.4.4 Interprétation graphique des données 163
6.5 Discussion des résultats 166
6.6 Conclusion 169

7 Conclusion générale 172
7.1 Conclusions 173
7.2 Perspectives 176

8 Publications de l'auteur 178

9 Références 180

1 Introduction Générale

Sommaire

1.1 Contexte 15
1.2 Objectif de recherche 16
1.2.1 Critères/Indicateurs 18
1.2.2 SMA 18
1.2.3 Approche méthodologique/Validation 18
1.3 Plan de thèse 19

1.1 Contexte

En 1994, Sir Geffrey Holland, Vice-Chancellor de l'université Exeter, a déclaré que « D'ici l'an 2020 tous les programmes d'éducation et de formation menant à une qualification ou d'un crédit seront disponibles selon trois modes : à temps plein, à temps partiel et à travers l'enseignement à distance. » [Wild, 1994].

L'apprentissage à distance tient une place importante dans le domaine des Technologies de l'Information et de la Communication (TIC). La E-formation est aujourd'hui une réalité grâce au développement de la technologie du web, réseaux informatiques, de l'internet et surtout des environnements virtuels d'apprentissage souvent appelés plates-formes de formation à distance. En quelque sorte, c'est le savoir constamment actualisé qui va vers l'apprenant. Nos travaux de recherche se situent dans le domaine des Environnements Informatiques pour l'Apprentissage Humain (EIAH), plus précisément dans le domaine de recherche sur L'Apprentissage Collaboratif Assisté par Ordinateur (ACAO), en anglais Computer-supported collaborative learning (CSCL).

L'intérêt pour ces environnements est relativement récent. Lipponen [Lipponen, 2002, P.72] donne la définition suivante: « *Put briefly, CSCL is focused on how collaborative learning supported by technology can enhance peer interaction and work in groups, and how collaboration and technology facilitate sharing and distributing of knowledge and expertise among community members* ». Ces environnements relatifs au monde de l'éducation sont inspirés du monde du travail qui se concentre sur la façon et la manière dont l'apprentissage collaboratif assisté par la technologie peut améliorer l'interaction entre pairs « Apprenant-Apprenant » « Tuteur-Apprenant » ainsi que le travail de groupe, et comment la combinaison de la collaboration et la technologie peuvent faciliter la diffusion et le partage d'expertise et connaissances entre les membres d'une communauté.

Les EIAH s'intéressent aux outils informatiques pour favoriser les interactions et le travail collaboratif à distance, permettent l'échange d'informations et le partage de connaissances entre apprenants. Parmi les objectifs de ces environnements, on peut citer : aider les apprenants dans leur démarche d'apprentissage, par exemple en leur donnant des conseils qui devraient être significatifs et pertinents pour l'apprenant.

Les développements technologiques liés à l'arrivée de ces environnements ont permis progressivement d'intégrer des outils favorisant les interactions : forums, chats, tableaux blancs, wikis, blogs, etc. Ces outils sont potentiellement intéressants pour supporter des formes d'apprentissage collectif.

Toutefois, même si les environnements d'ACAO peuvent aider le processus d'apprentissage, une formation à distance de qualité repose souvent sur les enseignants ou les tuteurs qui accompagnent les apprenants. L'importance du tutorat dans la réussite d'un dispositif de formation à distance est une réalité qui s'impose de plus en plus. Ce principe de tutorat distant ouvre un nouvel axe de recherche en EIAH, dans lequel nous nous situons, celui du suivi des apprenants. Actuellement, de nombreuses équipes de recherche se sont engagées dans l'élaboration et l'instrumentation des activités du tuteur pour aider celui-ci à soutenir les apprenants dans leur processus d'apprentissage collaboratif. Toutefois, il existe encore de nombreux problèmes. En effet, les difficultés résident dans la mauvaise gestion de l'accompagnement des apprenants dans ces dispositifs à cause, d'une part d'un manque de rétroaction, et, d'autre part, d'une perte de perception de la part du tuteur vis à vis des activités de l'apprenant.

Pour contourner ces lacunes, le travail proposé dans cette thèse vise à offrir un outil qui aide le tuteur et les concepteurs pédagogiques à interpréter les parcours des apprenants dans le cadre de pédagogie de projet. Pour cela, nous proposons une approche qui se base sur la modélisation des traces d'activités des apprenants dans une formation à distance et son usage au sein d'un EIAH : la difficulté pour l'exploiter et l'appropriation de la trace modélisée.

Dans la suite de cette introduction, nous présenterons d'abord le contexte général de nos travaux de recherche et les objectifs escomptés de ces travaux, ainsi que les contributions apportées dans le cadre de cette thèse, avant de terminer par présenter l'organisation du présent document.

1.2 Objectif de recherche

Ce travail vise à supporter le processus d'apprentissage à distance. En particulier, l'objectif est d'aider le tuteur à mieux encadrer les apprenants. Pour cela, il faut prendre en compte la nature distribuée et asynchrone de ce processus d'apprentissage. Le mécanisme de support va donc utiliser les traces produites par les apprenants au cours du processus d'apprentissage et les analyser afin d'aider le tuteur dans sa tâche.

Cet objectif principal peut, de manière évidente, se décliner en plusieurs sous-objectifs. Par ailleurs, le domaine de l'apprentissage à distance étant trop large pour être traité de manière complète et pertinente, nous avons opéré des restrictions. Dans ce travail, nous nous intéressons aux forums de discussions asynchrones qui occupent une place centrale dans le cadre de l'ACAO. En effet, l'accent est mis sur le rôle des forums de discussions dans les

activités d'apprentissage collaboratif [Calvani, 2010]. Ces forums permettent le développement des discussions où les apprenants peuvent accéder à des messages, réviser ou réinterpréter, construire, et affiner les idées, ainsi que stimuler la réflexion plus profonde [Calvani, 2010]. Du point de vue de l'enseignant, ces types d'activités donnent un aperçu sur la qualité de l'apprentissage et de l'enseignement. Il/elle peut alors effectivement avoir une meilleure compréhension de ce qui a été appris et quelles sont les difficultés d'apprentissage rencontrées pendant le processus d'apprentissage. L'enseignant peut ainsi bénéficier des informations accumulées au cours du temps sur les interactions des apprenants est par la suite construire un modèle d'évaluation qui permettrait de diriger à la fois l'enseignement et l'apprentissage.

Nous avons également choisi la pédagogie de projet comme fondement de notre travail qui vise à renforcer l'apprentissage collectif dans un contexte de distance. Cette pédagogie consiste à faire travailler les apprenants sur des projets pour réaliser un travail collaboratif.

Le premier sous-objectif repose sur l'identification des besoins des tuteurs dans le contexte d'un travail collaboratif en ligne. Cette identification repose principalement sur la définition d'indicateurs qui vont permettre de caractériser le processus d'apprentissage.

Un deuxième sous-objectif consiste en la proposition d'un modèle de système d'assistance aux tuteurs. Ce système doit satisfaire les contraintes : de s'intégrer à un environnement d'ACAO et de ne pas surcharger le tuteur dans ses tâches quotidiennes. Pour satisfaire ces contraintes nous avons fait le choix d'utiliser le paradigme des Systèmes Multi-Agents (SMA) [Hilaire, 2000 ; Rodriguez, 2005]. En effet, les approches multi-agents nous semblent particulièrement bien indiquées pour surveiller les activités des utilisateurs, tout en interagissant avec des outils de communication asynchrones, à savoir les « forums ». De plus les SMA exhibent des caractéristiques intéressantes telles que : modularité, facilité d'évolution, possibilité de répartition des agents dans l'environnement distribué et de prise en compte de la nature distribuée et asynchrone du problème à traiter.

Un troisième sous-objectif consiste à valider le modèle afin de s'assurer que les résultats produits de manière automatique par le système ont un sens et qu'on peut leur accorder un degré de confiance significatif.

Les sous-sections suivantes décrivent chaque sous-objectif.

1.2.1 Critères/Indicateurs

Dans le cadre de l'analyse automatique des activités de collaboration de l'apprenant, nous proposons une approche pluridisciplinaire entre les sciences humaines et sociales et les sciences de l'ingénieur pour l'analyse qualitative et quantitative des conversations textuelles asynchrones d'apprenants pour déterminer leur comportement social [Pléty, 1998; Chaabi, 2014], les rôles sociaux attribués à chaque apprenant et l'organisation qu'ils établissent, pour mener à bien le travail collaboratif. Cette analyse utilise le contenu des messages et des données stockées dans l'environnement de l'apprentissage humain (EHL), tels que le volume d'intervention, fréquence de message, type d'intervention et les réactions entraînées, etc., pour déterminer le comportement social de chaque apprenant. L'assistance au tuteur permet de situer le travail des apprenants par rapport à d'autres apprenants ou des groupes à d'autres groupes, mais prend également en charge les apprenants dans leur processus d'apprentissage, et les apprenants qui sont isolés pour leur éviter l'abandon [Sujana, 2012].

1.2.2 SMA

Notre système intègre l'analyse automatique des interactions et le contenu textuel des échanges grâce à des outils de communication dans les systèmes de gestion de l'apprentissage (LMS). Les indicateurs utilisés dans cet ouvrage sont une combinaison d'indicateurs rapportés dans la littérature et constituent un support à la réalisation de l'objectif principal de ce travail. Cet objectif est la spécification, la conception et le développement d'un système intelligent avec une architecture ouverte qui devrait permettre l'intégration d'autres indicateurs et d'autres techniques pour déterminer les comportements sociaux des apprenants dans le cadre d'une pédagogie de projet. Par conséquent, il est impératif de proposer des nouveaux outils pour visualiser et surveiller le travail et l'activité des apprenants. Le système devrait fournir des indicateurs d'un niveau d'abstraction élevé pour aider le tuteur à apprécier le travail des apprenants.

1.2.3 Approche méthodologique/Validation

L'approche méthodologique choisie consiste à apporter une réponse aux différents éléments de la problématique dégagés plus haut et à remplir nos objectifs de recherche. Pour cela, nous avons mise en place une expérimentation réelle de formation à distance. Cette expérimentation a été menée dans le cadre du master « Qualité du logiciel » en 2013/2014 à l'université Ibn Tofail de Kenitra, Maroc. Durant cette expérience, nous avons recueilli un

corpus de messages d'interactions et nous avons aussi vécu au quotidien la progression de la session durant les quatre mois de formation. Les données recueillies nous ont servi de support de développement et de test de notre système d'assistance au tuteur.

D'un point de vue conceptuel, notre démarche consiste à intégrer des tuteurs dans le processus de conception, à affiner progressivement des spécifications de notre système. Les entretiens que nous avons eus avec les tuteurs de la formation nous ont permis de définir la nature des indicateurs dont ils ont besoin afin de les aider dans leurs tâches de suivi et de régulation. Après la finalisation technique de notre système nous l'avons mis à la disposition des tuteurs pour évaluer les spécifications et les modifier au besoin.

Cette expérimentation nous a permis de valider notre système d'un point de vue fonctionnel et ergonomique, et d'affiner les choix que nous avons adoptés. Elle nous a permis aussi d'évaluer l'apport des différents indicateurs intégrés à notre système et de mesurer l'impact de notre système sur les activités des tuteurs.

1.3 Plan de thèse

L'approche proposée visait à apporter une réponse aux différents éléments de la problématique dégagée plus haut et à remplir nos objectifs de recherche. Après l'introduction du chapitre 1 dans laquelle nous avons introduit la problématique de notre thèse, ses objectifs et ses contributions, nous avons présenté, dans le chapitre 2, le contexte d'étude. Celui-ci est divisé en deux sections. Section 1, il s'agit des fondements théoriques de notre travail. Dans cette section, nous présentons un état de l'art sur le domaine de l'apprentissage collaboratif assisté par ordinateur en commençant par un travail terminologique qui précise le sens de certains termes employés dans cette thèse. Dans la deuxième section, Nous nous intéressons à l'utilisation des systèmes multi-agents dans le domaine de la formation à distance comme solution pour la mise en œuvre de notre système d'assistance au tuteur. Pour cela, quelques définitions sur les systèmes multi-agents seront énoncées. En second lieu, nous exposons le méta-modèle CRIO pour l'analyse et la conception de systèmes complexes. En troisième lieu, nous présentons le processus de développement ASPECS pour la modélisation et l'implantation de systèmes complexes. En dernier lieu, nous présentons la plate-forme Janus et ses principales caractéristiques.

Le chapitre 3 est réservé à la présentation des différentes phases du processus de classification automatique de textes. Nous présentons les différents objectifs et intérêts de la classification puis nous décrivons le processus général de la catégorisation des textes avec

toutes ses étapes. Dans la deuxième section de ce chapitre nous présentons la théorie de la logique floue, sa définition, ses caractéristiques essentielles de raisonnement approximatif, et le système d'inférence floue. Cette théorie permet de modéliser l'expertise de l'enseignant dans l'analyse des interactions entre apprenants pour déterminer ces comportements sociaux.

Le chapitre 4 présente notre approche pour l'analyse automatique des conversations textuelles dans les forums de discussions asynchrones pour déterminer les comportements sociaux des apprenants dans le cadre d'un travail collaboratif à distance. Ce chapitre constitue le cœur de notre contribution, il présente de manière approfondie les bases de notre travail. Une étude sur les comportements d'apprenants travaillant en groupe est décrite pour commencer. Ensuite, les possibilités d'analyse automatique de conversations par des actes de langages. Enfin, nous présentant la définition des modèles agents pour la satisfaction des buts du système et leur contribution dans les architectures des systèmes d'enseignement assisté par ordinateur.

Le chapitre 5 donne la définition des mécanismes de base des différents agents. Dans celui-ci nous présentons les besoins informatiques en termes d'architecture pour le système et les choix techniques des solutions utilisées. Puis le développement ainsi que les différents traitements réalisés par le système, sont alors exposés.

Le chapitre 6 décrit une expérimentation réalisée pour le test et l'évaluation de notre système. L'expérimentation a été menée dans le cadre du master « qualité du logiciel » à l'Université Ibn Tofail de Kenitra, Maroc. Leur déroulement ainsi que leurs résultats sont ensuite présentés et discutés.

Pour conclure nous présentons dans le chapitre 7 un bilan de ce travail en mettant en avant, les réalisations et leur apport en matière de soutien et d'assistance au tuteur pour la réussite d'une formation en ligne. Nous présenterons également les perspectives et les opportunités de recherches qui en découlent.

2 Etat de l'art

Sommaire

2.1 Apprentissage collaboratif Assisté par Ordinateur (ACAO) 23
2.1.1 Introduction générale....23
2.1.2 L'apprentissage collaboratif à distance assisté par ordinateur 24
2.1.2.1 Apprentissage collaboratif et approche traditionnelle....24
2.1.2.2 Les Outils du CSCL 25
2.1.3 Les Traces d'interactions dans les formations à distances 25
2.1.3.1 Introduction....25
2.1.3.2 Définition des traces....26
2.1.3.3 Types de traces....27
2.1.3.4 Traitement et l'analyse des traces....30
2.1.4 Les indicateurs....30
2.1.4.1 Définition 30
2.1.4.2 Types d'indicateurs....31
2.1.4.3 Les indicateurs dans les EIAH : quelques exemples dans la littérature EIAH....33
a. La division du travail....33
b. Indicateur de repérage d'interaction....34
c. Indicateur Agent Actif....34
d. Indicateur associé aux actions collaboratives....34
e. Indicateur sur les actions non verbales....35
f. Indicateur sur la contribution d'un acteur 35
g. Indicateur sur le pourcentage de participation 36
h. Un indicateur pour les réseaux sociaux....36
2.1.5 Analyse des forums de discussion....37
2.1.5.1 Introduction....37
2.1.5.2 Une revue de la littérature....38
2.1.5.3 Forum de discussion dans un cadre d'évaluation 39
2.1.5.4 L'analyse automatique des forums 40
2.1.5.5 La théorie des actes de discours et analyse des conversations....41
a. Acte de langage 42
b. Types d'actes de parole....42
2.1.6 Profils Apprenants....44
2.1.6.1 Introduction....44
2.1.6.2 Profils d'apprenants....45
a. Définition de Profil des apprenants....45
b. Représentation de Profils des apprenants....45
c. Les dimensions des profils d'apprenants....45
2.2 Systèmes Multi-Agents (SMA)....47
2.2.1 Introduction 47

2.2.2 Concepts de base de la technologie multi-agents 48
2.2.2.1 Définition d'un agent 48
2.2.2.2 Définition de Système Multi-agents 49
2.2.3 Typologie des agents 50
2.2.3.1 Agent cognitif 51
2.2.3.2 Agent réactif 52
2.2.3.3 Agents à réflexes simples 53
2.2.3.4 Agents conservant une trace du monde 53
2.2.3.5 Agents délibératifs 54
2.2.4 Propriétés des systèmes multi-agents 54
2.2.4.1 Interactions entre agents 54
a. Coopération 55
b. Coordination 55
c. La négociation 56
2.2.4.2 Communication 57
2.2.5 Modélisation et Implémentation 58
2.2.5.1 ASPECS : un processus d'ingénierie logicielle pour la modélisation et l'implantation de systèmes complexes [Cossentino et al, 2010] 58
2.2.5.2 CRIO : Un Metamodèle pour l'analyse et la conception de systèmes complexes 60
2.2.5.3 Janus : une plate-forme organisationnelle pour l'implantation et la simulation de systèmes complexes 61
a. Métamodéle de Janus : Domaine de la solution de CRIO 61
b. Architecture et noyau de la plateforme Janus 63
c. Caractéristiques de Janus 65
2.3 Conclusion 66

2.1 Apprentissage collaboratif Assisté par Ordinateur (ACAO)

2.1.1 Introduction générale

« Même si la salle de classe exerce encore sa suprématie dans nos universités, l'apprentissage en ligne est porteur de nombreux espoirs, dont celui de provoquer une véritable révolution éducative. Cette forme d'apprentissage modifierait à la fois la transmission du savoir et la façon de le recevoir. » Louise Marchand, « Expérimenter l'e-formation », Sciences humaines, Hors-Série n°40, mars-mai 2003, p° 46.

Une des exigences de base en matière de l'éducation dans l'avenir est de préparer les apprenants pour participer à un travail collaboratif en ligne, dans une société en réseau ou la connaissance sera la ressource la plus importante pour le développement économique et social.

L'apprentissage collaboratif Assisté par ordinateur (ACAO) est l'une des innovations les plus prometteuse pour l'amélioration de l'enseignement et de l'apprentissage. Historiquement c'est lors d'un Workshop d'une conférence en 1989 [J. Bannon, 1995] que le terme CSCL est apparu pour la première fois, or souci de clarté, il apparaît tout d'abord nécessaire de préciser la terminologie qui sera utilisée ici : L'apprentissage collaboratif à distance assisté par ordinateur (ACAO) dans des environnements médiatisés, est d'abord connu sous l'abréviation anglo-saxonne CSCL (Computer supported collaborative learning). Il s'agit de situations d'apprentissage collaboratif dans des environnements médiatisés. Le choix de ce sigle n'est pas innocent.

Le « CSCL », permet d'englober des préoccupations multiples : l'apprentissage humain (Learning), l'apprentissage collaboratif (Collaborative Learning), le soutien à cet apprentissage (Support Collaborative Learning) et l'utilisation de l'outil informatique comme instrument de soutien à cet apprentissage collaboratif (Computer-Supported Collaborative Learning).

Au cours de ces dix dernières années, la plupart des recherches se sont concentrées sur l'utilisation de l'information et la communication dans l'éducation et la possibilité de considérer la technologie comme moyen pour faciliter l'interaction sociale entre l'enseignant et les élèves, et entre élèves. La collaboration et la communication est certainement une idée centrale pour les environnements d'apprentissage basés sur les réseaux, mais l'interaction sociale a également été de plus en plus prise en considération dans la conception et la mise en œuvre des systèmes d'apprentissage en ligne [De Corte., 1996].

Ce chapitre introduit en sous-section 2.1.2 l'apprentissage collaboratif assisté par ordinateur. C'est un domaine vaste qui a donner lieu à des recherches multiples. Dans cette thèse nous nous sommes focalisés sur le support aux tuteurs et apprenants dans un contexte d'apprentissage à distance. Les sous-sections suivantes décrivent donc : les traces stockées lors de l'ACAO (sous-section 2.1.3), les indicateurs qui permettent de caractériser les activités d'ACAO (sous-section 2.1.4), les techniques d'analyse des forums de discussion (sous-section 1.5) et une typologie des profils d'apprenants (sous-section 2.1.6).

2.1.2 L'apprentissage collaboratif à distance assisté par ordinateur

2.1.2.1 Apprentissage collaboratif et approche traditionnelle

L'apprentissage coopératif et collaboratif diffère des approches traditionnelles d'enseignement car les apprenants travaillent ensemble plutôt que de se concurrencer les uns avec les autres.

L'apprentissage collaboratif peut avoir lieu en tout temps, les apprenants travaillent ensemble - par exemple, quand ils s'entraident à faire leurs devoirs. L'apprentissage coopératif a lieu lorsque les élèves travaillent ensemble dans le même lieu sur un projet structuré dans un petit groupe. Les groupes de compétences mixtes peuvent être particulièrement utiles pour les étudiants à développer leurs habiletés sociales.

Les compétences nécessaires pour le travail collaboratif dans des groupes diffèrent de celles utilisées pour l'écriture d'un document ou des devoirs ou tout autre "travail en classe" [Dillenbourg, 1999].

La recherche suggère que l'apprentissage coopératif et collaboratif apporte des résultats positifs, comme une meilleure compréhension du contenu, l'amélioration du rendement global dans les classes, l'amélioration de l'estime de soi, et une plus grande motivation pour rester sur la tâche.

Néanmoins des situations de CSCL admettent tout à fait que des actions coopératives soient entreprises au sein de la collaboration dans l'élaboration de sous-tâches.

Le CSCL s'intéresse aux outils informatiques pour soutenir et favoriser le travail collaboratif en ligne. L'ordinateur devient garant de la mémoire collective, il aide à la focalisation du discours et de l'action, représente les éléments d'une discussion en étant le principal médium de la communication. Ainsi les plates-formes collaboratives intègrent des outils de communication destinés à favoriser les interactions favorables et à créer des interdépendances positives dans une communauté de travail en développant parallèlement une certaine autonomie du côté des apprenants.

2.1.2.2 Les Outils du CSCL

Les outils pour soutenir les activités d'apprentissage ont évolué à partir du courrier électronique, des conférences électroniques, des contenus hypermédias [J. Bannon,. 1995]. Ces exemples montrent clairement que l'ordinateur joue un rôle important, car il facilitent à la fois le processus d'enseignement et l'apprentissage par le biais d'interactions.

Plusieurs CSCL ont été développés pour permettre aux étudiants de collaborer au sein d'un même environnement de travail. Par exemple, CoVis, WebICL, VLE, et GREWPtool. Le projet CoVis est spécifiquement conçu pour apprendre à être utilisé par les étudiants et les enseignants dans un nouvel environnement « groupe de travail » pour soutenir l'apprentissage collaboratif [R.D. Gomez 1998]. Comme décrit par O'Leary [O'Leary, 2007].

Aujourd'hui la majorité des plates-formes utilise un modèle d'enseignement basé sur des ressources pédagogiques à disposition des apprenants, proposent d'avantage d'activités collaboratives en fournissant des outils de communication synchrone ou asynchrones (chat, forum, mails, etc) et des espaces de travail partagés permettant la réalisation d'activités en groupes.

Une plate-forme collaborative propose aux utilisateurs un espace virtuel accessible depuis n'importe quel ordinateur connecté à internet. Après identification, l'utilisateur peut accéder à son espace personnel, sur lequel peuvent, par exemple être stockés des documents, et à un espace collaboratif où il est possible d'échanger, d'argumenter, d'interagir avec les autres apprenants, soit de manière synchrone, soit de manière asynchrone.

2.1.3 Les Traces d'interactions dans les formations à distances

Fondamentalement une trace d'activité numérique est le résultat d'un enregistrement systématique des interactions des utilisateurs qui constituent partiellement la réalisation d'une activité.

Les actions des élèves dans un environnement CSCL peuvent être enregistrées et stockées en tant que données d'interaction. Ces données peuvent être consultées et analysées automatiquement. Les enseignants, les étudiants et les chercheurs peuvent bénéficier de ces analyses.

2.1.3.1 Introduction

Une partie de la recherche dans le domaine des environnements informatiques pour l'apprentissage humain (EIAH) est fortement intéressé par le suivi des activités des

utilisateurs, tout en interagissant avec des outils de communication fournis par chaque plate-forme d'apprentissage.

Dans les situations d'apprentissage à distance, le suivi des activités des apprenants et l'exploitation de leurs traces ont attiré l'attention des chercheurs. La trace de l'activité de l'apprenant est reconnue comme une source d'information qui révèle non seulement les activités elles-mêmes, mais aussi leurs résultats (produits par les apprenants).

Dans le cadre de formation à distance, il y a toujours une séparation dans le temps et l'espace, le tuteur ne se trouve pas face à face avec ses apprenants pour déterminer, à partir de leurs manières de parler, de poser des questions et interactions, s'ils ont bien assimilé les connaissances présentées. Dans le but d'aider le tuteur à établir une telle assistance ou bien de guider les apprenants, il faut comprendre et interpréter les résultats des activités et des événements ainsi que les obstacles à l'effectivité de l'apprentissage. Le recueil et l'interprétation en cours de session, d'informations appelées traces. L'exploitation de ces traces, définies comme une séquence temporelle d'observation pour fournit des « connaissances sur l'activité » que nous appelons indicateurs d'apprentissage.

Pour atteindre ces objectifs, de nombreuses recherches sur l'analyse des traces et l'interprétation des activités réalisées par les apprenants, durant leurs interactions avec l'environnement de formation à distance ont été menées [Lemieux, 2013 ; Armentano, 2012].

2.1.3.2 Définition des traces

La définition de la notion de trace diffère selon son rôle et son utilisation dans chaque domaine de recherche.

Actuellement, il y a plusieurs points de vue sur ce que pourrait être la définition d'une trace.

Selon Jermann et ses collègues [Soller, 2001], une trace est une observation ou enregistrement d'un apprenant en interaction avec le système d'analyse.

[Champin et al., 2004] définissent la trace comme étant une séquence d'états et de transitions représentant l'activité de l'utilisateur : « la séquence temporelle des objets et opérations mobilisés par l'utilisateur lorsqu'il utilise le système est appelée trace d'utilisation».

Dans le cadre spécifique des EIAH, [Champin et al., 2004] définissent la trace d'apprentissage comme « une observation ou un enregistrement de l'interaction de l'apprenant avec un système en vue d'une analyse ».

Pour [Pernin, 2005], la trace représente « un indice de l'activité des acteurs d'une situation d'apprentissage qu'elle soit ou non instrumentée ». Il précise aussi qu'il s'agit « d'un résultat obtenu au cours ou au terme d'une activité, d'un événement ou d'un ensemble d'évènements relatif au déroulement de la situation d'apprentissage ».

Dans la même veine, [Cram et al., 2007] définissent une « trace d'interactions » comme « tout objet informatique dans lequel s'accumulent des données à propos des interactions entre un système informatique et son utilisateur ».

D'un autre côté, dans leurs travaux liés à la réingénierie des EIAH, [Choquet et Iksal, 2007] considèrent la trace comme « un objet pédagogique au même titre que les ressources ou les scénarios pédagogiques ».

Une définition plus générale, est celle donnée par [Settouti et al., 2006] : « la trace est définie comme une séquence temporelle d'observés ». Le terme « séquence temporelle » dénote l'existence d'une relation d'ordre organisant les données de la trace par rapport à un repère de temps. Le terme « observé » dénote que les traces sont issues d'une observation.

Dans notre travail, nous considérerons cette même définition de la trace car elle nous semble la plus complète. Cependant, nous nous intéressons uniquement aux traces numériques.

2.1.3.3 Types de traces

Les traces se présentent sous des formes de plus en plus variées, incluant des relations riches et des possibilités de navigation poussées : allant des simples clicks de la souris, URLs visitées et temps de consultation jusqu'à la voix enregistrée, le parcours chronologique au sein d'une activité, l'usage des services de communication dans l'espace pédagogique et les réponses aux exercices et questionnaires.

[Lejeune et al., 2005] et [Gwenegan, 2005] regroupent ces traces suivant le contenu de l'information qu'elles portent. En conséquence, quatre groupes de traces se dégagent :

- Les traces informatives (les informations personnelles ex. nom, prénom, etc.) et les informations techniques (ex. adresse IP, navigateur, etc.) ;
- Les traces liées à l'exploitation d'une ressource (référence de la ressource, origine et historique des accès, etc.) ;
- Les traces associées à l'activité d'apprentissage (temps de réponse, résultats de tests, etc.) ; et
- Les traces attachées à l'activité de communication (contenu d'un message, destinataire, etc.).

La définition de ces quatre groupes permet d'envisager un traitement unique pour l'ensemble des traces associées à un groupe.

Quatre classes de traces qui fournissent des données objectives sur les apprenants et leurs activités [Gwenegan, 2005 ; Branovic, 2002]. Les classes sont regroupée dans le tableau 2.1 suivant [Rozé, 2005] :

TYPE	TRACES	Indiv	Coll	Qual	Quant	Comp	Conn	Exploitation possible
INFORMATIONS	Informations Personnelles (âge, genre, ..)	++	+					Identifier l'apprenant
	Informations techniques (ID, browser, SE, …)	++	+					Identifier l'apprenant
RESSOURCES	Nom (référence) de la ressource traitée par un apprenant à un moment donnée	++	+	+			+	Informer/tenir au courant le tuteur
	Nombre d'accès	++			++	+	+	Dégager des informations sur l'apprenant (capacités, compétences, …)
	Durée de consultation (connexion)	++			++	++	+	Evaluer le niveau d'activité
	Origine de l'accès à la ressource	++		++			++	Dégager le cheminement conceptuel
	Historique du parcours des ressources	++		++			++	Dégager le cheminement conceptuel/profil d'apprenant
Activité d'APPRENTISSAGE	Qualité d'une production	++		++			++	Evaluation de l'assimilation de concepts/connaissances
	Temps de réponse à une question	++			++	+	+	Evaluation
	Tests antérieurs	++		++			++	Capitalisation pour permettre une adaptation
	Réalisations d'actions (ex : exécution, débogage)	++		+		+	+	Suivi du traval de l'apprenant/ Respect du scénario
Activité de COMMUNICATION	Nombre messages envoyés (mail)		++		++	+		Améliorer la régulation d'un groupe
	Nombre messages reçus (mail)		++		++	+		Améliorer la régulation d'un groupe
	Nombre messages postés (forum)		++		++	+		Evaluer l'interaction dans un groupe
	Nombre messages lus (forum)		++		++	+		Evaluer l'interaction dans un groupe
	Fréquence des messages (sur une période donnée)		++		++	+		Améliorer la régulation d'un groupe/Evaluer la réactivité
	Destinataire des messages		++			+		Evaluer l'interaction dans un groupe
	Message de communication (contenu)		++	+		+	+	Evaluer la qualité des interactions
	Demande d'aide en ligne	++	+	+	+		+	Evaluer l'adéquation d'une ressource/Fournir une aide
	Demande d'assistance au tuteur	++	+	+	+		+	Régulation de la situation

Tableau 2-1 : Le choix des données à observer dépend de l'environnement d'interaction, et des objectifs de l'observation [Gwenegan, 2005].

Légende :

Indiv: Individuelle; **Coll**: Collective; **Qual**: Qualitative ; **Quant**: Quantitative ; **Comp**: Comportement; **Conn**: connaissance.

A partir de plusieurs lectures d'articles de recherche dans les champs des EIAH [Cram, 2007; Pernin, 2005], les articles abordant les notions de traces et d'exploitation des traces, également les articles présentant des objectifs différents d'exploitation des traces tels que la visualisation de l'activité de l'apprenant pour le tuteur [Mazza, 2003], l'élaboration d'un modèle de connaissance [Biuk-Aghai, 2001], la construction du modèle cognitif de l'apprenant [Heller, 2004], l'assistance à la régulation [Heraud, 2004], l'évaluation de l'activité [Zaïane, 2001] ou encore l'analyse de l'interaction entre apprenants dans des environnements d'apprentissage collaboratifs [Siebra, 2005].

Les traces sont regroupées suivant le contenu de l'information qu'elles portent.

- Le premier groupe contient les traces portant des informations d'identification : les informations personnelles (nom, prénoms, âge, classe, etc.) et les informatiques techniques liées au support utilisé (adresse IP, navigateur, système d'exploitation, etc.).
- Le type informative, permet de caractériser la situation d'apprentissage dans laquelle la trace est utilisée : situation d'apprentissage collectif (par exemple : le nombre de messages postés sur un forum) ou situation d'apprentissage individuel (par exemple : le temps de réponse à une question).
- Dans un deuxième groupe, représente les traces liées à l'exploitation d'une ressource : nom, référence de la ressource, nombre d'accès, durée de consultation, origine de l'accès et historique du parcours des ressources.
- Le type ressources, détermine la valeur d'usage de la trace : la trace porte une valeur d'usage plutôt qualitative (par exemple : le résultat d'un test) ou plutôt quantitative (par exemple : la durée de connexion à une ressource).
- Le troisième groupe contient les traces relatives à l'activité d'apprentissage : qualité d'une production, temps de réponse, résultats de tests et réalisation d'actions.
- Le type Activité d'apprentissage, identifie le cadre usuel d'exploitation de la trace : la trace apporte une information utilisée dans le cadre d'une étude de la connaissance de l'apprenant (par exemple : la qualité d'une production) ou dans le cadre d'une étude de son comportement (par exemple : la fréquence des messages postés).
- Quant au dernier groupe, il réunit les traces liées à l'activité de communication : nombre de messages envoyés et lus dans un forum, contenu d'un message, destinataires, fréquence des messages sur une période donnée. On trouve également dans ce groupe, l'affectation d'un rôle à un acteur de la situation, la demande d'aide en ligne et la demande d'assistance au tuteur.

2.1.3.4 Traitement et l'analyse des traces

Après la structuration des traces, la prochaine étape est le traitement de ces traces, qui peut être fait par différentes méthodes. Mais avant de pouvoir être exploitées, il faut franchir plusieurs étapes. On peut distinguer dans le processus de traitement des traces, trois phases principales : la phase de collecte, la phase d'analyse et la phase d'exploitation. Les principaux problèmes soulevés par cette étape d'analyse des traces sont liés à l'efficacité et la qualité des résultats retourné à partir de l'analyse [Lejeune et al., 2005]. Les méthodes de traitement et d'analyse utilisées doivent être en mesure d'extraire les traces des informations de synthèse qui non seulement correspondent aux exigences spécifiques des utilisateurs, mais aussi répond explicitement pour ce qui a été défini comme objectifs de l'analyse (c'est à dire ce qui se passe exactement après que l'étudiant soit connecté sur les plates-formes de formation à distance. Comment savoir si un message n'a pas été lu par les utilisateurs après son affichage ? etc). Il est également nécessaire d'examiner à ce stade, les moyens d'enrichir les traces enregistrées : le fait que les traces sont modifiées par l'ajout de données plus descriptifs (informations complémentaires) à sa représentation originale, permettant aux traces à restructurer, transformés en un autre format de représentation réutilisable pour d'autres types d'analyse.

Les techniques d'analyse des traces d'interaction dépendent de la nature des traces et des objectifs poursuivis. L'objectif principal des stratégies de visualisation et d'analyse de traces est d'élaborer des indicateurs pour le suivi, l'assistance, le soutien de l'apprenant et régulation des activités.

2.1.4 Les indicateurs

2.1.4.1 Définition

L'analyse des traces est un processus en plusieurs étapes qui consiste à : (i) La collection des données d'observation ; (ii) Le traitement de ces données ; (iii) L'interprétation des données obtenues. Le traitement et l'interprétation des traces d'apprentissage peuvent être effectuées par la production d'indicateurs pédagogiques qui ont pour principal objectif, l'amélioration de la perception de l'activité de l'apprenant. Ces indicateurs aident l'enseignant à interpréter facilement le niveau de l'apprenant ou le niveau du groupe et évaluer les sessions effectuées.

Les indicateurs sont dédiés à aider les tuteurs dans la réingénierie de leurs scénarios pédagogiques. En fait, cette réingénierie signifie l'amélioration, le contrôle et la refonte du processus d'apprentissage en fonction des besoins de l'apprenant et les variations de

l'environnement d'apprentissage. Les indicateurs sont susceptibles d'apporter une information pertinente sur l'exécution de scénarios pédagogiques.

Selon [Dimitracopoulou et al., 2004] les indicateurs dans le cadre d'apprentissage sont des variables qui peuvent décrire : le mode, le processus ou la qualité du système cognitif de l'activité d'apprentissage ; les fonctionnalités ou la qualité de l'interaction ; le mode, le processus ou la qualité de la collaboration. Leurs valeurs dérivent du traitement d'une ou de plusieurs traces.

Chaque indicateur possède un nom, une définition, éventuellement une description en langage naturel, un objectif, un domaine de valeurs, une durée de validité ainsi qu'une interprétation qui naturellement dépend du contexte dans lequel il est calculé. Ainsi, nous pouvons distinguer plusieurs classes d'indicateurs que nous détaillons dans la section suivante.

Selon [Dimitracopoulou et Bruillard, 2006], « un indicateur est une variable au sens mathématique à laquelle est attribuée une série de caractéristiques. C'est une variable qui prend des valeurs de forme numérique, alphanumérique ou même graphique. La valeur possède un statut : elle peut être brute (sans unité définie), calibrée ou interprétée. Le statut identifie une caractéristique bien précise : celle du type d'assistance offerts aux utilisateurs. Chaque indicateur peut dépendre d'autres variables comme le temps, ou même d'autres indicateurs.

2.1.4.2 Types d'indicateurs

D'après les travaux des recherches de [Dimitracopoulou et al., 2004], on peut constater que pour interpréter l'indicateur, il est nécessaire de tenir compte de :

- L'environnement et l'activité d'apprentissage : activité individuelle ou collaborative ;
- L'aspect traité de l'interaction : les caractéristiques ou la qualité de l'activité liée à une tâche d'apprentissage spécifique « indicateur cognitif » ; les caractéristiques ou la qualité de la collaboration « indicateur social » ;
- Le profil des acteurs et leurs rôles : à qui fait référence cet indicateur (l'apprenant, un groupe d'apprenants, toute la communauté ou à l'enseignant) et qui va l'utiliser (l'apprenant, l'enseignant ou le système) ;
- La dépendance du temps et/ou du contenu, ainsi que d'autres variables ;

Au-delà des aspects liés au contexte, les indicateurs peuvent être classés selon leur type de valeurs (quantitative ou qualitative), ou selon le niveau d'interprétation associé [Dimitracopoulou et al., 2004] :

- Bas niveau, n'ayant pas une valeur interprétative autonome et généralement déduit directement à partir des données brutes ;
- Atomique ou composite ;
- Haut niveau, ayant une valeur interprétative, souvent déduite par un processus complexe de traitement des traces.

Suivant ces classes d'indicateurs rapportés dans la littérature, le tableau 2.2 présente un descriptif d'un exemple d'indicateurs tirés des travaux existants [Merceron, 2007], [Dimitracopoulou et al., 2004].

Aspect traité	Type	Exemples d'indicateurs
Environnement de l'activité d'apprentissage	Individuel	Catégorie d'envoi par participant [Chen, 2004]
	Collaboratif	Niveau d'interaction du groupe [Schummer et al., 2005] Niveau d'activité de la communauté [Jermann, 2004]
Caractéristique traitée de l'activité	Cognitif	Profondeur de fil de discussion [Gerosa, 2005] Cohérence des thèmes de discussion [Michozuki et al., 2005] Distribution des catégories des messages [Gerosa, 2005]
	Social	Le degré de présence [Dringus et Ellis, 2005] Niveau de coordination, argumentation et d'initiative [Barros et al., 2002] La cohésion du groupe [Reyes et Tchounikine, 2005] La cohésion et la centralité dans les réseaux sociaux [Reffay et Lancieri, 2006]
	Affectif	L'état de motivation [Reimann, 2003] L'état social [Reyes et Tchounikine, 2005] et [Vassileva et al., 2004]
Type de valeur	Quantitatif	Le degré d'implication de chaque apprenant dans la formation [Santos et al., 2003]
	Qualitatif	Densité du réseau social [Martinez et al., 2003]
Niveau d'interprétation	Haut niveau	Qualité d'argumentation [Barros e al., 2002]
	Intermédiaire	Durée moyenne de consultation d'une page de cours.
	Bas niveau	Durée de consultation d'une page Fréquence de consultation d'une page

Tableau 2.2 : Classification des indicateurs

D'après les informations regroupées dans le tableau 2.2, nous remarquons que la majorité des indicateurs définis dans la littérature sont des indicateurs de bas niveau, quantitatifs, et sociaux (ex. Le taux de participation [Dimitracopoulou et Bruillard, 2006]). Peux d'indicateurs permettent d'analyser des aspects cognitifs. Ces indicateurs nécessitent des années de recherche pour être définis [Diagne, 2009]. Par exemple, l'indicateur cognitif identifiant les mis-conceptions des élèves lors de la résolution d'exercices en électricité [Michelet et al., 2007]. Partant de ce constat, l'idée de réutilisation d'indicateurs est née. Dans ce sens, plusieurs modèles d'indicateurs ont été proposés dans la littérature [Dimitracopoulou et Bruillard, 2006, Michelet et al., 2007].

2.1.4.3 Les indicateurs dans les EIAH : quelques exemples dans la littérature EIAH

Dans le cadre d'EIAH, le calcul d'indicateurs est d'une autre dimension. En effet, l'indicateur pertinent doit permettre de synthétiser des statistiques sur les traces de l'espace de travail commun des apprenants (du groupe) et les traces individuelles de chacun. La définition des indicateurs dans ce contexte tente de décrire le comportement des apprenants dans la collaboration.

Dans cette section nous détaillons quelques indicateurs issus du domaine des EIAH [Dimitracopoulou, 2004]. Le 1er proposer dans la thèse de [Jermann, 2004] et les autres ont été proposés par [Avouris et al., 2003].

a. La division du travail

Cet indicateur défini et implémenté dans la plateforme COTRAS "COllaborative TRAffic Simulator" [Jermann, 2004] identifie la division du travail adoptée par deux utilisateurs qui agissent sur un ensemble de ressources. Cet indicateur permet d'identifier le rôle pris par chaque participant dans le processus de collaboration pour l'apprentissage. On peut distinguer trois types de division du travail : la division basée sur la tâche où chaque acteur agit séparément sur des ressources différentes. Une division basée sur le rôle, où un seul des deux acteurs agit sur toutes les ressources. Une division du travail concourant où les deux acteurs agissent (en même temps) sur l'ensemble des ressources.

La caractérisation de ces trois types de division du travail peut se calculer à partir de la somme des différences SD et celle des différences absolues SAD (équation 1).

$$SD = \frac{\sum_i (S1Ai - S2Ai)}{S1A + S2A} \quad SAD = \frac{\sum_i |S1Ai - S2Ai|}{S1A + S2A} \qquad (1)$$

Avec $S1Ai$ (respectivement $S2Ai$) : total des actions faites par le sujet S1 (respectivement S2) sur la ressource Ai et $S1A$ (respectivement $S2A$) : total des actions faites par le sujet S1 (respectivement S2) sur toutes les ressources. Le SAD indique la symétrie des actions. La valeur 0 signifie que les deux acteurs font le même nombre d'actions, alors que la valeur +1 signifie que toutes les actions sont faites par le même acteur, identifiable grâce au signe de SD.

b. Indicateur de repérage d'interaction

L'indicateur Interaction implémenté dans la plateforme ModellingSpace [Avouris et al., 2003] est utilisé par les enseignants pour mesurer le taux d'activité lors de la résolution d'un problème donné. Il calcule le nombre d'actions faites dans un module d'activité, dans un intervalle de temps. Si on considère l'intervalle [t0, tm] associé à une session de collaboration, le temps est quantifié de la façon suivante : $t_i = t_0 + (i * d)$, avec $\alpha = \frac{t_m - t_0}{n}$. L'interaction se calcule de la façon suivante (équation 2):

$$Interactions\ (k, t_i) \tag{2}$$

L'indicateur interaction représente le nombre d'actions effectuées dans un module k durant l'intervalle de temps [ti-1, t_i] avec les valeurs k varie de $[1, k_{max}]$. k correspondant aux outils d'interaction { k =1 =>chat, k =2=>forum,…}.

Si $Interactions\ (k, t_i)$ = 0 alors aucune action n'est faite sur le module d'activité k dans l'intervalle de temps [t_i-1, t_i]. Alors que si la valeur des Interactions (k ,t_i) est grande, il y a peut-être plus de collaboration entre les acteurs dans ce module.

c. Indicateur Agent Actif

Cet indicateur implémenté dans la plateforme ModellingSpace [Avouris et al., 2003] est utilisé par les enseignants pour mesurer l'activité au cours de la résolution d'un problème donné. Un acteur n'est actif que s'il interagit dans un module d'activité et dans un intervalle de temps donnée, l'indicateur Agent actif, présente le nombre des acteurs qui ont interagi au moins une fois dans le module k dans un intervalle de temps donné. Nous pouvons calculer cet indicateur pour chaque module d'activité. Ainsi, $Agents\ (k, t_i)$ donne le nombre des acteurs qui ont interagi dans le module k, et dans l'intervalle de temps [ti-1, t_i].

d. Indicateur associé aux actions collaboratives

Cet indicateur implémenté dans la plateforme ModellingSpace [Avouris et al., 2003] est utilisé pour représenter à la fois le nombre d'actions et le nombre d'agents actifs au cours de la résolution d'un problème donné. On calcule cet indicateur à partir de deux indicateurs précédents : $Agents\ (k, t_i)$ et $Interactions\ (k, t_i)$, sur un intervalle de temps

t_i. L'indicateur CA est une somme sur l'ensemble des modules, dont les termes sont précisés par la formule suivante (équation 3) :

$$CA(t_i) = \sum_{k=1}^{k=k\,max} Agents(k, t_i) * Interactions(k, t_i) \quad (3)$$

e. Indicateur sur les actions non verbales

L'indicateur NVA implémenté dans la plateforme ModellingSpace [Avouris et al., 2003] est utilisé par les enseignants et les chercheurs durant et après la session de collaboration et représente le pourcentage de toutes les actions non verbales sur les différents outils d'interaction. L'idée est de calculer d'abord toutes les actions verbales (e.g. message chat/forum, etc.), et ensuite de considérer le reste des actions comme actions non verbales. La formule de calcul suivante considère que les seules actions verbales sont les actions de type chat. Mais nous pouvons élargir la règle pour le reste des actions verbales (équation 4) :

$$NAV(t_i) = \frac{\sum_{k=1}^{k=k\,max} Interactions\ (k, t_i) - Interactions('Chat', t_i)}{\sum_{k=1}^{k=k\,max} Interactions\ (k, t_i)} \quad (4)$$

Si NVA f ≈ 0 alors les acteurs ont tendance à n'utiliser que des actions verbales. NVA f ≈ 1 alors les acteurs ont tendance à n'utiliser aucune action verbale dans les outils dans l'intervalle de temps sélectionné.

f. Indicateur sur la contribution d'un acteur

L'indicateur SAC implémenté toujours dans la plateforme ModellingSpace [Avouris et al., 2003] est utilisé par les enseignants, les chercheurs et les tuteurs pour évaluer la participation d'un acteur donné au cours de la résolution synchrone d'un problème posé. Il s'agit du taux d'interaction d'un acteur donné (Agent) par rapport à l'ensemble des interactions (de tous les acteurs) pour un module d'interaction k, et sur un intervalle de temps [t_i-1, t_i]. La formule de calcul est la suivante (équation 5) :

$$SAC(agent, t_i) = \frac{\sum_{k=1}^{k=k\,max} Interactions\ (k, agent, t_i) - Interactions('Chat', t_i)}{\sum_{k=1}^{k=k\,max} Interactions\ (k, t_i)} \quad (5)$$

Interactions (k, agent, ti) étant le nombre des actions faites par un acteur visé (agent) dans le module k durant l'intervalle de temps [t_i-1, t_i]. SAC appartient à [0, 1]. Si SAC = 0

alors l'acteur n'a pas agi dans l'intervalle de temps considéré. Si SAC =1 alors l'acteur est l'unique intervenant dans l'intervalle de temps considéré.

g. Indicateur sur le pourcentage de participation

L'indicateur PART implémenté dans la plateforme ModellingSpace [Avouris et al., 2003] représente la proportion des agents ayant interagi dans au moins un module dans l'intervalle de temps considéré. La formule de calcul est la suivante (équation 6) :

$$PART(t_i) = \frac{Agent(t_i)}{TotalAgents} \quad (6)$$

Par exemple, si PART(ti)=0,5 alors la moitié des agents ont interagi dans au moins l'un des modules sur l'intervalle de temps [t_i-1, t_i]. Si $PART(t_i)$=0 alors aucun acteur du groupe n'agir durant cette période.

h. Un indicateur pour les réseaux sociaux

Le travail de [Reffay et Lancieri, 2006] sert à calculer la cohésion et la centralité dans les réseaux sociaux à partir des forums de discussion. On trouve dans ce travail une classification des indicateurs quantitatifs. Les cinq indicateurs distingués sont les suivants :

- Des indicateurs déjà disponibles sur les LMS (Learning Management System) liées à un acteur comme : le nombre de connexions, le nombre de pages demandées, le nombre de messages postés dans le forum, le nombre de messages postés et reçus par courriel,
- Le nombre de messages postés et reçus par courriel. Les messages ouverts ne sont pas forcément lus, mais les messages non ouverts sont forcément non lus. Dans ce cas, il est intéressant de supprimer ces messages non lus de la liste des communications.
- Le nombre de personnes ayant émis des messages vers cet acteur et le nombre de personnes ayant reçu au moins un message émis par cet acteur.
- Le nombre de messages : la valeur associée à chaque message + la source + les destinataires sont précisés.
- Des indicateurs qui permettent de préciser la position et le rôle de chacun des individus dans le groupe.
- C'est au sein de cette dernière classe d'indicateurs qu'il est possible de distinguer les trois indicateurs proposés par la littérature des Réseaux sociaux : la cohésion, la centralité et l'intermédiarité.

2.1.5 Analyse des forums de discussion

2.1.5.1 Introduction

Les forums de discussions dans les milieux de travail collaboratif jouent un rôle important dans les processus d'apprentissage collaboratifs entre apprenants. Grâce à l'interaction, l'apprenant devient plus actif et s'engage dans le partage d'informations et de points de vue [Erlin, 2009].

Les forums de Discussion constituent une excellente plate-forme pour l'apprentissage par interaction entre apprenants. Elles sont basées sur l'apprenant qui est le centre de l'environnement d'apprentissage et orientée vers le développement des compétences et la pensée critique [Desai, 2008]. Les interactions entre apprenants fournissent un large spectre d'apprentissage, en plus de l'orientation de tuteur, les apprenants peuvent partager des idées et des points de vue des autres apprenants. La force de ces forums est basée sur les apprenants impliqués dans l'activité de l'apprentissage, encourager l'interaction entre eux, ce qui permet à l'apprenant de jouer un rôle majeur dans le processus d'apprentissage et l'enseignant qui est l'animateur de l'activité [Desai, 2008 ; Silva, 2004].

Il y a quelques désaccords concernant l'utilisation des forums de discussions comme outil d'apprentissage et de leur influence sur la réalisation d'un apprentissage significatif. Certains chercheurs le considèrent comme un très bon outil de collaboration et d'autres pensent que leur utilisation effective et significative est minime [Kay, 2006a]. La raison principale derrière cette divergence pourrait être l'absence d'un outil solide et complet pour évaluer les forums, qui devrait être cohérent, intégré et basé sur des concepts théoriques [Kay, 2006b]. Basé sur une révision approfondie de la recherche précédente, Kay utilise les variables suivantes pour évaluer les forums traditionnels de discussion: fil de discussion, la position du message dans le fil de discussion, l'auteur (l'apprenant par rapport à l'éducateur), la clarté du domaine, moment de la publication, de déchéance temps écoulé depuis le précédent message, le nombre de fois qu'un message a été lu, nombre de mots, le but principal, la qualité du message, le niveau de complexité de la matière, le niveau de connaissances et le niveau de l'utilisation et la transformation des ressources externes. Ces variables ont prouvé leur efficacité dans l'évaluation des 12 domaines clés du forum de discussion. Kay affirme que ce type de mesure est essentiel pour améliorer notre compréhension de la ligne des forums de discussion.

2.1.5.2 Une revue de la littérature

Les chercheurs classent les forums dans un cadre éducatif en trois grandes catégories :

- Mise en place de nouvelles fonctionnalités ;
- Développement d'outils d'analyse (a posteriori) pour les chercheurs ;
- Calcul et visualisations de différentes caractéristiques pour les participants (apprenants et ou tuteurs) en cours d'activité.

Une première catégorie de recherches s'intéresse aux fonctionnalités susceptibles de faciliter la collaboration et l'apprentissage sur les forums. George [George, 2005] étudie comment favoriser les discussions autour d'une activité en y rapprochant le forum. Le projet CONFOR (CONtextual FORum) vise à déterminer comment rapprocher les activités de discussion des activités d'apprentissage avec des outils de communication adaptés [Géorge, 2003]. CONFOR intégré ainsi automatiquement la partie du forum concernant un cours en ligne, sur la même page que le cours. Reyes [Reyes, 2003] cherche à résoudre les incohérences interactionnelles lors des discussions. Son application permet à un utilisateur voulant répondre, de sélectionner précisément à quelle partie de quel message veut faire référence dans sa réponse. La plate-forme collaborative FLE3 [Kligyte, 2001] est basée autour d'un forum amis chaque apprenant possède aussi une zone personnelle, une sorte de portfolio ou il peut classifier les liens, messages ou fils de discussion qu'il souhaite conserver et se créer ainsi un extrait personnel du forum. Enfin, le Knowledge Forum de Scardemalia et Bereiter [Scardemalia et Bereiter, 1994] donne la possibilité aux utilisateurs de restructurer les messages en les classant, les commentant où les modifiant. Les fonctionnalités du forum se rapprochent alors de celle d'un Wiki et le forum sert autant de lieu de discussion que de base structurée de connaissances élaborées de manière collaborative.

Une seconde catégorie de recherche se penche sur des analyses a posteriori des forums, portant notamment sur le contenu des messages. Les analyses linguistiques que nous avons trouvées, portent sur la catégorisation des messages d'après des grilles a posteriori (par exemple, [Rourkeet al.,2001]). Mais la diversité des forums et de leurs contextes semble poser des problèmes lors de l'analyse de contenu pour généraliser les résultats et pour suivre une méthodologie de recherche indépendante du contexte [Henri et Charlier, 2005]. De plus, si ces grilles peuvent être utiles pour le chercheur, leur remplissage ne peut être que difficilement automatisé [Desjardins, 2002].

Une dernière catégorie de recherches s'intéresse à l'analyse automatique des forums, souvent par l'étude de la participation ou des interactions. Par exemple, le projet Netscan [Fiore et al., 2002] analyse les newsgroups de USENET et fournit de nombreux calculs de

participation par utilisateur ou par newsgroup. Parmi ces recherches, certaines se focalisent sur l'intérêt qu'il peut y avoir à présenter le résultat des analyses automatiques directement aux participants des forums analysés. L'hypothèse suivi, est de fournir aux apprenants les différents indices caractérisant leur participation (au plan quantitatif et qualitatif) qui peut les aider [Bratitsis & dimitracopoulou, 2005]. Le groupe de recherche ICALTS (interaction & collaboration analysis supporting Teacher & students' Self-regulation) présente, dans son état de l'art, plusieurs outils pour analyser les interactions sur les plates-formes collaboratives comme par exemple les réseaux sociaux qui représentent chaque acteur par un point puis relient les points entre eux selon l'intensité ou la fréquence des interactions [Martinez et al., 2003 ; Reffat & Chanier, 2003].

Les difficultés d'analyse que posent les forums ne sont pas résolues, surtout si l'on souhaite une analyse automatique pour traiter de grands volumes de contributions [Véronis et Guimier, 2004]. C'est le cas des forums libres, encore peu étudiés, et des forums s'étendant sur une longue durée.

2.1.5.3 Forum de discussion dans un cadre d'évaluation

Les forums de discussion ont souvent été utilisées avec succès comme outil de communication dans des environnements d'apprentissage en ligne pour faciliter l'interaction entre les étudiants et le partage de leurs connaissances [Rovai, 2002 ; Bradshaw et Hinton, 2004 ; Berner, 2003]. Il existe différents niveaux de participation à des forums de discussion. Tout d'abord certains sont "indépendant" c'est à dire qu'il suffit de lire les messages et ne participent pas. Ils peuvent apprendre en lisant les messages et en intégrant les idées dans leurs missions [Guzdial et Carroll, 2002]. Deuxièmement certaines personnes lisent les messages est les traitent comme un panneau d'affichage leur propre position et avoir une interactivité limitée. Troisièmement, la participation et l'interactivité [Ho, 2002] pour l'apprentissage où la collaboration facilite la réalisation de bons résultats d'apprentissage. Sur ce point, quelques auteurs, y compris Berner [Berner, 2003] et Laurillard [Laurillard, 2002], notent que la participation est plus active si une sorte d'évaluation est lié. Klisc, McGill et Hobbs [Klisc, McGill et Hobbs, 2009] suggérée que l'incorporation de l'évaluation de la participation a un impact positif sur les résultats d'apprentissage. En effet, si les cours sont entièrement ou partiellement en ligne, Burkett, Leard et Spector [Burkett, Leard et Spector, 2004], Leh (Leh, 2002] et Seo [Seo, 2007] ont tous indiqué comment les points de qualité pourraient être utilisés comme une incitation à accroître la participation chez les apprenants.

Toutefois, l'évaluation de la participation à des discussions asynchrones des élèves est un défi majeur et difficile pour les instructeurs [Liu, 2007]. La principale préoccupation est de savoir comment évaluer pour prendre en compte l'évaluation. Cette question de l'évaluation de participation des élèves à la discussion en ligne a été un sujet controversé parmi les éducateurs et les chercheurs en éducation en ligne [Bonk et Dennen, 2003]. Il y a une certaine littérature à cet égard ; et il y a un manque d'études empiriques [Ho, 2002].

Pour l'évaluation de la participation de forum de discussion pour travailler efficacement, il doit y avoir un cadre globalement défini qui peut aider clairement les évaluateurs et les étudiants. Avoir un cadre global peut aussi agir comme un guide pour les participants et les tuteurs. Brannon et Essex [Brannon et Essex, 2001] ont affirmé la nécessité pour les protocoles et les exigences pour une communication claire, et a suggéré que la poursuite du développement d'un cadre d'évaluation innovant est nécessaire pour améliorer la qualité des contributions à une discussion en ligne. Une rubrique qui décrit explicitement les niveaux de réponses va stimuler l'apprentissage en remettant en cause les élèves à réfléchir et penser de façon critique, plutôt que de soumettre des déclarations de base de la compréhension et de la simple opinion [Rourke, 2001].

2.1.5.4 L'analyse automatique des forums

L'analyse automatique de forums a fait l'objet de plusieurs études, cette analyse pose de grandes difficultés dans les approches classiques qui sont fondées sur la consultation de dictionnaires [Véronis et Guimier 2004; Véronis et Guimier, 2006]. Les approches d'analyse actuelles basées sur l'exploitation du lexique pour extraire des mots-clés en vue d'une indexation ou du résumé se heurtent à la néologie, la néographie et les erreurs associées [Klaas, 2005]. Si l'on cherche à mener une analyse syntaxique classique nécessitant la reconnaissance des mots dans la phrase, les fautes d'orthographe et de ponctuation omniprésentes dans les messages rendent la tâche complexe [Torzek, 2004].

Une revue de la littérature suggère que même si le traitement est réalisé dans un format lisible par une machine, des chercheurs ont prouvé qu'il n'y avait que très peu de preuves de l'utilisation des ordinateurs pour aider à l'analyse des forums de discussions [McKlin et al., 2001]. Il existe d'abord un certain nombre d'outils logiciels qui peuvent être utilisés pour faciliter la tâche de l'analyse de texte. Ces outils d'analyse de texte comprennent Wordnet, WordStat, NUD * IST, HyperQual et général Inquirer [Rourke et al., 2001]. Ces outils sont principalement des systèmes de traitement de texte qui identifient les mots comme des unités

mais à l'exception des problèmes découlant de l'utilisation des alphabets spéciaux ont tendance à être indépendant de la langue. Les outils les plus puissants, permettent aux chercheurs de briser une unité de transcription et d'attribuer les unités à un certain nombre de catégories de codage. Une fois les transcriptions ont été codées, les résultats peuvent être importés dans des programmes statistiques pour une analyse quantitative plus détaillée. La majorité des outils d'analyse automatique de texte est générique et peut être appliquée à un certain nombre de situations d'analyse de texte. Les outils génériques ne viennent pas avec des catégories de mots intégrés qui pourraient être appliquée pour classer les activités cognitives, mais ils s'appuient sur les chercheurs pour créer des catégories de mots.

McKlin et al. [McKlin et al. 2001] ont décrit l'utilisation d'un outil automatisé qui utilise un logiciel de réseau neuronal pour classer les messages à partir d'une transcription de forum de discussion. Ils ont suggéré que l'outil peut finalement être utilisé pour évaluer, guider, diriger et manipuler l'environnement d'apprentissage. L'analyse a été basée sur Garrison et al. [Garrison et al. 2001, 2003] la communauté de modèle d'enquête. Les valeurs de facteur de fiabilité de 84% et 76% lorsque l'étude rapportée par rapport aux résultats de codeurs humains, ce qui suggère que le réseau de neurones a le potentiel de coder avec succès pour identifier des transcrits présence cognitive. L'étude a également suggéré que l'outil serait raffiné pour produire un système qui pourrait être utilisé pour classer les messages de manière fiable dans des catégories spécifiques.

George et al [George, 2006] ont décrit « Une discussion dans le forum peut être considérée comme une succession d'échanges composés d'interventions et constituées elles-mêmes d'actes de langage dont l'un est l'acte directeur ». Si l'on cherche à analyser automatiquement une conversation, il est alors important de repérer ces actes de langage directeurs. Nous présentons dans ce qui suit la théorie des actes de langage et l'analyse des conversations.

2.1.5.5 La théorie des actes de discours et analyse des conversations

L'analyse des conversations est devenue un centre d'intérêt pour la théorie des actes de langage. Ce champ d'application des actes de langages [Searle et al. 1992] doit être interprété comme un mouvement réactif plutôt que comme une extension naturelle du domaine de la théorie des actes de langage. L'objectif de cette section est d'expliciter les principes et contributions des actes de langage à l'analyse de discussion.

a. Acte de langage

L'expression « acte de langage » traduite en anglais speech act. Cette notion a été développée dans la seconde moitié du xxe siècle par les philosophes dits de l'école d'Oxford, tenants d'un courant également connu sous le nom de « philosophie du langage ordinaire ».

Le langage est un outil qui nous permet d'agir. Nous parlons ou écrivons dans le but d'informer, de faire plaisir, d'obtenir un renseignement, de donner un conseil, de convaincre, de séduire, d'impressionner une autre personn, absente ou présente, réelle ou imaginaire.

Donner un renseignement, faire une promesse, poser une question, donner un ordre, avertir, conseiller, sont des manières d'agir à l'égard d'un interlocuteur, réel ou possible, en utilisant le langage : on appelle ces réalisations des actes de langage.

Un acte de langage instaure une relation active entre les interlocuteurs. Les actes de langage sont gouvernés par les règles de la vie en société.

Ex. : Bonjour, Monsieur. Je voudrais…

Un acte de langage demande la participation de l'interlocuteur. L'acte de langage (poser une question, donner un conseil…) n'est efficace que si l'interlocuteur le reconnaît comme tel et l'accepte.

Ex. : *Je voudrais que vous me fassiez la clef du bonheur. — Volontiers, mon petit.* Si l'interlocuteur ne répond pas à la demande, l'acte de langage est manqué.

La pragmatique linguistique s'est développée à partir de la théorie des actes de langage. Cette théorie montre que la fonction du langage n'est pas essentiellement de décrire le monde, mais aussi d'accomplir des actions. L'initiateur de cette théorie est le philosophe britannique Austin dans son ouvrage : How to do things with words, elle est développée par J.-R. Searle dans deux ouvrages. Les Actes de Langage, et Sens et expression. Le développement le plus récent de la pragmatique linguistique est la pragmatique cognitive (issue de la théorie de la pertinence de Sperber et Wilson) qui réduit l'importance des actes de langage et qui simplifie la théorie.

b. Types d'actes de parole

Pretheoretically, nous pensons à un acte de communication, linguistique ou autre, comme un acte de s'exprimer. Cette idée plutôt vague peut être rendue plus précise si nous obtenons des résultats plus précis sur ce qui est exprimé. Prenons le cas d'une excuse. Si vous prononcez, « [Je suis] désolé, je n'ai pas rappelé » et l'intention ceci comme une excuse, vous exprimez des regrets pour quelque chose, dans ce cas, pour ne pas retourner un appel

téléphonique. Une excuse est communicative, car il est destiné à être pris comme l'expression d'une certaine attitude, dans ce cas regret. Il succède en tant que telle si elle est ainsi prise. En général, un acte de communication réussit si elle est prise comme prévu. Autrement dit, il doit être entendu ou, selon les mots d'Austin, « produire absorption ». Avec des excuses, cette question du destinataire reconnaissant l'intention de l'orateur soit exprimant son regret pour quelque acte ou d'une omission. L'utilisation d'un dispositif spécial, comme le performatif « Je m'excuse » peut bien sûr faciliter la compréhension (compréhension est corrélatif de la communication), mais en général ce n'est pas nécessaire. Le succès de la communication est réalisé si le locuteur choisir bien ses mots de manière à ce que l'auditeur apprécie la discussion.

En disant quelque chose qu'on entend généralement plus que juste de communiquer- se faire comprendre est destiné à produire un certain effet sur l'auditeur. Cependant, notre vocabulaire des actes de langage peut occulter ce fait. Quand on s'excuse, par exemple, on a l'intention non seulement d'exprimer des regrets, mais aussi de demander pardon. Cherchant le pardon est, à proprement parler, distinct de s'excuser, même si un énoncé est l'accomplissement d'un acte de deux types. Comme une excuse, l'énoncé réussit si elle est prise en exprimant son regret pour l'acte en question ; comme un acte de la demande de pardon, il réussit si le pardon est ainsi obtenu. Les actes de discours, étant perlocutoire ainsi qu'illocutoire, ont généralement un but inavoué, mais ils se distinguent principalement par leur type illocutoire, comme affirmant, en demandant, prometteur et s'excuser, qui à son tour se distinguent par le type d'attitude exprimée. L'acte perlocutoire est une question d'essayer d'obtenir l'auditeur pour former une attitude corrélative et dans certains cas, à agir d'une certaine façon.

Déclarations, demandes, des promesses et des excuses sont des exemples des quatre grandes catégories d'actes illocutoires de communication : constatifs, directives et remerciements. Il s'agit de la nomenclature utilisée par Kent Bach et Michael Harnish, qui développent une taxonomie détaillée dans laquelle chaque type d'acte illocutoire est individuée par le type d'attitude exprimé (dans certains cas, il existe des contraintes sur le contenu). Ils adoptent le terme "reconnaissance", sur "la behabitive 'Austin et Searle « expressive », pour les excuses, salutations, félicitations, etc, qui expriment une attitude en ce qui concerne l'auditeur qui est occasionnée par un événement qui est étant ainsi reconnu, souvent dans la satisfaction d'une attente sociale. Voici des exemples variés de chaque type :

- **Constatifs :** Affirmer, alléguant, annonçant, en répondant, en attribuant, en prétendant, de classement, concordante, confirmant, conjecturant, niant, en désaccord,

la divulgation, en contestant, identifier, d'informer, en insistant, la prévision, le classement, l'information, indiquant, stipulant, etc.

- **Directives :** Conseiller, admonesté, demander, mendier, rejetant, excusant, interdisant, instruire, commande, permettant, demandant, exigeant, suggérant, exhortant, avertissement.
- **Remerciements :** excuses, féliciter, salutation, remercier, accepter (reconnaissant un accusé de réception) Bach et Harnish précisent la corrélation entre le type d'acte illocutoire et le type d'attitude exprimée. Pour les types d'actes qui se distinguent par le type d'attitude exprimé, il n'est pas nécessaire d'invoquer la notion de convention d'expliquer comment il peut réussir. L'acte peut réussir que si l'auditeur reconnaît l'attitude est exprimée, comme la croyance dans le cas d'une déclaration et d'un désir dans le cas d'une demande.

A partir du nombre de messages postés dans les forums relatifs à chaque catégorie (question, clarification, argumentation, etc). Cette information permettra au système d'identifier l'acte de langage le plus utilisé par chaque apprenant ce qui renseigne le tuteur sur le profil comportemental de l'apprenant.

2.1.6 Profils Apprenants

2.1.6.1 Introduction

Dans le domaine de la recherche sur les EIAH, le terme le plus général utilisé pour représenter des renseignements personnels sur un apprenant est le « modèle de l'apprenant ». Cependant, nous préférons d'utiliser "profil de l'apprenant » dans le même sens que [Keenoy, 2004 ; Vassileva, 2003] ou [Villanova-Oliver, 2002]. Nous définissons un profil de l'apprenant, des informations concernant un apprenant ou un groupe d'apprenants, collectées ou déduites d'une ou plusieurs activités pédagogiques, informatisés ou non. Les informations contenues dans le profil de l'apprenant peuvent concerner ses connaissances, ses capacités aux conceptions ou son comportement [Daubias, 2005]. Si la notion de profils d'apprenants est maintenant bien connue et intégrée dans de multiples EIAH (Environnements Informatiques pour l'Apprentissage Humain), la prise en compte de l'évolution dans ces profils ne bénéficie pas des mêmes avancées.

Un profil d'apprenant permet de rassembler des informations sur les connaissances, les compétences ou encore le comportement d'un apprenant, informations rassemblées à l'issue d'une activité pédagogique, qu'elle soit informatisée ou non.

2.1.6.2 Profils d'apprenants

Il existe de nombreux profils des apprenants dans le processus de l'éducation, de l'école primaire, à l'université et à la formation continue. Ces profils peuvent provenir de sources différentes comme les EIAH ou la source papier - crayon, ils concernent tous les sujets et tous les niveaux scolaires. Nous vous proposons ici la définition de profil d'un apprenant, avant de montrer la diversité de cette notion.

a. Définition de Profil des apprenants.

Le profil d'un apprenant peut être définie comme l'information concernant un élève ou un groupe d'apprenants, collectées ou déduites d'une ou plusieurs activités pédagogiques, informatisés ou non. Les informations contenues dans le profil de l'apprenant peuvent concerner ses connaissances, ses capacités, aux conceptions ou son comportement.

Nous préférons parler du profil de l'apprenant que le modèle de l'apprenant, parce que les informations contenues dans le profil de l'apprenant sont une vue partielle et subjective de la connaissance de l'apprenant et non pas une vue exacte. Les profils des apprenants peuvent provenir de différentes sources, des personnes différentes peuvent être à l'origine de leur création et ces profils contient différents types d'informations.

b. Représentation de Profils des apprenants.

Plusieurs observations peuvent être utilisées pour les mêmes informations contenues dans le profil de l'apprenant. Par exemple, une des informations associant des marques avec des capacités peuvent être représentés sous forme d'un texte, d'un graphique numérique, d'une forme hiérarchique [Hartley, 2002], ou sous forme d'un graphe conceptuels [Zapata-Rivera, 2001 ; Dimitrova, 2002].

c. Les dimensions des profils d'apprenants

Les profils d'apprenants peuvent tout d'abord concerner différents sujets (cf. 1 sur le tableau 2.3) : un unique apprenant ou un groupe d'apprenants (un binôme, une classe) [Kosba et al., 2005; Upton et al., 2009]. Ils peuvent provenir de l'analyse d'activités de différents types du point de vue de la place de la collaboration (cf. 2 sur le tableau 2.3) [Upton et al., 2009] et de la distance dans ces activités (cf. 3 sur le tableau 2.3) : collaboratives ou non, présentielles, semi-distancielles ou distancielles. Ces activités peuvent être de disciplines variées (cf. 4 sur le tableau 2.3) et correspondre à différents niveaux (cf. 5 sur le tableau 2.3),

scolaires, universitaires ou de formation continue. Les profils d'apprenants peuvent être créés à la demande de différents initiateurs (cf. 6 sur le tableau 2.3) : l'enseignant, afin de suivre l'évolution de l'apprentissage de ses élèves dans l'année ; l'institution, pour suivre celle de l'ensemble des apprenants (par exemple le collège ou le ministère [M.E.N._EducEval, 2008] dans un contexte scolaire, l'entreprise dans un contexte de formation continue); ou encore l'apprenant lui-même afin de suivre l'évolution de ses connaissances. De plus, les profils d'apprenants peuvent être constitués par différents créateurs (cf. 7 sur le tableau 2.3) : certains sont constitués de façon automatique ou semi-automatique par un logiciel comportant un modèle de l'apprenant, d'autres sont issus des pratiques des enseignants, pouvant s'appuyer sur des documents de référence (programmes scolaires ou référentiels de compétences), le plus souvent sous forme papier, mais parfois sous forme numérique (tableur par exemple), d'autres encore sont établis par l'apprenant lui-même, en situation d'autoévaluation, que ce soit sous forme papier-crayon ou logicielle [Cogne et al., 1998; Eyssautier et al., 2004].

	Dimension	**Definition**	**Exemples**
1	Sujet	Acteur humain concerné par le profil	Apprenant, binôme, groupe d'apprenants, classe
2	Collaboration	Place de la collaboration dans les activités servant de base au profil	Activités individuelles, collaboratives
3	Distance	Place de la distance dans les activités servant de base au profil	Activités présentielles, semi-distancielles,
4	Discipline	Discipline des informations contenues dans le profil	Mathématiques, français, conjugaison, musique
5	Niveau	Niveau scolaire du sujet	Maternelle, primaire, collège, lycée, université, formation continue
6	Initiateur	Acteur humain à l'origine de la décision de création du profil	Enseignant, institution, apprenant
7	Créateur	Acteur constituant le profil	Enseignant, EIAH, apprenant
8	Destinataire	Acteur exploitant le profil	EIAH, enseignant, apprenant, famille, Institution, chercheur
9	Temps	Prise en compte du temps	Période concernée, durée de l'activité
10	Evolution	Evolutivité des profils	Profil figé, modifiable, évolutif
11	Type	Type d'informations contenues dans le profil	Profil d'utilisation, d'apprentissage, de caractéristiques
12	Nature	Nature des informations contenues dans le profil	Connaissances, méta-connaissances, compétences, stratégies, conceptions, liens entre informations, motivation, comportement d'apprentissage
13	Evaluation	Type d'échelle utilisée pour évaluer	Note, taux de réussite, indication textuelle, appréciation
14	Représentation interne	Représentation utilisée par le système pour manipuler les profils	Couples attribut/valeur, faits, règles, réseau bayésien
15	Représentation externe	Représentation utilisée par le système pour stocker le profil	Liste de valeurs, couples attribut/valeur, texte structuré, texte balisé
16	Visualisation	Représentation utilisée pour présenter le profil à ses destinataires	représentation textuelle, numérique, graphique, symbolique
17	Norme	Norme ou standard pédagogique	LOM, SCORM, PAPI, IMS-LIP

		respecté	
18	Format	Format de stockage du profil	Texte, tableur, base de données, XML
19	Plateforme	Plateforme informatique compatible	Web, Unix, Windows, MacOs, platform mobile
20	Dispositif	Type de dispositif de visualisation du profil	Papier, ordinateur, téléphone intelligent, tablette tactile

Tableau 2.3 : Les 20 dimensions des profils d'apprenants [Cogne et al., 1998 ; Eyssautier et al., 2004].

2.2 Systèmes Multi-Agents (SMA)

2.2.1 Introduction

Rappelons que l'objectif de nos travaux de recherche est de développer un système d'assistance au tuteur qui permet d'analyser automatiquement les conversations pour déterminer les profils comportementaux des apprenants. Ces profils permettront au tuteur de jouer son rôle cognitif, social et technique, au sein d'un système générique adaptable que nous pourrons greffer aux plates-formes existantes.

Ce contexte, dans lequel se situent nos travaux, peut être vu comme un système qui exhibe des caractéristiques de distribution (géographique), de fortes interactions et de complexité des tâches. Ces éléments nous ont amené à utiliser le paradigme des Systèmes Multi-Agents (SMA) qui a montré sa capacité à gérer ce type de problèmes grâce à ses caractéristiques telles que : la réactivité, la poursuite autonome et intentionnelle de buts, l'habileté sociale qui donne lieu à une résolution collective et distribuée de problèmes. De plus, les SMA facilitent la conception des systèmes nécessitant : l'adaptation aux changements de l'environnement, la capacité à intégrer à la fois des agents humains et/ou artificiels.

Les SMA contiennent un ensemble d'agents qui coopèrent, négocient, coexistent et interagissent afin de résoudre un problème. Le principe est que la résolution de problèmes s'opère par un réseau faiblement couplé d'agents qui interagissent et qui individuellement ne pourrait satisfaire les objectifs globaux du système [Durfee, 1989].

Les SMA forment un paradigme qui fédère un grand nombre d'approches. Ce chapitre a pour objectif de présenter les bases de ce paradigme et parmi l'ensemble d'approches possibles celle que nous avons choisi.

Nous commencerons dans cette section par définir les concepts de base des systèmes multi-agent, puis nous examinerons les différentes méthodologies de développement, particulièrement la méthodologie ASPECS que nous avons retenue pour la conception de

notre système. Nous présenterons ensuite la plate-forme JANUS de développement des SMA, qui répond au mieux à nos besoins.

2.2.2 Concepts de base de la technologie multi-agents

Cette section présente quelques définitions sur les concepts de la technologie multi-agents. Pour le concept d'agent, il n'existe pas encore un consensus sur la définition d'un agent. A cause de la relative jeunesse du domaine, les chercheurs du domaine ayant fourni de nombreuses définitions, chacune visant généralement à mieux expliquer leur propre utilisation du concept agent. Nous reprenons dans la suite, certaines de ces définitions pour donner un aperçu de comment les agents et les systèmes multi agents sont considérés par les différents auteurs.

2.2.2.1 Définition d'un agent

Il n'existe pas de consensus sur la définition d'un agent. Les définitions qui suivent sont toutefois couramment utilisées :

- Un agent est une entité logicielle ou physique à qui est attribuée une certaine mission qu'elle est capable d'accomplir de manière autonome et en coopération avec d'autres agents [Briot et all, 2002].
- Un agent logiciel est une entité autonome capable de communiquer, disposant de connaissances et d'un comportement privés ainsi que d'une capacité d'exécution propre. Un agent agit pour le compte d'un tiers (un autre agent, un utilisateur) qu'il représente sans être obligatoirement connecté à lui [Szyperski, 1997 ; Oussalah, 2005].
- Un agent est un programme en interaction avec son environnement, capable de percevoir des informations et de réagir en fonction de celles-ci. La raison de vivre d'un agent est de satisfaire donc à une fonction objective que celle-ci soit personnelle ou issue d'une collaboration avec d'autres agents. En effet, chaque agent peut non seulement communiquer avec les autres agents de son environnement mais aussi offrir ses services à ces derniers.
- Partant de la définition de Ferber, on peut définir un agent logiciel comme système informatique situé dans un environnement, capable de mener de manière autonome (il entreprend des actions en fonction de son état interne sans intervention humaine), des actions sur cet environnement en vue d'accomplir ses objectifs (objectifs qui lui ont été assignés au moment de sa conception) [Ferber, 1995].

Les principales caractéristiques des agents sont :

- **Autonomie** : un agent est capable de prendre des décisions et d'agir sans l'intervention directe d'un tiers (humain ou agent), de contrôler ses propres actions ainsi que son état interne ;

- **Réactivité** : un agent est situé dans un environnement et est capable de percevoir et de réagir par rapport aux changements de l'environnement ;
- **Pro-activité** : un agent est capable d'agir en réponse à son environnement mais également de prendre des initiatives propres pour atteindre son objectif ;
- **Sociabilité** : un agent interagit avec d'autres agents (éventuellement humains) et typiquement est capable de s'engager dans des activités sociales (comme la résolution coopérative d'un problème ou la négociation) ;
- **Sécurité :** La sécurité est une propriété importante, notamment dans le contexte de ce travail, car elle permet de garantir que lorsque l'on interagit avec un agent, que cet agent n'a pas été corrompu par un virus, par de fausses croyances ou par des connaissances qui n'ont pas de sens [Oliveira, 1998] ;
- **Apprentissage :** L'apprentissage est une propriété qui fournit aux systèmes la capacité d'acquérir la compréhension de certains comportements au cours du temps, sans nécessiter que ces comportements soient programmés manuellement [Guessoum & all, 2005].
- **Adaptabilité :** L'adaptabilité est la capacité d'un agent de s'adapter à l'environnement dans lequel il est situé [Oliveira, 1998]. Un agent adaptatif est un agent capable de contrôler ses aptitudes (communicationnelles et comportementales) selon l'environnement dans lequel il évolue et selon l'agent avec lequel il interagit [Guessoum, 1996].

2.2.2.2 Définition de Système Multi-agents

Un système multi-agents (avec son acronyme SMA, et MAS pour « multi-agent system » en anglais) est plus immédiate : « Un ensemble d'agents qui agissent (et interagissent) dans un environnement commun ». La définition usuelle du terme système : « un ensemble organisé d'éléments ». Cela signifie que dans un système multi-agent, il existe une ou plusieurs organisations qui structurent les règles de cohabitation et de travail collectif entre agents (définition des différents rôles, partages de ressources, dépendances entre tâches, protocoles de coordination, de résolution de conflits, etc.). Dans un même système, il existe en général plusieurs organisations et un même agent peut appartenir à plusieurs simultanément.

[Moulin, 1996] définit un SMA comme un ensemble faiblement interconnecté d'agents qui travaillent ensemble pour résoudre un problème en s'appuyant sur les capacités individuelles de chaque entité. Dans la plupart des situations réelles, l'agent n'est pas seul dans son environnement, il y a d'autres agents présents autour de lui. Ces agents doivent interagir entre eux pour effectuer leurs tâches. De tels systèmes sont appelés « systèmes multiagents » » et ils possèdent les caractéristiques principales [Jennings & all, 1998] suivantes :

- Chaque agent a des informations ou des capacités de résolution de problèmes incomplètes,
- Donc chaque agent a un point de vue limité ;
- Il n'y a pas de contrôle global du système ; les données sont décentralisées ;
- Les calculs sont asynchrones.

Selon [Ferber, 1995] : On appelle un système multi-agents (SMA) comme étant un système composé des éléments suivants :

1) Un environnement E, c'est-`a-dire un espace disposant généralement d'une métrique.
2) Un ensemble d'objets O. Ces objets sont situés, c'est-`a-dire que, pour tout objet, il est possible, à un moment donné, de lui d'associer une position dans E. Ces objets sont passifs, c'est-à-dire qu'ils peuvent être perçus, crées, détruits et modifiés par les agents.
3) Un ensemble A d'agents, qui sont des objets particuliers (A _ O), lesquels représentent les entités actives du système.
4) Un ensemble de relations R qui unissent des objets (et donc des agents) entre eux.
5) Un ensemble d'opérations Op permettant aux agents de A de percevoir, produire, consommer, transformer et manipuler des objets de O.
6) Des opérateurs chargés de représenter l'application de ces opérations et la réaction du monde à cette tentative de modification, que l'on appellera les lois de l'univers.

Comme définition de travail, nous dirons qu'un SMA est un système composé d'un ensemble d'agents, dont les agents communiquent entre eux par le biais de messages. Ces agents travaillant ensemble (coopérativement ou parallèlement) pour remplir les objectifs du système.

2.2.3 Typologie des agents

L'architecture d'un agent désigne l'ensemble des structures de données et des processus internes à un agent lui permettant de prendre une décision consistant à choisir une action en vue de modifier l'environnement.

Les agents sont des entités intelligentes, c'est-à-dire capables de résoudre certains problèmes par eux-mêmes [Ferber, 1995], ou bien faut-il les assimiler à des êtres très simples réagissant directement aux modifications de l'environnement ?

La communauté multi-agents distingue deux grandes catégories d'architectures d'agents : celles des Agents cognitifs et celles des agents réactifs :

- Les systèmes délibératifs, dits aussi cognitifs,
- Les systèmes réactifs.

Ces deux conceptions ont donné lieu à deux écoles de pensée. La distinction qu'on peut faire entre cognitif et réactif tient essentiellement du processus décisionnel chez l'agent et de la représentation du monde dont il dispose. Si l'individu est doté d'une représentation symbolique du monde à partir de laquelle il est capable de formuler des raisonnements, on parle d'agent cognitif tandis que s'il ne dispose que d'une représentation sub-symbolique, c'est-à-dire limitée à ses perceptions, on parle d'agent réactif [Ferber, 1995].

Cette distinction cognitif/réactif correspond à deux écoles de pensée des SMA. La première, l'école « cognitive » est la plus représentée dans le domaine que l'on appelle intelligence artificielle distribuée (IAD) car elle trouve son origine dans la volonté de faire communiquer et coopérer des systèmes experts classiques, avec une perspective plus sociologique [Bond & Gasser, 2014 ; Demazeau & Müller, 1991 ; Chaib-Draa & all, 1992 ; Briot & all, 2006]. La deuxième est l'école « réactive » prétend au contraire qu'il n'est pas nécessaire que les agents soient intelligents individuellement pour que le système ait un comportement global intelligent [Deneubourg & all, 1991 ; Ferber & Drogoul, 1992]. Des mécanismes de réaction aux événements, ne prenant en compte ni une explicitation des buts, ni des mécanismes de planification, peuvent alors résoudre des problèmes qualifiés de complexes [Deneubourg & all, 1991 ; Ferber & Drogoul, 1992].

2.2.3.1 Agent cognitif

Un agent cognitif dispose d'une base de connaissances comprenant l'ensemble des informations et des savoir-faire nécessaires à la réalisation de sa tâche et à la gestion des interactions avec les autres agents et avec son environnement. On dit aussi que les agents sont « intentionnels », c'est-à-dire qu'ils possèdent des buts et des plans pour décider de leurs actions.

L'une des architectures cognitives les plus connues est l'architecture BDI : Belief (Croyance), Desire (Désir) et Intention (Intention) [RAO & Georgeff, 1991].

Les agents se basent donc sur ces trois aspects pour choisir leurs actions. Dans ce cadre, [Wooldridge, 1999] propose une architecture ayant sept composantes, telles que présentées sur la figure 2.1.

- Un ensemble de croyances courantes, représentant les informations que l'agent possède à propos de son environnement courant ;
- Une fonction de révision des croyances, qui prend les entrées des capteurs et les croyances actuelles de l'agent et qui détermine un nouvel ensemble de croyances;

- Une fonction de génération des options, qui détermine les options disponibles pour l'agent (i.e. ses désirs), en se basant sur les croyances courantes de l'agent à propos de son environnement et sur ses intentions courantes ;
- Un ensemble de désirs, représentant les options disponibles à l'agent ;
- Une fonction de filtre, qui représente le processus de délibération de l'agent et qui détermine les intentions de l'agent en se basant sur ses croyances, ses désirs et ses intentions courantes ;
- Un ensemble d'intentions courantes, représentant le centre d'attention actuel de l'agent, c'est-à-dire les buts envers lesquels il s'est engagé et envers lesquels il a engagé des ressources ;
- Une fonction de sélection des actions, qui détermine l'action à effectuer en se basant sur les intentions courantes de l'agent.

En résumé, un agent BDI "Croyance-Désir-Intention" doit donc mettre à jour ses croyances avec les informations qui lui proviennent de son environnement, décider quelles options lui sont offertes, filtrer ces options afin de déterminer de nouvelles intentions en se basant sur ses intentions.

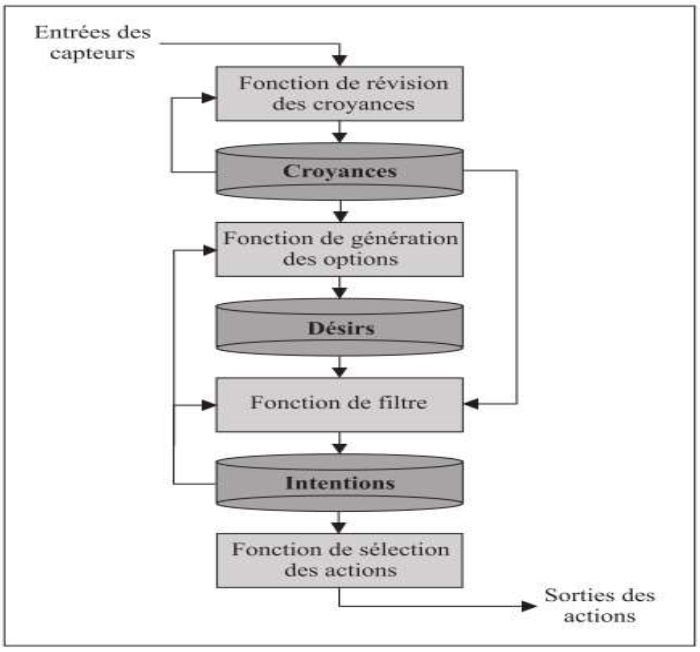

Figure 2-1 : Diagramme d'une architecture BDI [Wooldridge, 1999]

2.2.3.2 Agent réactif

Comme son nom l'indique, un agent est réactif s'il répond de manière opportune aux changements de son environnement. Autrement dit, un tel agent ne fait ni délibération ni planification, il se contente simplement d'acquérir des perceptions et de réagir à celles-ci en

appliquant certaines règles prédéfinies (Figure 2.2). Étant donné qu'il n'y a pratiquement pas de raisonnement, ces agents peuvent agir et réagir très rapidement.

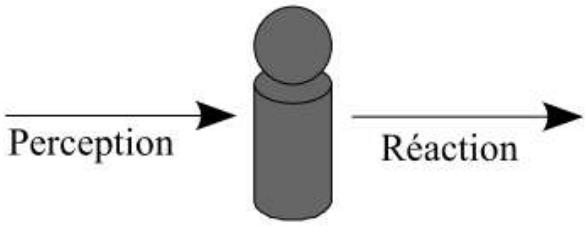

Figure 2-2 : Agent réactif [Drogoul, 1993]

On trouve de nombreux agents réactifs dans la littérature, mais peu le sont purement. Pour la plupart, la réactivité n'est qu'une caractéristique et n'interdit pas l'action intentionnelle.

Un exemple d'architecture couramment utilisée pour concevoir des agents réactifs est celui de l'architecture proposée par [Brooks & Connell, 1986]. Le principe de cette architecture est de concevoir le processus décisionnel en considérant un ordre de priorité entre des modules qui représentent les tâches pouvant être accomplies par l'agent. Les conditions d'activation de chaque module sont alors évaluées suivant cet ordre, de manière à sélectionner la tâche à effectuer en priorité.

2.2.3.3 Agents à réflexes simples

Ce type d'agent agit uniquement en se basant sur ses perceptions courantes. Il utilise un ensemble de règles prédéfinies, du type Si condition Alors action, pour choisir ses actions. Comme on peut le constater, ces règles permettent d'avoir un lien direct entre les perceptions de l'agent et ses actions. Le comportement de l'agent est donc très rapide, mais peu réfléchi.

À chaque fois, l'agent ne fait qu'exécuter l'action correspondant à la règle activée par ses perceptions.

2.2.3.4 Agents conservant une trace du monde

Ce type d'agent doit maintenir des informations internes sur l'état du monde dans le but d'être capable de distinguer deux situations qui ont des perceptions identiques, mais qui, en réalité, sont différentes. Pour que l'agent puisse faire évoluer ses informations internes sur l'état du monde, il a besoin de deux types d'information. Tout d'abord, il doit avoir des informations sur la manière dont le monde évolue, indépendamment de l'agent. Il doit avoir

ensuite des informations sur la manière dont ses propres actions affectent le monde autour de lui.

2.2.3.5 Agents délibératifs

Un agent délibératif est un agent capable de résoudre des problèmes complexes. Il est ainsi capable de raisonner sur une base de connaissances, de traiter des informations diverses liées au domaine d'application et des informations relatives à la gestion des interactions avec d'autres agents et avec l'environnement [Guessoum, 1996].

Ces agents maintiennent une représentation interne de leur monde et un état mental explicite qui peut être modifié par un raisonnement symbolique [Müller, 1996].

Les SMAs délibératifs ont deux problèmes majeurs [Guessoum, 1996]:

- La traduction de l'univers de l'agent en une description symbolique ;
- La représentation symbolique des informations de l'univers complexe des entités et des processus ainsi que la manière de raisonner sur ces informations.

Les agents délibératifs sont des agents qui effectuent une certaine délibération pour choisir leurs actions. Une telle délibération peut se faire en se basant sur les buts de l'agent ou sur une certaine fonction d'utilité. Elle peut prendre la forme d'un plan qui reflète la suite d'actions que l'agent doit effectuer en vue de réaliser son but.

2.2.4 Propriétés des systèmes multi-agents

2.2.4.1 Interactions entre agents

Les systèmes multi-agents ont surtout l'avantage de faire intervenir des schémas d'interaction sophistiqués. Les agents doivent interagir ensemble pour réaliser leurs tâches. Jacques Ferber donne la définition suivante de l'interaction : « Une interaction est la mise en relation dynamique de deux ou plusieurs agents par le biais d'un ensemble d'actions réciproques. Les interactions sont non seulement la conséquence d'actions effectuées par plusieurs agents en même temps, mais aussi l'élément nécessaire à la constitution d'organisations sociales. » [Ferber, 1995].

Les principales propriétés de l'agent dans SMA est en interaction avec d'autres agents. Ces interactions sont généralement définies comme toute forme d'action exécutée dans le système et qui provoque la modification du comportement d'un autre agent.

Une interaction est une liaison dynamique entre deux ou plusieurs agents à travers un ensemble d'actions réciproques. Les interactions sont exprimées à partir d'une série d'actions. L'interaction peut être décomposée en trois phases :

- La réception d'informations ou la perception d'un changement,
- Le raisonnement sur les autres agents à partir des informations acquises,
- Une émission de message(s) ou plusieurs actions (plan d'actions) modifiant l'environnement. Cette phase est le résultat d'un raisonnement de l'agent sur son propre savoir-faire et celui des autres agents.

Les types d'interactions comprennent la coopération (travailler ensemble pour résoudre un but commun), la coordination (organiser la résolution d'un problème de telle sorte que les interactions nuisibles soient évitées ou que les interactions bénéfiques soient exploitées) ; et la négociation (parvenir à un accord acceptable pour toutes les parties concernées).

Nous allons donc voir maintenant les différentes propriétés qui sont impliquées dans la sociabilité.

a. Coopération

Pour Ferber [Ferber, 1995], pour que plusieurs agents soient dans une situation de coopération, il faut que l'une des deux conditions suivantes soit vérifiée :

1. L'ajout d'un nouvel agent permet d'accroître différentiellement les performances du groupe
2. L'action des agents sert à éviter ou à résoudre des conflits potentiels ou actuels.

La coopération entre agents peut être [Oliveira, 1998] :

- Soit **implicite** et dans ce cas les agents ont un but commun implicite qu'ils doivent atteindre en exécutant des actions indépendantes. Dans ce type de coopération, la communication entre les agents est facultative.
- Soit **explicite** et dans ce cas, les agents exécutent des actions qui leur permettent d'achever non seulement leurs propres buts mais aussi les buts des autres agents.

Ce type de coopération nécessite une communication entre les agents pour prendre connaissance des buts des autres agents.

b. Coordination

Dans un système multi-agents, la coordination des actions des différents agents permet d'assurer une cohérence du système. Il existe plusieurs mécanismes de coordinations parmi lesquels, nous retrouvons : l'organisation, la planification et la synchronisation [Guessoum, 1996 ; Oliveira, 1998]. Nous définissons ces trois mécanismes.

➢ **L'organisation :**

Une organisation représente un groupe d'agents qui travaillent ensemble afin de réaliser une ou plusieurs tâches [Guessoum, 1996]. De nombreux travaux sur l'organisation ont été

proposés. T. Bouron [Bourron, 1992] les a regroupés en deux classes où l'organisation est définie :

- Soit comme une structure externe aux agents et elle est représentée par un objet ou un agent.
- Soit comme un objet abstrait dont la représentation est distribuée parmi les membres de l'organisation.

➢ **La planification :**

La planification multiagents est une autre approche de coordination dans les systèmes à base d'agents. Afin d'éviter des actions conflictuelles ou inconsistantes, les agents construisent un plan multi-agents qui détaille toutes les actions futures et les interactions nécessaires pour atteindre leurs buts [Oliveira, 1998].

➢ **La synchronisation :**

La synchronisation est le "bas niveau" de la coordination où sont implémentés les mécanismes de base permettant aux différentes actions de s'articuler correctement. Elle permet de synchroniser l'enchaînement des actions des différents agents [Ferber, 1995]. J. Ferber répertorie deux types de synchronisation :

- La synchronisation par mouvement utilisée lorsque plusieurs éléments doivent se déplacer ensemble. Dans ce type de synchronisation, il s'agit de coordonner le rythme et le positionnement dans le temps d'actions en fonction des événements qui se produisent ;
- La synchronisation d'accès à une ressource utilisée lorsque plusieurs agents doivent partager une ressource.

c. La négociation

La négociation joue un rôle important dans les activités de coopération en permettant aux personnes de résoudre des conflits qui pourraient mettre en péril des comportements coopératifs.

Durfee et ses collègues [Durfee, 1989] définissent la négociation comme le processus d'améliorer les accords (en réduisant les inconsistances et l'incertitude) sur des points de vue communs ou des plans d'action grâce à l'échange structuré d'informations pertinentes. En général, les chercheurs en intelligence artificiel distribuée utilisent la négociation comme un mécanisme pour coordonner un groupe d'agents. Différentes approches ont été développées en s'appuyant sur la riche diversité des négociations humaines dans divers contextes [Conry, 1991 ; Davis,1983 ; Sycara, 1989].

2.2.4.2 Communication

Dans un SMA, les agents communiquent entre eux en s'échangeant des informations via un langage de communication agent [Wooldridge & Jennings, 1994]. Le langage, le plus utilisé aujourd'hui est KQML (Knowledge Query and Manipulation Language) qui est un langage de haut niveau utilisant une liste de types de messages, appelés performatifs [Ferber, 1995]. Les agents peuvent également communiquer en utilisant d'autres mécanismes tels que le mécanisme de tableau noir [Guessoum 1996].

L'intérêt des langages d'interaction entre agents est de réduire les communications en évitant une description exhaustive des messages ad hoc et une gamme inutilement étendue de protocoles. Ces langages se focalisent essentiellement sur la manière de décrire exhaustivement des actes de communication d'un point de vue syntaxique et sémantique supportant un langage de représentation des connaissances. Toutefois, l'aspect ontologique et l'utilisation de conventions garantissant un comportement collectif cohérent du système et l'aspect conversationnel n'est pas facile à garantir [Drogoul, 1993]. Plusieurs tentatives de normalisation de la communication inter-agents ont été effectuées au sein de la communauté multi-agents ces dernières années.

- **Langage KQML**

KQML (Knowledge Query and Manipulation Language) [Finin & al. 1994] est un langage de communication de haut niveau pour l'échange de l'information, orienté messages et indépendant de la syntaxe du contenu et de l'ontologie applicable. Il est fondé sur la théorie des actes de langage dans le but de permettre aux agents cognitifs de coopérer.

Il est basé sur le fait de pouvoir coder explicitement dans les messages des actes illocutoires en termes de type de message ou " performatives " et repose sur les états mentaux des agents. Le contenu du message échangé est une expression spécifiée en kif (Knowledge Interchange Format) qui utilise le formalisme de la logique de premier ordre.

- **Langage FIPA-ACL**

Sur la base du langage KQML et pour répondre aux critiques faites à KQML, la FIPA (Foundation for Intelligent Physical Agents) a développé un nouveau langage de communication agent appelé FIPA-ACL (Fipa Agent Communication Language) (fIPA00a). Extérieurement, FIPA-ACL ressemble beaucoup à KQML. Il utilise la même syntaxe mais pour la description d'un message, on retrouve quelques attributs de plus comme par exemple:

- Reply-to : indique que les messages de réponse à ce message doivent être adressés à l'agent introduit dans ce paramètre plutôt qu'à l'agent expéditeur.
- Protocol : dénote le protocole d'interaction que l'agent expéditeur a choisi d'utiliser pour le contrôle de la conversation.
- Conversation-id : introduit une expression (un identificateur de conversation) qui permet ensuite d'identifier la séquence des actes de communication en cours composant une conversation.

2.2.5 Modélisation et Implémentation

Un système complexe peut être décomposé en un ensemble de parties en interaction. En adoptant une approche organisationnelle, le comportement global associé au « tout » sera modélisé à l'aide de la notion d'organisation, et ses composantes avec le concept de rôle. Ce système peut être décomposé niveau par niveau, chaque niveau du système pourra à son tour être décomposé sous forme d'organisations et de rôles.

Pour mener à bien cette activité de décomposition hiérarchique d'un système complexe. Il est nécessaire de disposer d'un guide méthodologique pour assister le travail de l'ingénieur. Ce guide est fourni par le processus d'ingénierie logicielle ASPECS (agent-oriented software process for engineering complex systems).

2.2.5.1 ASPECS : un processus d'ingénierie logicielle pour la modélisation et l'implantation de systèmes complexes [Cossentino et al, 2010]

ASPECS est un processus de développement de logiciel basé sur le métamodèle CRIO. Elle s'appuit sur un processus itératif et incrémental comme de nombreuses méthodes d'analyse et de conception SMA [Kolp et al, 2006 ; Seidita, 2010]. Le métamodèle fournit les abstractions et la philosophie utilisées pour modéliser un problème et sa solution. Ce processus expose les étapes à suivre pour obtenir cette solution. Les résultats intermédiaires, associés à chacune des étapes, doivent également être décrits.

ASPECS fournit les moyens nécessaires pour décomposer un système complexe, niveau par niveau, et étudier les relations entre eux. Différents points de vue peuvent être adoptés pour décrire un système, ASPECS tente également de les intégrer dans la conception d'un modèle unique.

D'après [Dastani and Gomez-Sanz, 2005], les applications multi-agents ne pourront effectivement convaincre les industriels que si on comble le fossé qui sépare d'une part l'analyse et la conception des SMA et d'autre part leur implantation. Combler ce fossé passe

d'abord par le développement d'outils qui facilitent une implantation simple, voire directe, des abstractions utilisées dans la phase d'analyse et de conception.

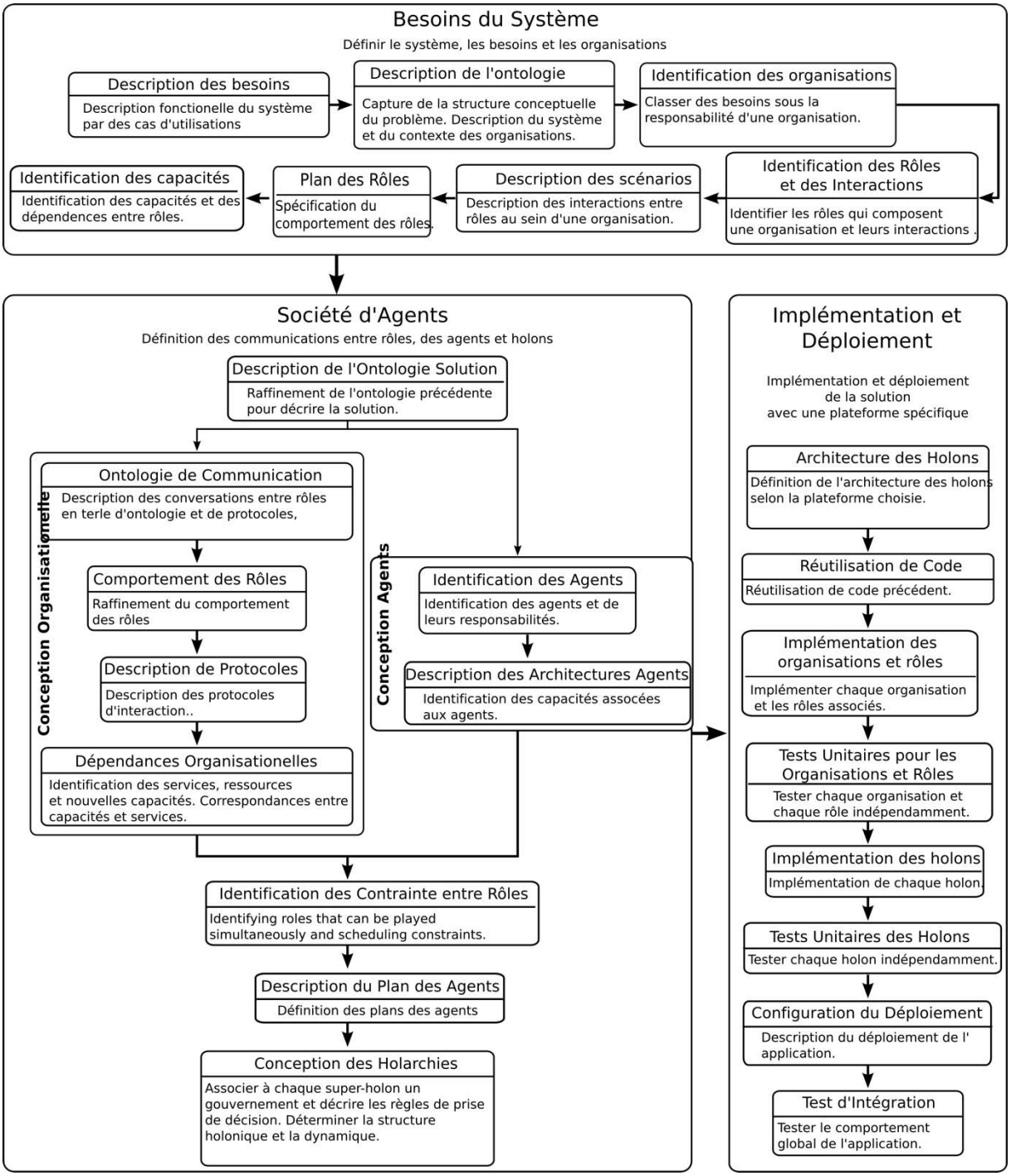

Figure 2-3 : Aperçu du processus ASPECS [Dastani and Gomez-Sanz, 2005]

La figure 2.3 ci-dessus présente l'ensemble des activités du processus ASPECS [Dastani and Gomez-Sanz, 2005]

2.2.5.2 CRIO : Un Metamodèle pour l'analyse et la conception de systèmes complexes

CRIO est un métamodéle dédiée à l'analyse et la conception des systèmes complexe sous une perspective holonique [Cossentino, 2007]. Il cherche à exploiter les propriétés hiérarchiques des systèmes complexes. CRIO fournit un ensemble approprié d'abstractions pour formuler, décrire et identifier un problème et sa solution. Le métamodéle CRIO adopte une approche organisationnelle dont la notion de comportement est l'abstraction qui en constitue la base.

L'identification des organisations doit établir une première décomposition organisationnelle du système et définir les objectifs de chaque organisation. Chacun des buts, identifiés dans la première activité, se voit associer une organisation incarnant le comportement global en charge de le satisfaire ou de le réaliser. Le contexte de chacune de ces organisations est défini par un sous-ensemble des concepts de l'ontologie du problème. Le résultat de cette activité est ainsi incarné par un ensemble de triplets associant des concepts de l'ontologie, un ou plusieurs buts et une organisation. Les organisations, ainsi identifiées, sont directement ajoutées au diagramme de buts, sous la forme de paquetages stéréotypés englobant les buts que ces organisations sont en charge de satisfaire. Cette étape de l'activité d'identification des organisations permet de fixer les objectifs d'un premier ensemble d'organisations.

ASPECS se basant sur un processus itératif, cet ensemble d'organisations peut ensuite être complété au fur et à mesure des itérations, afin de déterminer la hiérarchie organisationnelle représentant le système. Cette décomposition hiérarchique du système se poursuit jusqu'à un niveau où la complexité des comportements (rôles) est suffisamment faible pour être exécutée par des entités considérées comme atomiques et facilement implémentables. Les liens de composition (comportementale et hiérarchique) entre les organisations sont décrits par des contraintes dans les diagrammes de classe (Figure 2.4).

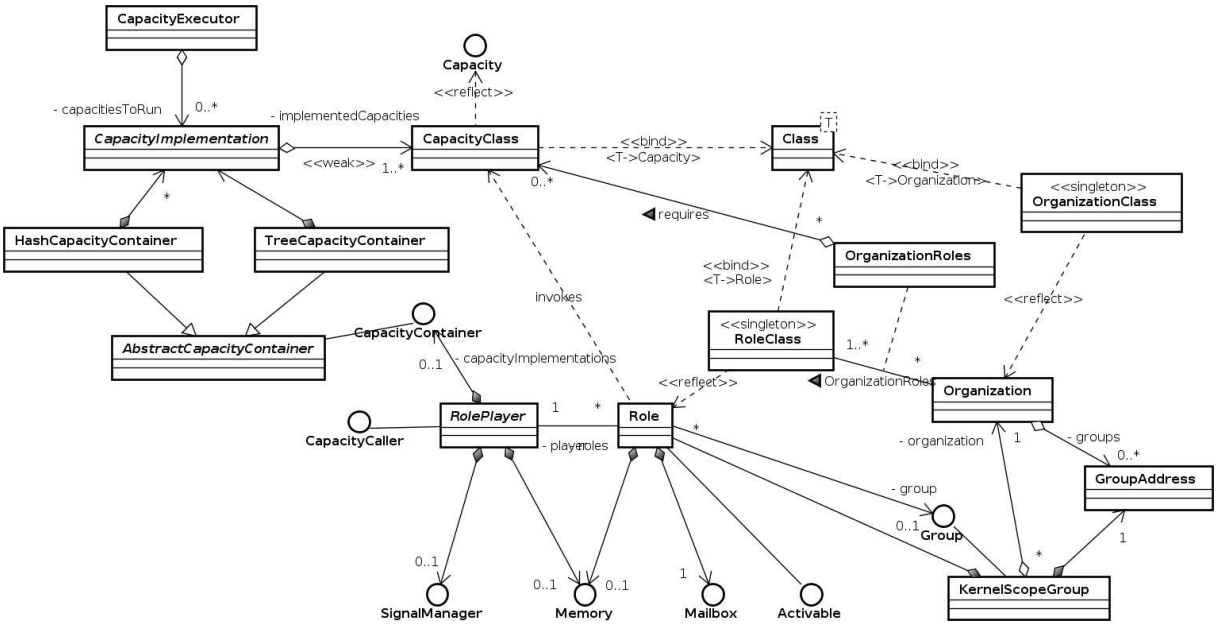

Figure 2-4 : Diagramme de classes UML de Métamodèle CRIO relatif à la Plateforme Janus 0,5 [Cossentino, 2007].

2.2.5.3 Janus : une plate-forme organisationnelle pour l'implantation et la simulation de systèmes complexes

La plate-forme Janus a été implantée de manière à fournir directement les concepts à la base de la conception de systèmes multi-agents holoniques, notamment ceux résultant de la méthodologie ASPECS. En particulier, les concepts d'organisation, de rôle et de holon sont considérés comme des entités de premier ordre.

Le métamodèle de JANUS s'intègre dans le troisième et dernier domaine du métamodèle d'ASPECS : le domaine de la solution. Ce domaine correspond au modèle PSM1 défini dans l'approche MDA, car il est dépendant d'une plate-forme d'implantation particulière.

a. Métamodéle de Janus : Domaine de la solution de CRIO

Les concepts introduits dans ce domaine sont partiellement décrits dans le diagramme UML de la figure 2.5. Le domaine de la solution représente une partie des concepts fournis par la plate-forme logicielle Janus [Lin & all, 2012].

L'organisation est considérée comme une classe (au sens du métamodèle objet) qui regroupe un ensemble de classes de rôles. Une organisation peut être instanciée sous forme de groupes, chaque groupe contenant un ensemble d'instances des différentes classes de rôles, associées à l'organisation qu'il implémente. Le nombre d'instances autorisées pour chaque rôle est spécifié dans l'organisation. Un rôle est local à un groupe et fournit les moyens de communiquer au sein de ce groupe.

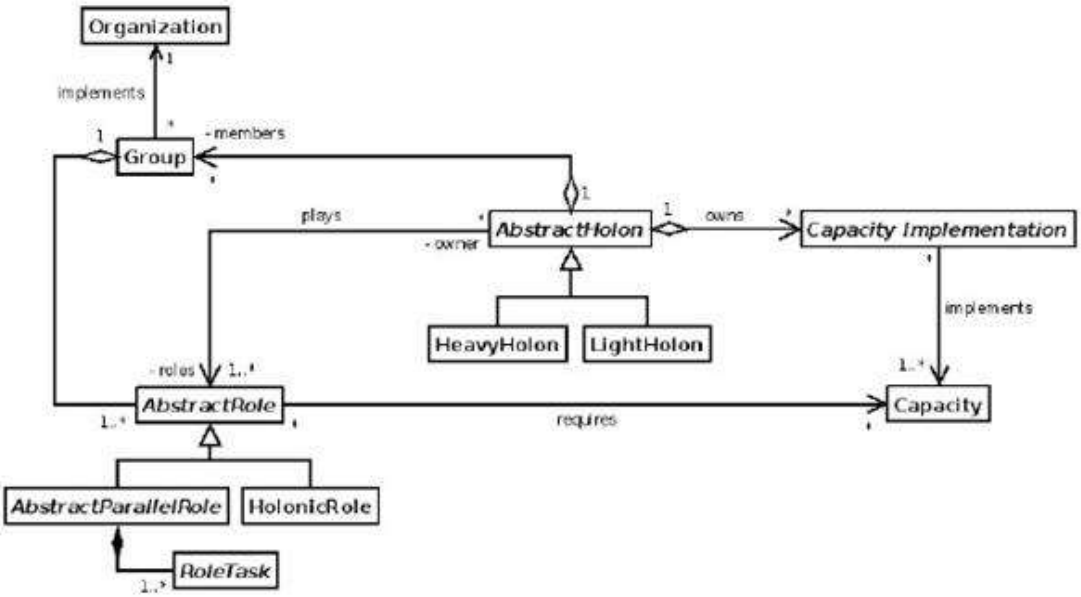

Figure 2-5 : Diagramme UML simplifié de la plate-forme Janus et du domaine de la solution (PSM « *Platform-Specific Model* ») du métamodèle CRIO

Le rôle est considéré comme une classe à part entière, et les rôles sont implantés indépendamment des entités qui les jouent. Une telle implantation facilite la réutilisation des organisations dans d'autres solutions, mais autorise également une grande dynamique pour les rôles.

Dans Janus, un agent est représenté par un holon atomique (ou non composé). Janus définit deux principaux types de holon : "HeavyHolon" et "LightHolon". Un "HeavyHolon" dispose de sa propre ressource d'exécution et peut donc fonctionner de manière indépendante.

Le "LightHolon" est associé aux mécanismes d'exécution synchrone. Cette architecture s'intègre dans un modèle de synchronisation qui définit les différentes politiques d'exécution du système et des holons en charge de leur mise en œuvre.

Un holon peut jouer simultanément plusieurs rôles définis dans plusieurs groupes. Il peut accéder dynamiquement à de nouveaux rôles et libérer des rôles dont il n'a plus l'usage. Lorsqu'un holon accède à un rôle, il obtient une instance de la classe de ce rôle qu'il stocke dans son conteneur de rôles ; respectivement, lorsqu'il libère un rôle, l'instance correspondante est supprimée. La taille, en terme de code, d'un holon est toujours minimale, car il ne contient que le code des rôles qu'il joue à un instant donné. Pour accéder ou libérer un rôle, un holon doit satisfaire les conditions d'obtention ou de libération du rôle et du groupe correspondant ("obtainConditions" et "leaveConditions"). Ce mécanisme offre de

nombreux avantages en terme de sécurité, puisqu'un holon ne pourra accéder au comportement d'un rôle, et donc obtenir le code exécutable correspondant, que s'il valide les conditions d'accès du rôle et celles de son groupe. Pour une même organisation, chacune de ses instances (ou groupes) peut disposer de droits d'accès et de libération spécifiques.

Les conditions d'accès et de libération des rôles sont en revanche définies au niveau de l'organisation.

La notion de capacité permet de représenter les compétences d'un holon. Tout holon dispose, dès sa création, d'un ensemble de compétences de bases, dont la possibilité de jouer des rôles (et donc de communiquer), d'obtenir des informations sur les organisations et les groupes existants au sein de la plate-forme, de créer des holons, et d'obtenir de nouvelles capacités. Tout comme dans le domaine agent du métamodèle CRIO, la capacité permet de faire l'interface entre le holon et les rôles qu'il joue. Le rôle requiert certaines capacités pour définir son comportement, lesquelles peuvent ensuite être invoquées dans l'une des tâches qui composent le comportement du rôle. L'ensemble des capacités requises par un rôle sont spécifiées dans les conditions d'obtention du rôle. Une capacité peut être réalisée de diverses manières, et chacune de ces réalisations concrètes est modélisée par la notion d'implantation de capacité (ou "Capacity Implementation"). Ce concept correspond à la représentation implantatoire de la notion de service au sein du domaine agent.

En plus de ces différents concepts, Janus fournit tout une gamme d'outils pour faciliter le travail du développeur. Les différentes fonctionnalités offertes par Janus seront décrites dans la section suivante.

b. Architecture et noyau de la plateforme Janus

L'architecture complète de la plate-forme Janus est présentée dans la figure 2.6. Janus est développée en Java. Le cœur de la plate-forme est constitué par son noyau qui fournit l'implantation du modèle organisationnel et de la notion de holon. Le noyau a ensuite été étendu pour intégrer le module de simulation et les holons en charge du fonctionnement de la plate-forme et de son intégration avec les applications [Buisson, 2013 ; Rodriguez, 2011].

Les diverses fonctionnalités fournies par le noyau Janus sont décrites ci-dessous :

- **Gestion du modèle organisationnel** Ce module gère les organisations et leurs instanciations sous forme de groupes. Il fournit également les mécanismes d'acquisition, d'instanciation et de libération des rôles, ainsi que les mécanismes d'acquisition et d'exécution des capacités. Cette gestion de l'information organisationnelle est gérée au plus bas niveau afin que tout holon, y compris ceux de la plate-forme, aient accès à ces fonctionnalités. Ce

module est en relation avec le holon noyau en charge du maintien de cette information avec les autres noyaux distants.

- **Gestion des holons** Le noyau fournit également l'ensemble des outils nécessaires à la gestion du cycle de vie des holons : adressage, lancement, arrêt, etc. Chaque classe de holon (HeavyHolon et LightHolon) propose, en natif, un ensemble de politiques d'exécution pour les rôles qu'il joue, ainsi que différentes politiques pour la gestion des messages.

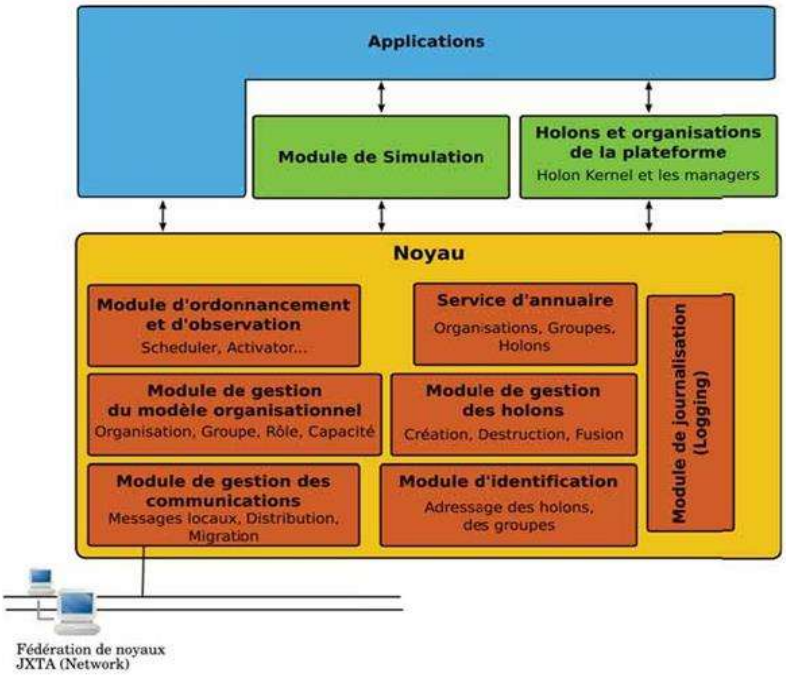

Figure 2-6 : L'architecture complète de la plate-forme Janus [Buisson, 2013 ; Rodriguez, 2011].

- **Gestion des communications** Le noyau assure l'acheminement des messages en local et au sein de la fédération de noyaux lorsque l'application est distribuée. Le fait de considérer le rôle comme une classe à part entière affecte également cet aspect de la plate-forme.
- **Gestion de l'identification** Le noyau fournit l'ensemble des mécanismes nécessaires pour l'attribution d'une adresse unique (GUID) à tous les éléments du modèle qui le nécessitent. Ainsi, les holons, les groupes, et les rôles disposent d'une adresse unique au sein d'une fédération de noyaux.
- **Service d'annuaires** : Un service d'annuaire est également disponible au sein de la plate-forme. Il référence notamment l'ensemble des groupes, des holons et des organisations définis dans le noyau.
- **Gestion de l'ordonnancement** des holons Janus fournit deux politiques de base pour l'ordonnancement des holons : un modèle pour l'exécution concurrente et un moteur d'exécution synchrone inspiré de celui de Madkit. Ce module fournit également une instrumentation à base de sonde permettant à un rôle d'observer un autre rôle. Contrairement à Madkit qui gère les droits d'observation au niveau de l'agent (agent qui implante ou non l'interface

Referencable Agent), Janus les gère au niveau du rôle. Elle permet ainsi une gestion plus fine des droits d'observation. Un holon peut donc autoriser l'observation de l'un de ses rôles et l'interdire pour les autres.

- **Gestion de la journalisation** Toutes les applications basées sur Janus ont accès à un système de journalisation intégré à la plate-forme, ce qui facilite le processus de débogage ("debugging"). Le journal peut être directement affiché ou stocké dans un fichier. Ce système peut être changé et intégré à des systèmes existants. L'implantation actuelle de cette fonctionnalité est basée sur le système log4j2 fourni par l'Apache Software Foundation.
- **Comme Madkit**, Janus exploite son propre modèle dans la conception même de la plate-forme. Tous les services, excepté ceux assurés directement par le noyau, sont assurés par des agents ou des holons. Le noyau est ainsi associé à un holon KernelHolon, qui contient 'ensemble des organisations locales en charge de gérer la plate-forme, et qui représente son noyau dans la fédération distribuée sur le réseau. Une fédération est, en fait, une organisation en charge de gérer les différents échanges entre noyaux et de propager l'information de gestion du modèle organisationnel : création d'une nouvelle organisation ou d'un nouveau groupe, migration d'un holon, etc.
- **L'architecture de Janus respecte** globalement l'architecture abstraite définie par le FIPA pour les plates-formes multi-agents3. Seule la partie ACL4 n'est pas encore complètement implantée. Pour pallier cette lacune, il est prévu, à court terme, d'intégrer la partie correspondante du "middleware" JADE afin d'assurer une compatibilité complète de Janus avecle standard FIPA. À plus long terme Janus devra être associée à un AGL spécifique supportant le métamodèle CRIO, afin d'assurer une transition rapide entre les phases de conception et d'implantation.

c. Caractéristiques de Janus

Cette section est consacrée à la présentation des principales caractéristiques de la plate-forme Janus. Les aspects liés à l'implantation du concept de holon et aux mécanismes de communication entre rôles, seront notamment détaillés.

Pour que deux holons puissent communiquer, ils doivent appartenir à un groupe commun et jouer un rôle dans ce groupe. Si un groupe est distribué parmi plusieurs noyaux, une instance de ce groupe existe dans chaque noyau ; et toutes les instances possèdent le même nom.

La communication dans Janus est basée sur les rôles. Les messages sont délivrés aux agents en ne connaissant que leurs rôles. Ce mode d'interaction permet l'implantation des communications entre un émetteur et plusieurs receveurs ("one-to-many communication", [Gouaich et al., 2004]). L'adresse des agents receveurs est découverte dynamiquement en fonction des rôles qu'ils jouent. Lorsque plusieurs agents jouent le même rôle au sein d'un

même groupe, un mode de communication basé sur l'identifiant des agents peut également être utilisé. Bien que l'identifiant explicite de l'agent soit connu, la communication passe toujours par un rôle donné dans un groupe donné. Ce type d'interaction permet l'implantation de la communication entre deux agents ("one-to-one communication").

Chaque holon dispose ainsi d'une boîte à lettres personnelle pour l'émission et la réception de messages. Un holon peut jouer simultanément plusieurs rôles et en acquérir dynamiquement. Le rôle constitue le moyen d'interaction du holon dans le contexte d'interaction particulier représenté par le groupe. Les rôles sont à l'origine de toute interaction.

Chaque rôle dispos donc de sa propre boîte à lettres et la boîte à lettres d'un holon n'est que le regroupement de l'ensemble des boîtes à lettres des rôles qu'il joue. Un holon ne peut donc recevoir de messages que par l'intermédiaire de ses rôles.

2.3 Conclusion

Nous avons consacré la première partie de ce chapitre à l'étude de l'Apprentissage Collectif Assisté par Ordinateur (ACAO) plus connu sous sa dénomination anglaise Computer-Supported Collaborative Learning (CSCL). Dans cette partie nous avons présenté une définition général à L'ACAO et une synthèse des travaux des chercheurs (Dillenbourg, 1999 ; O'Leary, 2007, J. Bannon,. 1995) sur les outils ACAO, leur définition et leur processus de fonctionnement.

Nous avons présenté par la suite, les traces d'interactions dans les formations à distances. Ces traces sont connues comme une source d'information qui révèle non seulement les activités elles-mêmes, mais aussi leurs sorties (les résultats des activités). C'est une des raisons pour lesquelles chaque trace acquise dans des environnements d'apprentissage, a de toute évidence devenu plus avantageux de différents acteurs, comme les chercheurs, les développeurs, les apprenants, et les tuteurs.

Dans le troisième paragraphe, et afin de mieux positionner nos travaux de recherche, nous avons adressé un état de l'art des indicateurs d'analyse automatique des interactions en se référant aux travaux de (Pléty, 1998 ; Dimitracopoulou, 2008). Notre démarche a consisté en l'étude des propriétés de chaque indicateur ; sa nature (Cognitif ou Social), son point de vue (Groupe, Individu).

Dans le quatrième paragraphe, nous avons réalisé une étude sur l'analyse des forums de discussion. Une partie de la recherche dans le domaine des environnements informatiques

pour l'apprentissage humain est fortement intéressé par le suivi des activités des utilisateurs, tout en interagissant avec des outils de communication fournis par chaque plate-forme d'apprentissage. Ces outils, tels que les forums, les chats, etc, sont devenues obligatoires pour les situations d'apprentissage à distance, en facilitant les communications entre les apprenants et entre apprenants et tuteur. Pendant ce temps, le suivi des communications sur les outils CMC est devenu l'un parmi d'autres intérêts de recherche et mis en place un certain nombre de défis et difficultés : l'exploitation des traces recueillies, fournissant des informations de synthèse des activités de communication via l'analyse de traces, la réutilisation des traces, etc.

Le cinquième paragraphe, présente une étude sur les profils des apprenants dans le cadre à distance.

Nous avons consacré la deuxième partie de ce chapitre à l'étude de la technologie multi-agents, technologie qui répond le plus à nos besoins informatiques en termes de démarche de conception, d'évolution et d'adaptation. Après avoir justifié le choix de cette technologie, nous avons défini les concepts de base des agents : leurs définitions, architecture et langage de communication.

Dans le deuxième paragraphe nous avons brièvement présenter la plate-forme Janus destinée à l'implantation et au déploiement de systèmes multi-agents holoniques. Janus permet véritablement de considérer l'organisation comme un module Java à part entière. En outre Janus, fournissant une implantation directe des quatre concepts à la base de métamodèle CRIO et méthodologie d'ASPECS (capacité, rôle, organisation et holon). Cette plate-forme s'intègre ainsi dans la perspective consistant à fournir une suite complète d'outils logiciels pour le développement d'applications complexes à vocation industrielle.

L'étude réalisée dans ce chapitre sur les systèmes multi-agents nous a permis de bien cerner notre problématique de point de vue technique et d'affiner les choix technologiques qui répondent le mieux à nos besoins informatique.

Cet état de l'art nous permettra de recenser les réponses qui sont actuellement données aux questions que nous nous posons.

3 Classification automatique de texte et logique floue

Sommaire

3.1 Classification automatique de texte 69
3.1.1 Introduction 69
3.1.2 L'apprentissage automatique 69
3.1.3 L'apprentissage supervisé 70
3.1.4 La catégorisation est un problème de classification supervisée 70
3.1.5 Comment classer ? 70
3.1.6 Codage des textes 71
3.1.6.1 Introduction 71
3.1.6.2 Le texte 71
3.1.6.3 Prétraitements 72
3.1.6.4 Définition de descripteurs 73
a. Représentation en « Sac de mots » « bag of words » 73
b. Représentation des textes par des phrases 74
c. Représentation des textes avec des racines lexicales (stemming) 74
d. Représentation des textes avec des lemmes (lemmatisation) 75
e. Représentation des textes avec la méthode des n-grammes 75
3.1.6.5 Codage des termes (Calcul du poids) 76
a. Le codage TFxIDF 76
b. Codage TFC 77
c. L'entropie 77
3.1.6.6 Réduction de la dimension 77
3.1.7 Différents Modèles de classifieurs 78
3.1.7.1 Réseaux bayésiens classifieurs 79
3.1.7.2 Machines à Vecteurs Support – SVM 79
3.1.7.3 Rocchio 80
3.1.7.4 K plus proches voisins - kPPV 80
3.1.7.5 Arbres de décision 81
3.1.7.6 Les approches neuronales 81
3.1.8 Avantage de la méthode adoptée (Naïve Bayes) 82
3.2 Logique floue 83
3.2.1 Introduction 83
3.2.2 Logique classique et logique floue 84
3.2.3 Concepts et définitions 85
3.2.3.1 Sous-ensembles flous ou Fuzzification 86
3.2.3.2 Variables linguistiques 87
3.2.3.3 Le raisonnement en logique floue 88
a. Modus ponens généralisé 89
b. Méthodes d'inférence floue 89
3.2.3.4 La défuzzification 90
3.3 Conclusion 91

3.1 Classification automatique de texte

3.1.1 Introduction

Dans cette section, nous exposant les principes de base et les particularités de chaque méthode de classification de textes pour la catégorisation et classification automatique de conversations textuelles, nous commençant par un petit rappel sur le principe de base de l'apprentissage automatique et plus particulièrement l'apprentissage supervisé, nous pourrons conclure facilement que la problématique de catégorisation de textes se situe bien dans le cadre de l'apprentissage supervisé, nous présentant par la suite la façon avec laquelle les messages sont classés. Pour en finir, nous donnerons un aperçu sur le choix de méthode naïve bayes.

3.1.2 L'apprentissage automatique

Les approches traditionnelles et manuelles de classification de textes sont peu génériques, coûteuse en temps de traitement, et relativement peu efficaces. Au contraire, les autres solutions de classification automatique ont été admises, qui consiste à faire apprendre automatiquement à l'ordinateur, sur la base d'un corpus de textes qui servent d'exemples « base d'apprentissage », les paramètres de la fonction de classement.

L'apprentissage automatique depuis une vingtaine d'années a été considéré comme un problème d'apprentissage automatique et est rapidement devenue un centre d'intérêt pour les chercheurs et un champ d'essai sollicité par les différentes techniques de classification. De toute façon, quelle que soit l'approche retenue, une des particularités de cette tâche est la très grande dimensionnalité de l'espace dans lequel les textes sont représentés, qui comprend généralement plusieurs centaines de termes.

L'apprentissage automatique s'intéresse particulièrement aux méthodes inductives permettant d'acquérir des connaissances à partir d'observations d'un phénomène. Cette connaissance peut être exploitée pour des tâches de décision ou de prévision : c'est le cadre de l'apprentissage supervisé ; ou à des fins d'analyse exploratoire ou de structuration d'un ensemble de données : c'est le cadre de l'apprentissage non supervisé [Stroppa, 2007].

Le contexte de notre étude se situe dans le premier cas.

3.1.3 L'apprentissage supervisé

L'apprentissage supervisé est un type d'algorithme d'apprentissages automatique qui utilise un ensemble de données déjà connues (appelé l'ensemble de données de formation) pour faire des prévisions. L'ensemble de ses données de formation comprend des données d'entrée et les valeurs de réponse. Au-delà, l'algorithme d'apprentissage supervisé cherche à construire un modèle qui peut faire des prévisions sur les valeurs de réponse pour un nouvel ensemble de données.

Dans ce cadre plusieurs algorithmes ont été utilisés dans le domaine de catégorisation de texte pour résoudre ce problème : réseaux de neurones, Formule de Bayes, arbres de décisions, k-plus proche voisins, machines à vecteurs de support (SVMs) et bien d'autres.

3.1.4 La catégorisation est un problème de classification supervisée

Le couple (Message, Classe) est nécessaire pour construire un filtre relatif à une classe donnée, ces exemples de chaque classe, préalablement étiquetés constituent le corpus d'apprentissage.

On fait appel aux méthodes d'apprentissage supervisées pour ajuster un modèle qui crée une association entre les messages d'entrée et les classes de sortie. Ainsi, par ces méthodes d'apprentissage, il est possible de construire un modèle de classification, à partir de ces exemples connus a priori (Message, Classe) [Jalam, 2003].

Ce qui affirme clairement que la catégorisation de textes est bien un problème de classification supervisée.

3.1.5 Comment classer ?

Classer les messages revient en réalité à déterminer les paramètres de la fonction de classement. Les étapes pour classer les messages seront comme suit :

- Il faut disposer d'un corpus d'apprentissage, qui va servir d'entrée à un algorithme d'apprentissage.
- On sélectionne un autre corpus qui sert pour l'évaluation (corpus de test)
- Il faut d'abord déterminer les descripteurs (variables de la fonction)
- Il faut fournir à l'ordinateur un type de fonctions de classement lui permettant d'associer une catégorie à un texte.
    - SVMs
    - Naïve Bayes
    - Règles de décision
    - Arbres de décision
    - Réseaux de neurones

    - Autres fonctions ...
- On infère, à partir des données, et par des méthodes mathématiques complexes, les paramètres de la fonction de classement utilisé, qui peuvent être :
    - Coefficients de l'hyperplan dans les SVMs
    - Distributions de probabilité dans les classificateurs probabilistes
    - Règles dans les règles de décision
    - Conditions et branchements dans les arbres de décision
    - Poids dans les réseaux de neurones
    - ...
- On se fonde sur la connaissance préalable des bonnes catégories pour les messages du corpus d'apprentissage (apprentissage supervisé).

3.1.6 Codage des textes

3.1.6.1 Introduction

La grande quantité d'informations textuelles produite par les apprenants dans les forums de discussions asynchrones, fait sentir le besoin de recherche de mécanismes et outils de traitement automatique des quantités d'informations diffusées sur ces forums.

Le manque de structure au sein de ces collections volumineuses de messages rend difficile le traitement et l'accès à l'information qu'elles contiennent, d'où la nécessité aujourd'hui, de chercher comment structurer automatiquement ces corpus pour les rendre utilisables d'une façon rapide et optimale pour y faciliter leurs traitements automatiques et notamment la classification et la catégorisation.

Pour appliquer les différents algorithmes et techniques d'apprentissage, une transformation et codage de ces messages non structurés sont nécessaires. Le codage ou la transformation de ces messages est une préparation à « l'informatisation » de ces derniers. Plusieurs approches de représentation des messages textuels ont été proposées dans ce contexte, la plupart étant des méthodes vectorielles.

Les principales méthodes de représentation de textes n'utilisent pas d'information grammaticale ni d'analyse syntaxique des mots : seule la présence ou l'absence de certains mots est porteuse d'informations.

3.1.6.2 Le texte

Le texte est considéré aujourd'hui dans la plupart des cas, comme une séquence des caractères. Ces caractères peuvent être représentés dans différents espaces de codage, le plus courant étant le codage ASCII admettant 256 caractères différents, mais en dépit ce codage ne prenait pas en charge les langues comme l'arabe ou le chinois. Afin de pouvoir représenter

ces langues, différentes normes de codages sont créées et plus largement utilisées aujourd'hui comme la norme UNICODE qui permet la représentation de 65536 caractères.

Ainsi la représentation informatique de ces textes nécessite un traitement spécifique.

Les méthodes se sont heurtées au fait qu'un texte n'est pas un sac dans lequel seraient mélangées en vrac ses propres éléments. Le moins qu'on puisse dire sur un texte qu'il est une chaîne linéaire, donc un espace ordonné. (« Voile du bateau » et « Bateau à voile » ont des sens complètement différents).

La composition d'un texte fait appel à deux définitions de la composition : il s'agit à la fois de déterminer les unités qui vont constituer le texte, tels les atomes qui composent les molécules, et de constituer un texte c'est-à-dire de distribuer, d'organiser ces unités afin d'atteindre certaines idées, comme une molécule qui possède certaines propriétés en raison de sa structure. Plusieurs approches de représentation dans cet espace sont proposées dans la littérature. Nous détaillons par la suite ses différentes représentations.

3.1.6.3 Prétraitements

Dans la section suivante, nous allons aborder les différentes méthodes de représentation de texte. Ses différentes représentations sont toutes effectuées à base de mots qui sont une séquence de caractères. Il est donc nécessaire d'effectuer au préalable du codage d'un document dans un espace de mots, une transformation permettant le passage de l'espace du caractère à un espace de mots.

Le prétraitement des textes est une phase importante et capitale dans le processus de classification et catégorisation, puisque la connaissance imprécise de la population peut faire échouer l'opération. Le prétraitement est généralement effectué en plusieurs étapes séquentielles :

- La reconnaissance des termes utilisés,
- Le processus de filtrage qui permet de supprimer les informations inutiles des messages afin que les connaissances gardées soient aussi pertinentes qu'il se peut. En effet dans les messages textuels de nombreux mots apportent peu (voir aucune) d'informations sur le document concerné. Les algorithmes dits de « mots vides » s'occupent de les éliminer,
- Un autre traitement nommé « Stemming » permet également de simplifier les textes tout en augmentant leurs caractères informatifs comme d'autres méthodes qui proposent de supprimer des mots de faible importance,
- Un traitement sémantique consiste à extraire la signification des expressions et traiter la polysémie à savoir les différents sens possibles d'un même mot. Par exemple, cette phrase permet de différencier le mot « opération » qui peut correspondre à une

opération chirurgicale ou à une opération mathématique. C'est une opération laborieuse, qui fait appel aux ontologies qui n'est pas bien maîtrisée dans les systèmes de classification.

Toutes ces étapes font partie de ce qu'on appelle le prétraitement. Plusieurs d'entre elles sont spécifiques à la langue des messages (on ne fait pas le même type de prétraitement pour des messages écrits en anglais qu'en français ou encore en arabe).

3.1.6.4 Définition de descripteurs

La représentation des textes est la phase la plus importante dans le processus de catégorisation de texte parce que les algorithmes d'apprentissage ne sont pas capables de traiter directement les textes. C'est pour cette raison, qu'on doit opter pour une extraction de caractéristiques au sein d'un texte qui est une phase décisive puisque la représentation doit conserver l'information contenue dans le texte.

La totalité des systèmes de catégorisation de texte représente les messages par la présence ou l'absence de termes dans le texte. La plus petite unité informationnelle étant le caractère, à un niveau supérieur on a les termes qui sont l'unité minimale constitutive d'un texte, puis à un niveau plus global nous pouvons définir les phrases, les paragraphes, et pour finir le message lui-même.

L'étape de définition de descripteurs consiste généralement à représenter chaque message par un vecteur, dont les composantes sont par exemple les termes contenus dans le texte, à ces termes on associe des poids pour rendre chaque vecteur exploitable par les algorithmes d'apprentissage.

Différentes méthodes sont proposées pour le choix des termes et les poids attribués à ces termes. Dans la section suivante, nous allons définir les différentes sortes de termes, utilisés dans la littérature, pour la représentation d'un document texte.

a. Représentation en « Sac de mots » « bag of words »

Le modèle de sac de mots « Bag-of-words » est une représentation simplifiée utilisée dans le traitement du langage naturel et la recherche d'informations. Le choix des mots comme descripteurs d'un message c'est le choix le plus intuitif. Les textes sont représentés dans l'espace par un vecteur dont chaque composant représente un terme. [Salton & McGill, 1986 ; Lewis, 1992 ; Sahami& all, 1998 ; Yang, 1999 ; Vinot & all, 2003 ; Pessiot & all, 2004; Pothin & Richard, 2007; Gotab, 2009; Jégou & all, 2010], et bien d'autres ont préféré l'utilisation des mots comme descripteurs pour le codage des messages et documents. Pour

clarifier la notion de mot, Y. Gilly dans son ouvrage « Texte et fréquence » [Gilly, 1988] l'a considéré comme étant une séquence de caractères appartenant à un dictionnaire, ou formellement, comme étant une suite de caractères séparés par des espaces ou des caractères de ponctuation (cette définition n'est pas valable pour toutes les langues).

b. Représentation des textes par des phrases

Un certain nombre de chercheurs proposent d'utiliser les phrases comme unité de représentation au lieu des mots comme le cas dans la représentation« sac de mot », puisque les phrases sont plus informatives que les mots seuls, par exemple« recherche d'informations », « world wide web », ont un degré plus petit d'ambiguïté que les mots constitutifs, et aussi que les phrases ont l'avantage de conserver l'information relative à la position du mot dans la phrase [Caropreso et al., 2000; Furnkranz et al, 1998].

L'utilisation de "sac de phrases" entraîne évidemment un problème de taille (pour n mots il y existe potentiellement nk combinaisons de longueur k). Ces techniques sont venues pour remédier à la déstructuration syntaxique causée par la représentation en "sacs de mots".

c. Représentation des textes avec des racines lexicales (stemming)

L'opération de stemming expliquée précédemment, vise de ramener un mot à sa forme canonique (racine). Par exemple le stem de numériseur serait numéris regroupant aussi : numériseur, numériseurs, numérisée, numériser, numérisation, numérisations, etc... l'intérêt du stemming puisque qu'il rassemble plusieurs mots de significations très proches dans un même groupe. Malheureusement, cette opération n'étant pas basée sur des contraintes linguistiques puissantes, peut conduire, à une amplification du bruit et des confusions sémantiques, en regroupant par erreur des mots de différentes significations, peuvent être générées.

La représentation des textes par ces stems peut apporter des résultats supérieurs à ceux obtenus par les lemmes (que nous allons voir dans ce qui suit), démontré et approuvé par de Loupy, et bien meilleur que le codage de type « sac de mots » ou chaque variation d'un mot est considérée comme une nouvelle composante du vecteur. Alors, on peut facilement imaginer combien on va gagner en question de dimensionnalité en optant pour les stems comme descripteurs.

d. Représentation des textes avec des lemmes (lemmatisation)

La lemmatisation décrite auparavant, consiste à utiliser l'analyse grammaticale afin de remplacer les verbes par leur forme infinitive et les noms par leur forme au singulier.

La lemmatisation est donc plus compliquée à mettre en œuvre que la recherche de racines, puisqu'elle nécessite une analyse grammaticale des textes.

La substitution des mots par leur lemme réduit également l'espace des descripteurs comme pour les racines, et permet de représenter par un même descripteur des termes de même signification. Par exemple, le remplacement des mots banking, bank, banks, par l'unique racine bank semble être rentable tout comme le remplacement des formes conjuguées rebondit et rebondi par le lemme rebondir. Depuis la fin des années 80, les lemmatiseurs sont capables d'associer à chaque mot d'un texte son lemme grâce à un étiquetage morpho-syntaxique (nom, verbe, adjectif, etc.) [Jaillet, 2004].

Cette représentation est simple mais présente plusieurs problèmes :

- La perte de l'information donnée par le contexte syntagmatique, nécessaire à la distinction des lemmes polysémiques (« prix » n'a pas le même sens dans « prix Goncourt », « grand prix » ou «prix d'une marchandise»).
- La présence de synonymes, considérés comme des lemmes différents même s'ils font référence au même concept (« mission » et « délégation » peuvent dénommer la même entité dans un article de journal) [Ignat, 2006].

e. Représentation des textes avec la méthode des n-grammes

Une autre approche pour coder les documents émerge : les n-grammes [Shannon, 1948]. On définit un n-gramme (n-gram) par est une séquence de n caractères : bi-grammes pour n=2, tri-grammes pour n=3, quadri-grammes pour n=4, etc. On n'a plus besoin de chercher les délimiteurs (les espaces ou les caractères de ponctuations) comme c'était le cas pour les mots.

Pour un texte quelconque, les n-grammes correspondants sont générés en faisant déplaçer un masque de n caractères sur tout le texte. Ce déplacement s'effectue caractère par caractère, à chaque déplacement la séquence de n caractères est enregistrée, l'ensemble de ces séquences constitue l'ensemble des n-grammes représentant le texte. [Miller & all, 1999] Par exemple, pour générer les 3-grammes de la phrase "Tu es libre", on obtient : "Tu", "u e", "es", "es", "s l", " li", "lib", "ibr", "bre".

Historiquement, les n-grammes étaient conçus pour la reconnaissance de la parole, pour prédire l'apparition de certains caractères en fonction des autres caractères mais par la suite, le concept n-grammes a été bénéfique, pour le domaine de recherche d'information et la

classification de textes, avec plusieurs travaux qui ont démontré que cette segmentation ne faisait pas perdre d'information.

Enfin, notons que la technique des lemmatisation est toujours très utilisée, pour représenter les textes, en raison de ses avantages qui nous ont incité nous aussi, d'ailleurs, pour l'adopter dans notre étude.

3.1.6.5 Codage des termes (Calcul du poids)

Une fois la liste des attributs déterminés, il reste à donner une pondération à chaque composant du vecteur. Le tableau (termes x messages) est constitué par le nombre d'apparitions du terme dans le message du corpus. Cette information de base doit être pondérée en fonction de divers paramètres liés au message lui-même (ex : le nombre de termes par message) ou au corpus en intégralité (ex : le nombre de termes du corpus). L'intérêt de cette pondération est de mieux exploiter l'information contenue dans le document pour améliorer les performances d'un système de classification de textes [Spaeck-Jones, 1972].

Plusieurs systèmes de pondération ont été développés dans la littérature, qui se reposent, tous sur les deux hypothèses suivantes :

- Plus le nombre d'apparitions d'un terme dans un texte est important, plus ce terme est discriminant pour la classe associée.
- Plus le nombre d'apparitions d'un terme dans le corpus est important, alors moins ce terme peut discriminer les textes.

Il existe plusieurs façons pour coder ces composantes, nous détaillons les codages les plus utilisés pour le domaine de la catégorisation automatique : le codage TFIDF, TFC, Lnu et l'entropie.

Voici, un petit aperçu sur les pondérations les plus habituellement utilisées signalées dans [Clech, 2004] :

a. Le codage TFxIDF

La pondération TFxIDF est largement utilisée pour calculer le poids d'un terme (acronyme pour « term frequency inverse document frequency »). Cette représentation se base principalement sur une certaine loi appelée loi de Zipf qui montre la façon dont les mots sont distribués dans un corpus. Ces mots sont la plupart du temps des mots outils. Par ailleurs, les mots les moins fréquents du corpus ne sont également par les plus porteurs d'informations [Béchet, 2008]. Issue du monde de la recherche d'information, celle-ci donne

plus d'importance aux mots qui apparaissent souvent à l'intérieur d'un même texte, ce qui correspond bien à l'idée intuitive que ces mots sont plus représentatifs du message.

Si on désire avoir des poids entre 0 et 1, on peut appliquer une normalisation, ce qui est souvent le cas. La fonction TFIDF a démontré une bonne efficacité dans des tâches de catégorisation de textes, et, en plus, son calcul est simple [Saldarriaga, 2010].

b. Codage TFC

Le TFC normalise le TFxIDF en fonction de l'ensemble des termes du message. Le codage TFxIDF ne corrige pas la longueur des messages. Pour ce faire, le codage TFC est similaire à celui de TFxIDF mais il corrige les longueurs des textes par la normalisation en cosinus, pour ne pas favoriser les messages les plus longs.

c. L'entropie

La pondération basée sur l'entropie est la plus performante comparée à six autres méthodes, approuvée par les expérimentations de [Dumais, 1991] et confirmée par d'autres auteurs récemment. Selon l'auteur l'entropie devance le TFxIDF sur les cinq corpus testés. Néanmoins, cette méthode est assez complexe du fait qu'elle fait intervenir l'ensemble des autres textes, sollicitant un temps de traitement plus grand.

3.1.6.6 Réduction de la dimension

La taille des données peut être mesurée selon deux dimensions, le nombre de variables et le nombre d'exemple. Ces deux dimensions peuvent prendre des valeurs très élevées, ce qui peut poser un problème lors de l'exploration et l'analyse de ces données. Pour cela, il est fondamental de mettre en place des outils de traitement de données permettant une meilleure compréhension de la valeur des connaissances disponibles dans ces données.

[Jalam ,2003] a rappelé ce qui a été évoqué par [Sebastiani, 2002] sur la réduction des dimensions qui peut être localement ou globalement :

- **Réduction locale :** Chaque classe est caractérisée par un profil composé d'un ensemble de termes, et chaque texte sera représenté par une liste de termes dépendante de la catégorie.
- **Réduction globale :** Contrairement au cas précédent, un texte est représenté par une seule liste de termes dans tous le corpus indépendamment des classes.

La réduction des dimensions est l'une des plus vieilles approches permettant d'apporter des éléments de réponse à ce problème. L'objectif principal est d'extraire ou sélectionner un sous-ensemble de caractéristique qui permet d'éliminer les informations non-pertinentes et

redondantes selon le critère utilisé, réduire le temps d'apprentissage et d'utilisation, faciliter la visualisation et la compréhension des données, réduire l'espace de stockage nécessaire et enfin identifier les facteurs pertinents.

Les techniques mathématiques de réduction de dimension sont classées en deux catégories :

1- La sélection d'attributs « feature selection » prend les attributs (ou mots) d'origine et conserve seulement ceux jugés utiles à la classification, selon une certaine fonction d'évaluation. Les autres sont rejetés.
2- D'autres méthodes effectuent plutôt une extraction d'attributs (« feature extraction »). À partir des attributs de départ, elles créent de nouveaux attributs, en faisant soit des regroupements ou des transformations. On constate que la sélection d'attributs est meilleure quant à l'élimination d'attributs réellement inutiles ou même d'attributs erronés « noisy » (mots mal orthographiés par exemple) tandis que l'extraction d'attributs est plutôt axée sur la réduction du nombre d'attributs redondants.

3.1.7 Différents Modèles de classifieurs

Historiquement, différentes générations d'algorithmes de classification automatique de textes se sont succédés. Une part d'amélioration a été apportée par chaque nouvelle génération par rapport aux antécédentes. Parmi les premières, nous trouverons certainement les approches sémantiques dont l'handicap principal résidait dans le coût excessif puisque plusieurs experts sont chargés dans des intervalles de temps importants, pour mettre à jour les plans de classement.

Pour apaiser à ces limitations, plusieurs boites ont conçues et commercialisées des technologies de catégorisation automatique basées sur des approches purement statistiques.

Plusieurs générations de techniques statistiques ont depuis été développées et qui ont confirmé leurs performances en obtenant de meilleurs résultats. Actuellement, plusieurs approches de catégorisation coexistent. Nous citons parmi les plus utilisées le modèle probabiliste Naïve Bayes, ou les modèles vectoriels de Machine à Vecteur de Support, les k-plus-proches voisins, Rocchio, ainsi que des modèles à base de règles ou d'arbre de décision ou encore des approches fondées sur les réseaux de neurones qui ont été proposées. Chacun de ces modèles possède certains avantages et certains inconvénients. Dans ce qui suit une liste plus ou moins exhaustive des différents modèles et de leur intérêt respectif sera

présentée. Une description plus détaillée sera accordée à l'approche naïve bayésienne dans le chapitre 6 qui fera l'objet des modèles de classification dans notre approche proposée.

3.1.7.1 Réseaux bayésiens classifieurs

Classificateurs bayésiens naïfs, une famille de classificateurs qui sont basés sur la probabilité du théorème de bayes. Le classifieur naïf de bayes est l'une des méthodes les plus simples en apprentissage supervisé. Un avantage de cette méthode est la facilité d'estimation des paramètres et sa rapidité (même sur de très grandes bases de données), la simplicité de programmation. L'adjectif naïve provient de l'hypothèse que les caractéristiques dans un ensemble de données sont mutuellement indépendants. Dans la pratique, l'hypothèse d'indépendance est souvent violé, mais le classifieur naïfs bayes ont tendance à bien performer dans cette hypothèse irréaliste [Rish, 2001]. Surtout pour les échantillons de petite taille [Domingos, 1997].

Naïve bayes est utilisées dans de nombreux domaines différents. Quelques exemples comprennent le diagnostic des maladies et des décisions sur es processus de traitement [Kazmierska, 2008], la classification des séquences dans les études taxonomiques (Wang, 2007), et le filtrage des E-mail (spam) [Dumais, 1998].

3.1.7.2 Machines à Vecteurs Support – SVM

L'algorithme SVM (Support Vector Machine) proposées par Vapnik [Cortes 1995]. SVM est une méthode d'apprentissage supervisée relativement récente introduite pour résoudre un problème de reconnaissance de formes à deux classes. Ce sont des méthodes de classification binaire par apprentissage supervisé. Le principe de SVM a été proposé par Vapnik à partir de la théorie du risque empirique.

La méthode SVM est un classificateur linéaire utilisant des mesures de distance. Elles reposent sur l'existence d'un classificateur linéaire dans un espace approprié. Cette méthode fait appel à un jeu de données d'apprentissage pour apprendre les paramètres du modèle. Elle est basée sur l'utilisation de fonction dite noyau (kernel) qui permet une séparation optimale des données [Mohamadally, 2006].

L'approche par SVM permet donc de définir, par apprentissage, un hyperplan dans un espace vectoriel qui sépare au mieux les données de l'ensemble d'apprentissage en deux classes, minimisant le risque d'erreur et maximisant la marge entre deux classes. La qualité de l'hyperplan est déterminée par son écart avec les hyperplans parallèles les plus proches des points de chaque classe. Le meilleur hyperplan est celui qui a la marge la plus importante.

3.1.7.3 Rocchio

La méthode Rocchio parue dans [Rocchio, 1971], est un classifieur linéaire basée sur le calcul des mesures de distance, il fait partie des premières techniques de classification supervisée. Ce classificateur a été largement utilisé dans la communauté de la catégorisation textuelle. Plusieurs améliorations ont été apportés sur le modèle mais la définition présentée ici est celle de la version initiale.

Dans la phase d'apprentissage de la méthode, les représentations vectorielles, vont permettre le codage de chaque catégorie par un vecteur dont lequel figure tous les termes générés avec leur nombre d'occurrence. Le processus représente donc les classes par des profils prototypiques correspondants à des vecteurs dans un espace vectoriel similaire aux documents. Ces profils sont donc des listes de termes pondérés générées pendant l'apprentissage de même pour les vecteurs correspondants aux textes qui sont aussi générés durant cette phase. Le profil d'une catégorie doit contenir tous les termes qui caractérisent cette catégorie par rapport aux autres.

3.1.7.4 K plus proches voisins - kPPV

La méthode des k plus proches voisins ou The k-NN classification (k-Nearest Neighbors) est un classifieur à base d'instances qui fait partie des méthodes géométriques utilisant des mesures de distance.

k-NN est un algorithme classique d'apprentissage qui a été longtemps à la base des algorithmes de catégorisation des documents, elle a été employée avec succès dans de domaine de classification et a engendré toute une famille de classifieurs connus sous le nom de classifieurs paresseux (lazy learns). Dans ces systèmes, le seul traitement effectué au cours de la phase d'apprentissage est la mémorisation des exemples sous une forme optimale de façon à pouvoir les extraire ensuite rapidement. Chaque texte est représenté dans un espace vectoriel, dont chacun des axes représente un descripteur. Tous les calculs sont reportés à la phase de classification (d'où le terme de paresseux).

k-NN est un algorithme de catégorisation dans lequel les classes ne sont pas représentées sous forme de texte "prototype" (profil de catégorie).

La partie classification est en contrepartie plus coûteuse en temps : Le classifieur calcule la similarité du nouveau texte à catégoriser avec l'ensemble des autres exemples du corpus d'apprentissage, dont les catégories sont déjà connues, puis il sélectionne les k documents les plus proches du document à classer. Ensuite, pour affecter la catégorie, les relations entre ces

k documents et les catégories sont évaluées et un score est calculé par catégorie afin d'évaluer la pertinence de la catégorie au document. La catégorie (ou les catégories) ayant le score le plus élevé (celle qui contient le plus de textes voisins) est affectée au document [Yang, 2002].

3.1.7.5 Arbres de décision

Les arbres de décision, utilisés dès les années 60 en statistique, sont des outils supervisés. Ils sont devenus, depuis une vingtaine d'années des outils très populaires pour générer des règles de classification et plus généralement des règles de prédictions. On parle ainsi d'arbre de classification lorsque la variable à prédire est catégorielle et que ses valeurs représentent donc des classes.

Les arbres de décision ont été utilisés pour permettre la catégorisation des documents dans un certain nombre de classes prédéfinies. L'apprentissage des probabilités d'attribution d'un texte à une classe est réalisé sur des textes étiquetés manuellement. L'arbre de décision généré, sur l'ensemble des documents du corpus d'apprentissage, permet de décider à quelle classe appartient chaque nouveau document du corpus de test. Chaque feuille de l'arbre contient la probabilité d'appartenance à l'une ou l'autre des classes. Suivant les réponses aux questions posées au document à classer, celui-ci est « dirigé » vers telle ou telle feuille de l'arbre. Le document est alors attribué à la classe de plus forte probabilité. [Bellot, 2001] Selon Gilbert Ritschard, Simon Marcellin et Djamel A. Zighed dans [Ritschard & all, 2009], un partitionnement récursif des données est le principe de base sur lequel s'appuie la méthode pour avoir des ensembles cohérents par rapport à la variable à prédire. Un enchaînement hiérarchique de règles est généré. Chaque feuille (nœud terminal) de l'arbre caractérise une règle, de type « Si condition Alors résultat », dont la prémisse est définie par les conditions d'embranchement le long du chemin menant du nœud initial à la feuille, la conclusion de la règle correspondant à la classe assignée à la feuille. L'ensemble de toutes ces règles représente le modèle de prédiction.

3.1.7.6 Les approches neuronales

Plusieurs travaux, dans le domaine de classification automatique de textes, à base des approches neuronales ont été évoqués.

Une approche fondée sur les réseaux de neurones a été proposée par Wiener dont les résultats ont été repris dans [Wiener & all, 1995]. Deux architectures neuronales sont proposées et testées sur le corpus Reuters.

Les approches neuronales furent les premières à être utilisées afin de réaliser un apprentissage. Ces approches s'inspirent du fonctionnement du système nerveux humain. Ainsi, elles se fondent sur l'utilisation des "neurones" artificiels qui vont effectuer la tâche d'apprentissage. Un neurone est une fonction algébrique non linéaire, paramétrée, à valeurs bornées [Dreyfusetall, 2002].

Les premières approches de ce type furent les "réseaux de neurones monocouches ensuite des algorithmes multicouches (l'architecteur du perceptron multicouches) Le principe général d'une approche neuronale est présenté ci-dessous figure 3.1.

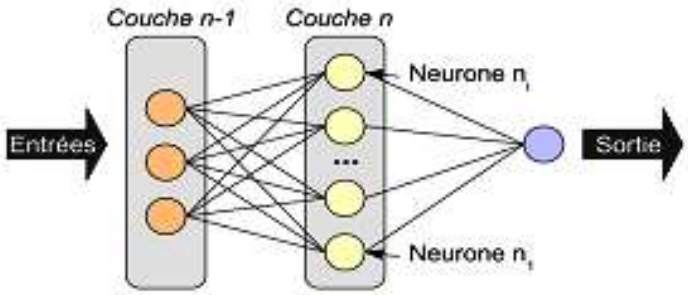

Figure 3-1 : Architecture générale d'un réseau de neurones artificiels.

Un réseau de neurones artificiels est composé d'une ou de plusieurs couches se succédant dont chaque entrée est la sortie de la couche qui la précède comme illustré sur la figure 3.1. Ainsi, une couche est composée d'un certain nombre de neurones artificiels dont les entrées sont constitués des sorties des neurones de la couchen-1. Les neurones humains reçoivent en entrée les signaux provenant des autres neurones par des synapses. Appliquée aux réseaux de neurones artificiels, la connexion neuronale s'effectue par le biais de liaisons pondérées (les synapses) unidirectionnelles. Nous pouvons ainsi voir un réseau de neurones artificiels comme un réseau ou graphe orienté dont les nœuds sont les neurones artificiels. Le but va être d'attribuer des poids synaptiques à chaque neurone afin d'obtenir le résultat voulu en sortie [Dziczkowski, 2008].

3.1.8 Avantage de la méthode adoptée (Naïve Bayes)

L'algorithme Naïve bayes est équivalent à un classifieur linéaire, dans sa rapidité d'application, est connu par son efficacité et sa simplicité. Cet algorithme est utilisé dans de nombreux autres domaines que le texte. David et Hilali [Lewis, 2004 ; Hilali, 2009] présentent un ensemble d'avantages du classifieur bayésien naïf, parmi lesquelles :

Le classificateur bayésien naïf est une technique basée sur une théorie mathématique précise. Cette technique est particulièrement adaptée lorsque la dimensionnalité des entrées est élevée. Malgré sa facilité et simplicité à implémenter, Naïve Bayes peut souvent surperformer méthodes de classification plus sophistiquées.

L'avantage du classifieur bayésien naïf est qu'il requiert relativement peu de données d'entraînement pour estimer les paramètres nécessaires à la classification, à savoir moyennes et variances des différentes variables, au contraire à d'autres techniques.

Le choix de cet algorithme répond à de nombreux critères plus généraux dont nous pouvons citer les suivants : L'efficacité et la rapidité dans l'apprentissage et la classification, facile à mettre à jour avec de nouveaux exemples d'apprentissage. Le point fort de cet algorithme est qu'il résiste au bruit existant dans les données d'entrée et très efficace avec des petits corpus d'apprentissage. Le plus important c'est que les méthodes Naïve Bayes donnent de bons résultats

Cette méthode est très adaptée à notre situation de classification des messages textuels pour la détermination des comportements sociaux d'apprenant, au contraire à d'autres techniques plus sophistiquées gourmandes en ressources (gestion de mémoire vive) et en temps d'exécution avec des taux d'amélioration des résultats très minimes.

3.2 Logique floue

3.2.1 Introduction

Dans le monde réel, l'information est souvent ambigüe ou imprécise. Le raisonnement humain interprète l'information afin d'arriver à des conclusions ou décision. Les machines ne peuvent pas encore traiter une information imprécise de la même manière que les humains, les programmes informatiques avec la logique floue deviennent très utiles lorsque le volume des tâches défie l'analyse et l'action humaines.

Une méthode paganisée pour faire face à des données imprécises est appelé la logique floue [Bouchon, 1995]. Les données sont considérées comme des ensembles flous. Les jeux traditionnels comprennent ou ne comprennent pas un élément individuel, il n'y a pas d'autres cas que vrai ou faux. La logique floue est fondamentalement une logique à valeurs multiples qui permet aux valeurs intermédiaires d'être définie entre les évaluations classiques comme oui/non, vrai/faux, noir/blanc, etc.

Le souci d'automatiser ou d'assister de façon automatique les actions humaines, naturellement empiriques et empreintes d'imprécisions, dans le cadre d'une aide à la décision

ou du contrôle par exemple, renforce l'intérêt des scientifiques pour l'approche floue et justifie son intense développement au cours de ces dernières années. C'est pour les possibilités qu'elle offre de gérer l'incertitude et l'imprécision, que nous nous sommes intéressés à la logique floue.

Zadeh est un chercheur de l'Université de Berkeley en Californie [Zadeh, 1973, 1968]. Il a introduit la notion de sous-ensemble flou pour répondre aux problèmes auxquels sont confrontés de nombreux systèmes complexes, qui doivent traiter des informations qui sont de nature imparfaite, son concept de base est de graduer l'appartenance à un ensemble, c'est un moyen efficace pour prendre en compte l'imprécision dans la connaissance et de formaliser le processus de raisonnement humain.

La motivation de la logique floue a été exprimée par Zadeh [Zadeh, 1984] de la manière suivante : « La capacité de l'esprit humain à la raison en termes flous est en fait d'un grand avantage même si une énorme quantité d'informations est présenté aux sens humains dans. Une situation donnée - un montant qui serait étouffer un ordinateur typique -. En quelque sorte l'esprit humain a la capacité de jeter plus de cette information et se concentrer uniquement sur l'information qui est la tâche pertinente Cette capacité de l'esprit humain à traiter uniquement avec les informations qui est une tâche pertinente est reliée à sa possibilité de traiter l'information floue. En se concentrant uniquement sur les informations de tâche pertinente, la quantité d'informations que le cerveau doit traiter est réduite à un niveau gérable » [Zadeh, 1984].

La logique floue a émergé comme une technique puissante pour les commandes des processus industriels, les systèmes de diagnostic et d'autre systèmes experts. La croissance rapide de cette technologie a effectivement commencé au Japon et ensuite propagée aux Etats-Unis et en Europe. La majorité des applications de la logique floue sont dans la zone de contrôle.

3.2.2 Logique classique et logique floue

Dans le cadre de la logique classique, il n'admet pour les variables que les valeurs 0 et 1, une proposition est soit vraie, soit fausse. Par exemple, la logique classique peut facilement partitionne les personnes par âge, les personnes âgées de moins de 30 ans sont systématiquement jeunes et les plus de 50 ans sont âgées, sans pour autant que, cette classification soit logique. Alors que la logique floue, dont les variables peuvent prendre n'importe qu'elle valeur comprise entre 0 et 1, permet de tenir compte du passage progressif de l'individu d'un âge à un autre, on parle alors, de fonction d'appartenance u. La figure 3.2

(1) montre une classification possible ; une personne de 25 ans appartient à l'ensemble (jeune) avec une valeur u = 0,75 de la fonction d'appartenance, et à l'ensemble (entre deux âges) avec u = 0,25. Par contre une personne de 70 ans appartient avec une valeur μ = 1 de la fonction d'appartenance à l'ensemble (âge).

La logique floue est une extension de la logique classique qui utilise des ensembles flous plutôt que des ensembles classiques. Il y a quelques différentes explications de ce qui est logique floue, nous coterons une des explications avancées par Lotfi Zadeh, Zadeh dit : « Dans son sens étroit, la logique floue est une logique de raisonnement approximatif qui peut être considérée comme une généralisation de la logique à valeur multiples. Mais dans un sens plus large et beaucoup plus importante, la logique floue est coextensive à la théorie des ensembles flous. Dans son sens le plus large, la logique floue a beaucoup de branches allant de l'arithmétique floue, automates floue à la reconnaissance floue de modèle, langues floues et les systèmes experts flous ».

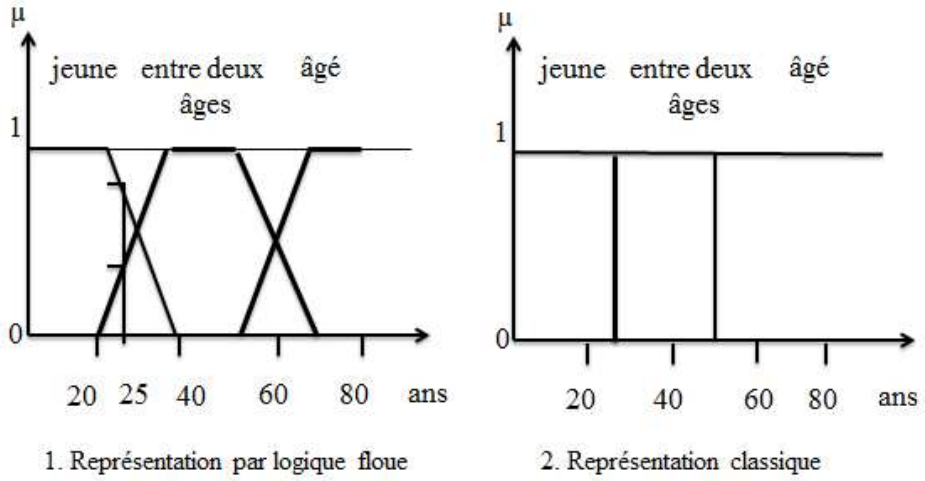

Figure 3-2 : Classification des personnes par âge en trois sous-ensembles [Zadeh, 1984].

3.2.3 Concepts et définitions

La logique floue trouve ses applications dans de nombreux domaines, comme modèle de représentation et de manipulation des connaissances. Elle est particulièrement utilisée pour la mise au point de contrôleurs et de systèmes expert flous, pour l'aide à la décision, pour l'interrogation flexible des bases de données et dans filtrage flou.

Il y a deux grands types de système dont le fonctionnement est basé sur la logique floue :

- **Les systèmes de commande floue :** De façon générale, un système de commande a pour objectif de piloter l'entrée d'un processus afin d'obtenir un fonctionnement correct de ce dernier. Lorsqu'un opérateur humain commande manuellement un système, les actions qu'il réalise sont dictées par une connaissance subjective du

fonctionnement de ce système. Par exemple, s'il fait « froid » dans une pièce, on « augmente » le chauffage ; s'il fait « très froid », on « chauffe plus ». Ce principe est à la base de la commande floue.

- **Les systèmes experts flous :** Ces systèmes peuvent être composés de plusieurs couches de règles avec des opérateurs ET et Ou. Les variables de sortie résultats de l'exécution d'un nombre de règles peuvent être appliquées aux entrées d'autres règles. Les applications de ces systèmes sont diverses : le contrôles linéaire et non linéaire, l'aide à la décision, la reconnaissance de forme, le système financière, la recherche opérationnelle, etc.

La mise en œuvre de ces deux systèmes s'effectue à travers un Système d'Inférence Flou composé généralement d'une base de connaissance avec les trois modules suivants (figure 3.3):

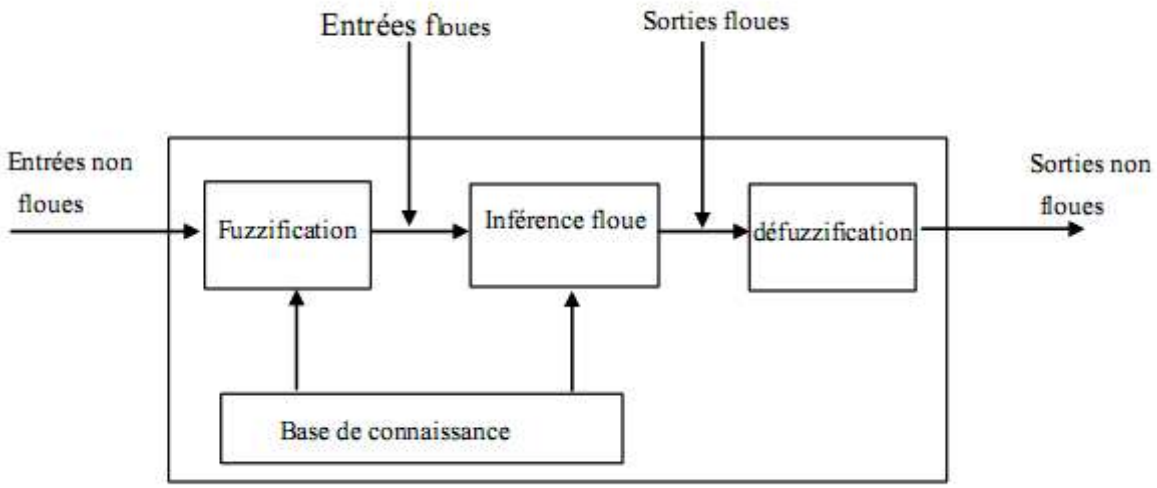

Figure 3-3 : Structure d'un système flou [Zadeh, 1984].

3.2.3.1 Sous-ensembles flous ou Fuzzification

En mathématiques un ensemble, par définition, est une collection de choses qui appartiennent à une certaine définition. Tout élément appartient soit à cet ensemble ou ne fait pas partie de cet ensemble.

La notion de sous-ensembles flous a pour but de permettre des gradations dans l'appartenance d'un élément à une classe. C'est-à-dire d'autoriser un élément à appartenir plus ou moins à une classe prédéfinie.

Un ensemble flou est un ensemble qui est définie par une fonction d'appartenance. Une fonction d'appartenance attribue à chaque élément dans l'ensemble un degré d'appartenance qui est une valeur dans l'intervalle [0, 1]. Dans les ensembles classiques, un élément soit

appartenir à un ensemble ou non ; il n'y a pas d'autre choix. En définissant un ensemble en utilisant une fonction d'appartenance, il est possible pour un élément d'appartenir partiellement à un ensemble. Par exemple, si une porte est Entr'ouverte, on pourrait dire que la porte est ouverte, avec un degré d'appartenance de 0,2 pour indiquer que la porte est Entr'ouverte. Nous pourrions également dire que la porte est fermée, avec un degré d'appartenance de 0,8. En utilisant un ensemble flou, nous sommes en mesure d'indiquer que la porte est partiellement ouverte ou partiellement fermé. Contrairement à la logique classique, la porte serait considérée comme ouverte ou fermée.

Cette présentation évite l'utilisation arbitraire de limites rigides à des classes et répond au besoin de présenter des connaissances imprécises, soit parce qu'elles sont exprimées en langage naturel « on utilise aussi le terme variable linguistique », soit parce qu'elles sont obtenues avec des instruments d'observations qui produisent des erreurs de mesure.

Définition

Étant donné un ensemble de référence X, on peut indiquer les éléments de X qui appartiennent à une certaine classe de X et ceux qui n'y appartiennent pas. Cette classe est alors un sous-ensemble de X. Si l'appartenance de certains éléments de X à une classe n'est pas absolue, on peut indiquer avec quel degré chaque élément appartient à cette classe.

Celle-ci est alors un sous-ensemble flou de X [Bouchon, 1995].

Un sous-ensemble flou A dans un univers du discours X est caractérisé par sa fonction d'appartenance $\mu_A(x)$ qui associe à chaque élément x de X une valeur dans l'intervalle des nombres réels [0, 1].

$$\mu_A : X \rightarrow [0, 1]$$

Ainsi un sous-ensemble flou A dans X peut être représenté par un ensemble de couples ordonnés (équation 7):

$$A = \{(x, A(x));\ x \in X, A(x) \in [0, 1]\} \quad (7)$$

3.2.3.2 Variables linguistiques

Toutes les variables numériques prennent des valeurs numériques, en logique floue, les variables linguistiques représentent, avec des mots (termes linguistique) avec des degrés d'appartenance dans l'ensemble X dans une situation donnée [Zadeh, 1996]. Cependant, les conditions d'observations ne permettent pas toujours de connaître parfaitement cette valeur unique. Donc, une variable linguistique sert à modéliser les connaissances imprécises ou vagues sur une variable dont la valeur précise peut être inconnue.

Une variable linguistique est généralement décomposée en un ensemble de termes linguistiques.

Exemple : Afin de décrire la température d'une pièce par une variable linguistique, on peut utiliser l'ensemble des termes suivants : T(V) = {froid, tiède, chaud}. En considérant que l'univers du discours est l'intervalle [0,45], on peut définir les termes linguistiques comme suit : « froid » est «une température environ inférieure à 10 degrés», «tiède» est «une température d'environ 17 degrés» et «chaud» est «une température environ supérieure à 24 degrés». Ces termes peuvent être caractérisés par les fonctions d'appartenance représentées sur la figure 3.4. N'importe quelle fonction d'appartenance, par exemple la fonction « froid », définit un nombre flou.

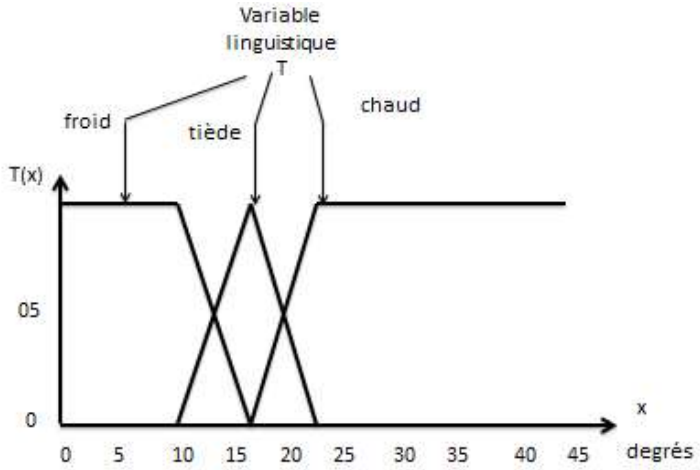

Figure 3-4 : Variables linguistiques (V,T(V),X) pour décrire la température [Zadeh, 1996].

3.2.3.3 Le raisonnement en logique floue

La formalisation des méthodes de déduction est l'une des apports principaux de la logique standard, qui sont en quelque sorte un outil de raisonnement. Les méthodes de déduction utilisées en logique standard permettent de définir une nouvelle certitude à partir d'autres connaissances certaines. Dans le cadre de la logique floue, il est possible de généraliser les méthodes de raisonnement lorsqu'on dispose de connaissances incertaines ou imprécises. Les méthodes d'inférence, utilisées habituellement en logique standard, peuvent être généralisées dans le cadre de la logique floue pour permettre de raisonner lorsque les règles ou les faits sont connus de façon imparfaite. La méthode d'inférence la plus connue est le modus ponens, qui permet de déduire une nouvelle connaissance en se basant sur la connaissance d'un seul fait et d'une seule règle.

a. Modus ponens généralisé

En logique booléenne traditionnelle, la règle d'inférence essentielle est appelée modus ponens, et elle est basée sur l'implication p → q, où p et q sont deux propositions binaires, définie de la manière suivante : Si (p → q) et (p est vrai) alors (q est vrai).

En logique floue, les sous-ensembles flous en entrée d'une règle floue ne correspondent pas exactement aux sous-ensembles flous des prémisses de cette règle, initialement définies. Pour obtenir une conclusion à partir d'une règle dont la prémisse est différente de la prémisse réelle, il est nécessaire d'appliquer la règle du modus ponens généralisé, introduite par Zadeh [Zadeh, 1968]. A partir d'une règle floue et d'un fait observé A′ pour V, on en déduit une valeur B′ pour la conclusion :

R : si V est A alors W est B
Observation : V est A′

Conclusion : W est B′

La fonction d'appartenance de la conclusion B′ est construite à partir de la fonction d'appartenance de l'observation A′ et de la règle floue (équation 8) :

$$\forall\, y \in Y,\ \mu_{B'}(y) = \sup_{x \in x} T\,(\mu_{A'}(x),\ \mu_{A \to B}(x, y)) \qquad (8)$$

La t-norme T utilisée, appelée operateur de Modus Ponens Généralisé, est choisie pour conserver les propriétés d'un Modus Ponens classique (si A′ = A alors, on doit en déduire que B′ = B). Elle doit dépendre de l'implication floue utilisée (et inversement).

Lukasiewicz Tl (u,v) = max (u + v − 1,0)
Probabiliste Tp(u,v) = u. v
Zadeh Tz (u,v) = min (u, v)

b. Méthodes d'inférence floue

Nous distinguons une variété importante de méthodes d'inférence floue, mais nous nous contentons d'en présenter les deux méthodes les plus utilisés. L'inférence de Mamdani (Min-Max) et l'inférence par la méthode additive. Les deux inférences diffèrent par la manière de déterminer les sorties.

Supposons que la base de connaissances contient n règles d'inférence contenant chacune m prémisses et une même conclusion y est Bi.

Règle 1 : Si (X_1 est A_{11}) et · · · et (X_m est A_{1m}) ; alors (Y est B_1)

Règle 2 : Si (X_1 est A_{21}) et · · · et (X_m est X_{2m}) ; alors (Y est B_2)
· · ·
Règle n : Si (X_1 est A_{n1})) et · · · et (X_m est A_{nm}) ; alors (Y est B_n)

La méthode Min-Max : se décompose en deux étapes, pour chaque règle appliquée, le minimum de degré d'appartenance est retenu dans le résultat. Par contre, si plusieurs règles donnent un même résultat, le maximum de ces résultats est retenu [Cox, 1994]. Ces opérations sont expliquées par :

$$\mu_{Ri}(y) = \min (\mu_{A11} (X_1), \mu_{A12} (X2) \ \ldots, \mu_{Aim} (X_m)).$$
$$\mu_{Résultat\ final} (y)= \mu_{B1} (y) \ = \min (\mu_{R1} (x),\mu_{R2} (x) ,\ldots, \mu_{Rn})$$

La méthode additive : Contrairement à la méthode Min-Max, dans la méthode additive le résultat final d'inférence est le minimum entre un «1», et l'addition de toutes les appartenances individuelles. Cette méthode est représentée par :

$$\mu_{i1}(y) = \min (\mu_{A11} (X_1) \ , \mu_{A12} (X_2) \ ,\ldots, \mu_{A1m} (X_m).$$

3.2.3.4 La défuzzification

La défuzzification est la dernière étape pour avoir un système floue opérationnel. Lors de l'inférence floue, on a généré un tas de commandes sous la forme de variables linguistiques (une commande par règle). L'objectif de la défuzzification est de réunir ces commandes et de transformer les paramètres résultants en donnée numérique.

Le résultat de l'inférence en utilisant une des méthodes d'implication floue, comme Max-min, est une valeur floue. Cette information ne peut être utilisée directement. Une transformation doit être prévue à la sortie du bloc d'inférence pour la convertir en grandeur fixe, cette transformation étant connue par le terme défuzzification (concrétisation).

Il existe plusieurs méthodes pour défuzzifier. Parmi les plus utilisés, on peut citer la méthode de la moyenne des maximas et la méthode du centre de gravité. [Cox, 1994].

- **La méthode du centre de maximum**

Dans cette méthode, la valeur de sortie est estimée par l'abscisse du point correspondant au centre de l'intervalle pour lequel la fonction d'appartenance est maximale (Figure 3.5) [Zadeh, 1973, 1968].

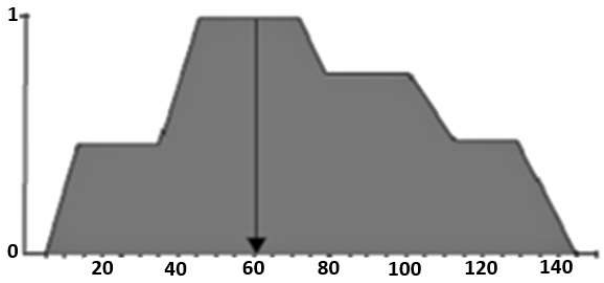

Figure 3-5 : La défuzzification par la méthode du centre de maximum [Zadeh, 1973, 1968]

- **La méthode du centre de gravité (COG)**

La seconde méthode du centre de gravité donne des meilleurs résultats et est largement utilisée dans les systèmes flous. Cependant, elle possède l'inconvénient d'être très couteuse. En effet, pour appliquer cette méthode de défuzzification, il faut calculer le centre de gravité de la surface sous la fonction d'appartenance et de prendre l'abscisse de ce centre de gravité. Pour cela, il faut décomposer la fonction d'appartenance en petits morceaux et intégrer sur chacun des morceaux [Zadeh, 1973, 1968].

La défuzzification par centre de gravité consiste à calculer l'abscisse du centre de gravité de la fonction d'appartenance (Figure3.6) :

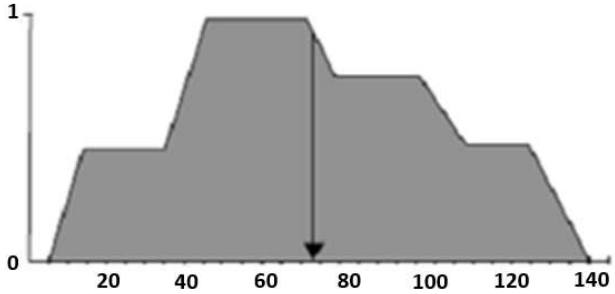

Figure 3-6 : La défuzzification par la méthode du centre de gravité

3.3 Conclusion

Pour pouvoir appliquer les différents algorithmes d'apprentissage sur les messages de type textuels, un ensemble de techniques ont été développé pour montrer comment l'information textuelle est habituellement prise en compte pour la représentation « informatique » de ces messages.

Ainsi avant la codification des messages, un ensemble d'opérations préliminaires doivent être faites pour épurer le texte de tous les mots inutiles et conserver seulement ceux qui sont porteurs d'informations et utiles pour le processus de classification. Mais malgré tous les prétraitements appliqués sur le message, l'espace des descripteurs, qui peuvent être des n-

grammes, des stems, des phrases, des concepts ou tout simplement des mots, reste très grand et très creux, d'où la nécessité d'une diminution préalable de cet espace.

Une fois la liste des descripteurs arrêtés, un degré d'importance ou poids est attribué à tous les termes présents dans la représentation vectorielle ou probabiliste puisque chaque terme possède un certain nombre d'occurrences dans le message ou dans le corpus qui est différents aux autres.

Finalement on peut qualifier notre texte, par fichier « informatique » apte à être employé dans les différentes méthodes d'apprentissage automatique.

La phase de catégorisation de textes et le choix de technique d'apprentissage, c'est le cœur du processus de classification automatique de textes, elle est située entre une phase primordiale, de préparation des messages et catégories à l'informatisation (codage des messages), pour le bon fonctionnement du processus, et une autre phase d'évaluation du ou des classifieurs utilisés.

Beaucoup d'approches différentes ont été utilisées pour la catégorisation de textes. Une des questions récurrentes est : quelle est la meilleure méthode pour la catégorisation de textes ? Il existe, en pratique, plusieurs méthodologies, pour tenter de répondre à cette question. Habituellement plusieurs paramètres peuvent influencer le choix de la technique de classification qui est relatif aux attentes des utilisateurs du système.

Nous avons constaté que les réseaux Bayésiens possèdent tous les avantages des modèles probabilistes standards, ainsi que des avantages supplémentaires liés à leur représentation graphique. De même, leur représentation facilite la compréhension dans un domaine de connaissances (elle représente directement les connaissances du domaine et non des procédures de raisonnement) et modélise explicitement tous les liens de dépendances entre variables. Nous avons montré que la difficulté dans le réseau bayésien consiste dans la construction de la structure en utilisant les différents algorithmes d'apprentissage de structure, ainsi l'estimation de ses paramètres, que ce soit les tables de probabilités s'il s'agit de données discrètes ou encore des probabilités des distributions gaussiennes s'il s'agit de données ou variables continues.

La deuxième partie de ce chapitre présente la logique floue. La logique floue est une technique qui représente l'expertise et la connaissance d'un expert humain. Cette méthode est très utile lorsque l'on se trouve confronté à des systèmes qui ne sont pas, ou difficilement modélisable. La logique floue permet de résoudre tous les problèmes dans lesquels on dispose de connaissances imprécises, soumises à des incertitudes de nature non probabiliste. Il permet de raisonner non pas sur des variables numériques, mais sur des variables linguistiques, c'est-

à-dire, sur des variables qualitatives (froid, petit, moyen, chaude, fort, etc) dans le but d'imiter au mieux le réel. De même, cette méthode est très avantageuse si l'on possède un bon niveau d'expertise humaine. En effet, il faut fournir au système flou toute une base de règles exprimé en langage naturel pour permettre de raisonner et de tirer des conclusions. Plus l'expertise humaine d'un système est importante et plus on est capable d'ajouter des règles d'inférences au système.

Le fait de raisonner sur ces variables linguistiques va permettre de pouvoir manipuler des connaissances en langage naturel. Tout ce que l'on a à rentrer au système, ce sont des règles d'inférences exprimées en langage naturel.

4 Modélisation du système d'aide au tuteur

Sommaire

4.1 Introduction 95
4.2 La problématique du suivi des apprenants 95
4.3 Analyse des conversations textuelles asynchrones d'apprenants pour déterminer leur comportement social 98
4.3.1 Les activités de collaboration 98
4.3.2 Objectifs recherchés dans la détermination de profils de comportements sociaux 100
4.3.3 L'analyse qualitative et quantitative des conversations textuelles asynchrones 101
4.3.4 Choix des indicateurs 103
4.3.5 Classification par approche Bayésienne 104
4.3.6 Niveau logique floue 106
4.4 Approche méthodologique 108
4.5 Conclusion 109

4.1 Introduction

L'importance croissante de la formation en ligne a mis l'accent sur le rôle du tutorat à distance. Un nouveau domaine de recherche, dédié à l'environnement pour l'apprentissage humain (EIAH), est en train d'émerger. Nous sommes préoccupés par ce domaine. Plus précisément, nous allons nous concentrer sur le suivi des apprenants dans le cadre d'un travail collaboratif en ligne.

Dans ce travail de recherche, nous nous intéressons à l'apprentissage collaboratif assisté par ordinateur (ACAO) (voir chapitre 2 - Etat de l'art). Une partie des recherches dans le domaine des ACAO est fortement intéressé par le suivi des activités de l'utilisateur, tout en interagissant avec les outils de communication fournis par chaque plate-forme d'apprentissage tel que des forums, email, etc (voir chapitre 2, sous-section 2.1.5). Ces outils sont devenus indispensables pour les situations d'apprentissage à distance. En effet, ils facilitent les communications apprenant-apprenant et tuteur-apprenant. De façon quasi-simultanée, le suivi des communications sur des outils de communication médiatisée par ordinateur (CMO), est devenu un sujet d'intérêt majeur de recherche et a mis en évidence un certain nombre de défis et de difficultés : l'analyse des traces recueillies, analyse des échanges, l'analyse et la classification de textes, la détermination de profils comportementaux et la définition (ou la réutilisation) d'indicateurs pertinents aptes à supporter l'analyse de traces, etc.

Dans le cadre de l'analyse des activités collaboratives entre apprenants, nous proposons une approche pour l'analyse automatique des conversations textuelles asynchrones entre apprenants pour déterminer leurs comportements sociaux [Pléty, 1998 ; Chaabi & all, 2014; Chaabi & all, 2015], les rôles sociaux attribués à chaque apprenant ainsi que l'organisation qu'ils mettent en place pour mener à bien un travail collaboratif dans le cadre d'une pédagogie de projet. L'assistance au tuteur permet de situer les travaux des apprenants par rapport aux autres apprenants et de situer leur groupe par rapport aux autres groupes d'une part et d'autre part permettre au tuteur d'accompagner les apprenants dans leur processus d'apprentissage, repérer et soutenir les individus en difficulté pour leur éviter l'abandon.

4.2 La problématique du suivi des apprenants

Le suivi des apprenants est une tâche complexe et est confronté à plusieurs problèmes. Le tuteur doit remédier au manque de rétroaction pour assurer la motivation des apprenants et d'éviter l'abandon. Les apprenants en ligne peuvent facilement se sentir isolés et rencontrer de

nombreuses difficultés lors des travaux de collaboration en ligne. D'autre part, le temps passé par le tuteur pour suivre et analyser les traces d'interaction et l'activité de chaque apprenant, temps qui peut vite devenir rédhibitoire. Pour économiser ce temps, nous proposons une approche multidisciplinaire s'inspirant entre autres des sciences humaines et sociales. Le principe de l'approche est basé sur l'analyse des échanges textuels entre apprenants. Cette analyse utilise le contenu des messages pour déterminer le comportement social de chaque apprenant.

Dans le cadre d'un travail collaboratif à distance les apprenants et tuteur sont séparés dans le temps et dans l'espace. Ce travail devrait se faire exclusivement en ligne en utilisant des moyens de communications qui peuvent être synchrones (visioconférence par exemple) ou asynchrones (mails, forums, …). Dans nos recherches, nous souhaitons nous intéresser plus particulièrement aux forums asynchrones que nous avons créé sur la plateforme Moodle. Il est important de rappeler que la communication sur ces forums permet à chaque apprenant de partager et de penser aux problèmes posés lors des discussions mises en jeu et que le forum de discussion asynchrone présente la caractéristique de préserver le fil des idées énoncées (les nouveaux sujets de discussion sont placés à la racine de l'arbre, les autres se raccrochant aux messages existants) (voir chapitre 2, sous-section 2.1.5).

Nous avons choisi l'analyse de communication entre apprenant sous forme textuelle (écrites) pour plusieurs raisons (figure 4.1). Premièrement, la possibilité d'analyser les contenus automatiquement. Deuxièmement, une liaison internet normale (premier niveau de vitesse de l'ADSL ou équivalent suivant les pays) est largement suffisante pour un outil de communication textuel. Troisièmement, une conversation textuelle est beaucoup plus exploitable qu'une communication audiovisuelle ou audio. Enfin, un avantage pédagogique : la communication en mode textuel permet davantage de réflexion de la part des apprenants. Les apprenants participant au travail ont été contraints de poser des questions, de produire des arguments et de se positionner par rapport aux avis de leurs pairs.

Au contraire, l'analyse des conversations orale ou audiovisuelle est très difficile lors d'un travail collaboratif en ligne à cause des problèmes de limites de bande-passante, qui ne permet pas d'avoir des qualités de son et de vidéo satisfaisante. De plus, la qualité des images et du son est très faible et ne permet pas de tenir une réelle conversation asynchrone à distance. Même si des progrès ont été réalisés dans le domaine d'analyse de conversations orales et audiovisuelles durant ces dernières années [Traverso, 2003 ; Wu, 2013; Hung, 2007], il est difficile de transposer ces travaux dans notre contexte d'analyse automatique de conversations asynchrones.

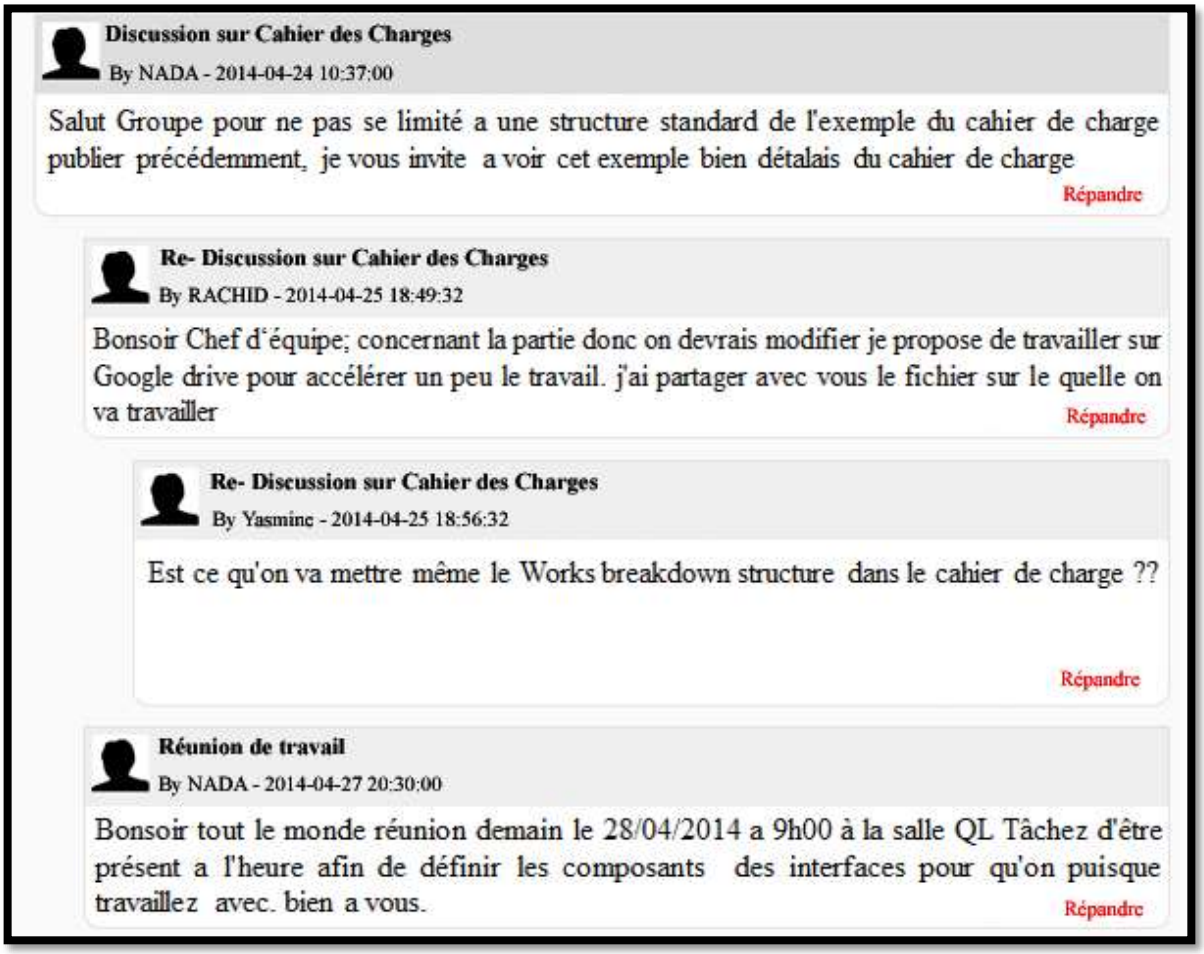

Figure 4-1 : Exemple de messages pendant une discussion

Il existe déjà des systèmes d'analyse de traces écrites. Ces systèmes supposent fréquemment que les messages soient semi-structurés [Oumaira, 2011 ; George, 2006; George, 2004]. Cela signifie que les apprenants doivent suivre un modèle et une syntaxe prédéfinie. Dans notre cas, les messages ne sont pas sous de telles contraintes. Le système analyse automatiquement les messages afin d'en extraire l'information pertinente.

Dans notre travail de recherche, nous proposons de lever certaines contraintes liées à la saturation de la communication et automatiser la détermination des comportements sociaux dans les forums libres (non structurés), à travers un système d'analyse automatique des conversations textuelles asynchrones entre apprenants. Concrètement, il s'agit d'un système multi-agent (SMA) qui permet de réaliser une séquence complète de traitements afin de déterminer les profils sociologiques des apprenants dans le cadre du processus de collaboration à distance. A partir de profils définis et adaptés à partir des travaux de Pléty [Pléty, 1998] (Chapitre 2, sous-section 2.1.6), nous calculons des indicateurs de nature

quantitative et qualitative, qui couplés aux classifications de messages, permettent d'attribuer un profil sociologique à chaque apprenant.

L'objectif principal de ce chapitre est de souligner l'originalité de nos travaux et de motiver notre choix d'utiliser la logique floue et naïve bayes avec les SMA. Ce chapitre est divisé en trois parties. La première partie s'attache à décrire les objectifs recherchés dans la détermination de profils de comportements sociaux. Puis, nous présentons les indicateurs qui peuvent nous aider à déterminer les comportements sociaux des apprenants lors d'un travail collaboratif en ligne. La deuxième partie de ce chapitre présente la définition des modèles agents pour la satisfaction des buts du système et leur contribution dans les architectures des systèmes d'enseignement assisté par ordinateur. Nous avons intégré à notre système, les approches « naïve bayes » et « logique floue » dans le souci de reproduire la flexibilité du raisonnement humain (tuteur) quant à sa capacité de prise en compte des imperfections des données. En dernier lieu, nous présentons l'approche méthodologique qui vise à apporter une réponse aux différents éléments de problématique dégagés plus haut et à remplir nos objectifs de recherche. Pour cela, nous avons participé à la mise en place d'une expérimentation réelle de formation à distance. Cette expérimentation nous a permis d'évaluer l'apport des différents indicateurs intégrés à notre système et de mesurer l'influence de notre système sur les activités des tuteurs.

4.3 Analyse des conversations textuelles asynchrones d'apprenants pour déterminer leur comportement social

4.3.1 Les activités de collaboration

Dans un travail en groupe le taux d'abandon est très élevé à cause des mauvais comportements sociaux et d'un manque d'encadrement de qualité. Ces comportements peuvent influencer la qualité de la collaboration et de travail. Selon les études de [Bernatchez, 2003 ; Rodet, 2003 ; Lièvre, 2006], de l'ordre de 25% à 50% du taux d'abandon est généralement pointé comme la conséquence d'un isolement de l'apprenant. Ses études ont été réalisées à partir de tous types de publics, niveaux, et toutes formations confondues (internes, universitaires, professionnelles) et puisent dans différentes sources. Pour réduire le risque d'apparition de tels comportements, une solution pour résoudre ce problème et pour améliorer la persistance serait de mettre en place un encadrement de qualité qui englobe toutes les formes de soutien offerts à l'apprenant engagé dans un travail collaboratif à distance.

Pendant un travail en groupe, une organisation se met en place spontanément. Les apprenants se positionnent les uns par rapport aux autres en fonction des comportements sociaux. Des bons et des mauvais comportements peuvent apparaitre. Par exemple :

- « ***L'autostoppeur*** » se dit que les autres membres du groupe en font déjà assez et n'est donc pas motivé pour contribuer ;
- « ***Animateur*** » Fait des propositions, poste des messages d'organisation et/ou encouragement. Intervient pour calmer un conflit.
- Le « ***Paresseux social*** » va faire moins d'efforts quand il est en groupe, tentant de profiter du groupe sans faire l'effort de contribuer ;
- Le « ***Sucker effect*** » est au contraire la décision des personnes pouvant amener le groupe à faire moins (souvent en conséquence à la découverte « d'autostoppeurs » ou de « paresseux sociaux » dans le groupe) pour ne pas devenir celles qui travaillent pendant que les autres ramassent les bénéfices sans rien faire.

L'observation de ces comportements dans le cadre d'une pédagogie de projet fournit à l'enseignant encadrant des indicateurs pour réagir, comprendre et intervenir auprès des groupes. Pour résoudre ces problèmes, nous pensons qu'il est utile d'avoir un système informatique capable d'analyser automatiquement les conversations textuelles asynchrones entre apprenant et par la suite associer un profil comportemental à chaque apprenant. D'après les études de Pléty, dans chaque groupe les élèves jouent les rôles de l'animateur, vérificateur, quêteur et indépendant. Pléty indique qu'une organisation se met en place spontanément, et qu'« au sein de cette organisation chacun des partenaires sait trouver, suivant ses capacités et des connaissances personnelles, sa propre place » [Pléty, 1998]. Ses études lui ont permis d'identifier des profils de comportement qui caractérisent les rôles que les apprenants jouent dans le groupe. Trois sortes d'observations sont effectuées sur chaque apprenant.

- Le volume d'intervention
- Le type d'intervention
- Les réactions des autres aux interventions.

Les caractéristiques de ces quatre profils sont résumées dans le tableau 4.1 suivant :

profil	Volume des interventions	Type des interventions	Réactions entrainées
Animateur	Important	Fait une proposition, poste un message d'organisation et/ou encouragement. Intervient pour calmer un conflit.	Suivi de réactions.
Vérificateur	Assez Important	(Propose éventuellement), Réagi à une proposition. Evalue une proposition.	Peu suivi de réactions
Quêteur	Peu Important	(Ne fait pas de proposition) Pose des questions ou il exprime ses doutes sur une démarche ou proposition (esprit plutôt négatif).	Ses questions sont très bien Acceptées
Indépendant	Faible	Ses interventions semblent décalées par rapport à la discussion en cours et souvent non suivies de réactions des autres membres du groupe.	Ses interventions restent en suspens

Tableau 4.1 : Caractéristiques des profils comportementaux d'apprenants travaillant en groupe – Adaptation à partir des travaux de [Pléty, 1998]

Nous nous sommes alors posés la question de savoir si, à distance, il est possible de retrouver ces mêmes caractéristiques de groupes d'apprenants. Nous avons donc émis l'hypothèse que les caractéristiques des profils comportementaux relevés par Pléy dans des groupes classiques se retrouvent pour ceux qui travaillent à distance.

4.3.2 Objectifs recherchés dans la détermination de profils de comportements sociaux

Une des difficultés qui se posent dans le cas d'apprentissage en ligne est de comment déterminer le comportement des apprenants sachant qu'il y a séparation dans l'espace et le temps. Dans le contexte de formation en ligne, le tuteur perd une certaine perception de ce qui se déroule dans sa classe par rapport à une formation en présentiel. En effet, les données récupérées tendent à être plus imparfaites que celles obtenues lors d'une interaction directe. Ce processus soulève la question du temps passé par le tuteur chargé de suivre et d'analyser l'activité de chaque apprenant.

Le comportement social présenté par les apprenants au sein de la classe, détermine principalement comment les apprenants interagissent entre eux dans un groupe. Bien que le comportement social soit souvent modélisé pour créer un environnement social confortable, un comportement antisocial, comme l'indépendance, peut également être défini comme un comportement social négatif, ces comportements peuvent influencer les comportements des

autres apprenants et sur la qualité de collaborations dans un travail en groupe. Seules les interactions positives entre les apprenants, aident les groupes à créer un environnement agréable pour partager des informations et réaliser des travaux de qualité.

Dans le contexte de pédagogie de projet, il est intéressant de déterminer automatiquement les profils comportementaux identifiés par Pléty. Ces profils permettraient au tuteur de mieux comprendre les relations sociales dans l'équipe, lui donner une vue sur les comportements des apprenants, lui donner ainsi un autre regard sur le processus de travail et donc de mieux suivre l'activité de projet, dans le but d'éviter l'abandon de ses apprenants et d'équilibrer les équipes. De même, pour les apprenants, la perception des comportements des individus du groupe permet de mieux réguler le travail collectif.

Notre approche, consiste à analyser automatiquement le contenu des messages pour ensuite en déduire des profils comportementaux comme étant une étape pour déterminer le comportement social des apprenants, ou l'interactivité entre tuteurs-apprenants ou apprenants-apprenants se fait essentiellement à travers l'échange textuel. Le défi consiste alors à déterminer des profils comportementaux des apprenants à travers l'analyse automatique du contenu des messages échangés.

Le problème est de savoir comment transposer les profils définit par Pléty dans notre contexte de conversation asynchrone. Pour ce faire, il a été question de procéder à une analyse automatique de contenu de conversation textuelle entre élèves.

4.3.3 L'analyse qualitative et quantitative des conversations textuelles asynchrones

Rappelons qu'un des sous-objectifs de notre travail est l'analyse automatique des contenus des conversations textuelles dans le but de déterminer les comportements sociaux des apprenants et par la suite d'améliorer la perception du tuteur quant à la nature des échanges entre apprenants. Nous avons vu précédemment qu'une conversation pouvait être considérée comme une succession d'échanges constitués d'actes de langage dont l'un est l'acte directeur. Si l'on cherche à analyser automatiquement une conversation, il est alors important de repérer ces actes de langage directeur par une analyse du contenu des interventions (Chapitre 2, sous-section 2.1.5).

Pour commencer, nous souhaitons préciser le contexte d'extraction des actes de langage, nous avons travaillé régulièrement avec un certain nombre de tuteurs sur des échanges entre apprenants, nous en sommes venus à aborder un ensemble d'actes de langage spécifique pour chaque profil.

L'analyse intuitive des messages montre que ces derniers peuvent être classés comme suit : des messages qui ont comme but d'initier une interaction et d'amorcer un sujet de discussion proposition, des messages où on demande des informations et où on attend une réponse de la part d'autrui, des messages où on répond aux sollicitations des autres, où on répond aux questions et aux interrogations des autres et enfin des messages antérieurs qui clarifient ou approfondissent un sujet actuel de discussion. Les actes de langage proposés par chaque catégorie dépassent la quarantaine.

La plus grande difficulté dans l'analyse des conversations par des actes de langage est sans aucun doute celle de choisir l'acte approprié à ce que l'on veut dire. Si ce choix est trop contraignant, notre système risqué de choisir la première venue. Afin de faciliter cette sélection, les actes proposés doivent avoir une grande distance sémantique.

Les actes de notre système sont choisis relativement aux indicateurs de type d'intervention relevés par Pléty (colonne « type d'intervention » du tableau 4.1) à savoir : proposition, réaction, évaluation, interrogation, réponse et question. En se basant sur les propositions des tuteurs et sur la littérature relative au sujet [Pléty, 1998 ; George, 2004], nous avons défini des actes de langage répartis en quatre catégories (Animateur, Vérificateur, quêteur et indépendant).

Nous choisissons de présenter les actes de langage sous forme de verbes (tableau 4.2). En effet, un verbe est plus proche de l'action de l'utilisateur qui veut exprimer quelque chose. Notre choix s'oriente vers un ensemble de verbes qui expriment des forces illocutoires de chaque profil d'apprenant.

Catégories	Actes de Langage
Animateur	Proposer, demander, affirmer, soumettre, présenter, conseiller, démontrer, calmer, etc.
Vérificateur	Approuver, agréer, adopter, consentir, valider, supporter, rejeter, exclure, renvoyer, comparer, confirmer, justifier, contrôler, essayer, etc.
Quêteur	Comment, pourquoi, quand, quoi, quel, est-ce-que, qu'est-ce que, ?, etc.
Indépendant	Soumettre, présenter, offrir, conseiller, etc.

Tableau 4.2 : Exemple des actes de langage proposés.

Sur la base des travaux réalisés par [Pléty 1998], notre système calcule trois indicateurs : volume des interventions, réactions entrainées (ce que les interventions provoquent) et le type des interventions.

4.3.4 Choix des indicateurs

En nous fondant sur les travaux de Pléty concernant la détermination de profils comportements sociaux d'apprenants travaillant en groupe. Les indicateurs permettant de déterminer ces profils « peuvent être classés en deux natures : d'une nature quantitative en ce qui concerne le volume d'intervention et réaction entrainée (réaction directe et indirecte) et d'une nature qualitative pour ce qui est du type d'intervention. Nous rappelons que Pléty a identifié quatre profils [Pléty, 1998] : Animateur, Vérificateur, Quêteur et Indépendant. Nous présentons ici les indicateurs utilisés par le système pour calculer les profils comportementaux des utilisateurs pendant une conversation asynchrone.

Volume d'intervention : indicateur qui représente l'activité des étudiants, une forte activité est prouvée par le nombre de messages postés sur le forum. Ces messages montrent qu'outre les connexions sur le forum, les apprenants font un travail personnel chez eux. Les apprenants ne se limitent pas à poster des messages en réponse à d'autres envoyés par leurs pairs mais continuent à les consulter afin de se constituer une idée plus globale sur le sujet de discussion grâce aux discussions en jeu. C'est ainsi qu'ils se réfèrent à ce qui a été dit (écrit) pour relancer une discussion, s'informer, échanger et collaborer.

Type d'intervention : Lors d'échange, l'intervention d'un apprenant en travail collaboratif vise, d'une part, à accompagner une personne, calmer un conflit, …, etc, d'autre part à obtenir le plus grand nombre d'information et ressource afin qu'elle puise participer activement dans le travail et au devenir collectif et social dans son groupe.
L'indicateur « type d'intervention », d'après l'analyse de contenu de message, correspond au nombre de messages de type animateur, vérificateur, quêteur et indépendant envoyés par l'apprenant pour identifier de quelle catégorie de profil il s'agit. Comme exemple l'animateur concentre ses interventions sur la productivité du groupe. Il intervient alors pour faire expliquer le problème, structurer les échanges et coordonner les apports dans le but d'engager le groupe dans un processus consensuel pouvant le mener au final à une prise de décision commune. Par contre l'indépendant ses interventions semblent décalées par rapport à la discussion en cours et souvent non suivies de réactions des autres membres du groupe.

Réactions Entrainées : Il est important de rappeler que la communication asynchrone permet à chaque apprenant de partager et de penser aux problèmes posés lors des discussions mises en jeu et que le forum de discussion présente la caractéristique de préserver le fil des idées énoncées. Les apprenants ont été contraints de poser des questions, de produire des arguments et de se positionner par rapport aux avis de leurs pairs. Ils ont été obligés à collaborer, à produire des écrits et à s'organiser en groupes.

Selon les caractéristiques des profils définis (tableau 4.1), le volume de réactions entraînées par un message permet de caractériser un profil comportemental. Par exemple, un profil d'animation nécessite un suivi de réactions très important par rapport à celle d'un vérificateur. Une intervention d'un animateur se caractérise par le fait qu'elle entraîne souvent des réactions. D'une manière plus précise les sujets qui ont un grand nombre de réponses et interactions sont des sujets d'intérêt pour le groupe. Le système calcule ensuite, pour chaque message, les réactions directes (les premières réactions à un message) et les réactions indirectes (nombre d'interventions après le lancement du message).

Dans les soucis d'épargner un temps précieux passé à visualiser et analyser les traces d'interactions de chaque apprenant. Nous proposons de lever certaines contraintes liées à la saturation de la communication et automatiser la détermination des comportements sociaux dans les forums libres (non structurés), à travers un système d'analyse automatique des conversations textuelles asynchrones entre apprenants. Concrètement, il s'agit d'un système multi-agent (SMA) qui permet de réaliser une séquence complète de traitements afin de déterminer les profils sociologiques des apprenants dans le cadre du processus de collaboration à distance.

4.3.5 Classification par approche Bayésienne

De nombreuses recherches dans le domaine des environnements de l'apprentissage humain sont fortement intéressées par le suivi des activités de l'utilisateur, tout en interagissant avec des outils de communication fournis par chaque plate-forme d'apprentissage. Ces outils de communication asynchrones tels que les e-mails et les forums, permettent aux apprenants de communiquer entre eux et d'organiser leurs activités en vue d'atteindre l'apprentissage en ligne.

Face à la grande quantité d'information enregistrée dans les messages déposés dans les forums, les messages postés par les différents utilisateurs via courrier électronique et les

discussions sur l'espace du chat, les tuteurs d'une communauté de pratiques se sentent souvent démunis pour construire une représentation synthétique de l'activité des individus et des groupes. Ils doivent souvent se satisfaire de l'idée qu'ils s'en font à partir du souvenir de leur lecture. Le tuteur risque de manquer d'objectivité quand il s'en sert pour évaluer l'implication et la place des apprenants dans les échanges, ce qui rend plus difficile son soutien auprès des individus.

Une conversation peut être considérée comme une succession d'échanges composés d'interventions de locuteurs et constituées elles-mêmes d'actes de langage dont l'un est l'acte directeur [Roul et Al, 1985] qui détermine la force illocutoire de l'intervention. Si l'on cherche à analyser automatiquement une conversation, il est alors important de repérer ces actes de langage directeurs.

Dans la littérature, de nombreux systèmes existants basés sur l'analyse des conversations par des actes de langage afin de faire ressortir des profils comportementaux des utilisateurs [Chaabi et al., 2015 ; George, 2004]. Ces approches semi-structurées par des actes de langage présentent des inconvénients ; les résultats ne sont pas toujours fiables, les étudiants sont freinés lors de la rédaction d'un message, et elles ne permettent pas aux apprenants de se concentrer sur leurs tâches. Il y a un autre inconvénient de l'utilisation des actes de langage, dans le cadre d'analyse des forums libre « non structuré ». L'absence d'une normalisation des actes de langage pour déterminer les comportements sociaux des apprenants, rend la tâche d'analyse et de classification très difficile.

D'après ces inconvénients, pour tenter de répondre à ces besoins nous proposons ici l'utilisation d'une approche probabiliste capable de modéliser la notion d'incertitude au sens d'un choix parmi un ensemble des classes. Ce type d'approche aborde l'imprécision de manière quantitative. Il s'agit d'estimer la fréquence (les chances) d'occurrence d'un message respectivement a une classe (animateur, vérificateur, quêteur et indépendant).

Dans sa forme le plus simple, la catégorisation de messages entre apprenants consiste à assigner à un texte une ou plusieurs étiquettes permettant d'indexer le message dans un ensemble prédéfini de catégories (animateur, vérificateur, quêteur et indépendant). Il s'agit d'un processus qui consiste à mesurer la similarité textuelle permet d'identifier la ou les catégories parmi une liste prédéfinie les plus proches du message à classer. Ce processus est basé sur une phase d'apprentissage. Un ensemble de messages bien choisis est donc fourni au départ comme représentatif des catégories voulues. Cet ensemble permet ensuite de catégoriser de nouveaux textes (voir chapitre 3, section 3.1).

Afin de comprendre les techniques de classifications présentées, le concept de message doit être examiné (voir chapitre 3, sous-section 3.1.6).

Dans ce travail, nous poursuivons un double objectif, d'une part, en proposant une classification des conversations textuelles par un algorithme d'apprentissage supervisé « Naïve bayes ». D'autre part, une comparaison entre les deux méthodes de classification, celle basée sur les actes de langage et l'autre sur naïve bayes pour identifier du type d'intervention. Ainsi, l'utilisation de naïve bayes pour l'outil proposé, plutôt que d'être un obstacle, doit permettre l'amélioration des résultats d'analyse.

Naïve bayes est l'une des techniques de classification de texte les plus élémentaires ayant diverses applications en catégorisation des documents, la détection spam, tri des emails et détection de la langue [Qiang, 2010 ; Meena, 2009] (voir chapitre 3, sous-section 3.1.7.1).

Le classificateur Bayésien donner de bons résultats dans de nombreux problèmes complexes. Même si elles sont souvent surpassées par d'autres techniques telles que l'arbre de décision, Support Vector Machines, etc. les classifieurs Bayésiens sont très efficaces car ils demandent moins de calcul [Huang.J, 2003] et nécessitent une petite quantité de données de formation.

4.3.6 Niveau logique floue

Il est difficile de tenter de modéliser le comportement de l'être humain dans le cadre d'un travail à distance. Notre objectif est d'avoir une image du comportement social de chaque apprenant lors de conversation en groupe.

Maintenant que nous avons vu les indicateurs proposés pour notre système, leur analyse à la lumière des caractéristiques des profils définis (voir tableau 4.1) est réalisée pour permettre aux tuteurs d'associer un profil à chaque apprenant. La présence d'information imparfaite est un facteur important qui peut conduire à des erreurs durant l'évaluation de l'apprenant. Ces imperfections sont la conséquence des approximations impliquées dans la collecte de données en raison de la nature de la connaissance humaine. Le processus d'évaluation est bien évidemment fortement influencé par cette imprécision. Afin de traiter ce problème, nous proposons de mettre en œuvre une approche fondée sur la logique floue pour gérer l'incertitude et exprimer les connaissances qualitatives des experts de manière lisible. La logique floue représente la connaissance des enseignants-experts sous forme linguistique et comprend les caractéristiques de l'apprenant. En d'autres termes, une décision est prise par

une combinaison d'un ensemble de faits flous, chacun contribuant à un certain degré à une relation floue et donc à la décision finale.

La plupart des problèmes rencontrés sont modélisable mathématiquement mais ces modèles nécessitent des hypothèses parfois trop restrictives rendant délicate l'application au monde réel. Les problèmes du monde réel comme notre cas doivent tenir compte d'informations imprécises, incertaines. Les connaissances dont disposent les tuteurs sur les comportements sociaux des apprenants ne sont presque jamais parfaites. Les connaissances sur lesquelles se base le raisonnement humain sont presque toujours entachées d'une quantité d'incertitudes et d'imprécisions. Nous ne parlons pas ici du raisonnement scientifique, dont l'objectif est justement de se débarrasser de toute imperfection, mais de tous les autres raisonnements que nous faisons tous les jours, sans cesse, sur les choses, les personnes et les pensées nous environnant.

Exemple de modélisation floue : considérons un client qui consulte, via Internet des offres de location de voitures. Le client souhaite trouver une voiture de préférence ayant une consommation de carburant « moyenne » et un loyer « modéré » avec un nombre de places entre 4 et 6. D'autre part, le client souhaite que les données retournées soient ordonnées et présentées à lui selon ces préférences (consommation non forte, loyer non grand, et nombre de places plus de 5).

Pour représenter le besoin du client nous utilisons les variables linguistiques suivants :

Consommation de carburant : pour décrire la consommation d'une voiture on utilise l'ensemble des termes suivants : {faible, moyenne, forte}. Ces termes peuvent être caractérisés par les fonctions d'appartenance (voir chapitre 3, section 3.2). De même pour le loyer : pour décrire le loyer d'allocation d'une voiture on utilise l'ensemble des termes suivants : {petit, moyen, grand}. Et pour le nombre de places on utilise l'ensemble des termes suivants : {proche, loin, très-loin}.

Lorsqu'un propriétaire d'une agence de voiture entré dans un marché, le système multi-agent crée un agent vendeur, qui lui permet de présenter ces voitures dans un marché. Dans le cas où le vendeur demande l'estimation de son bien, l'agent vendeur envoie les caractéristiques de son bien vers l'agent flou. L'agent flou calcule le prix approximatif par inférence floue et retourne le résultat ver l'agent vendeur.

La logique floue semble donc reproduire la flexibilité du raisonnement humain quant à sa prise en compte des imperfections des données accessibles. Il serait donc intéressant de l'utiliser au cœur des systèmes experts, systèmes dont le but est de reproduire les mécanismes cognitifs d'un expert, dans un domaine particulier qui est l'analyse automatique de

conversations pour la détermination des comportements sociaux. La logique floue peut également servir pour un système décisionnel, lors de la phase d'analyse des données par exemple et peut s'avérer utile pour la prise de décision, soit pour découvrir des règles ou inférences floues permettant de mieux comprendre les données et ainsi éclairer les décisions, soit pour effectuer des requêtes dites floues en se basant sur les connaissances des experts.

4.4 Approche méthodologique

Nous avons demandé à des étudiants du master de former des groupes de quatre apprenants et de mettre en place des projets mobiles dans le cadre de leurs cours de développement mobile. Le service pédagogique est basé sur un tutorat où le tuteur est considéré comme un accompagnateur encadrant, suffisamment disponible pour guider et répondre aux questions des apprenants lorsque ceux-ci rencontre des difficultés dans la gestion et la réalisation de leurs projets. Dans un dispositif de formation à distance, la mise en ligne des ressources pédagogiques, assortie d'un simple échange de questions-réponses entre apprenant et tuteur n'est qu'une forme très réductrice du processus d'apprentissage.

C'est la raison pour laquelle des expérimentations ont été mises en place pour mettre à l'épreuve des modèles pédagogiques notamment l'Apprentissage Collectif Assisté par Ordinateur (ACAO).

La démarche pédagogique consiste à faire travailler les apprenants par groupes de quatre sur des projets planifiés sur des périodes (phases) de un à quatre mois, (du 02 mars 2013 au 02 juin 2013), organisée en 6 phases, correspondant à des tâches différentes, basés exclusivement sur le travail collaboratif à distance. Pour chaque phase du projet, le nombre de contributions échangées par les apprenants de chaque groupe dépasse cependant largement la centaine (100 et 110), ce sont donc des forums actifs.

Pour soutenir leurs travaux, nous avons mis en place un learning management system LMS de formation à distance permettant aux membres du même groupe de communiquer, d'organiser et de partager des idées, des contributions et des informations. Les étudiants n'étaient pas au courant que nous voulons analyser leurs messages écrits et à travers ceux-ci nous allons attribuer un profil approprié à chacun d'entre eux.

Cet environnement se présente sous la forme d'un ou plusieurs forums de discussion, intégrant également un service de publication de documents. Cette analyse de conversation textuelle se base sur les caractéristiques des profils comportementaux définis plutôt (Tableau 4.1). Le but est de normaliser l'analyse des tuteurs, pour comparer les résultats d'analyse.

Cette expérimentation nous a permis de valider notre système d'un point de vue fonctionnel et ergonomique, le tuteur a apprécié la qualité de système et l'a jugé simple d'utilisation et facile à prendre en main.

Les résultats de cette expérience ont montré que les tuteurs, en utilisant notre système d'analyse automatique des conversations entre apprenant, sur la base d'informations quantitatives et qualitatives, arrive à dériver une information qualitative sur les comportements sociaux des apprenants et des groupes en corrélant les indicateurs entre eux.

4.5 Conclusion

Le travail collaboratif en ligne pose des problèmes spécifiques. Au nombre de ces problèmes, on note l'isolement éprouvé par les utilisateurs et le déficit d'accompagnement des apprenants ; tout ceci conduit à un grand taux d'abandon et déperdition. Devant le fort taux d'abandon observé, il serait particulièrement intéressant de déterminer les comportements sociaux des apprenants engagés dans un travail collaboratif en ligne.

Nos travaux se situent dans le champ de l'apprentissage collaboratif à distance. Nous nous intéressons à la construction d'outils informatiques intelligents capables d'analyser automatiquement les traces d'interactions entre apprenant pour déterminer ses comportements sociaux dans l'objectif de fournir un soutien et une assistance aux tuteur pour surveiller les apprenants engagés dans un travail à distance.

Pour le tuteur responsable de l'encadrement des apprenants lors d'une formation à distance, des informations sur des profils de comportements sociaux lui permettraient de mieux comprendre les relations sociales existantes. Dans le cas d'activités pédagogiques collectives, le tuteur pourrait alors mieux suivre l'activité qui se déroule et donc mieux encadrer les groupes d'apprenants. D'une manière générale, nous pensons que des indications sur les comportements sociaux des apprenants peuvent contribuer à aider le tuteur dans son suivi pédagogique, suivi qui constitue un point crucial pour la bonne réussite d'une formation à distance. En particulier, le suivi d'un tuteur aide à limiter les abandons liés au sentiment d'isolement des apprenants. Pouvoir repérer rapidement un apprenant isolé est un enjeu important. Notre proposition est donc de soutenir informatiquement le suivi du tuteur. Le tuteur ne peut pas être présent à chaque connexion des apprenants. Un système informatique l'assistant dans sa tâche de suivi sera, dans ce contexte, d'une aide précieuse. L'objectif est de concevoir des systèmes présentant des vues synthétiques des informations pertinentes pour le tuteur. Toujours du côté des enseignants, nous voyons d'autres perspectives possibles à

l'utilisation des profils de comportement. Par exemple, ces profils peuvent constituer une aide à la formation de groupes d'apprenants. Un enseignant aurait la possibilité d'utiliser les profils en cherchant à « équilibrer » un groupe ou, au contraire, en cherchant à créer des conflits intéressants d'un point de vue pédagogique.

En ce qui concerne les apprenants, leur fournir des informations sur leurs profils de comportement social permettrait de leur donner une vue sur leurs actions et leur donner ainsi un autre regard sur leur processus de travail. Cet « effet miroir » est l'une des aides possibles d'un environnement support d'apprentissage collectif. Nous pouvons même aller au-delà et utiliser les profils analysés pour adapter les environnements aux apprenants. Par exemple, un apprenant ayant un comportement d'animateur se verrait attribuer des outils particuliers d'animation.

Dans notre travail, nous nous intéressons au problème de conception de systèmes multi agents dotés des caractéristiques flous pour réponde aux besoins d'analyse des conversations textuelles pour la détermination des comportements sociaux. Un système complexe moderne qui doit traiter avec des informations imparfaites. Pour résoudre ce problème nous avons proposé dans ce mémoire une nouvelle approche pour la conception de ce système multi-agents. Notre système s'appuie sur l'idée d'intégrer la logique floue dans un niveau multi-agent, par l'utilisation des agents flous. En effet, Notre système bénéficié des avantages de l'approche multi-agent et des capacités de la logique floue, tel que sa capacité de représenter et de manipuler des connaissances imparfaites.

5 Définition des mécanismes de base des différents agents

Sommaire

5.1 Introduction 112
5.2 Architecture multi-agents pour l'analyse et l'assistance 112
5.3 Les agents du système 114
5.3.1 Agent Récupération 114
5.3.2 Agent Filtrage 116
5.3.3 Agent Lemmatisation 119
5.3.4 Agent Classification de messages 123
5.3.4.1 Classification par actes de langage 123
5.3.4.2 Classification Bayésienne 125
5.3.5 Agent indicateur 135
5.3.5.1 Volume d'interventions : 135
5.3.5.2 Type d'interventions 136
5.3.5.3 Réactions Entrainées 137
5.3.6 Agent Logique floue 139
5.3.6.1 Première étape : Fuzzification 140
5.3.6.2 Deuxième étape : Le moteur d'inférence 143
5.3.6.3 Troisième étape : défuzzification 144
5.4 Conclusion 144

5.1 Introduction

Les Systèmes Multi-Agents, au départ issu de travaux menés en intelligence artificielle, se sont révélés comme un nouveau paradigme pour la conceptualisation, la conception et la mise en œuvre de systèmes logiciels. Les agents sont des programmes informatiques sophistiqués qui agissent de manière autonome, dans des environnements ouverts et distribués, pour résoudre un nombre croissant de problèmes complexes. Dans la plupart des cas, ces applications nécessitent plusieurs agents qui travaillent ensembles à la résolution collective de problèmes.

Les systèmes multi-agents (SMA) ont montré leur pertinence pour la conception d'applications distribuées (logiquement et/ou physiquement), complexes et robustes. Le concept d'agent est aujourd'hui plus qu'une technologie efficace, il représente un nouveau paradigme pour le développement de logiciels dans lesquels l'agent est un logiciel autonome qui possède un objectif, évolue dans un environnement dynamique et interagit avec d'autres agents au moyen de langages et de protocoles. Souvent, l'agent est considéré comme un objet « intelligent » ou comme un niveau d'abstraction au-dessus des objets et des composants. Les méthodes de développement orientées objet – au vu des différences entre les objets et les agents – ne sont pas directement applicables au développement de SMA.

Il existe plusieurs approches pour l'ingénierie de SMA. Parmis ces approches, notre choix s'est porté sur ASPECS. En effet, cette méthode permet de couvrir l'ensemble du cycle de vie d'un SMA de son analyse à son développement et son déploiement en utilisant une des plateformes dédiées. De plus, les concepts organisationnels à la base de cette méthode nous paraissent adapté au traitement du problème soulevé dans cette thèse.

5.2 Architecture multi-agents pour l'analyse et l'assistance

L'objectif principal de notre système est l'analyse des interactions et des contenus en vue de supporter les apprenants et les tuteurs dans le cadre d'un travail collaboratif en ligne. Cette analyse se fait par le biais des conversations textuelles asynchrones.

Les objectifs du système sont :

- Récupérer les données d'interactions,
- Analyser ces données en calculant un certain nombre d'indicateurs,
- Évaluer les profils comportementaux des apprenants.

Nous avons défini, selon la méthodologie ASPECS, un ensemble de comportements collaboratifs qui contribuent à la satisfaction de ces objectifs.

Ces comportements sont décrits par la figure 5.1. La notation utilisée est celle définie au sein d'ASPECS. Dans la partie supérieure du schéma est décrite l'organisation (ensemble de rôles en interactions). Cette organisation définit six rôles : *prétraitement, filtrage, Lemmatisation, classification, indicateur et logique floue*. Ces rôles représentent des comportements abstraits qui sont mis en œuvre par des agents du système.

Nous avons six agents simples : un agent (1) prétraitement, (2) filtrage, (3) lemmatisation, (4) classification, (5) indicateurs et (6) logique floue en utilisant des rôles (six rôles) dans un contexte organisationnel.

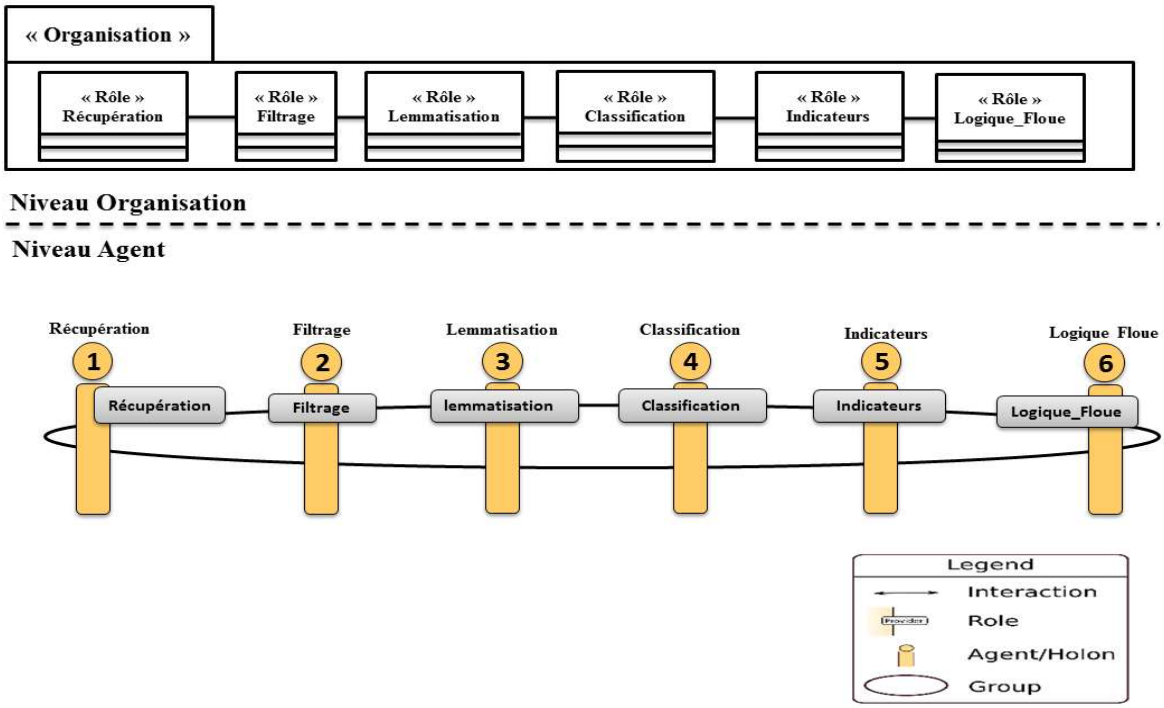

Figure 5-1 : Architecture de système

Le rôle récupération est une sorte d'interface du système avec l'environnement d'EIAH. Ce rôle a pour objectif d'observer les traces d'interactions et de nourrir le système avec des nouvelles traces. Ce rôle est donc à la frontière du SMA et il sera mis en œuvre de manière spécifique pour chaque EIAH. De plus, ce rôle procède à une étape de correction orthographique des messages.

Les rôles de filtrage et lemmatisation constituent une phase de prétraitement du corps des messages. En effet, de part le fait que les messages ne suivent pas un schéma prédéfini la langue naturelle est employée ce qui rend leur analyse directe complexe. Afin de pallier à ce problème, les deux rôles simplifient les messages afin de les rendre exploitables dans le cadre d'une analyse automatique. Il s'agit, en particulier, de diminuer voire d'éliminer les incohérences lexicales et les ambiguïtés syntaxiques contenus dans les textes. Le premier

rôle, filtrage, applique un filtre qui supprime certains mots des messages. Ces mots sont supposés n'apporter aucune information ou ne pas être exploitable par le processus d'analyse. Le deuxième rôle, lemmatisation, donne au message une forme canonique dont l'objectif est de faciliter l'analyse. Cette transformation se fait sur la base des champs lexicaux des mots employés dans le message.

Le rôle classificateur a pour objectif de trouver la classe d'appartenance d'un message en fonction de son contenu. Ces classes d'appartenance sont définies par les profils sociaux utilisés dans cette thèse : Animateur, Vérificateur, Quêteur et Indépendant. Il existe de nombreuses possibilités pour réaliser cet objectif de classification. Deux techniques différentes ont été abordées dans ce manuscrit et sont décrites dans la section suivante.

Le rôle indicateur est en charge du calcul des indicateurs sur la base des messages prétraités et dont la catégorie a été déterminée. Ces indicateurs peuvent être de natures multiples. Nous allons pour la suite nous baser sur ceux définis dans le chapitre précédent.

Enfin, le rôle expert à l'aide d'un mécanisme adéquat prend une décision sur la nature des apprenants et des échanges.

5.3 Les agents du système

En premier, un module de récupération permet l'extraction des messages d'interaction de la base de données LMS, et le préparer aux différents traitements ultérieurs.

5.3.1 Agent Récupération

En premier, un module de récupération permet l'extraction des messages d'interaction de la base de données LMS de Moodle, et leur préparation aux différents traitements ultérieurs.

D'après notre expérimentation, un premier traitement du corpus réside dans la correction d'orthographe. Pour le filtrage et lemmatisation, nous avons entrainé un outil d'étiquetage morphosyntaxique sur des corpus de messages.

Correction d'orthographe

L'orthographe française est complexe : la marque du pluriel est abstraite, redondante, inaudible ; le principe alphabétique est irrégulier (polyvalence de la correspondance phonie et graphème), etc.

A l'aide du clavier et de la souris, un apprenant saisit une première phrase pour enrichir la conversation. Celle-ci comporte plusieurs fautes d'orthographe. La correction automatique des fautes d'orthographe est l'un des axes les plus importants dans le domaine du traitement

automatique des langues naturelles. La recherche dans ce domaine a débuté dans les années 60 [Kukich, 1992].

La correction automatique des erreurs orthographique consiste à trouver le mot le plus proche des mots d'un lexique d'une langue donnée, en se basant sur la similarité et la distance inter mots.

- Les premières études ont été consacrées à déterminer les différents types d'erreurs orthographiques élémentaires, appelées opérations d'éditions [Damerau, 1964] notamment : insertion (ajout d'un caractère), suppression (omission d'un caractère), permutation (changement de position entre deux caractères), remplacement (remplacer un caractère par un autre).
- Damerau, Levenshtein [Levenshtein, 1966] n'a considéré que trois opérations d'éditions (insertion, suppression, permutation) et il a définit une distance qui compare deux mots en calculant le nombre d'opérations d'édition qui transforme le mot erroné au mot juste. Cette distance est appelée aussi distance de Damerau-Levenshtein.
- K.Oflazer a proposé une nouvelle approche appelée « Reconnaissance tolérante des erreurs » basée sur l'utilisation d'un dictionnaire représenté sous forme d'automate à états finis. Selon cette approche, la correction d'un mot erroné consiste à parcourir un automate à états finis en calculant pour chaque transition une distance appelée cut-off edit distance [Oflazer, 1996].
- Oflazer, A.Savary [Savary, 2003] a proposé une variante de cette méthode avec des modifications majeures, du fait qu'elle écarte le calcul du cut-off edit distance. Le principe étant d'explorer un dictionnaire-automate en largeur sans dépasser un seuil d'erreurs, et dans le cas de blocage, l'algorithme tolère un retour en arrière (back tracking) pour un parcours en profondeur en se basant sur des suppositions liées aux opérations d'éditions (insertion, inversion, omission, et remplacement).
- Pollock et Zamora [Pollock et Zamora, 1984] ont défini une autre manière pour modéliser les erreurs orthographiques et ceci par le calcul de ce qu'on appelle l'alpha-code (skeletton Key), d'où la nécessité d'avoir deux dictionnaires : un dictionnaire des mots et un pour les alpha-codes. Ainsi pour corriger un mot erroné, on extrait tout d'abord son alpha-code (abréviation des consonnes constituant le mot erroné) et on cherche les alpha-codes les plus proches. Cette méthode est efficace pour le cas des erreurs de permutation.

Malgré la disponibilité d'un ensemble de méthodes pour la correction orthographique, on constate qu'on ne dispose pas encore de correcticiels robustes capables de traiter de façon pertinente la totalité des erreurs commises dans un texte écrit, voir aussi un mauvais ordonnancement des solutions suggérées lors de la correction. Une analyse critique faite par A. Souque, [Souque, 2000] et Mitton [Mitton, 2009], confirme que les éditeurs de traitement

de texte commerciaux tels WinWord et OpenOffice, présentent des limitations dans la suggestion des solutions de quelques types de mots erronés.

Les fautes d'orthographe dans les textes des corpus posent parfois des problèmes d'analyse. Un mot où l'accent est mal mis peut changer complètement l'analyse. Prenons la phrase « Cet article me parait inacceptable car extrêmement confus et peu compréhensible, en tout cas par quelqu'un ne disposant que des informations contenues dans l'article. » Ici, il manque l'accent circonflexe sur le i de parait, le verbe. L'analyse retourne l'infinitif parer au lieu de paraître. La relation d'objet prédicatif entre me et paraître ne peut pas être extraite. C'est une condition indispensable pour l'extraction de la relation de sentiment entre article et inacceptable. Heureusement, la grammaire permet d'intercepter certains des cas les plus fréquents, par exemple par l'ajout de mots mal orthographiés au lexique (intéressant, sans accent et avec deux r, est apparu plusieurs fois dans les corpus.).

Les règles syntaxiques sont souvent assez souples pour faire l'accord entre un nom et un adjectif, ou un nom et un verbe même si le e ou le s manque. Mais malheureusement, il y a aussi des textes tellement mal écrits dans les corpus que l'analyse peut échouer. D'après ce que nous avons pu observer ces textes sont heureusement en minorité dans les corpus et ne modifient pas trop les résultats.

5.3.2 Agent Filtrage

Une fois la tâche de récupération réalisée, un second processus se charge d'effectuer le traitement de pré filtrage. On supprime automatiquement des termes qui n'apportent aucune information, en utilisant des « stop-words » propres à chaque langue.

Les mots qui apparaissent le plus souvent dans un corpus sont généralement les mots grammaticaux, mots vides (empty words) ou mots outils (stop words) : les articles, les prépositions, les mots de liaisons, les déterminants, les adverbes, les adjectifs indéfinis, les conjonctions, les pronoms et les verbes auxiliaires etc.., qui constituent une grande part des mots d'un texte, mais malheureusement sont faiblement informatifs, sur le sens d'un texte puisqu'ils sont présents sur l'ensemble des textes.

A titre d'exemple on peut citer en dans la langue Française, le cas des articles « le », « la », « les » ou de certains mots de liaison « ainsi », « toutefois » etc. L'élimination systématique du corpus des mots vides peut se faire par l'intermédiaire d'une liste prédéfinie de mots pour chacune des langues étudiées. Cependant, l'établissement d'une telle liste peut poser des problèmes. D'une part, il n'est pas facile de déterminer le nombre de mots exacts

qu'il faut inclure dans cette liste. D'autre part, cette liste est intimement liée à la langue utilisée et n'est donc pas transposable directement à une autre langue.

Par exemple Sahami.M dans sa thèse de Phd [Sahami, 1999] définit une liste de 570 mots courant en anglais, plus une liste de 100 mots très fréquents sur le web (tableau 5.1). Comme on peut les écarter en fixant un seuil maximal de fréquence, ce seuil est défini par l'expert de domaine, pour ne pas sélectionner les mots présents dans une grande partie du corpus.

Exemple : Liste des mots vides
alors, au, aucuns, aussi, autre, avantavec, avoir, bon, car, ce, cela, ces, ceux, chaque, ci, comme, comment, dans, des, du, dedans, dehors, depuis, deux, devrait, doit, donc, dos, droite, début, elle, elles, en, encore, essai, est, et, eu, fait, faites, fois, font, force, haut, hors, ici, il, ils, je, justela, le, les, leur, là, ma, maintenant, mais, mes, mine, moins, mon, mot, même, ni, nommés, notre, nous, nouveaux, ou, où, par, parce, parole, pas, personnes, peut, eu, pièce, plupart, pour, pourquoi, quand, que, quel, quelle, quelles, quels, qui, sa, sans, ses, seulement, si, sien, son, sont, sous, soyez, sujet, sur, ta, tandis, tellement, tels, tes, ton, tous, tout, trop, très, tu, valeur, voie, voient, vont, votre, vous, vu, ça, étaient, état, étions, été, être.

Tableau 5.1 : Exemple d'une liste des mots vides

Ces termes très fréquents peuvent être écartés du corpus pour en réduire la dimension. Cette possibilité de réduire la taille des entrées de l'index en éliminant les mots vides s'explique par le fait que ces termes sont présents dans la quasitotalité des messages et ont donc un pouvoir discriminant faible en comparaison avec d'autres termes.

D'après la loi de Zipf [W. Li, 1992]. Leur élimination lors d'un prétraitement du document permet par la suite de gagner beaucoup de temps lors de la modélisation et l'analyse des messages.

Ces mots doivent être supprimés de la représentation des textes pour deux raisons :

- D'un point de vue linguistique, ces mots ne comportent que très peu d'informations. La présence ou l'absence de ces mots n'aident pas à deviner le sens d'un texte. Pour cette raison, ils sont communément appelés « mots vides ».
- D'un point de vue statistique, ces mots se retrouvent sur l'ensemble des textes sans aucune discrimination et ne sont d'aucune aide pour la classification.

Une autre manière d'éliminer les mots qui n'apportent aucune information d'un texte passe par l'utilisation d'un étiqueteur syntaxique (Part of Speech Tagger) – Les mots sont écartés en fonction de leur étiquette syntaxique sans avoir besoin de liste prédéfinie.

- **Suppression des mots rares**

En général, les auteurs cherchent également à supprimer les mots rares (nom des personnes, des logiciels, des horaires, etc), qui n'apparaissent qu'une ou deux fois sur un corpus, afin de réduire de façon appréciable la dimension des vecteurs utilisés pour représenter les textes, puisque, d'après la loi de Zipf, ces mots rares sont très nombreux.

Nous nous intéressons à un outil qui s'appelle TreeTagger pour l'annotation du texte avec une partie du discours et de l'information lemme. Cet outil a été développé par Helmut Schmid dans le projet de TC à l'Institute for Computational Linguistics de l'Université de Stuttgart, Il a été utilisé avec succès pour marquer français, anglais, espagnol, bulgare, russe, portugais et les anciens textes français et est adaptable à d'autres langues. Le TreeTagger a été utilisée avec succès pour l'étiquetage morpho-syntaxique qui consiste à associer une étiquette morpho-syntaxique à chaque mot, il repose sur la segmentation en mots et en phrases effectuée préalablement.

Exemple d'étiquetage morpho-syntaxique (tableau 5.2).

TreeTagger – Les étiquettes pour le français			
Etiquettes	**Descriptif**	**Etiquettes**	**Descriptif**
ABR	Abreviation	NAM	Nom Propre
ADJ	Adjectif	NOM	Nom
ADV	Adverbe	NUM	Numéral
DET:ART	Article	PRO	Pronom
DET:POS	Pronom Possessif (ma, ta, ...)	PRO:DEM	Pronom Démonstratif
INT	Interjection	PRO:IND	Pronom Indefini
KON	Conjunction	PRO:PER	Pronom Personnel
PRO:POS	Pronom Possessif (mien, tien, ...)	PRP	Préposition
PRO:REL	Pronom Relatif	PUN	Ponctuation
PRP:det	Préposition + Article (au,du,aux,des)	PUN:cit	Ponctuation de citation
SENT	Balise de phrase	VER:infi	Verbe à infinitif
SYM	Symbole	VER:pper	Verbe au participe passé
VER:cond	Verbe au conditionnel	VER:ppre	Verbe au participe présent
VER:futu	Verbe au futur	VER:pres	Verbe au présent
VER:impe	Verbe à l'impératif	VER:simp	Verbe au passé simple
VER:impf	Verbe à l'imparfait	VER:subi	Verbe à l'imparfait du subjunctif
VER:subp	Verbe au présent du subjunctif		

Tableau 5.2 : Les étiquettes pour la langue française

Après cet étiquetage des corpus nous avons supprimé les mots rares notamment les noms des logiciels, noms des personnes, les horaires et les noms des jours, etc.

Voici un exemple d'étiquetage d'une phrase « TreeTagger permet d'annoter plusieurs langues. » :

```
TreeTagger    NAM       <unknown>
permet        VER:pres  permettre
```

d'	PRP	de
annoter	VER:infi	annoter
plusieurs	PRO:IND	plusieurs
langues	NOM	langue
.	SENT	.

avec

```
ABR : abréviation
ADJ : adjectif
VER : verbe
DET:ART :article
...
```

D'un point de vue linguistique, la suppression de ces mots n'est pas nécessairement justifiée : certains mots peuvent être très rares, mais très informatifs. Néanmoins, ces mots ne peuvent pas être utilisés par des méthodes à bases d'apprentissage du fait de leur très faible fréquence ; il n'est pas possible de construire des statistiques fiables à partir d'une ou deux occurrences ; Une des méthodes communément retenues pour supprimer ces mots consiste à ne considérer que les mots dont la fréquence totale est supérieure à un seuil fixé préalablement.

Notons enfin, que les mots ne contenant qu'une seule lettre sont généralement écartés pour les mêmes raisons précédentes, comme par exemple le mot « D » dans la « Vitamine D » ou le mot « C » dans le « langage C ».

5.3.3 Agent Lemmatisation

L'analyse linguistique est l'une des étapes clé dans le processus du traitement automatique du langage. Il s'agit d'extraire à partir du texte les éléments marquants tels que les termes candidats et les relations lexicales.

Après l'étape de filtrage, le message subit une analyse linguistique. Il s'agit d'une normalisation des mots, technique qui permet de ramener un mot à sa forme canonique (lemmatisation). Ce traitement peut entrainer une réduction importante du lexique.

Les règles de lemmatisation sont appliquées sur les différents mots du message pour réduire les variantes morphologiques à une forme commune (mettre les verbes à l'infinitif, supprimer les formes plurielles, etc).

La technique de lemmatisation consiste à effectuer un traitement au niveau de chacun des mots en fonction de leurs variations morphologiques : flexion, dérivation, composition afin de rassembler les mots de sens identiques. Donc, le but est de regrouper par exemple les termes « manger » et « mangent » ou les termes « cheval » et « chevaux » car ils ont la même

signification. L'intérêt de cette opération est la réduction des dimensionnalités de l'espace de codage des textes afin d'améliorer davantage la performance du système de classification en matière d'espace mémoire et vitesse de traitement.

Plusieurs traitements morphologiques existent :

- **Le stemming ou la désuffixation** regroupe sous un même terme (stem) les mots qui ont la même racine. L'extraction des stems se fait par la technique de racinisation (ou stemming) qui utilise à la place des dictionnaires, des algorithmes simples basées sur des règles de remplacement de chaînes de caractères pour supprimer les suffixes les plus utilisés.

  Le stemming est un traitement linguistique moins approfondie que la lemmatisation, ayant deux avantages : Plu rapide que la lemmatisation (algorithmes simples ne faisant pas référence aux dictionnaires et règles de dérivation) et la possibilité de traiter les mots inconnus sans traitement spécifique. [Clech, 2004] Néanmoins, sa précision et sa qualité sont naturellement inférieures, du fait qu'elle ne gère que les règles principales et ne peut pas prendre en compte les nombreuses exceptions des règles de dérivations. Par exemple, en français l'une des règles préconise de supprimer le « e » final de chaque mot, le mot « fraise » est alors transformé en « frais » ce qui suppose une relation entre les deux mots qui n'existe pas. Qui fait de cette opération dépendante de la langue, nécessitant une adaptation pour chaque langue utilisée.

  Plusieurs stemmers ont été développés pour déterminer les racines lexicales, l'algorithme le plus couramment utilisé pour la langue anglaise est celle de PORTER [Porter, 1980].

- **La lemmatisation** conserve, non pas les mots eux-mêmes, mais leur racine ou lemme. Ce principe permet de prendre en compte les variations flexionnelles (singulier/pluriel, conjugaisons, etc) ou dérivationnelles (substantifs, verbes, adjectifs, etc) en regroupant sous le même terme tous les mots de la même famille et donc d'améliorer la classification.

  La lemmatisation est donc une tâche plus compliquée à mettre en œuvre que la recherche de racines, puisqu'elle elle s'appuie sur des outils de Traitement Automatique de Langues Naturelles, ce qui nécessite beaucoup de ressources linguistiques (dictionnaires, règles de dérivation, etc..). De plus les résultats contiennent encore des erreurs à cause des problèmes de polysémie (ambiguïté) et d'incomplétude des dictionnaires.

Un algorithme efficace, nommé TreeTagger [Schmid, 1994] a été développé pour les langues anglaise, française, allemande et italienne. Cet algorithme utilise des arbres de décision pour effectuer l'analyse grammaticale, puis des fichiers de paramètres spécifiques à chaque langue. Toutes les études montrent que les performances des systèmes de classification, après lemmatisation, sont plus nettement supérieures à celles avant lemmatisation.

- **Etiquetage morpho-syntaxique**

L'étiquetage morphosyntaxique d'un texte consiste à identifier pour chaque mot sa classe morpho-syntaxique à savoir sa catégorie grammaticale, genre, nombre, temps, etc. (figure 5.2). Sur la toile existe une multitude d'outils d'étiquetage tels que Cordial, Brill et Tree Tagger, Multext, Genia Tagger, MBT et Tnt. Chacun de ces outils est fondée sur des hypothèses linguistiques distinctes.

Notre objectif n'étant pas de mener une étude comparative des différents outils d'étiquetage morphosyntaxique qui existent dans la littérature, ni d'évaluer qualitativement ou quantitativement les sorties étiquetées, le choix d'une solution d'entrainement parmi d'autres repose sur certain critère de sélection :

- Avoir un outil open-source simplement téléchargeable sur la toile.
- Permettre l'étiquetage des textes en français puisque le corpus que nous allons utiliser est écrit en langue française.
- Pouvoir effectuer la lemmatisation et être pourvu de découpeur (tokenizer).
- Disposer d'un fichier de sortie peu complexe et facile à exploiter.

En se basant sur ces critères, c'est la version française de tree tagger qui a été retenue. En effet, MBT et Tnt ne permettent pas l'étiquetage des textes français. La version de Brill n'est pas pourvue d'un découpeur et Cordial n'est pas un outil open source.

Tree tagger [Schmid, 1994] est un outil qui permet d'annoter un texte avec des informations sur les parties du discours (genre de mots : noms, verbes, infinitifs et particules) et des informations de lemmatisation. Il a été développé dans le cadre du projet « Textual Corpora and tools for their exploration TC » dans le « Institute for Computational Linguistics of the University of Stuttgart ICLUS ». TreeTagger permet l'étiquetage de l'Allemand, l'Anglais, le Français, l'Italien, l'Espagnol, le Bulgare, le Russe, le Grec, le Portugais, le chinois et les textes français anciens. Il est adaptable à d'autres langages si des lexiques et des corpus étiquetés manuellement sont disponibles.

L'étiquetage morpho-syntaxique d'un texte (Part-of-Speech tagging ou POS tagging en anglais), est souvent l'une des premières étapes en traitement automatique des langues

(TAL). Il consiste à identifier pour chaque mot sa classe morpho-syntaxique (catégorie grammaticale, genre, nombre, temps, etc) à partir de son contexte. La lemmatisation désigne l'analyse lexicale consistant à retrouver la forme canonique d'un mot fléchie appelée le lemme.

TreeTagger permet d'effectuer les opérations décrites ci-dessus et peut également être utilisé pour identifier des parties du discours et délimiter des groupes syntaxiques.

La lemmatisation consiste à associer un lemme à chaque mot du texte. Si le mot ne peut pas être lemmatisé (nombre, mot étranger, mot inconnu), aucune information ne lui est associée. La lemmatisation suppose que l'analyse morpho-syntaxique a déjà été faite (figure 5.2).

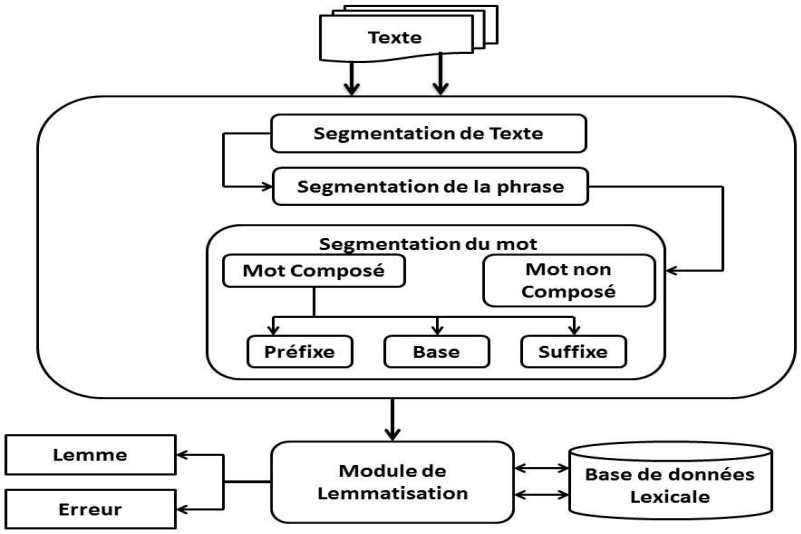

Figure 5-2 : Architecture du Lemmatiseur

Voici un exemple de lemmatisation d'une phrase :

```
TreeTagger        NAM        <unknown>
permet            VER:pres   permettre
d'                PRP        de
annoter           VER:infi   annoter
plusieurs         PRO:IND    plusieurs
langues           NOM        langue
.                 SENT
```

Par exemple, le mot « **jouer** », verbe à l'infinitif ni accordé, ni conjugué est un lemme. Il possède différentes flexions qui correspondent à ses formes conjuguées à diverses personnes et temps : « il **jouera** », « nous **jouons** », « ils ont **joué** », etc.

Un autre exemple, avec un nom cette fois. Le mot « **garage** » correspond à la forme non accordée et donc au lemme des flexions accordées « un **garage** » et « des **garages** ».

Nous avons cependant aussi étudié l'effet de ne pas lemmatiser les corpus. La lemmatisation (même si elle est plus coûteuse en temps) et non la simple radicalisation s'avère plus adaptée pour le français, qui est une langue latine à fort taux de flexion.

5.3.4 Agent Classification de messages

Plusieurs travaux de recherche ont porté sur l'analyse textuelle basés sur la théorie des actes de langage [George, 2004 ; Chaabi, 2014]. Si vous essayez d'analyser automatiquement une conversation, il est alors important d'identifier les actes de langage. Cette technique présente des inconvénients, comme le manque de flexibilité ou de contraintes ajout et de la discipline pour l'utilisateur [Suchman, 2015], en l'absence d'une normalisation des actes de langage pour déterminer les comportements sociaux des apprenants, nous poursuivons un double objectif, d'une part, en proposant une classification des conversations textuelles par un algorithme d'apprentissage supervisé "Naïve Bayes". Et d'autre part, nous avons effectué une comparaissant entre deux méthodes de classifications (par actes de langage et Naïve Bayes).

5.3.4.1 Classification par actes de langage

Nous nous intéressons aux conversations textuelles asynchrones, une conversation peut être considérée comme une succession de tours de parole c'est-à-dire une suite d'échanges composés d'interventions de locuteurs et constituées elles-mêmes d'actes de langage dont l'un est l'acte directeur.

Une intervention alors d'un apprenant dans un groupe de travail lors d'une conversation est constituée de plusieurs actes de langage. Si l'on cherche à analyser automatiquement une conversation, il est alors important de repérer ces actes de langage directeurs. Qui permettent de faire la différence entre un message textuel d'un apprenant à un autre.

Les actes de langage que nous proposons, se répartissent en quatre catégories selon les quatre profils associés aux travaux de Pléty [pléty, 1998].

- Les actes de langages de catégorie Animateur.
- Les actes de langages de catégorie Vérificateur.
- Les actes de langages de catégorie Quêteur.
- Les actes de langages de catégorie Indépendant.

Pour commencer, nous souhaitons préciser le contexte d'extraction des actes de langage. Nous avons travaillé régulièrement avec un certain nombre de tuteurs sur des échanges entre apprenants, nous en sommes venus à aborder un ensemble des actes de langage spécifique à chaque profil.

En se basant sur les propositions des tuteurs et sur la littérature relative au sujet [George, 2004]. L'analyse intuitive des messages montre que ces derniers peuvent être classés comme suit : des messages qui ont comme but d'initier une interaction et d'amorcer un sujet de discussion proposition, des messages où on demande des informations et où on attend une réponse de la pat d'autrui, des messages où on répond aux sollicitations des autres, où on répond aux questions et aux interrogations des autres et enfin des messages antérieurs qui clarifient ou approfondissent un sujet actuel de discussion.

Nous choisissons de présenter les actes de langage sous forme de verbes. En effet, un verbe est plus proche de l'action de l'utilisateur qui veut exprimer quelque chose [Scapin, 1986]. Notre choix s'oriente vers un ensemble de verbes qui expriment des forces illocutoires de chaque profil d'apprenant (tableau 5.3).

Catégories	Actes de Langage
Animateur	Proposer, demander, affirmer, soumettre, présenter, conseiller, démontrer, calmer, etc.
Vérificateur	Approuver, agréer, adopter, consentir, valider, supporter, rejeter, exclure, renvoyer, comparer, confirmer, justifier, contrôler, essayer, etc.
Quêteur	Comment, pourquoi, quand, quoi, quel, est-ce-que, qu'est-ce que, ? , etc.
Indépendant	Soumettre, présenter, offrir, conseiller, etc.

Tableau 5.3 : Exemple des actes de langage proposés.

Quatre expressions permettent de calculer la présence des actes de langage dans le message pour identifier la catégorie de profil : présence d'actes de langage animateur, vérificateur, quêteur et indépendant. La présence des actes de langage dans le message lancé par un apprenant est calculée comme suit (équation 9) :

$$PActeAnimateur = \frac{NbrActesAnimateur}{NbrTotaleActes(A,V,Q,I)} * 100 \quad (9)$$

$$PActeVérificateur = \frac{NbrActesVérificateur}{NbrTotaleActes(A,V,Q,I)} * 100$$

$$PActeQuêteur = \frac{NbrActesQuêteur}{NbrTotaleActes(A,V,Q,I)} * 100$$

$$PActeIndépendant = \frac{NbrActesIndépendant}{NbrTotaleActes(A,V,Q,I)} * 100$$

5.3.4.2 Classification Bayésienne

« *Il ne s'agit pas d'atteindre une résultat vrai ou faux, probable ou improbable, mais seulement profitable ou non profitable* ». [Lance et Willams, 1967].

La classification bayésienne naïve de textes est une approche probabiliste de classification simple. Cette approche est basée sur un modèle probabiliste dérivant du théorème de bayes. On obtient toutefois des résultats satisfaisants dès lors que le classificateur est bien entrainé.

Dans le domaine de la catégorisation de textes, il existe deux types de structures produits par les algorithmes de classification : celles qui sont hiérarchiques (hierarchical clustering) et celles qui sont non hiérarchiques (non hierarchical clustering). Une classification hiérarchique est souvent préférable lorsqu'on désire effectuer une analyse détaillée des données tandis que les méthodes non hiérarchiques sont plus rapides et choisies lorsque le nombre de données est grand.

Les méthodes de classification hiérarchiques correspondent à un hard clustering, c'est-à-dire qu'elles n'acceptent qu'une possibilité d'appartenance à une classe contrairement aux méthodes non hiérarchiques correspondant au soft clustering qui acceptent plusieurs possibilités.

Une distinction importante existe entre les algorithmes dits de soft clustering et ceux dits de hard clustering. Si les seconds n'autorisent, que l'appartenance à une seule classe (cluster), les premiers font correspondre pour chaque élément, leurs degrés d'appartenance à chacune des catégories.

Le soft clustering apparaît donc comme un modèle probabiliste pour lequel ce degré d'appartenance est la probabilité qu'un élément appartienne à une classe ou une catégorie. La méthode de classification bayésienne que nous exposons dans ce rapport est une méthode de type hard clustering et dont les principes sont présentés dans la section suivante (figure 5.3).

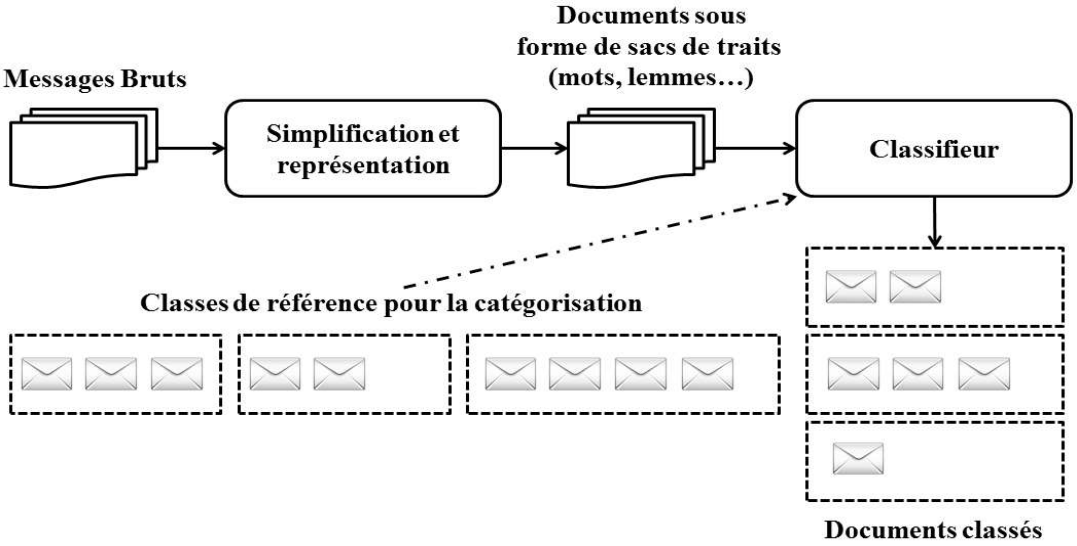

Figure 5-3 : Les étapes de classification

- **Description du modèle Bayésien**

De manière abstraite, le modèle probabiliste pour un classificateur bayésien est un modèle conditionnel. Il se base sur la règle de bayes qui s'énonce de la manière suivante (équation 10) :

$$P(A \mid B_1,..., B_n) = \frac{P(B_1,..., B_n \mid A) * P(A)}{P(B_1,..., B_n)} \tag{10}$$

La probabilité d'avoir l'évènement A étant donné $B_1,.... B_n$ est donné par le rapport entre la probabilité d'avoir les évènements $B_1,.... B_n$ étant donné A et la probabilité que $B_1,.... B_n$ Se soient produits. Tant que le dénominateur ne dépend pas de l'évènement A, on peut Considérer la probabilité P$(B_1,.... B_n)$ comme étant constante. Le numérateur peut être écrit Encore de la manière suivante (équation 11) :

$$\begin{aligned} & P(B_1,..., B_n \mid A) * P(A) \\ & = P(A.B_1,..., B_n) \\ & = P(A) * P(B_1 \mid A) * P(B_2,....., B_n \mid A, B_1) \\ & = P(A) * P(B_1 \mid A) * P(B_2 \mid A, B_1) \\ & = P(B_3,....., B_n \mid A, B_1, B_2) \end{aligned} \tag{11}$$

La décomposition de P(A, $B_1,.... B_n$) se termine lorsqu'on a parcouru l'ensemble des classes $B_1,.... B_n$ Le caractère "naïf" de ce théorème vient du fait qu'on suppose

l'indépendance des différentes classes B_i,.... B_j .Ce qui en d'autres termes se traduit par (équation 12) :

$$P(B_i \mid A, B_j) = P(B_i \mid A) \tag{12}$$

Cette hypothèse permet également d'écrire (équation 13) :

$$\begin{aligned} &P(A.B_1,..., B_n) = P(A) * P(B_1 \mid A) * P(B_2 \mid A) \\ &* ..... * P(B_n \mid A) \\ &= P(A) \prod_{i=1}^{n} P(B_i \mid A) \end{aligned} \tag{13}$$

Ce théorème a beaucoup d'applications dans le domaine du traitement de l'information notamment en traitement de la parole, traitement des images et bien d'autre. Nous nous limiterons ici à l'application de ce théorème à la catégorisation de messages pour la détermination des comportements sociaux d'apprenant.

- **Application à la classification de messages.**

Illustrons les équations du théorème vu à la section précédente au problème de la classification de textes Supposons que nous disposons de **n** catégories de Messages, déterminer à quelle catégorie **Ci** sera associée un Message **M** revient à calculer la probabilité d'appartenance du Message **M** à la catégorie **Ci**. En se basant sur le théorème précédemment énoncé. On peut calculer cette probabilité de la façon suivante (équation 14) :

$$P(C_i \mid M) = \frac{P(M \mid C_i) * P(C_i)}{P(M)} \tag{14}$$

Dans cette formule, $P(C_i \mid M)$ représente la probabilité d'appartenance du message M à la catégorie C_i qui peut être également déterminée en évaluant la fréquence d'apparition des mots du message M qui sont associés à la catégorie C_i.

$P(M \mid C_i)$ est la probabilité selon laquelle, pour une catégorie donnée, les mots du message M sont associés à la catégorie C_i. $P(C_i)$ est la probabilité qui associe le message M à la catégorie C_i indépendamment du contenu du message. $P(M)$ est la probabilité propre du message M.

Pour déterminer réellement à quelle catégorie un message appartient, il faut calculer $P(C_i \mid M)$ pour chacune des catégories. Etant donné que $P(M)$ reste constant pour toutes les catégories, déterminer $P(C_i \mid M)$ se résume juste au calcul de $P(M \mid C_i) * P(C_i)$.

- **Comment calcule-ton ces probabilités ?**

En considérant que le message M est composé d'un ensemble de mots que nous noterons W1,...,Wm, calculer $P(M \mid C_i)$ reviendrait à calculer le produit des probabilités d'apparition de chaque mot Wi dans la catégorie C_i. Ce calcul se justifie par l'hypothèse selon laquelle tous les Mots apparaissent indépendamment les uns des autres dans un message. Ce qui permet finalement d'écrire (équation 15)

$$P(M \mid C_i) = P(W_1, C_i) * P(W_2, C_i) * ..... * P(W_m, C_i) \tag{15}$$

Pour chacune des catégories, $P(W_i, C_i)$ est le rapport entre le nombre de fois que le mot W_i apparaît dans la catégorie C_i et le nombre total de mots que comprend la catégorie Ci. $P(C_i)$ est calculé en divisant le nombre total de mots pour la catégorie Ci par la somme du nombre total de mots dans toute les catégories. D'où la formulation suivante (équation 16):

$$P(C_i \mid M) = P(W_1, C_i) * ..... * P(W_m, C_i) * P(C_i) \tag{16}$$

Ce calcul est effectué pour chaque catégorie et on considère la probabilité la plus élevée pour choisir à quelle catégorie sera associée le message qu'on souhaite classier.

- **Description de modèle bayésien**

Classificateurs de texte n'utilisent pas souvent tout type de représentation profonde de la langue : souvent un message est représenté comme un sac de mots. (Un sac est comme un ensemble qui permet de répéter les éléments). C'est une représentation extrêmement simple : il ne sait quels mots sont inclus dans le message (et combien de fois chaque mot se produit), et jette l'ordre des mots !

Considérons un message M, dont la classe est donnée par C. Dans notre cas, il y a quatre classes

C= A(Animateur) etC=V(Vérificateur) et C = Q (Quêteur) et C = I (Indépendant). Nous classons M que la classe qui a la plus forte probabilité postérieure P (C | M), qui peut être ré-exprimé en utilisant le théorème de Bayes (équation 17) :

$$P(C \mid M) = \frac{P(M \mid C)P(C)}{P(M)} \alpha P(M \mid C)P(C) \tag{17}$$

Nous allons examiner un modèle probabiliste de message en utilisant l'hypothèse Naïve Bayes. Un modèle représente des messages en utilisant des vecteurs de caractéristiques dont les composantes correspondent à des types de mots. Si nous avons un vocabulaire V, contenant |V| type mots, alors le vecteur de caractéristique de dimension d = |V|.

Bernoulli model de message : Un message est représenté par un vecteur de caractéristique avec des éléments binaires prennent la valeur 1 si le mot correspondant est présent dans le message et 0 si le mot n'est pas présent.

Exemple : Considérez le vocabulaire suivant :

V= {**Recherche, Proposer, Effectué, assembler, Intéressante, confirmer**}.

Dans ce cas, |V| = D = 6. Considérons maintenant le message court "**La recherche que tu as effectué youness est très intéressante ça va nous aider à mieux comprendre".** Si d^B est le vecteur de caractéristiques Bernoulli pour ce message, alors nous aurions :

$$M^B=(1,0,1,0,1,0)^T$$

Pour classer un message nous utilisons l'équation (1), ce qui nécessite l'estimation des probabilités de message de la classe donné P(M|C) et les probabilités a priori de classe P(C). Pour estimer la probabilité P(M|C) nous utilisons l'hypothèse Naïve Bayes appliqué aux modèles des messages que nous utilisons.

- **Le modèle Bernoulli de messagel**

Comme mentionné ci-dessous, dans le modèle de Bernoulli un message est représenté par un vecteur binaire, ce qui représente un point dans l'espace des mots. Si nous avons un vocabulaire V contenant un ensemble de |V| mots, alors la tiéme dimension d'un vecteur de message correspond à des mots w_t dans le vocabulaire. Soit b_i le vecteur de caractéristiques

pour un message M^i, puis le tiéme élément de b_i, est soit 0 ou 1 représentant l'absence ou la présence de mot w_t dans le i iémé message.

Soit P(w_t | C) la probabilité du poids des mots survenant dans un message de la classe C; la probabilité de ne pas se produire dans un message de cette classe est donnée par (1 − P(w_t | C)). Si nous faisons l'hypothèse de Naïve Bayes que la probabilité de chaque mot apparaissant dans le message est indépendante des occurrences des autres mots, nous pouvons écrire la probabilité de message P(M^i | C) en termes de probabilités individuels des mots P(w_t|C) (équation 18):

$$P(M^i \mid C) \approx P(b_i \mid C) = \prod_{t=1}^{|V|} [b_{it} P(w_t \mid C) + (1 - b_{it})(1 - P(w_t \mid C))] \tag{18}$$

Ce produit va sur tous les mots dans le vocabulaire. Si le mot wt est présent, alors b_{it}=1 et la probabilité est nécessaire P(w_t|C) ; si le mot w_t n'est pas présent, alors b_{it} = 0 et la probabilité requis est 1-P(w_t|C).

Nous pouvons imaginer cela comme un modèle pour générer une fonction vecteur de message vecteurs de classe C, dans lequel la fonction vecteur de message est modélisée comme une collection de |v|, le t ayant une probabilité de succès égale à P(w_t|C). Les paramètres des vraisemblances sont les probabilités de chaque mot compte tenu de la classe de message P(w_t|C) ; le modèle est également paramétré par les probabilités a priori, P (C). Nous pouvons apprendre (estimation) ces paramètres à partir d'un ensemble de messages étiquetés avec la classe de formation C = k. Soit $n_k(w_t)$le nombre de messages de la classe C = k dans laquelle est observé ; et soit N_k le nombre total de messages de cette catégorie. Ensuite, nous pouvons estimer les paramètres de probabilités de mot comme (équation 19) :

$$P(w_t \mid C = k) = \frac{n_k(w_t)}{N_k} \tag{19}$$

La fréquence relative des messages de la classe C = k qui contiennent le mot w_t. S'il existe N messages au total dans l'ensemble d'apprentissage, alors la probabilité a priori de la classe C = k peut être estimée comme étant la fréquence relative des messages de classe C= k (équation 20) :

$$P(C = k) = \frac{N_k}{N} \tag{20}$$

Ainsi étant donné un ensemble des messages de formation (chacun marqué avec une classe), et un ensemble de classes K, nous pouvons estimer un modèle de classification de texte Bernoulli comme suit :

Définir le vocabulaire V ; le nombre de mots du vocabulaire, définit la dimension des vecteurs de caractéristiques.

1. Comptez le vocabulaire dans l'ensemble de la formation :
   - N Le nombre total demessages.
   - N_k Le nombre de messagesmarqués avecclasseC =k,pour k =1,. ., K.
   - $n_k(w_t)$ le nombre de messagesde la classe C=kcontenantle motpour chaqueclasse etpour chaque motdans le vocabulaire.
2. Estimer les probabilités P (w_t | C = k) en utilisant l'équation (3).
3. Estimer les probabilités a priori P (C = k) en utilisant l'équation (4).

Pour classer un message non marqué, nous estimons la probabilité a posteriori pour chaque classe en combinant les équations (17) et (18) (équation 21) :

$$\begin{aligned} &P(C \mid M^J) = P(C \mid b_j) \\ &\alpha P(b_j \mid C) P(C) \\ &\alpha P(C) \prod_{t=1}^{|V|} [b_{jt} P(w_t \mid C) + (1 - b_{jt})(1 - P(w_t \mid C))] \end{aligned} \tag{21}$$

La méthode générique du traitement que nous utilisons se compose de cinq étapes décrites ci-dessous.

- Étape 1 : prétraitement de la langue : recherche dans le corpus sur des unités linguistiques.
- Étape 2 : prétraitement linguistique et représentation mathématique du corpus des textes.
- Étape 3 : Sélectionnez les caractéristiques des unités linguistiques de corpus.
- Étape 4 : choix de la méthode de classification.
- Étape 5 : Performance de la classification en coupant tout apprentissage/test.

➢ Application à la classification de message

Considérons un ensemble de messages, dont chacun est lié soit à Animateur (A), vérificateur (V), quêteur (Q) ou indépendant (I). Étant donné un ensemble de 27 messages de formation, nous tenons à estimer un classificateur Naïve Bayes, en utilisant le modèle de message Bernoulli, de classer les messages non marqués comme A ou V.

On définit un vocabulaire de huit mots « vocabulaires » d'extraction est faite automatiquement par le système à partir des données de formation” :

$$V = \begin{bmatrix} w_1 = \text{proposer} \\ w_2 = \text{approuver} \\ w_3 = \text{valider} \\ w_4 = \text{demander} \\ w_5 = \text{affirmer} \\ w_6 = \text{conseiller} \\ w_7 = \text{contrôler} \\ w_8 = \text{présenter} \end{bmatrix}$$

Ainsi, chaque message est représenté par un vecteur binaire de huit dimensions.
Les données de formation sont présentées ci-dessous comme une matrice pour chaque classe, dans lequel chaque ligne de huit dimensions représente un vecteur de message.

$$B^{Animateur} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad B^{Vérificateur} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$B^{Quêteur} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad B^{Indépendant} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Classer le suivant dans Animateur ou vérificateur ou quêteur ou indépendant à l'aide d'un classificateur Naïve Bayes.

$$B_1 = (1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1)^T$$
$$B_2 = (0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0)^T$$

- **Solution :**

Le nombre total de messages dans l'ensemble de formation est N=27 ; NA=7, NV=6, NQ=8, NI=6. En utilisant l'équation (20), on peut estimer les probabilités a priori à partir des données de formation :

$$P(A)=\frac{7}{27} \qquad P(V)=\frac{6}{27} \qquad P(Q)=\frac{8}{27} \qquad P(I)=\frac{6}{27}$$

Le nombre de mots dans les données de formation sont :

$$\begin{array}{llllllll}
n_A(w_1)=3 & n_A(w_2)=1 & n_V(w_1)=1 & n_V(w_2)=3 & n_Q(w_1)=4 & n_Q(w_2)=4 & n_I(w_1)=2 & n_I(w_2)=3 \\
n_A(w_3)=2 & n_A(w_4)=3 & n_V(w_3)=3 & n_V(w_4)=1 & n_Q(w_3)=3 & n_Q(w_4)=1 & n_I(w_3)=3 & n_I(w_4)=4 \\
n_A(w_5)=3 & n_A(w_6)=4 & n_V(w_5)=1 & n_V(w_6)=1 & n_Q(w_5)=2 & n_Q(w_6)=3 & n_I(w_5)=4 & n_I(w_6)=1 \\
n_A(w_7)=4 & n_A(w_8)=4 & n_V(w_7)=3 & n_V(w_8)=1 & n_Q(w_7)=1 & n_Q(w_8)=4 & n_I(w_7)=2 & n_I(w_8)=2
\end{array}$$

On peut estimer les vraisemblances de mot en utilisant l'équation (19)

Pour la classe d'animateur (A) :

$$\begin{array}{ll}
P(w_1|A)=\frac{3}{7} & P(w_2|A)=\frac{1}{7} \\
P(w_3|A)=\frac{2}{7} & P(w_4|A)=\frac{3}{7} \\
P(w_5|A)=\frac{3}{7} & P(w_6|A)=\frac{4}{7} \\
P(w_7|A)=\frac{4}{7} & P(w_8|A)=\frac{4}{7}
\end{array}$$

Et pour la classe de vérificateur (V) :

$$\begin{array}{ll}
P(w_1|V)=\frac{1}{6} & P(w_2|V)=\frac{1}{2} \\
P(w_3|V)=\frac{1}{2} & P(w_4V)=\frac{1}{6} \\
P(w_5|V)=\frac{1}{6} & P(w_6|V)=\frac{1}{6} \\
P(w_7|V)=\frac{1}{2} & P(w_8|V)=\frac{1}{6}
\end{array}$$

Pour la classe Quêteur (Q) :

$$\begin{array}{ll}
P(w_1|Q)=\frac{1}{2} & P(w_2|Q)=\frac{1}{2} \\
P(w_3|Q)=\frac{3}{8} & P(w_4|Q)=\frac{1}{8} \\
P(w_5|Q)=\frac{1}{4} & P(w_6|Q)=\frac{3}{8} \\
P(w_7|Q)=\frac{1}{8} & P(w_8|Q)=\frac{1}{2}
\end{array}$$

Et pour la classe d'Indépendant (I) :

$$\begin{array}{ll}
P(w_1|I)=\frac{1}{3} & P(w_2|I)=\frac{1}{2} \\
P(w_3|I)=\frac{1}{2} & P(w_4|I)=\frac{2}{3} \\
P(w_5|I)=\frac{2}{3} & P(w_6|I)=\frac{1}{6} \\
P(w_7|I)=\frac{1}{3} & P(w_8|I)=\frac{1}{3}
\end{array}$$

Nous utilisons l'équation (5) pour calculer les probabilités postérieures des deux vecteurs de test et donc les classer. $b_1 = (1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1)^T$

$$P(A|b_1)\alpha P(A)\prod_{t=1}^{8}[b_{it}P(w_t|A) + (1-b_{it})(1-P(w_t|A))]$$
$$\alpha \frac{7}{27}(\frac{3}{7}\times\frac{5}{7}\times\frac{4}{7}\times\frac{4}{7}\times\frac{1}{7}\times\frac{3}{7}\times\frac{3}{7}\times\frac{4}{7}) = \frac{60480}{155649627} = 3.8\times10^{-4}$$

$$P(V|b_1)\alpha P(V)\prod_{t=1}^{8}[b_{it}P(w_t|V) + (1-b_{it})(1-P(w_t|V))]$$

$$\alpha \frac{6}{27}(\frac{1}{6}\times\frac{1}{2}\times\frac{5}{6}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{6}\times\frac{5}{6}\times\frac{1}{6}) = \frac{150}{1679619} = 8.9\times10^{-5}$$

$$P(Q|b_1)\alpha P(Q)\prod_{t=1}^{8}[b_{it}P(w_t|Q) + (1-b_{it})(1-P(w_t|Q))]$$
$$\alpha \frac{8}{27}(\frac{1}{2}\times\frac{5}{8}\times\frac{3}{4}\times\frac{1}{8}\times\frac{1}{2}\times\frac{1}{8}\times\frac{5}{8}\times\frac{1}{2}) = \frac{600}{3538944} = 1.6\times10^{-4}$$

$$P(I|b_1)\alpha P(I)\prod_{t=1}^{8}[b_{it}P(w_t I) + (1-b_{it})(1-P(w_t|I))]$$
$$\alpha \frac{6}{27}(\frac{1}{3}\times\frac{1}{2}\times\frac{1}{3}\times\frac{1}{3}\times\frac{1}{2}\times\frac{2}{3}\times\frac{5}{6}\times\frac{1}{3}) = \frac{60}{157464} = 3.8\times10^{-4}$$

Classer ce message en tant que *Quêteur*.

$$b_2 = (0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0)^T$$

$$P(A|b_2)\alpha P(A)\prod_{t=1}^{8}[b_{it}P(w_t|A) + (1-b_{it})(1-P(w_t|A))]$$
$$\alpha \frac{7}{27}(\frac{4}{7}\times\frac{2}{7}\times\frac{3}{7}\times\frac{3}{7}\times\frac{1}{7}\times\frac{4}{7}\times\frac{4}{7}\times\frac{3}{7}) = \frac{6048}{155649627} = 3.8\times10^{-5}$$

$$P(V|b_1)\alpha P(V)\prod_{t=1}^{8}[b_{it}P(w_t|V) + (1-b_{it})(1-P(w_t|V))]$$
$$\alpha \frac{6}{27}(\frac{5}{6}\times\frac{1}{2}\times\frac{1}{6}\times\frac{1}{2}\times\frac{1}{2}\times\frac{5}{6}\times\frac{1}{6}\times\frac{5}{6}) = \frac{750}{1679616} = 4.4\times10^{-4}$$

$$P(Q|b_2)\alpha P(Q)\prod_{t=1}^{8}[b_{it}P(w_t|Q) + (1-b_{it})(1-P(w_t|Q))]$$
$$\alpha \frac{8}{27}(\frac{1}{2}\times\frac{3}{8}\times\frac{1}{4}\times\frac{7}{8}\times\frac{1}{2}\times\frac{7}{8}\times\frac{3}{8}\times\frac{1}{2}) = \frac{3528}{3538944} = 9.9\times10^{-4}$$

$$P(I|b_1)\alpha P(I)\prod_{t=1}^{8}[b_{it}P(w_t|I) + (1-b_{it})(1-P(w_t|I))]$$
$$\alpha \frac{6}{27}(\frac{2}{3}\times\frac{1}{2}\times\frac{2}{3}\times\frac{2}{3}\times\frac{1}{2}\times\frac{1}{3}\times\frac{1}{6}\times\frac{2}{3}) = \frac{96}{157464} = 4.4\times10^{-3}$$

Classer ce message comme Indépendant.

5.3.5 Agent indicateur

L'analyse automatique des Interactions (AAI) dans le but de déterminer les comportements sociaux des apprenants est une direction de recherche qui a émergé ces dernières années, elle a comme but d'aider les acteurs (Les tuteurs, les apprenants et les chercheurs, etc). Le processus consiste à capter, filtrer et traiter les données afin de produire des indicateurs d'analyse des activités d'interactions dans le cadre d'un forum de discussion éducatif.

Plusieurs systèmes d'AAI ont été développés, qui s'adresse aux apprenants qui discutent via un forum de façon asynchrone et synchrone [Oumaira, 2011 ; George, 2006; George, 2004], Citons par exemple l'outil I-Bee, Analytic Tool, reliés à l'environnement de collaboration asynchrone 'Knowledge Forum', l'outil d'AAI 'DIAS' offre un grande nombre d'indicateurs, afin d'assister les étudiants participant à un forum.

Les indicateurs sociaux indiquent des informations sur les comportements sociaux par l'interaction entre les états mentaux et les situations sociales immédiates lors des activités d'apprentissage collaboratif. Ces indicateurs concernent les interactions des participants liées à la tâche et au contenu de l'activité. Ainsi, nous avons défini deux catégories d'indicateurs de nature quantitative et qualitative.

En se basant sur les travaux réalisés par Pléty [Pléty, 1998] on a calculé 4 indicateurs : le volume d'intervention, le type d'intervention et réaction entérinées (réaction directe et indirecte).

5.3.5.1 Volume d'interventions :

L'indicateur volume d'intervention a permis au tuteur de mieux apprécier les productions des apprenants dans les groupes. En effet, les résultats de l'évaluation des productions vont dans le sens des valeurs retournées par l'indicateur volume d'intervention : les groupes qui ont le mieux réussi leurs travaux sont ceux qui présentent un volume d'intervention les plus élevés. Résultat que nous interprétons comme étant le signe d'une bonne coordination dans le groupe et de niveau d'échange et de concertation apprenants [Serge, 2004], critères nécessaires à la réussite de tout travail collaboratif. Si dans l'ensemble l'indicateur « volume d'intervention » a permis de conforter les observations du tuteur à travers une comparaison entre les résultats retournés par cet indicateur et la qualité et cohérence de la production des apprenants dans le groupe.

Ce type d'indicateurs rend compte de la qualité de la collaboration au cours d'une séquence. L'indicateur « degré d'interactivité des apprenants dans le forum et l'email » qui calcule respectivement le nombre de messages échangés dans chaque groupe par période dans le forum et l'email. Cet indicateur permet de repérer les membres isolés, ainsi que ceux qui dominent les interactions [Martinez & al. 2003].

Une première formule permet de calculer tout d'abord le ratio de participation d'un apprenant en divisant le nombre de messages envoyés par celui-ci, par le nombre de messages envoyés par les apprenants du même groupe.

Nous avons donc l'expression suivante « Ratio d'intervention RI » d'un apprenant (P) (équation 22) :

$$RI = \frac{nbrMsgApprenant}{NbrTotaleMessagesGroupe} * 100 \tag{22}$$

5.3.5.2 Type d'interventions

Quatre expressions permettent de calculer le type d'intervention de chaque message pour identifier la catégorie de profil : animateur, vérificateur, quêteur et indépendant. Le type d'intervention est calculé comme suit (équations 23) :

$$\begin{aligned} TIAnimateur &= \frac{MessagesAnimateur}{TotaleMessages\ (A,V,Q,I)} * 100 \\ TIVérificateur &= \frac{MessagesVérificateur}{TotaleMessages\ \ (A,V,Q,I)} * 100 \\ TIQuêteur &= \frac{MessagesQuéteur}{TotaleMessages\ \ (A,V,Q,I)} * 100 \\ TIIndépendant &= \frac{MessagesIndépendant}{TotaleMessages\ \ (A,V,Q,I)} * 100 \end{aligned} \tag{23}$$

Dans cette formule "$\boldsymbol{TIV}$é$\boldsymbol{rificateur}$", "MessagesAnimator" est le nombre de messages de type animateur (exemple proposer et encourager, etc.) envoyé par un apprenant divisé par "TotalMessages" est le nombre total de messages envoyés par cet apprenant. Les calculs d'autres pourcentages (Animateur, vérificateur, quêteur et indépendants) sont obtenus de manière similaire en prenant les informations fournies par Pléty et de les transformer en équations.

5.3.5.3 Réactions Entrainées

Selon les caractéristiques des profils définis (Tableau 4.1), le volume de réactions entrainées par un message permet de caractériser un profil comportemental. A titre d'exemple, un profil animateur exige un suivi de réactions très important par rapport à celui d'un vérificateur.

Un forum de discussion (en anglais « bulletin board») est un espace web dynamique permettant à différentes personnes de communiquer. Le forum de discussion est généralement composé de différents fils de discussion (le terme « fil de discussion » est parfois remplacé par sujet de discussion, post, thread, enfilade ou topic) correspondant chacun à un échange sur un sujet particulier. Le premier message d'un thread définit la discussion, et les messages suivants (situés généralement en dessous) tentent d'y répondre. Nous allons présenter l'arborescence de discussion sous forme d'arbres n-aire afin de faciliter le calcul des indicateurs (figure 5.4).

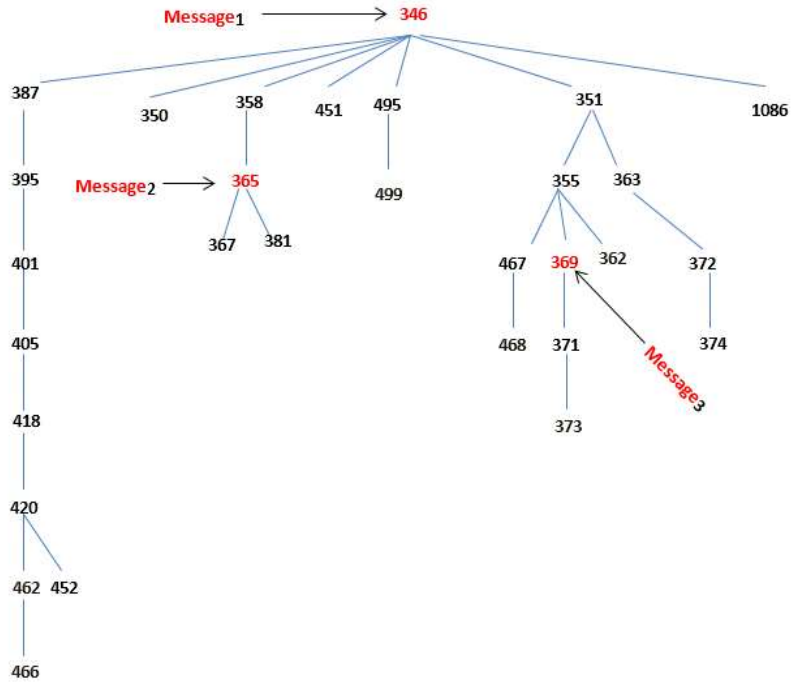

Figure 5-4 : Arborescence de discussion correspondent à un fil de discussion

En effet, les nœuds représentent les identifiants des messages envoyés par les apprenants et la taille de cet arbre est équivalente au nombre des interventions faites après le lancement du sujet de discussion. Cet indicateur permet d'avoir une traçabilité de la progression d'un fil de discussion dans un forum au cours du temps. Les résultats rapportés par cet indicateur montrent une certaine richesse dans les sujets de discussion (nombre d'arbres distincts) mais par contre une faible réactivité à ces sujets (longueur des fils de discussion).

La conversation est vue comme une co-construction dans laquelle les interlocuteurs interagissent de manière cohérente. D'une manière générale, nous retenons que « chaque interaction répond à une intervention précédente et en même temps impose des contraintes discursives sur l'interaction suivante. Dans une perspective d'assistance à la structuration de conversation, nous pensons qu'il est utile de prendre en compte ces contraintes discursives qui permettent de restreindre pour chaque énoncé l'ensemble des énoncés pouvant intervenir à la suite. D'une manière plus précise les sujets qui ont un grand nombre de réponses et interactions sont des sujets d'intérêt pour le groupe. Nous calculons alors pour chaque message les réactions directes (1ère réaction à un message) et les réactions indirectes (nombre des interventions faites après le lancement du message).

Conformément au modèle de profils sociaux adopté (tableau 4.1) ; le profil social est défini à partir de volume de réaction entrainées. A titre d'exemple, un profil animateur exige un suivi de réactions très important par rapport à celui d'un vérificateur. D'une manière plus précise les sujets qui ont un grand nombre de réponses et interactions sont des sujets d'intérêt pour le groupe. Nous calculons alors pour chaque message les réactions directes (1ère réaction à un message) et les réactions indirectes (nombre des interventions faites après le lancement du message).

- **Réaction directe (1ére réaction à un message)**

Nous appelons réaction directe la réaction A→B et réaction inverse la réaction B→A tout en remarquant que ce choix est arbitraire et que les termes de réactifs et de produits sont conventionnels et sont fixés par le choix des concentrations initiales en A et B.

$$Reaction_{Directe} = \frac{\sum_{i=1}^{n} reponse\ sur\ Message_i\ (APP)}{NbrReponseTotale\ (Groupe\ X)} * 100 \qquad (24)$$

On calcule l'indicateur de réaction directe (équation 24) comme le rapport entre le nombre de réponses directes sur le message de l'apprenant (reponse sur message) sur le nombre total de réponses directes sur les messages envoyés par les apprenants dans le groupe (NbrReponseTotale).

- **Réaction indirecte (Nombre des interventions faites après le lancement du message)**

$$Reaction_{Directe} = \frac{Profondeur - Reponse}{TotaleProfondeur} * 100 \qquad (25)$$

Le deuxième indicateur est calculé comme un rapport de profondeur de discussion moins le nombre de réactions directes, divisé par la somme des profondeurs de discussion.

5.3.6 Agent Logique floue

Le souci d'automatiser ou d'assister de façon automatique les actions humaines, naturellement empiriques et empreintes d'imprécisions, dans le cadre d'une aide à la décision ou du contrôle par exemple, renforce l'intérêt des scientifiques pour l'approche floue et justifie son intense développement au cours de ces dernières années. C'est pour les possibilités qu'elle offre de gérer l'incertitude et l'imprécision, que nous nous sommes intéressés à la logique floue. Dans un système d'autoformation à distance ou il n'y a aucune interaction entre le tuteur et l'apprenant, les données rassemblées tendent à être plus imparfaites que celles obtenues par l'interaction face à face traditionnelle. En outre, il est plus difficile de compter sur ce type de systèmes en tout ce qui concerne collecte de l'information de base et 'expérience appropriées', comme les tuteurs humains.

Maintenant que nous avons vu comment le système calcule les indicateurs, l'analyse de ces résultats à la lumière des caractéristiques des profils d'apprenant définis (Tableau 4.1), permet au tuteur d'associer un profil sociologique à chaque apprenant. Elle est formative lorsque le but de l'analyse est d'associer automatiquement un comportement social à chaque apprenant. Pour cela nous mettrons en œuvre une approche fondée sur la logique floue contribuerait à aborder des problèmes de manipulation de connaissance imparfaites qui ne pouvaient pas être posés en théorie des probabilités. En conséquence, l'adaptation dynamique sera déclenchée pour pouvoir associer automatiquement un comportement social à chaque apprenant. La parties floues (ou sous-ensembles flous) ont été introduites afin de modéliser la représentation humaine des connaissances, et ainsi améliorer les performances des systèmes qui utilisent cette modélisation.

La plupart des problèmes rencontrés sont modélisable mathématiquement. Mais ces modèles nécessitent des hypothèses parfois trop restrictives rendant délicate leur application au monde réel. Les problèmes du monde réel doivent tenir compte d'informations imprécises, incertaines. Les connaissances dont disposent les humains sur le monde ne sont presque jamais parfaites. Les connaissances sur lesquelles se base le raisonnement humain sont presque toujours entachées d'une quantité d'incertitudes et d'imprécisions. Nous ne parlons pas ici du raisonnement scientifique, dont l'objectif est justement de se débarrasser de toute imperfection, mais de tous les autres raisonnements que nous faisons tous les jours, sans cesse, sur les choses, les personnes et les pensées nous environnant (figure 5.5).

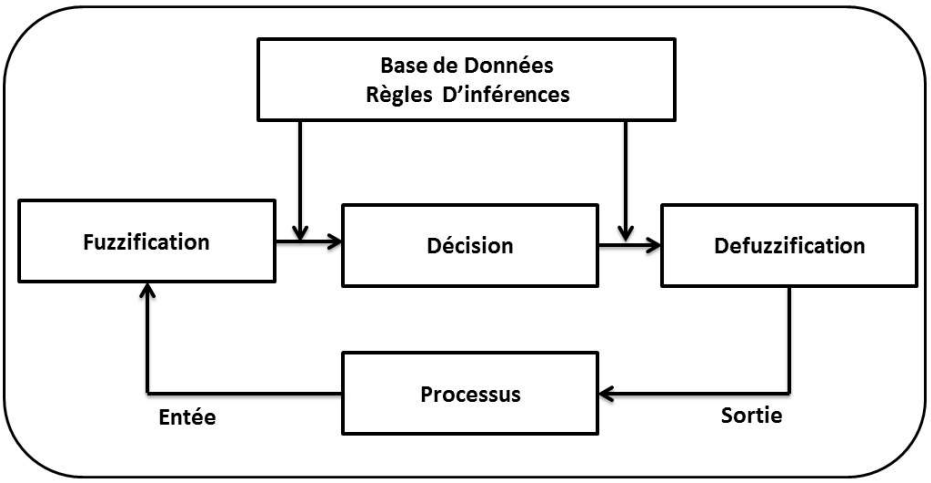

Figure 5-5 : Système flou

La logique floue semble donc reproduire la flexibilité du raisonnement humain quant à sa prise en compte des imperfections des données accessibles. Il serait donc intéressant de l'utiliser au cœur des systèmes experts, systèmes dont le but est de reproduire les mécanismes cognitifs d'un expert, dans un domaine particulier. La logique floue peut également servir pour un système décisionnel, lors de la phase d'analyse des données par exemple et peut s'avérer utile pour la prise de décision, soit pour découvrir des règles ou inférences floues permettant de mieux comprendre les données et ainsi éclairer les décisions, soit pour effectuer des requêtes dites floues en se basant sur les connaissances des experts.

En effet, l'algorithme flou se déroule en 3 étapes :

- Etape 1 : Transformation de variables quantitatives en variables logiques floues,
- Etape 2 : Utilisation de règles logiques pour évaluer de nouvelles variables floues en sortie,
- Etape 3 : Transformation de ces variables floues en variables qualitatives,

5.3.6.1 Première étape : Fuzzification

La fuzzification ou définition des fonctions d'appartenance des variables d'entrée et de sortie consiste à déterminer pour chaque variable les valeurs linguistiques ainsi que la forme des fonctions d'appartenance et le degré d'appartenance à différents états que l'on doit définir.

Un ensemble flou est caractérisé par une fonction d'appartenance $f : E \rightarrow [0, 1]$, qui positionne les membres de l'univers de discours E dans l'intervalle d'unité [0, 1]. La valeur 0 signifie que le membre n'est pas inclus dans l'ensemble donné, 1 décrit un membre entièrement inclus. Les valeurs entre 0 et 1 caractérisent les membres flous. L'univers de

discours d'une variable couvrira l'ensemble des valeurs prises par cette variable. Dans notre cas, l'univers du discours E correspond au pourcentage d'intervention, pourcentage de présence des actes de langage dans le message (type d'intervention), pourcentage de réactions directes et indirectes, discrétisé en 11 éléments {0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100}. Pour un élément x de E, la valeur f(x) représente le degré d'appartenance de x à un sous-ensemble flou.

Variables d'entrée :

Trois variables linguistiques {Faible, Moyen, Fort} qualifient nos variables d'entrées : (E1) Pourcentage d'intervention, (E2) Pourcentage type d'intervention, (E3) Pourcentage de réactions directes et (E4) Pourcentage de réactions indirectes.

Variable de sortie :

Niveau comportemental : caractérisation qualitative du comportement social de l'apprenant « Animateur, Vérificateur, Quêteur et Indépendant ».

- (S) Niveaux Comportementaux : {Faible, Moyen, Bon, Excellent}.

L'univers de discours de chaque variable d'entrée est divisé en trois sous-ensembles flous {Faible, Moyen, Fort}. En déterminant les plages de nos variables linguistiques, grâce aux intervalles de confiances (figures 5.6, 5.7, 5.8).

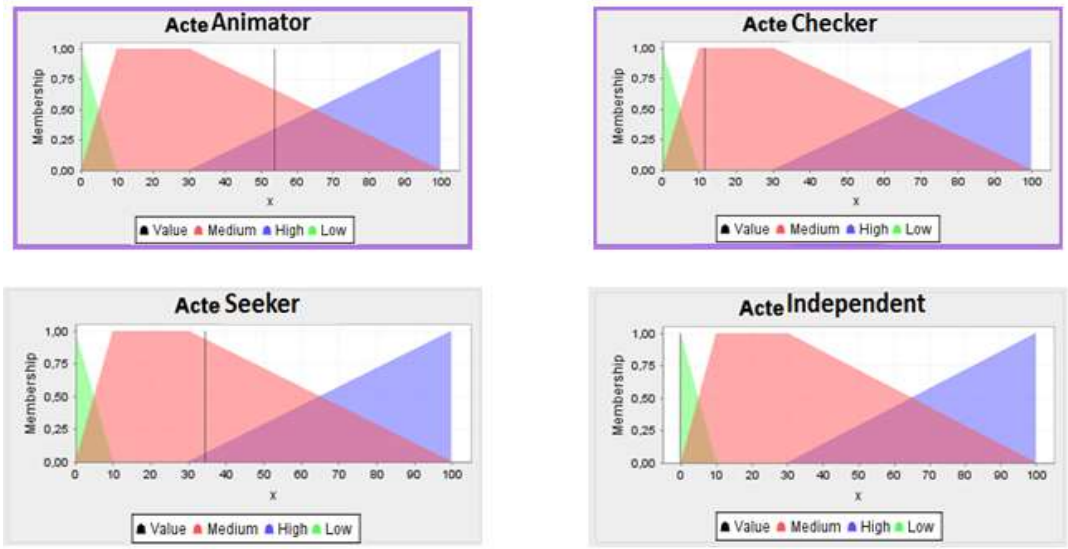

Figure 5-6 : Fonctions d'appartenance Actes catégorie Animateur, Vérificateur, Quêteur et Indépendant sont définis de manière similaire

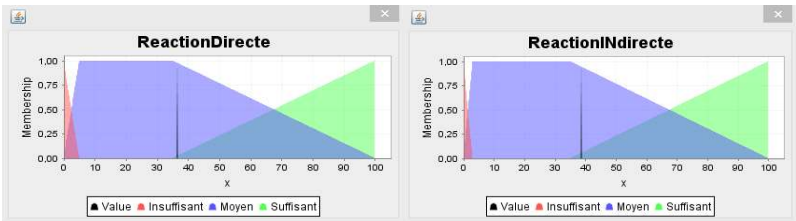

Figure 5-7 : Fonctions d'appartenance Réaction Directe et indirecte
Figure 5.7.

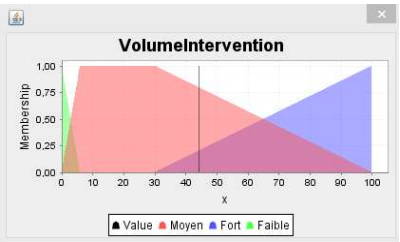

Figure 5-8 : Fonctions d'appartenance Volume d'intervention

Pour représenter les variables linguistiques des entrées, nous avons défini en collaboration avec l'enseignant expert la fonction d'appartenance de forme trapézoïdale. L'enseignant-expert spécifie les degrés d'appartenance des niveaux comportementaux de l'apprenant à chacun des sous-ensembles flous obtenus.

Les sous-ensembles flous associé à la variable de sortie, « Niveaux Comportementaux » sont {Faible, Moyen, Bon, Excellent} définis avec des fonctions d'appartenance en forme de raies (Figure 5.9).

La génération de la variable de sortie se fait par le système en utilisant la méthode du centre de gravité, selon laquelle le système calcule la variable de sortie arrondie au nombre entier le plus proche. Les plages de la variable de sortie ont été définies à partir des statistiques des analyses intuitives faites par les tuteurs.

Nous avons estimé les plages de résultats selon les statistiques de l'analyse intuitive par les tuteurs. Si le pourcentage se situe :

- Entre 0% et 20% le résultat est Insuffisant.
- Entre 20% et 50% le résultat est Moyen.
- Entre 50% et 70% (Bon) et entre 70% et 100% (Excellent).

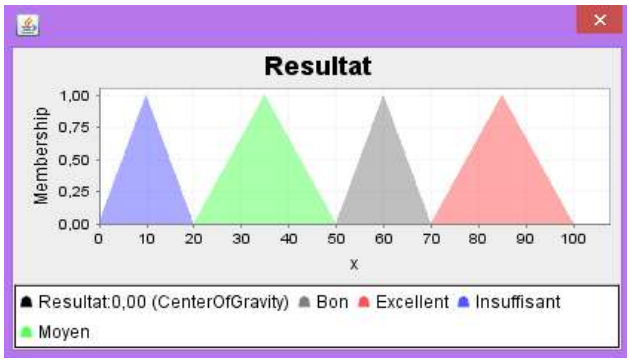

Figure 5-9 : Fonction d'appartenance Résultat

5.3.6.2 Deuxième étape : Le moteur d'inférence

Maintenant que l'on possède des variables linguistiques, on va pouvoir les utiliser dans le moteur d'inférence. Chaque règle du moteur d'inférence est écrite par le concepteur du système flou en fonction des connaissances qu'il possède. La conception d'une base de règles floues est un processus itératif. La plus grosse part de travail se trouve au niveau du recueil des connaissances expertes. Ainsi, en utilisant des données correspondant aux différentes entrées et sorties, l'enseignant-expert fournit une série de combinaisons basées sur les conditions (tableau 4.1) qui caractérisent chaque profil de comportement « Animateur, Vérificateur, Quêteur et Indépendant ».

Un des intérêts de la logique floue pour formaliser le raisonnement humain est que les règles soient énoncées en langage naturel.

Nous avons défini 60 ensembles flous. À titre d'exemple quelques règles pour déterminer le comportement social d'un apprenant.

```
SI
    (Ration d'intervention est fort) et
   (Type d'intervention catégorie Animateur est fort) et
   (Pourcentage Réaction Directe est fort) et
   (Pourcentage Réaction Indirecte est fort)
ALORS
   (Niveau Comportement Animateur est Excellent)


SI
    (Ration d'intervention est moyen) et
   (Type d'intervention catégorie Vérificateur est fort) et
   (Pourcentage Réaction Directe est moyen) et
   (Pourcentage Réaction Indirecte est moyen)
ALORS
   (Niveau Comportement Vérificateur est Excellent)
```

5.3.6.3 Troisième étape : défuzzification

La dernière étape pour avoir un système flou opérationnel s'appelle la défuzzification. A la fin de l'inférence, l'ensemble flou de sortie est déterminé, mais il n'est pas directement utilisable pour donner une information précise. Il est nécessaire de passer du « monde flou » au « monde réel ».

Il existe plusieurs méthodes (la moyenne des maximas et centre de gravité), dont la plus utilisée est le calcul du « centre de gravité » de l'ensemble flou. La valeur de la sortie (s) « Niveaux Comportementaux » (Animateur, vérificateur, Quêteur et Indépendant), une fois évaluée au moyen de la base des règles, puis « défuzzifiée », donne une estimation du profil de l'apprenant en fonction des indicateurs.

Finalement, notre Système sera doté d'une estimation qualitative des comportements sociaux de l'apprenant, lui permettant ainsi d'identifier ses lacunes et ses points faibles et d'équilibrer les groupes en fonction de leurs comportements sociaux (figure 5.10).

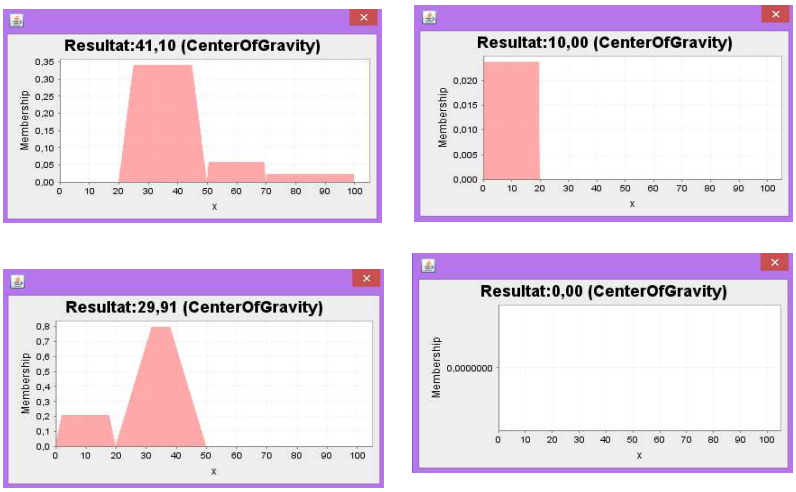

Figure 5-10 : Centres de gravité profil animateur, vérificateur, quêteur et indépendant

5.4 Conclusion

L'un des objectifs de notre travail de recherche consiste à proposer un système générique qui peut s'adapter à plusieurs environnements d'apprentissage. Le caractère générique de système impose un choix de données communes à la majorité des EIAH. Ainsi, les données que nous cherchons à analyser sont les données relatives aux échanges des apprenants à travers les outils de communication (forum et mail). Ces données sont présentes dans la majorité des EIAH existants, la seule différence réside dans leur structuration, qui peut

différer d'un EIAH à un autre.

Nous avons commencé par l'analyse des interactions entre apprenants dans le cadre d'un travail collaboratif à distance. Nous nous sommes proposés de développer un système multi agent qui pourra être greffé aux plates-formes existantes pour analyser les données d'interaction afin d'assister les apprenants et les tuteurs engagés dans une activité d'apprentissage en ligne. Le système permet de récupérer les données d'interaction des plates-formes, les analyser en calculant un certain nombre d'indicateurs et enfin entreprendre des actions en se basant sur le modèle de référence [2][3]. Ensuite, nous avons mené des études approfondies sur l'analyse des forums de discussion asynchrone semi-structurée et non semi-structurée (forum Libre). Nous avons proposés un système intelligent à base d'agent intègre des fonctionnalités d'analyse automatique des interactions et du contenu des échanges effectués à travers les différents outils de communication et d'organisation d'un LMS. Ce système est plus proche du résonnement humain basé sur des algorithmes intelligents pour l'analyse et catégorisations automatique de conversations textuelles asynchrones (Forum non semi-structurée) d'apprenants pour la détermination des profils comportementaux dans le cadre d'un travail collaboratif à distance. Les résultats obtenus sont concrétiser et intégrer dans un dispositif Tuteur Intelligent, qui apprendrait à partir de la base de connaissances constituée par toutes les traces d'interactions obtenues. Sur la base des règles d'inférences (logique floue) convenues, le système serait en mesure de proposer des rétroactions aux apprenants en fonction des rapports des indicateurs d'analyse de traces d'interactions.

6 Expériences & résultats

Sommaire

6.1 Introduction 147
6.2 Cadre théorique 147
6.2.1 Les interactions dans un forum de discussion éducatif 148
6.2.2 Problématique 149
6.3 Méthodologie 150
6.3.1 Contextes d'expérimentation 151
6.3.2 Organisation humaine dans le projet 151
6.3.3 Structuration du projet 152
6.3.4 Les outils mis à la disposition des apprenants 153
6.3.5 Le protocole de recueil de données 153
6.3.6 Description du corpus 154
6.4 Analyse et résultats de l'expérimentation 154
6.4.1 Analyse intuitive des tuteurs 154
6.4.2 Analyse de système 156
6.4.2.1 Analyse par actes de langage. 158
6.4.2.2 Analyse par naïve bayes 159
6.4.3 Comparaissant des deux méthodes de classification 159
6.4.4 Interprétation graphique des données 163
6.5 Discussion des résultats 166
6.6 Conclusion 169

6.1 Introduction

Ce travail a pour finalité d'automatiser certaines tâches habituellement effectuées par un tuteur humain. Dans ce sens, nous avons procédé à une étude comparative entre l'évaluation humaine, et celle obtenu par notre modèle.

Notre travail répond aux interrogations relatives à l'intégration des nouvelles technologies dans l'éducation et particulièrement l'analyse des forums de discussion éducatifs asynchrones utilisés pour l'apprentissage des apprenants.

L'un des objectifs de notre recherche est d'effectuer des expérimentations dans des contextes réels de formation pour ne pas nous limiter à l'observation de l'utilisation des fonctionnalités techniques mais pour observer les usages.

Dans ce chapitre, nous présentons des expérimentations que nous avons menées, notamment dans l'intention de vérifier que notre système peut être utilisée avec des publics variés et dans différents contextes pour déterminer les comportements sociaux des apprenants lors d'un travail collaboratif en ligne.

Avant de mettre en place une expérimentation, nous présentons le cadre théorique de notre expérience. La deuxième section sera consacrée à la description des interactions en classe et dans un forum de discussion éducatif. La troisième section sera consacrée à la description d'une expérimentation qui a été menée dans le cadre du master « Qualité logiciel » de l'université Ibn Tofail de Kenitra, Maroc. Nous décrirons dans ce paragraphe les choix organisationnels, pédagogiques et techniques mis en place. Ensuite, dans la quatrième section nous analyserons les résultats à l'issue d'une analyse intuitive des tuteurs. Enfin, la cinquième section, nous présentons un bilan général des expérimentations.

6.2 Cadre théorique

L'apprentissage collaboratif est une méthode d'enseignement permettant aux apprenants travaillent ensemble en équipes afin d'atteindre un but commun. Il présente les deux caractéristiques suivantes :

1. L'interdépendance positive entre les apprenants où chaque membre contribue au travail de l'équipe et est indispensable à l'accomplissement du travail de l'équipe.
2. L'interaction avec les apprenants et le partage d'informations où la production commune résulte d'un processus de mise en commun d'informations et des discussions entre les participants.

Nous nous appuyons sur deux théories, socioconstructivistes et interactionnistes qui prouvent que l'apprentissage est un acte social et culturel où l'interaction entre apprenants ou

tuteur joue un rôle important. Cette interaction, pouvant se réaliser à travers des forums de discussion asynchrone.

6.2.1 Les interactions dans un forum de discussion éducatif

Actuellement le forum de discussion et le courrier électronique font partie intégrante des plates-formes de formation en ligne. Ces outils de communication asynchrone sont chargés de nombreux rôles au sein d'une formation à distance : être le lieu d'interactions sociales, de co-construction de connaissance, de partage d'expérience, de coopération et de collaboration. D'où l'intérêt que nous avons accordé à l'analyse des données issues de ces deux outils de communication. Par ailleurs, les outils synchrones (chat) ont été exclus de notre analyse pour plusieurs raisons : ces outils de chat n'ont pas beaucoup évolué depuis une vingtaine d'années et demeurent peu adaptés pour des conversations soutenues [May& al. 2007]. Leur fonctionnement repose sur un empilement des messages des utilisateurs de façon temporelle. Nous pouvons formuler des critiques à ce fonctionnement. Tout d'abord, cet empilement de messages les uns à la suite des autres selon leur ordre d'arrivée pose deux problèmes majeurs dans un contexte de conversation synchrone. Le premier problème provient du temps nécessaire à la composition des textes (temps de frappe) qui ne permet pas de répondre de façon immédiate à un message. De ce fait, les temps de latence provoquent des imbrications des interventions qui rendent le suivi de la discussion difficile. Le second problème se situe ainsi dans la lecture des chats, deux messages pouvant se retrouver proches à l'interface alors qu'ils ne sont pas forcément liés et, à l'inverse, deux messages en relation peuvent être séparés par d'autres messages. Ce phénomène s'accentue quand le nombre de personnes participant à la discussion augmente. La conséquence de cette difficulté de lecture dans les chats classiques est que la conversation est délicate à suivre et à mener. Les utilisateurs tentent de s'adapter au médium par exemple en abrégeant au maximum les temps de composition des messages (au détriment de l'orthographe bien souvent) ou en se définissant des signes pour coordonner les tours de paroles [Herring, 1999]. Des résultats d'analyse de discussions médiatisées par des outils de chats ont montré que celles-ci étaient très souvent incohérentes et qu'on trouvait de nombreux messages de « réparation » de la discussion [Herring, 1999]. Ainsi nous avons privilégié le travail sur les outils de communication asynchrones.

6.2.2 Problématique

Notre travail se situe dans le domaine des Environnements Informatiques pour l'Apprentissage Humain (EIAH), Nous nous intéressons dans ce rapport à l'Apprentissage Collectif Assisté par Ordinateur (ACAO) plus connu sous sa dénomination anglaise Computer-Supported Collaborative Learning (CSCL).

Dans le contexte d'une formation collaborative à distance, la communication textuelle est d'une importance capitale. Les outils de communication asynchrones textuels tels que le courriel et le forum évitent les contraintes bloquantes de rendez-vous. Ces outils restent à ce jour le meilleur compromis entre souplesse et interactivité pour la réalisation d'un travail collaboratif en ligne [De Lievre, 2011]. Face à la grande quantité des messages déposés dans les forums, les tuteurs se sentent souvent démunis pour construire une représentation synthétique de l'activité des individus et des groupes. Le tuteur risque de manquer d'objectivité quand il s'en sert pour évaluer l'implication et la place des apprenants dans les échanges et de repérer leurs comportements sociaux [Pléty, 1998 ; George, 2004]. Dans le cadre de l'analyse automatique des activités collaboratives des apprenants [Pléty, 1998 ; George, 2004 ; Oumaira, 2011] nous proposons une approche basée sur l'analyse de conversations textuelles asynchrones entre apprenants pour déterminer leurs comportements sociaux.

Nous voulons étudier les comportements des étudiants, les rôles qu'ils s'attribuent ainsi que l'organisation qu'ils mettent en place pour mener à bien un travail collaboratif. Notre questionnement porte aussi sur la co-construction des connaissances en l'absence d'un tuteur et faisons l'hypothèse que des étudiants contraints à travailler en groupe dans un forum de discussion asynchrone se comportent en faisant référence aux situations d'apprentissage présentiels. Ils auraient donc tendance à adopter les mêmes rôles et à s'organiser spontanément pour avancer ensemble et co-construire leurs connaissances en l'absence de l'enseignant.

Dans le cadre de formation à distance où il n'y a aucune interaction entre le tuteur et l'apprenant, les données rassemblées tendent à être plus imparfaites que celles obtenues par l'interaction en présentiel. La présence d'informations imparfaites est un facteur important qui mène souvent aux erreurs dans la détermination des comportements sociaux de l'apprenant. Cette imperfection semble due aux erreurs et aux approximations impliquées lors du recueil de l'information, partiellement en raison de la nature abstraite de la connaissance humaine et de la perte d'information résultant de sa quantification. La théorie des sous-

ensembles flous [Zadeh, 1965] se présente comme un outil privilégié pour la modélisation des situations présentant des imprécisions. Une des motivations principales de l'usage de la logique floue est la manipulation améliorée de l'imperfection de l'information. En effet, le raisonnement d'un système de logique floue est considéré comme « facile », du point de vue compréhension et/ou modification par des concepteurs et utilisateur. Un des facteurs qui appuie cette considération est la similarité humaine. Elle peut fournir des descriptions de la connaissance comme un humain et imiter son modèle de raisonnement concernant les concepts vagues. Ceci est d'un intérêt particulier dans la conception d'un système modelant la connaissance interprétable de l'apprenant qui est basée sur le raisonnement et la conceptualisation de l'enseignant-expert.

6.3 Méthodologie

L'approche méthodologique choisie visait à apporter une réponse aux différents éléments de problématique dégagés plus haut et à remplir nos objectifs de recherche. Pour cela, nous avons participé à la mise en place d'une expérimentation réelle de formation à distance. Nous avons recueilli un corpus de données d'interactions et avons aux aussi vécu au quotidien la progression de la session durant la période de formation. Les données recueillies durant cette expérimentation nous ont servi de support de développement et de test pour notre système d'assistance au tuteur.

La méthodologie que nous avons adoptée se base sur l'étude des traces laissées dans les forums de discussion asynchrone et l'analyse de contenu dans le cadre d'une activité collaborative. L'analyse de contenu de message est définit comme : « une technique de recherche pour la description objective, systématique et quantitative du contenu manifeste de la communication » [Charaudeau & Mangueneau, 2002] et qui présuppose « deux opérations fondamentales » : « la pré-catégorisation thématique des données textuelles, et leur traitement quantitatif et qualitatif, généralement informatisé ».

Dans ce cadre, plusieurs indicateurs peuvent être construits [Pléty, 1998 ; Baron & Bruillard, 2003 ; Huberman & Miles. 1991] pour observer les processus sociaux, cognitifs et métacognitifs mis en œuvre. En effet, des variables comme le « volume d'interaction », le « type d'intervention » et même la « réaction entrainée » des interventions peuvent être mesurées pour rendre compte de la situation que nous étudions. Pour ce faire, il a été question de procéder à une analyse de contenu susceptible de caractériser l'interactivité virtuelle entre apprenants et répondre aux questions que nous nous sommes posées.

6.3.1 Contextes d'expérimentation

L'un des objectifs de notre recherche est d'effectuer des expérimentations dans des contextes réels de formation pour ne pas nous limiter à l'observation de l'utilisation des fonctionnalités techniques mais pour observer les usages. Une expérimentation a été menée avec les étudiants de master qualité logiciel. Le Master « Qualité logiciel » est un master dispensé par la faculté des sciences de l'université Ibn Tofail de Kenitra. Les responsables du master ont opté pour une modalité d'enseignement hybride ; ainsi les étudiants inscrits à ce master suivaient quatre modules en présentiel et un seul élément en présentiel et à distance, c'est l'élément « développement mobile » qui a fait l'objet de notre expérimentation. Au début du cours, les étudiants ont assisté à une séance d'initiation à l'utilisation de la plateforme utilisée accompagnée d'une présentation du scénario adopté. Ce travail devait se faire exclusivement en ligne en utilisant le forum que nous avons créé sur la plateforme Moodle.

Ce travail a été conditionné par une durée de 4 mois pour la réalisation des projets mobile organisés en 6 phases, correspondant à des tâches différentes. Cette période coïncide avec celle des vacances mais aussi par la constitution de quadrinômes formant ainsi 9 groupes de travail. Nous avons choisi cette période pour contraindre les étudiants de master à ne travailler qu'en ligne mais aussi parce qu'au-delà de cette durée, le corpus collecté aurait été très grand.

6.3.2 Organisation humaine dans le projet

La constitution des groupes n'est pas une chose facile. Collaborer avec trois autres personnes pour la mise en place d'un travail nécessite de faire des choix. Partager le même objectif que d'autres ne va pas de soi non plus. Il existe une volonté d'investissement qui guide ces choix. Les apprenants qui ont participé à ce travail sont des étudiants de première année « qualité du logiciel » de l'université Ibn Tofail de Kenitra. Ils ont un âge qui varie entre 22 et 23 ans. Les 32 apprenants engagés dans cette expérience ont été répartis en équipes de 4 membres. Dans notre expérience, l'enseignant a laissé aux apprenants eux-mêmes le choix des personnes avec lesquels ils veulent travailler. Aucune intervention de la part du professeur n'a été faite pour orienter la constitution des groupes. Les groupes se sont formés de manière spontanée mais après des temps relativement différents. Chaque équipe a été suivie par un tuteur. Les projets ont été définis en fonction du cahier des charges fourni par des chefs de projet « Tuteurs ».

Les apprenants qui ont participé à ce travail sont des étudiants qui ont une bonne connaissance des techniques de l'information et de la communication. Le travail que ces apprenants sont amenés à faire en ligne est comptabilisé dans leur note. Concernant le suivi des apprenants, nous avons mis à la disposition des tuteurs notre système afin d'analyser les données d'interactions.

6.3.3 Structuration du projet

Nous avons proposé de structurer les projets. Un projet se compose de phases. Chaque phase comprend un certain nombre d'activités. Une activité est définie par une ou plusieurs tâches à effectuer.

Etape 1 « Cahier des charges » : La phase d'élaboration de cahier des charges est une phase essentielle à la réalisation d'un projet. Il décrit précisément les besoins auxquels les apprenants doivent répondre : objectifs, cibles, concurrence, spécifications techniques, délais, etc.

Etape 2 « Analyse préliminaire » : Dans cette phase les apprenants ont pour objectif de faire une analyse du profil du porteur de projet et de la viabilité du projet à l'aide d'une fiche synthèse du projet.
Elle vise à :

- Présenter le promoteur et le projet en général.
- Examiner l'adéquation profil/activité.
- Dégager la pertinence et la valeur ajoutée de l'activité envisagée.

Etape 3 « Analyse détaillé » : L'étude approfondie consiste, à l'aide de la fiche d'évaluation du projet, en une analyse complète du projet dans ses aspects technique, commercial, stratégique, financier, institutionnel, organisationnel, environnemental, socioéconomique, etc.

Elle vise à analyser les dossiers selon la démarche suivante :

- Evaluation du projet
- Analyse de l'étude de faisabilité et de la rentabilité du projet

Etape 4 « Conception » : Chaque membre de l'équipe effectue une analyse des solutions techniques possibles, et des propositions pour résoudre des problèmes. Les apprenants doivent ensuite discuter de manière collaborative, en mode asynchrone pour formaliser les fonctionnalités de la solution retenue, l'architecture logiciel choisie, les interactions avec les différents systèmes SI extérieurs, etc.

Etape 5 « Implémentation » : La phase d'implémentation va toujours s'appuyer en premier lieu sur la capture des besoins initiaux. Encore une fois, chaque cas d'utilisation, chaque diagramme d'activité, d'état ou de séquences définies dans la phase de capture des besoins devra être réalisé dans la phase d'implémentation.
Les membres de l'équipe doivent ensuite discuter des sujets. Cette étape est réalisée en mode asynchrone à travers des forums.

Etape 6 « Intégration » : Les membres de l'équipe consultent les productions des coéquipiers et apportent leur remarque et suggestions. Les apprenants doivent ensuite travailler ensemble pour intégrer les sous projets sur lesquels ils se sont mis d'accord.

6.3.4 Les outils mis à la disposition des apprenants

Pour soutenir les travaux des apprenants, nous avons mis en place un environnement technique attaché à la plate-forme Moodle de l'université Ibn Tofail. Cet environnement se présente sous la forme d'un ou plusieurs forums de discussion, intégrant également un service de publication de documents.

Les apprenants avaient à leur disposition :

- Une description détaillée du cahier des charges de chaque projet (Le contexte et les enjeux du projet, les contraintes à prendre en compte, la description technique, le public visé, les objectifs généraux, les ressources nécessaires, le délai de réalisation, etc).
- Un espace de conversation asynchrone (Forum) pour chaque équipe, ou les apprenants peuvent discuter des différentes tâches.
- Des ressources mises en ligne sur la plateforme Moodle (cours, articles, lien Internet, etc).

6.3.5 Le protocole de recueil de données

Afin de recueillir les impressions du tuteur sur l'utilisation de système, nous avons réalisé un entretien direct avec lui. La première partie de l'entretien était consacrée à l'évaluation du système par les tuteurs d'un point de vue technique, fonctionnel et ergonomique. La seconde partie a été dédiée à la comparaison de la perception que les tuteurs avaient des états des apprenants et des groupes par rapport aux résultats produits par le système et l'apport du système dans l'assistance et le suivi des apprenants. Concernant les apprenants, un entretien avec les apprenants avait pour objectif de recueillir leurs impressions sur le déroulement de l'expérimentation, l'utilisation des outils "mail" et "forum libre" (non semi structurés par des actes de langage) et leur évaluation de la qualité du tutorat. Et finalement la troisième partie est consacrée aux recommandations d'amélioration proposés pour les tuteurs.

Les données recueillies lors de cette expérimentation sont de deux natures : les traces informatiques et les réponses à des questionnaires. L'environnement garde des traces des actions des utilisateurs dans le système. Nous avons ainsi la possibilité d'avoir des données sur les activités, le déroulement de travail, les communications effectuées ainsi que la collaboration entre les membres de l'équipe. Concernant le système il garde des traces sur les actions des apprenants dans l'outil de conversation asynchrone : liste des groupes, des apprenants et phase de projet, messages envoyés, actes de langage et fils de discussion.

6.3.6 Description du corpus

Notre corpus est constitué entre 80 et 120 messages envoyés par les apprenants au sein de leur même groupe par phase de projet. Ces messages sont enregistrés sur le forum et sont toujours consultables par les sujets participants au travail.

6.4 Analyse et résultats de l'expérimentation

Dans cette section, nous présentons les données des expériences menées. Plus précisément, les résultats de l'analyse intuitive de tuteur, ont été utilisés afin de vérifier performances de notre système. Pour rappel, et à titre d'exemple ; qualitativement un animateur fait des propositions, poste des messages d'organisation et/ou encouragement, intervient pour calmer un conflit, et quantitativement poste un volume de messages important, auxquels sont associés des suivis de réactions importants (Tableau 4.1). À cette fin, les résultats du système ont été comparés par l'évaluation de l'enseignant expert basé sur le calcul des indicateurs : le volume d'interventions, type d'interventions et de réactions entraînées.

Nous avons fait une comparaison entre deux méthodes de classification de texte pour calculer les types d'interventions sur la base des actes de langage et de la classification bayésienne de textes pour déterminer les comportements sociaux des apprenants.

6.4.1 Analyse intuitive des tuteurs

Les tableaux 1 et 2 illustrent les résultats de l'analyse intuitive faite par les tuteurs pendant les deux premières phases du projet, dont l'objectif est d'associer un profil comportemental à chaque apprenant. Pour plus de lisibilité, dans cette analyse, chaque tuteur va associer un profil à chaque apprenant selon son comportement social. En effet, après avoir identifié les profils des apprenants, nous avons demandé aux tuteurs, par analyse des contenus (de lire le contenu de tous les messages), de faire une classification des messages

(type : animateur, vérificateur, quêteur et indépendant) en identifiant les actes de langage qui les caractérisent : proposition, message d'organisation et/ou encouragement, intervention pour calmer un conflit, réaction à une proposition, expression de doute sur une démarche ou proposition, etc. (Tableau 4.1). Les résultats de l'analyse des tuteurs pour le groupe 1 durant les périodes 1 et 2 (tableau 6.1 et 6.2), est confirmée par les données relevées par le système à travers l'analyse de contenus des messages (voir tableau 6.4 et 6.5).

	Résultat Analyse intuitive pour le groupe 1 phase Rédaction cahier des charges						
	Groupe	profil	Classification intuitive de messages				Volume d'interventions
			Type des interventions				
		Impression de profildégagée par l'apprenant	Nombre de Messages catégorie "Animateur"	Nombre de Messages catégorie "Vérificateur"	Nombre de Messages catégorie "Quêteur"	Nombre de Messages catégorie "Indépendant"	
Apprenant 1	Groupe 1	Animateur	15 (34,88 %)	21 (48,83 %)	07 (16,27 %)	00 (00 %)	Important (44,32 %)
Apprenant 2		Animateur	11 (47,82 %)	11 (47,82 %)	01 (4,34 %)	00 (00 %)	Important (23,71 %)
Apprenant 3		Indépendant	3 (50 %)	3 (50 %)	00 (00 %)	00 (00 %)	Faible (06,1 %)
Apprenant 4		Animateur	13 (52 %)	11 (44 %)	01 (04 %)	00 (00 %)	Important (25,77 %)

Tableau 6-1 : Résultat d'analyse intuitive pour le groupe 1 phase rédaction de charge des charges

	Résultat Analyse intuitive pour le groupe 1 phase Analyse et Conception						
	Groupe	profil	Classification intuitive de messages				Volume d'interventions
			Type des interventions				
		Impression de profildégagée par l'apprenant	Nombre de Messages catégorie "Animateur"	Nombre de Messages catégorie "Vérificateur"	Nombre de Messages catégorie "Quêteur"	Nombre de Messages catégorie "Indépendant"	
Apprenant 1	Groupe 1	Animateur	22 (53,65 %)	12 (29,26 %)	05 (12,19 %)	02 (4,87 %)	Important (41,41 %)
Apprenant 2		Vérificateur	03 (15,78 %)	13 (68,42 %)	02 (10,52 %)	01 (5,26 %)	Assez Important (19,19 %)
Apprenant 3		Indépendant	1 (16,66 %)	3 (50 %)	00 (00 %)	02 (33,33 %)	Faible (06,06 %)
Apprenant 4		Animateur	16 (48,48 %)	12 (36,36 %)	03 (09,09 %)	02 (6,06 %)	Important (33,33 %)

Tableau 6-2 : Résultat d'analyse intuitive pour le groupe 1 phase analyse et conception

Exemple pour le groupe 1 durant la période de rédaction du cahier des charges, l'apprenant 1 dégage un profil animateur, qui correspond à l'analyse des données relevées par le tuteur : volume d'intervention élevé (44,32 %), type d'intervention élevé.

6.4.2 Analyse de système

Notre approche utilise la théorie des actes de langage et classification bayésien comme deux solutions pour l'analyse des messages textuels asynchrones et permet d'extraire un certain nombre d'indicateurs (qualitatifs et quantitatifs) qui permettent de définir les profils des apprenants (Tableau6.3).

Le tableau ci-dessous contient tous les indicateurs calculés par notre système par deux méthodes de classification, la présence des actes de langage dans les messages et classification bayésienne naïve au cours de la première phase du projet; (Le projet est divisé en cinq phases), les indicateurs calculés sont: le volume d'intervention, type d'intervention pour (Animateur, vérificateur, quêteur et indépendant), Pourcentage de réactions directes et indirectes de tous les apprenants d'un même groupe.

Résultat d'analyse de system												
Groupe	Apprenant	Volume d'intervention	Type d'intervention «Naive Bayes»				Type d'intervention «Acte de langage»				Réactionentrainées	
			Animateur	Vérificateur	Quêteur	Indépendant	Animateur	Vérificateur	Quêteur	Indépendant	Directe	Indirecte
Groupe 1	Apprenant 1	44,33 %	33,45 %	47,03 %	15,98 %	00,00 %	30,00 %	43,83 %	14,20 %	00,00 %	36,53 %	38,70 %
	Apprenant 2	23,71 %	46,82 %	46,12 %	03,33 %	00,00 %	45,11 %	46,67 %	02,89 %	00,00 %	26,92 %	35,23 %
	Apprenant 3	06,19 %	48,00 %	48,88 %	00,00%	00,00 %	45,00 %	46,90 %	00,00 %	00,00 %	04,80 %	09,25 %
	Apprenant 4	25,77 %	49,90 %	43,00 %	03,70%	00,00 %	49,12 %	43,11 %	03,11 %	00,00 %	28,84 %	39,50 %
Groupe 2	Apprenant 5	13,95 %	05,17 %	33,33 %	00,00 %	20,00 %	11,11 %	33,33 %	00,00 %	20,00 %	11,11 %	20,00 %
	Apprenant 6	00,00 %	00,00 %	00,00 %	00,00 %	39,75 %	14,81 %	25,00 %	50,00 %	39,75 %	14,81 %	39,75 %
	Apprenant 7	20,08 %	21,27 %	00,00 %	40,00 %	36,14 %	48,14 %	00,00 %	00,00 %	36,14 %	48,14 %	36,14 %
	Apprenant 8	47,13 %	36,17 %	18,18 %	09,09 %	26,50 %	07,40 %	00,00 %	00,00 %	26,50 %	07,40 %	26,50 %
Groupe 3	Apprenant 9	25,00 %	06,89 %	00,00 %	99,00 %	03,65 %	06,89 %	00,00 %	99,00 %	03,65 %	06,89 %	03,65 %
	Apprenant 10	38,75 %	55,17 %	10,00%	35,00 %	53,29 %	55,17 %	10,00%	35,00 %	53,29 %	55,17 %	53,20 %
	Apprenant 11	17,50 %	10,43 %	20,00 %	00,00 %	00,00 %	10,43 %	20,00 %	00,00 %	00,00 %	10,34 %	00,00 %
	Apprenant 12	18,18 %	05,17 %	33,33 %	00,00 %	13,95 %	05,17 %	33,33 %	00,00 %	13,95 %	05,17 %	13,95 %
Groupe 4	Apprenant 13	13,95 %	25,17 %	33,33 %	00,00 %	00,00 %	00,00 %	00,00 %	00,00 %	00,00 %	00,00 %	00,00 %
	Apprenant 14	43,00 %	00,00 %	00,00 %	00,00 %	20,08 %	21,27 %	00,00 %	40,00 %	20,08 %	21,27 %	20,08 %
	Apprenant 15	13,95 %	05,17 %	33,33 %	00,00 %	47,13 %	36,17 %	18,18 %	09,09 %	47,13 %	36,17 %	47,13 %
	Apprenant 16	18,75 %	23,40 %	00,00 %	57,14 %	19,67 %	23,40 %	00,00 %	57,14 %	19,67 %	23,40 %	19,67 %
Groupe 5	Apprenant 17	83,65 %	06,89 %	00,00 %	99,00 %	02,08 %	05,88 %	00,00 %	99,00 %	02,08 %	05,88 %	02,08 %
	Apprenant 18	53,29 %	55,17 %	10,00%	35,00 %	00,00 %	00,00 %	00,00 %	00,00 %	00,00 %	00,00 %	00,00 %
	Apprenant 19	20,00 %	10,43 %	20,00 %	00,00 %	17,72 %	52,94 %	00,00 %	20,00 %	17,72 %	52,94 %	17,72 %
	Apprenant 20	13,65 %	06,89 %	00,00 %	99,00 %	10,12 %	11,76 %	00,00 %	99,00 %	10,12 %	11,76 %	10,12 %

Tableau6-3 : Indicateurs calculés par le système

6.4.2.1 Analyse par actes de langage.

Les comportements des apprenants dans les forums est déterminé à partir des messages postés dans les forums relatifs à chaque catégorie (animateur, vérificateur, quêteur et indépendant). Cette information permettra au système d'identifier l'acte de langage la plus utilisé par chaque apprenant ce qui renseigne le tuteur sur le profil comportemental de l'apprenant.

Lorsque nous soumettons les mêmes données d'interaction entre apprenants, au système d'analyse automatique que nous proposons, nous obtenons les résultats exposés dans les tableaux 6.4 et 6.5, pour le même groupe et pour les mêmes périodes.

	Résultat Analyse système pour le groupe 1 phase Rédaction cahier des charges							
	Volume des interventions	Type des interventions				Réaction entrainées		Profil
		Animateur	Vérificateur	Quêteur	Indépendant	Directe	Indirecte	
Apprenant 1	44,33 %	30,00 %	43,83 %	14,20 %	00,00 %	36,53 %	38,70 %	Animateur
Apprenant 2	23,71 %	45,11 %	46,67 %	02,89 %	00,00 %	26,92 %	35,23 %	Animateur
Apprenant 3	06,19 %	45,00 %	46,90 %	00,00 %	00,00 %	04,80 %	09,25 %	Indépendant
Apprenant 4	25,77 %	49,12 %	43,11 %	03,11 %	00,00 %	28,84 %	39,50 %	Animateur

Tableau 6-4 : Indicateurs calculés par le système pour le groupe 1 phase rédaction cahier des charges

	Résultat Analyse système pour le groupe 1 phase Analyse et Conception							
	Volume des interventions	Type des interventions				Réaction entrainées		Profil
		Animateur	Vérificateur	Quêteur	Indépendant	Directe	Indirecte	
Apprenant 1	39,05 %	40,00 %	27,83 %	11,20 %	04,00 %	42,59 %	23,11 %	Animateur
Apprenant 2	18,10 %	14,11 %	62,67 %	09,89 %	05,00 %	20,37 %	15,16 %	Vérificateur
Apprenant 3	05,71 %	15,55 %	49,09 %	00,00 %	30,44 %	00,92 %	02,88 %	Indépendant
Apprenant 4	31,43 %	47,58 %	36,11 %	08,60 %	05,00 %	27,77 %	20,21 %	Animateur

Tableau 6-5: Indicateurs calculés par le système pour le groupe 1 phase analyse et conception

L'analyse de ces résultats à la lumière des caractéristiques des profils d'apprenants définis (Tableau 4.1), permet d'associer un profil sociologique à chaque apprenant. Vu les résultats des analyses statistiques des actes de langage (type d'intervention) ; deux profils se dégagent : animateur et vérificateur. Cependant, en analysant les volumes d'interventions et réactions entrainées associés, nous constatons qu'ils sont importants et caractérisent le profil animateur. Ainsi, pour l'exemple considéré, pour la période 1, l'apprenant (1) sera qualifié de « animateur ». La même approche a été utilisée pour définir les profils apprenants dans les périodes 1 et 2 pour les étudiants de groupe 1 (Tableau 6.4 et 6.5).

6.4.2.2 Analyse par naïve bayes

Lorsque nous soumettons les mêmes données d'interaction entre apprenants, au système d'analyse automatique que nous proposons qui utilise la naïve bayes pour la classification des messages pour calculé le type d'intervention, nous obtenons les résultats exposés dans les tableaux 6.6 et 6.7, pour le même groupe et pour les mêmes périodes.

	Résultat Analyse système pour le groupe 1 phase Rédaction cahier des charges							
	Volume des interventions	Type des interventions				Réaction entrainées		Profil
		Animateur	Vérificateur	Quêteur	Indépendant	Directe	Indirecte	
Apprenant 1	44,33 %	33,45 %	47,03 %	15,98 %	00,00 %	36,53 %	38,70 %	Animateur
Apprenant 2	23,41 %	46,82 %	46,12 %	03,33 %	00,00 %	26,92 %	35,23 %	Animateur
Apprenant 3	06,19 %	48,00 %	48,88 %	00,00 %	00,00 %	04,80 %	09,25 %	Indépendant
Apprenant 4	25,77 %	49,9 %	43,00 %	03,70 %	00,00 %	28,84 %	39,50 %	Animateur

Tableau6-6 : Indicateurs calculés par le système pour le groupe 1 phase rédaction cahier des charges

	Résultat Analyse système pour le groupe 1 phase Analyse et Conception							
	Volume des interventions	Type des interventions				Réaction entrainées		Profil
		Animateur	Vérificateur	Quêteur	Indépendant	Directe	Indirecte	
Apprenant 1	39,05 %	52,30 %	28,27 %	12,03 %	04,50 %	52,30 %	28,27 %	Animateur
Apprenant 2	18,10 %	14,80 %	67,80 %	10,00 %	05,12 %	14,80 %	67,80 %	Vérificateur
Apprenant 3	05,71 %	15,00 %	50,00 %	00,00 %	32,21 %	15,00 %	50,00 %	Indépendant
Apprenant 4	31,43 %	48,30 %	35,00 %	08,60 %	05,43 %	48,30 %	35,00 %	Animateur

Tableau6-7 : Indicateurs calculés par le système pour le groupe 1 phase analyse et conception

L'analyse de ces résultats à la lumière des caractéristiques des profils d'apprenants définis (Tableau 4.1), permet d'associer un profil sociologique à chaque apprenant. Vu les résultats des analyses statistiques de classification par classification Bayésienne (type d'intervention) ; deux profils se dégagent : animateur et vérificateur. Cependant, en analysant les volumes d'interventions et réactions entrainées associés, nous constatons qu'ils sont importants et caractérisent le profil animateur. Ainsi, pour l'exemple considéré, pour la période 1, l'apprenant (1) sera qualifié de « animateur ». La même approche a été utilisée pour définir les profils apprenants dans les périodes 1 et 2 pour les étudiants de groupe 1 (Tableau 6.6 et 6.7).

6.4.3 Comparaissant des deux méthodes de classification

Le corpus a fait l'objet de divers prétraitements linguistiques, filtrage et correction de grammaire et l'orthographe ainsi que la lemmatisation des mots. Deux méthodes de

classification ont été testées : Classification par la présence ou l'absence d'actes de langage dans les messages, et classification bayésienne naïf de texte.

Nous rappelant que les deux indicateurs volume d'intervention et réaction entrainées sont des indicateurs qualitatifs. D'apprêt les résultats d'analyse des tuteurs et des deux systèmes de classifications concernant ces deux indicateurs quantitatifs, nous remarquant que les résultats sont les même. L'idée donc est de compris l'indicateur qualitatif qu'est le type d'intervention.

	Résultat d'analyse intuitive des tuteurs :				**Résultat d'analyse de système « Classification de messages par acte de langage »**				**Résultat d'analyse de système: « Classification de messages par Naive Bayes »**			
	Animateur	Vérificateur	Quêteur	Indépendant	Animateur	Vérificateur	Quêteur	Indépendant	Animateur	Vérificateur	Quêteur	Indépendant
Apprenant 1	34,88 %	48,83 %	16,27 %	00,00 %	30,00 %	43,83 %	14,20 %	00,00 %	33,45 %	47,03 %	15,98 %	00,00 %
Apprenant 2	47,82 %	47,82 %	04,34 %	00,00 %	45,11 %	46,67 %	02,89 %	00,00 %	46,82 %	46,12 %	03,33 %	00,00 %
Apprenant 3	50,00 %	50,00 %	00,00 %	00,00 %	45,00 %	46,90 %	00,00 %	00,00 %	48,00 %	48,88 %	00,00 %	00,00 %
Apprenant 4	52,00 %	44,00 %	04,00 %	00,00 %	49,12 %	43,11 %	03,11 %	00,00 %	49,9 %	43,00 %	03,70 %	00,00 %

Tableau 6-8 : illustrent les résultats de l'analyse intuitive de tuteur et de l'analyse automatique, pour les premières phases du projet.

D'après les résultats d'analyse intuitive des tuteurs pour le groupe 1 au cours de la phase 1, l'impression des profils dégagés par les apprenants au cours de chaque période du projet est confirmée par les données enregistrées par le tuteur à travers l'analyse du contenu des messages.

L'analyse intuitive pour le groupe 1 montre que l'apprenant 1 présente le profil de l'animateur ; qui, correspond à l'analyse des données recueillies par le tuteur : Animateur (34,88 %), Vérificateur (48,83 %), Quêteur (16,27 %) et l'indépendant (00,00 %) (Tableau 6.8).

L'analyse de système basé sur les actes de langage pour le groupe 1 montre que l'apprenant 1 présente le profil de l'animateur ; qui, correspond à l'analyse des données recueillies par le système : Animateur (30,00 %), Vérificateur (43,83 %), Quêteur (14,20 %) et l'indépendant (00,00 %) (Tableau 6.8).

L'analyse de système basé sur classification bayésien pour le même groupe « groupe 1 » montre que l'apprenant 1 présente le profil de l'animateur ; qui, correspond à l'analyse des données recueillies par le système : Animateur (33,45 %), Vérificateur (47,03 %), Quêteur (15,98 %) et l'indépendant (00,00 %) (Tableau 6.8).

D'après les résultats du système obtenu par les deux méthodes de classification de texte pour calculer le type d'intervention (tableau 6.8). On peut observer que les résultats d'analyse sont à proximité. Tous ces résultats nous ont permis de vérifier que les estimations des tuteurs sont les mêmes que celles fondées sur ces deux méthodes.

La présentation des résultats sont résumés dans les Tableaux 6.9 et 6.10. Affiche les résultats du calcul de la marge d'erreur entre l'analyse intuitive de tuteur et l'analyse du système par deux méthodes de classification pour déterminer le comportement social des apprenants.

	Animateur	**Vérificateur**	**Quêteur**	**Indépendant**	**Résultat**
Apprenant 1	74,55 %	95,11 %	91,87 %	82,13 %	85,92 %
Apprenant 2	89,41 %	91,59 %	94,01 %	95,05 %	92,52 %
Apprenant 3	93,33 %	98,18 %	100 %	91,32 %	95,71 %
Apprenant 4	98,14 %	99,31 %	94,60 %	82,50 %	93,64 %
Résultat Totale :					**91,94 %**

Tableau 6-9 : Le calcul des *écarts* entre l'analyse intuitive des tuteurs et les résultats de système de la méthode de classification « Actes de langage »

Le tableau 6.9 montre la marge d'erreur entre l'analyse intuitive de tuteurs et l'analyse de système fondées sur des actes de langage. Cette marge d'erreur est une estimation de la mesure que les résultats de notre travail peuvent être validés. Plus la marge d'erreur est importante, moins on peut avoir confiance que les résultats sont proches des vrais résultats de l'enseignant d'experts, et ainsi, de la réalité. La marge d'erreur est calculée directement à partir des résultats de l'analyse intuitive des tuteurs et les résultats du système. Dans l'exemple ci-dessous. D'apprêt l'analyse intuitive de tuteurs l'apprenant émergent un profil d'animateur avec 34,88 % et 30,00% selon les résultats du système par rapport aux apprenants de groupe (tableau 6.8). Nous avons considéré le résultat d'analyse intuitive de tuteurs comme une référence ; nous pouvons voir que le taux d'erreur est de 14,00% par rapport aux résultats de l'analyse intuitive. Tous ces résultats nous ont permis de vérifier que le système a un taux d'erreur de 7,78% par rapport à l'enseignant expert.

	Animateur	**Vérificateur**	**Quêteur**	**Indépendant**	**Résultat**
Apprenant 1	97,48 %	96,61 %	98,68 %	92,40 %	96,29 %
Apprenant 2	93,78 %	99,09 %	95,05 %	97,33 %	96,31 %
Apprenant 3	90,03 %	100 %	100 %	96,63 %	96,66 %
Apprenant 4	99,62 %	96,25 %	94,60 %	89,60 %	95,02 %
Résultat Totale :					**96,07 %**

Tableau 6-10 : Le calcul des écarts entre l'analyse intuitive des tuteurs et les résultats de système de la méthode de classification « Naive bayes »

Le tableau 6.10 montre le calcul du taux d'erreur entre l'analyse intuitive de tuteurs et l'analyse de systèmes qui se fonde sur naïve Bayes. Toutes ces observations nous permettre de vérifier que notre système basé sur le système multi-agent, classification du texte avec Naïve Bayes et la logique floue, offre une amélioration significative des résultats de l'analyse du comportement social des apprenants par une analyse qualitative et quantitative des discussions. En effet, notre approche réduire la marge d'erreur par rapport à l'enseignant expert de (03,66%).

6.4.4 Interprétation graphique des données

L'expérimentation a permis de monter que les profils comportementaux calculés étaient fidèles à ce qui se passait dans les conversations. Cet outil a été conçu dans un objectif d'observation expérimental mais il pourrait être fourni au tuteur pour l'assister dans le suivi

des apprenants ou aux apprenants eux-mêmes pour leur donner une vue sur leurs propres comportements. Par ailleurs, nous avons représenté l'évolution dans le temps des profils (Animateur, Vérificateur, Quêteur et Indépendant) d'un apprenant présent sur le forum de discussion (Figure 6.1, 6.2). La visualisation par le biais des graphiques a permis aux tuteurs d'avoir une vue globale sur le comportement des apprenants et des groupes. Les profils calculés sur 100 sont représentés en ordonnée et le temps est en abscisse. Par exemple, le graphe (Figure 6.1) signale que l'apprenant 1 a majoritairement tenu un rôle animateur « leader du groupe » pendant la première phase du projet. Par ailleurs, nous pouvons remarquer que cet apprenant est devenu vérificateur à la deuxième phase.

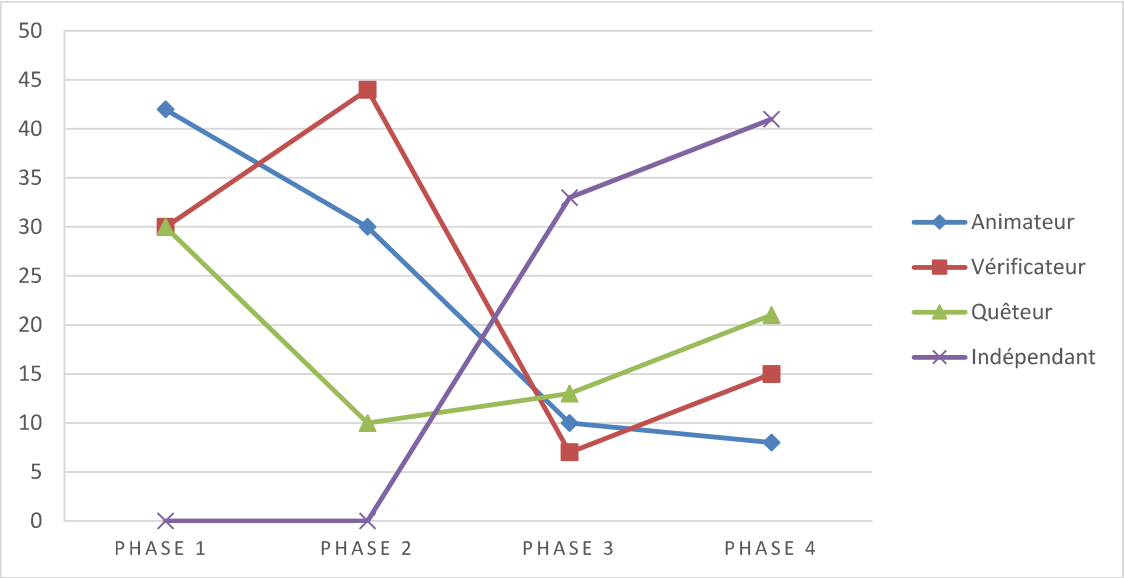

Figure 6-1 : Evolution du profil de l'apprenant (1) sur toute la durée du projet.

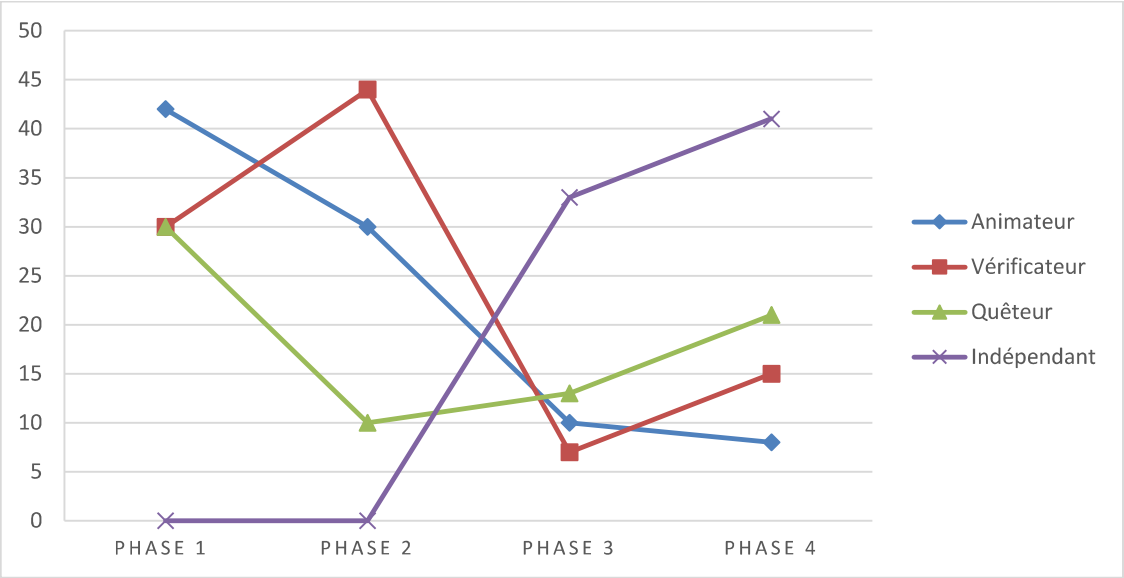

Figure 6-2 : Evolution du profil de l'apprenant (2) sur toute la durée du projet.

Concernant les analyses des interactions, les tuteurs ont déclaré que les résultats retournés par les indicateurs viennent conforter leurs observations de l'expérience. Les résultats d'analyse permettent aussi au tuteur de repérer aisément les apprenants isolés. C'est le cas l'apprenant 2 du groupe 1, le tuteur nous a expliqué qu'il ne s'est aperçu de l'isolement de cet apprenant qu'au moment de l'évaluation des livrables des activités qui étaient signés uniquement par les trois autres membres du groupe. Le graphe signale que l'apprenant (2) (Figure 6.2) a majoritairement tenu un rôle animateur dans la phase 1 du projet et devenu vérificateur durant la deuxième phase du projet, et indépendant durant la troisième et quatrième phase du projet.

Cette vue permet d'identifier le rôle joué par les apprenants dans leur groupe, à travers les différentes phases du projet. Cette variation est le résultat des préférences chez les apprenants par rapports aux tâches associées à chaque phase du projet. Par exemple, le graphique indique que l'apprenant (P) détient principalement un rôle indépendant au cours de la troisième et quatrième phase du projet. D'ailleurs, on peut remarquer que cet apprenant est isolé et n'est pas impliqué dans son travail.

Au vu de ces résultats, nous constatons que les observations faites par les tuteurs sont confirmées par l'analyse automatique faite par notre système, ce qui nous amène à affirmer que notre approche permet bien de retrouver les profils comportementaux dans les groupes d'apprenants travaillant à distance. Ces résultats démontrent que l'approche proposée est plus pertinente dans notre système ; ils indiquent que le système peut en effet effectuer l'évaluation et l'analyse du contenu de la discussion comme enseignant expert.

6.5 Discussion des résultats

La communication asynchrone permet aux apprenants de partager et de gérer les problèmes posés lors des discussions mises en jeux et que le forum de discussion asynchrone présente la caractéristique de conserver et préserver le fil des discussions et des idées énoncées et de permettre à des tuteurs et apprenant de les retrouvez facilement à tout moment.

Dans un dispositif d'apprentissage collaboratif à distance où les apprenants se fixent un projet commun à réaliser à distance, on s'axera surtout sur le soutien de la communication et de la communauté du groupe. Les apprenants qui ont participé à la réalisation de ce projet commun ont été contraintes de produire des arguments, proposer, collaborer et s'organiser en groupes, poser des questions et de se positionner par rapport aux avis des autres pour concevoir des travaux de qualité. L'incompréhension dans la réalisation de ces travaux se transforme en compréhensions partagées, cet objectif partagé qui semble être le moteur de ce type d'apprentissage.

A partir de cette expérimentation nous avons remarqué que pour l'établissement d'un lien qui affecte la collaboration de façon positive, la dimension affective des messages est importante. L'appartenance à un groupe naît le sentiment d'appartenance que Guertin décrit comme « un processus interactif par lequel les individus sont inter-reliés et se définissent en rapport les uns avec les autres en fonction de champs d'intérêts et d'affinités ». Ce sentiment d'appartenance à un groupe crée une solidarité, à forte cohésion interne, donnerait lieu à une coordination fluide et une efficacité accrue dans la mise en place de travaux de qualité.

Les résultats montrent que les groupes qui ont finalisé leurs projets et réaliser des projets de qualité se sont beaucoup entraidés et ont développé des solidarités qui ont renforcé leurs échanges et leurs actions. Les groupes qui ont une productivité élevée, ce sont les groupes qui travaillent dans un climat chaleureux où confiance et cohésion entre membres semblent acquises, qui ont bien gérer et collaboré et ont facilement gérer leurs conflits et communiqué dès le début.

L'appartenance au groupe semble renforcer l'apprenant qui voit son propre succès dans celui de son groupe. Dans certains groupes, par contre, c'est une seule personne qui est derrière la motivation de tous les autres membres « leader du groupe ». C'est lui qui propose et qui gère son groupe par l'envoi des messages d'encouragement et qui les pousse à s'impliquer dans l'exécution des tâches qui leur seront confiées.

Les participants à des forums de discussion apparaissent souvent à reproduire les comportements dégagés par les apprenants qui participent à une discussion en face-à-face

[Meyer, 2003]. Les apprenants qui sont confiants et extravertis sont contents de contribuer à des discussions actives tandis que les étudiants timides et introvertis sont réticents à communiquer et le faire avec parcimonie. Une des différences les plus significatives entre les discussions en présentiel (face à face) et à distance par un forum de discussion est l'absence d'indices de langage du corps et des sentiments qui indiquent la réaction de l'apprenant aux observations formulées par les autres apprenants. Ces indices visuels ont souvent l'effet de favoriser ou d'étouffer la participation des apprenants actifs. Quand un apprenant reçoit une réponse positive dans la forme d'indice positif de langage corporel, l'apprenant est plus susceptible de s'impliquer plus activement et de profiter des avantages de l'apprentissage que la participation active offre. De la même façon une réponse négative permettra de réduire la probabilité de participation.

Nonnecke et Preece [Nonneckr & Preece, 2001] décrivent les résultats d'une étude menée pour aider à identifier la raison pour laquelle de nombreux apprenants choisissent de ne pas participer activement aux forums de discussion. Les raisons invoquées pour faible niveau de contribution comprennent la timidité, l'anxiété, de faibles compétences en écriture, le manque de confiance dans le sujet discuté, le manque de temps, la surcharge d'information et le manque de motivation. Globalement, il existe trois niveaux de participation à des forums de discussion asynchrone :

(1) "Lurkers" qui lisent simplement les messages et ne participe pas. Ils peuvent apprendre en lisant les messages et en incorporant les idées dans leurs missions [Guzdial & Carroll, 2002].
(2) Les étudiants qui traitent le forum comme un tableau d'affichage, affichant leur propre position et ayant une interaction limitée.
(3) La participation est interactive et utilisé à son plein potentiel [Ho, 2002].

Une gestion prudente de forums de discussion peut aider à réduire un grand nombre des facteurs identifiés par Nonnecke et Preece [Nonnecke & Preece, 2001] comme contribuant à la contribution de discussion pauvres. Ils suggèrent que les participants timides sont plus susceptibles de contribuer s'ils peuvent rester anonymes. Bien qu'il y ait eu beaucoup de débats sur la valeur de l'attribution de grades pour la participation, la preuve suggère que motiver la participation des apprenants par des grades augmente de manière significative la contribution et l'activité d'apprenant [Campbell, 2002]. De même, l'intervention d'un tuteur, peut aider à désamorcer un débat houleux en cas de besoin et encourager la discussion entre les participants lorsque l'activité diminue [Campbell, 2002].

Une revue de la littérature suggère que des forums de discussion font beaucoup pour améliorer l'expérience des apprenants participant au e-learning. Des études ont été menées pour évaluer le niveau de participation et de mesurer la qualité de la discussion entre les apprenants [Hiltz, 1990 ; Mason 1991 ; Henri, 1992 ; Gunawardena, Lowe et Anderson, 1997]. Ces études reposent sur le codage manuel des activités pour recueillir des données relatives à l'activité de l'apprenant, un processus long et souvent fastidieux.

En termes de participation à des forums de discussion, la qualité de la participation est plus importante que la quantité [Hiltz, 1990]. Une des méthodes les plus généralement acceptées, de mesure la qualité des réponses et du processus d'apprentissage dans un forum de discussion par une analyse de contenu des conversations textuelles asynchrones, est soutenu par Hiltz [Hiltz, 1990] et Mason [Mason, 1991].

D'après la littérature qui est relative à l'utilisation des forums de discussion et leur valeur en termes de pédagogie de l'éducation, les chercheurs ont tenté d'enquêter sur comment les forums de discussion pourrait être amélioré pour automatiser la collecte de données relatives à l'activité de l'apprenant. Définir ce qu'il faut mesurer et comment déterminer les activités de collaboration au sein d'un forum de discussion s'est avéré être une tâche intéressante. Les données relatives à l'activité de l'apprenant en termes de temps passé en ligne, le nombre de messages lus, et le nombre et la qualité des sujets et messages créés.

- **Discussion critique**

L'étude que nous avons menée dans un forum de discussion libre avec les étudiants de 1er année master « qualité logiciel », peuvent être valorisée lors d'un travail collaboratif et peuvent être à l'origine de la construction et co-construction de connaissance. Les apprenants utilisent les forums pour demander de l'aide, partager, apprendre et expliquer en l'absence d'un tuteur qui les guide.

Nous avons remarqué que les apprenants travaillant via un forum de discussion, ils s'attribuent les mêmes rôles pour mener un travail collaboratif en ligne, s'organisent selon des affinités communes, collaborent et se comportent d'une manière similaire à une situation d'apprentissage classique. Les interactions dans les forums de discussion ne sont pas toutes idées à l'évolution au niveau des apprentissages. Elles sont conditionnées par une confiance en soi, par un bas niveau d'anxiété, mais aussi par un bon niveau de lecture et d'écriture. Ces derniers peuvent ne rien produire en l'absence d'une cohésion du groupe, les conflits entre les membres de l'équipe, les difficultés de communication et de perception du groupe sont autant

d'éléments qui affectent négativement le processus de discussion [Henri, 1992 ; Henri & Rigault, 1996].

6.6 Conclusion

L'apprentissage collaboratif dans des forums libre (non semi-structuré) a des avantages comme l'autonomie, esprit critique, flexibilité et convivialité sont autant de termes qui reviennent pour décrire l'apport des interactions pour la construction et la co-construction des connaissances.

Les apprenants qui ont participé à cette interaction en ligne, ont déclaré que pour avoir une meilleure productivité et un apprentissage de qualité, si le nombre de participants au forum ainsi que leur dispersion sont bien définis et si les membres du groupe ont une cohésion et motivation appréciables. Le tout étant conditionné par un suivi adéquat de la part d'un tuteur conscient des conditions de succès de l'apprentissage collaboratif et qui pourrait prendre en charge les apprenants isolés en leur dispensant des supports pouvant les aider à rejoindre le groupe et co-construire leurs connaissances. Les apprenants sont engagés dans un travail collaboratif à travers un forum de discussion éducatif non semi-structuré. En effet, nous avons montré que la communication en ligne via ce forum favorise la constitution des groupes, crée des liens sociaux et affectifs entre les apprenants et donne la possibilité de co-construire des connaissances.

Le volume d'intervention, type d'intervention et réaction entrainée peuvent être à l'origine d'une collaboration fructueuse pour des travaux de qualité. Le forum devient un lieu de collaboration où les interactions en ligne pourraient remplacer la présence de tuteur. Les conversations textuelles en ligne deviennent un outil didactique plus performant que la conversation face-à-face qui, en situation de classe classique, a la caractéristique d'être volatil. L'apprentissage collaboratif en ligne via des forums de discussion constitue l'apport pédagogique majeur de l'enseignement en ligne. Ce mode d'enseignement mais aussi d'apprentissage, considéré comme une innovation pédagogique.

Cette expérimentation nous a permis de valider notre système d'un point de vue ergonomique et fonctionnel, le tuteur a apprécié la qualité des résultats. L'évaluation de cette expérience a montré que les tuteurs, en utilisant le système, sur la base d'informations quantitatives et qualitatives, arrive à dériver une information qualitative sur les états des apprenants en corrélant les indicateurs entre eux.

Les résultats de ces indicateurs et l'évolution des profils dans le temps n'étaient pas à la disposition du tuteur durant l'expérimentation. Cependant, ayant vu a posteriori ces résultats, le tuteur estime que disposer de telles informations durant la formation, aurait certainement amélioré sa perception de l'organisation et des interactions dans les groupes et par conséquent améliorer la pertinence de leurs interventions auprès des apprenants.

Le tuteur a particulièrement apprécié les indicateurs sociaux retournés par le système qui leur ont permis de repérer aisément les apprenants isolés dans un groupe auprès desquels il a pu agir rapidement. L'indicateur « volume d'intervention », a permis au tuteur d'avoir une vue sur la fréquence des échanges par forum entre les membres de chaque groupe, et de suivre l'évolution des échanges par période. Cette vue permet d'identifier le rôle joué par les apprenants dans leur groupe respectifs, de repérer les apprenants entreprenants et ceux entrainés qui ne font que réagir aux actions des autres.

A partir de cette analyse nous avons eu recours au paradigme d'agent pour proposer et implémenter une architecture à base d'agents pour la modélisation du comportement de l'apprenant qui se trouve dans une situation d'apprentissage à distance, ce système a été analysé, conçu et spécifié suivant une méthodologie de développement de système multi-agents, qui nous énormément aidé à modéliser ce système grâce à la simplicité de ses concepts et utilisation.

L'ensemble de ces constats nous a permis de vérifier que le modèle basé sur la logique floue utilisé apporte une amélioration notable aux résultats de l'automatisation de détermination des comportements sociaux de l'apprenant. Ces résultats démontrent que l'approche proposée dans notre système est plus pertinente. Ils indiquent que le système peut effectuer l'évaluation « comme procède » un enseignant expert grâce à son module de détermination automatique des comportements sociaux.

On a pu apprécier, à travers cette étude, l'utilité de la notion de la logique floue dans l'évaluation des niveaux ssscomportementaux de l'apprenant. La théorie des sous-ensembles flous procure une méthode convenable pour incorporer la connaissance d'un enseignant expert en utilisant des termes qualitatifs et proches du raisonnement humain. Elle permet, ainsi de manipuler des informations imprécises et de modéliser des connaissances subjectives. En plus, l'utilisation des règles floues dans l'algorithme d'inférence du système fournit à l'utilisateur plus de souplesse et de facilité dans son processus jugement.

La remarque de Cerratto [Cerratto, 1999] qui signalait qu'à distance les échanges étaient souvent minimaux (initiatif-réactif), nous avons constaté beaucoup d'échanges complets

(initiatif-réactif-évaluatif) dans les conversations. Cet aspect évaluatif signale un engagement plus important des locuteurs.

7 Conclusion générale

Sommaire

7.1 Conclusions 173
7.2 Perspectives 176

7.1 Conclusions

Le travail présenté dans cette thèse propose un système d'analyse automatique des conversations textuelles asynchrones pour la détermination des comportements sociaux des apprenants dans le cadre d'un travail collaboratif à distance. L'analyse automatique des interactions est le bloc de construction de base pour le soutien des tuteurs qui jouent un rôle de tutorat à distance.

Plus spécifiquement, le système réalisé est basé sur le paradigme des Systèmes Multi-Agents. Un SMA réalise une séquence complète de traitements afin de déterminer les profils sociologiques de chaque apprenant. L'idée sous-jacente est de proposer un système qui consiste à mettre à la disposition du tuteur un système intelligent afin de l'aider à assurer les différentes tâches qui lui sont confiées. En particulier l'objectif est de repérer et soutenir les individus en difficulté pour leur éviter l'abandon.

L'idée d'avoir un système intelligent avec une architecture ouverte et évolutive nous a amené à considérer une approche orientée agent, des agents communiquent et collaborent entre eux afin de réaliser les tâches qui leurs sont confiées. Ainsi nous avons réalisé une étude sur la technologie multi-agents et particulièrement sur les méthodologies et les plates-formes de développement des SMA. Cette étude nous a permis de faire le choix de la méthodologie ASPECS que nous avons retenue pour la conception de notre système et la plateforme JANUS pour le développement de nos agents.

La mise en œuvre de ce type du système aborde souvent des problèmes de représentation et de manipulation d'informations imparfaites, ce qui implique la nécessité de fournir un SMA avec des techniques de représentation et de manipulation des connaissances, leur permettant de prendre en considération cette imperfection. L'approche que nous avons choisie pour remédier à ce problème est de fournir un SMA avec ces techniques. En fait, le but de notre démarche est de proposer un système d'analyse automatique des conversations textuelles à travers des outils de communication asynchrone (forum et mail). Un SMA basé sur des algorithmes d'analyse de texte, de classification Bayésienne et de logique floue, afin de répondre aux problèmes auxquels sont confrontés de nombreux systèmes complexes, qui doivent traiter des informations de nature imparfaite [Oumaira, 2011 ; George, 2006].

Cette procédure d'analyse se compose de plusieurs étapes : on commençant par une procédure complète d'analyse « full text » des échanges textuels. Cette analyse de procédure consiste en 3 étapes (Récupération, Filtrage et Lemmatisation) au bout desquels, nous définissons des actes de langages qui vont par la suite contribuer à la classification des

messages (type animateur, vérificateur, quêteur et indépendant). Par la suite nous avons appliqué une classification Bayésienne des conversations. L'objectif est de montrer comment les théories de Bayes peuvent être utilisées pour la classification de messages et donner des meilleurs résultats que la classification de messages en utilisant des actes de langage.

Nous avons vu qu'il était possible de déterminer des profils comportementaux en analysant « manuellement » les interactions et les conversations lors d'un travail collaboratif en ligne. Si l'on cherche à analyser automatiquement une conversation, il est alors important de repérer ces actes de langage directeurs. De nombreux systèmes existants supposent que les messages sont semi-structurés par des actes de langage. Cela signifie que les apprenants ont besoin de suivre un modèle et une syntaxe prédéfinie. Ces outils peuvent parfois entraver la communication pour la réalisation d'un travail collaboratif [George, 2006]. Par exemple, "le contenu du message est différent du sujet de discussion" [Oumaira, 2011 ; George, 2006], ce qui soulève des questions pour l'analyse des discussions. Dans notre approche, l'absence d'une normalisation des actes de langage pour déterminer les comportements sociaux des apprenants, rend la tâche d'analyse et de classification des conversations très difficile pour obtenir de bons résultats. Dans notre cas, les messages ne sont pas sous de telles contraintes. Le système analyse automatiquement les messages dans les forums libres (non structurés ni semi-structuré par des actes de langage) afin d'extraire l'information pertinente.

Comme il a déjà été précisé, c'est principalement par apprentissage automatique que nous avons résolu le problème de catégorisation automatique des messages dans le but de déterminer les comportements sociaux des apprenants en formation à distance. Dans cette optique, plusieurs algorithmes mis au point pour des problèmes quelconques en apprentissage automatique ont été adaptés et appliqués. Les résultats obtenus ont montré une nette amélioration des résultats en utilisant les réseaux Bayésiens comparativement à la classification par actes de langage.

Après une analyse et classification des conversations et à partir des profils définis et adaptés des travaux de Pléty [Pléty, 1998], nous calculons des indicateurs « volume d'intervention, type d'intervention et les réactions entraînées), qui couplés aux classifications de messages décrites ci-dessus, permettent d'affecter un profil sociologique à chaque apprenant. Le processus d'évaluation est entre autres influencé par imperfection semble due aux erreurs et aux approximations impliquées lors du recueil de l'information. Comme méthode de traitement de ces imperfections et incertitudes comme un expert humaine dans

l'évaluation de l'apprenant, nous utilisons une approche basée sur la théorie des ensembles flous.

Dans le cadre de la formation à distance où il n'y a aucune interaction entre le tuteur et l'apprenant, les données rassemblées tendent à être plus imparfaites que celles obtenues par l'interaction en présentiel. La présence d'informations imparfaites est un facteur important qui mène souvent aux erreurs dans la détermination des comportements sociaux de l'apprenant. Cette imperfection semble due aux erreurs et aux approximations impliquées lors du recueil de l'information. La théorie des sous-ensembles flous [Zadeh, 1965] se présente comme un outil privilégié pour la modélisation des situations présentant des imprécisions. Une des motivations principales de l'usage de la logique floue est la manipulation améliorée de l'imperfection de l'information. En effet, le raisonnement d'un système de logique floue est considéré comme « facile », du point de vue compréhension et/ou modification par des concepteurs et utilisateurs. Un des facteurs qui appuie cette considération est la similarité humaine. Elle peut fournir des descriptions de la connaissance comme un humain et imiter son modèle de raisonnement concernant les concepts vagues. Ceci est d'un intérêt particulier dans la conception d'un système modelant la connaissance interprétable de l'apprenant qui est basée sur le raisonnement et la conceptualisation de l'enseignant-expert.

Dans notre contribution, l'approche proposée est d'abord générique puisque nous avons constaté, dans la littérature, et contrairement aux travaux actuellement réalisés l'absence de la logique floue pour la conception de ce type des systèmes. Notre modèle de conception s'appuie sur l'idée d'intégrer la logique floue dans un SMA. Notre modèle va permettre d'associer un profil à chaque apprenant, pouvant bénéficier des avantages de l'approche multi-agent et des capacités de la logique floue, tel que sa capacité de représenter et de manipuler des connaissances imparfaites.

Nous avons présenté dans ce travail, le principe de fonctionnement de l'approche de décisions la plus utilisé avec les SMA, ainsi que leurs avantages et leurs limites notamment dans le contexte de représentation et de manipulation d'informations imparfaites. Nous avons vu que la logique floue est un moyen efficace pour créer des systèmes autonomes capable d'effectuer les tâches réalisées par les êtres humains.

Finalement pour tester les usages, les interfaces et le fonctionnement de notre environnement, nous avons mené une expérimentation dans le cadre du master « qualité du logiciel » de l'université Ibn Tofail. L'objectif de cette expérimentation était double ; nous souhaitions d'une part valider notre système d'un point de vue technique, fonctionnel et

ergonomique et d'autre part nous souhaitons comparer la perception que le tuteur avait des états des apprenants et des groupes par rapport aux résultats retournés par le système, sur la base d'informations quantitatives et qualitatives, arrive à dériver une information qualitative sur les états des apprenants et des groupes. Notre approche a été ensuite testée sur une situation réelle, qui a montré une parfaite concordance entre les résultats observés par des tuteurs humains et ceux déterminés automatiquement par notre système. Les résultats de l'analyse intuitive de tuteurs ont été considérés comme une référence. Nous avons pu constater que les résultats obtenus montrent que notre approche, en utilisant les actes de langage pour la classification de messages pour la détermination des comportement sociaux d'apprenant, a une marge d'erreur moyenne de 7,54% par rapport à l'enseignant expert. De plus, en utilisant le classificateur Bayésien, cette marge d'erreur a été réduite à 2, 54%.

7.2 Perspectives

Notre travail ne s'arrête pas là et des perspectives d'évolution de l'environnement, à court, moyen et long terme, s'offrent à nous.

A court terme, nous prévoyons plusieurs améliorations ; L'élaboration de modèles qui permettraient à partir de l'analyse des conversations et des interactions, de déclencher des mécanismes d'alerte en direction des tuteurs et/ou des apprenants est un point important de la suite de nos travaux de thèse. Une autre perspective de notre travail de recherche concerne des expérimentations en cherchant à observer l'évolution de profil d'apprenant pour chaque phase du projet en mettant en place un processus de développement en spirale. Ce genre d'expérimentation nécessite la mise en place d'une équipe pluridisciplinaire et un engagement de la part des tuteurs et des apprenants.

A moyen terme, nous envisageons de poursuivre le développement du SMA. L'objectif est d'enrichir notre système par l'analyse sémantique du texte. Une direction de recherche à la suite de notre travail implique une étude plus approfondie des algorithmes qui utilisent des mesures de similarité sémantique pour affecter les sens appropriés à des mots en fonction de leur contexte, à l'échelle d'un texte ou d'un corpus pour la classification automatique et la catégorisation des conversations textuelles, afin de déterminer le profil comportemental (sociologique) de chaque apprenant.

Le calcul de similarité sémantique entre les messages échangés au cours d'une session de communication, permet de renseigner sur le degré de cohérence dans un fils de discussion :

les messages qui s'écartent du sujet de discussion ; les apprenants susceptibles d'avoir dévié du sujet parent durant la discussion, etc.

A long terme, toujours dans le souci d'améliorer l'état de collaboration au sein d'une communauté d'apprentissage, il est important de compenser l'absence physique d'un interlocuteur en renforçant la prise de conscience de son état émotionnel (awareness) chez ses collaborateurs. Notre équipe travaille sur l'analyse des traits du visage lors de communications médiatisées dans un dispositif d'apprentissage à distance. Nous souhaitons intégrer ce système dans notre approche, afin d'intégrer la prise en compte de la gestualité dans la détermination des profils lors de rédaction de leurs messages.

8 Publications de l’auteur

Revues internationales

- ✓ Chaabi, Y., Messoussi, R., Hilaire, V., Ruichek, Y., Lekdioui, K., Touahni, R., "*Design of an Intelligent System to Support Tutors in Learning Communities Using Multi-Agent Systems and Fuzzy Logic*", (2015) International Review on Computers and Software (IRECOS), 10(8), pp. 845-855.
- ✓ Yahya Al-ashmoery , Rochdi Messoussi , Youness Chaabi, Raja Touahni, « *Monitoring And Visualizing Students Tracking Data Online Learning Activities (Tracking in E-learning Platforms) MVSA* », International association for engineering and management education (IAEME), ISSN 0976 – 6605, Volume 4 Issue 3 p:121 – 135 (September-December 2013).

Communications internationales avec actes et comité de lecture

- ✓ Chaabi.Youness, Rochdi.M, Vincent.H, Khadija.L, Yassine.R, Raja. T (2014) "*An automatic system for the determination of learner's sociological behavior from textual asynchronous conversations analysis in online collaborative learning*". IEEE Xplore Digital Library.
- ✓ Chaabi.Youness, Rochdi.M, Vincent.H, Khadija.L, Yassine.R, Raja. T (2014) "*Détermination des Comportements Sociaux dans un Contexte Collaboratif en Ligne, par Analyse Automatique de Conversations Textuelles Asynchrones D'apprenants*." In Proc. of SITA'14 - 9th International Conference on Intelligent Systems: Theories and Applications, 2014.
- ✓ Chaabi.Youness, Rochdi.M, Vincent.H, Khadija.L, Yassine.R, Raja. T (2014) "*Conception et Réalisation d'un Systéme Intelligent pour l'analyse automatique de conversation textuelles asynchrones d'apprenants pour la détermination de comportements sociaux*" In Proc. Of JDTIC'13 - 5éme édition des Journée Doctoral en Technique d'Information et de Communication, 2013.
- ✓ Khadija Lekdioui, Rochdi Messoussi, Youness Chaabi (2015) "*Etude et Modélisation des Comportements Sociaux d'Apprenants à Distance, à Travers l'Analyse des Traits du Visage*" Environnement Informatique pour l'Apprentissage Humain, Agadir, 2015.

9 Références

[Anderson, 2003] Anderson, T., Garrison, R., Archer, W., & Rourke, L. (2003). Critical thinking in a text based environment: Computer conferencing in higher education. Internet in Higher Education, 2(2), 87-105.

[Armentano, 2012] Armentano, M. G., & Amandi, A. A. (2012). Modeling sequences of user actions for statistical goal recognition. User Modeling and User-Adapted Interaction, 22(3), 281-311.

[Avouris, 2003] Avouris, N., Margaritis, M., Komis, V., Saez, A., & Meléndez, R. (2003). ModellingSpace: Interaction Design and Architecture of a collaborative modelling environment. Proc. of 6th CBLIS, 993-1004.

[Bannon, 1995] Bannon, L. J. (1995). Issues in computer supported collaborative learning. In Computer supported collaborative learning (pp. 267-281). Springer Berlin Heidelberg.

[Baron, 2003] Baron, G. L., & Bruillard, E. (2003, March). Forum et communautés d'enseignants. In Première conférence «Société de l'information. Saint-Rémy-lès-Chevreuse, France.

[Barros, 2002] Barros, B., Verdejo, M. F., Read, T., & Mizoguchi, R. (2002). Applications of a collaborative learning ontology. In MICAI 2002: Advances in artificial intelligence (pp. 301-310). Springer Berlin Heidelberg.

[Béchet, 2009] Béchet, N. (2009). Extraction et regroupement de descripteurs morpho-syntaxiques pour des processus de Fouille de Textes (Doctoral dissertation, Université Montpellier II-Sciences et Techniques du Languedoc).

[Bellot, 2001] Bellot, P., & El-Beze, M. (2001). Traitement automatique du langage naturel-Text classification and segmentation by means of decision trees for information retrieval. TSI-Technique et Science Informatiques-RAIRO, 20(3), 397-424.

[Bernatchez, 2003] Bernatchez, P. A. (2003). Vers une nouvelle typologie des activités d'encadrement et du rôle des tuteurs. DistanceS, 6(1), 5-25.

[Berner, 2003] Berner, R. T. (2003). The benefits of discussion board discussion in a literature of journalism course. The Technology Source.

[Biuk-Aghai, 2001] Biuk-Aghai, R. P., & Simoff, S. J. (2001, September). An integrative framework for knowledge extraction in collaborative virtual environments. In Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work (pp. 61-70). ACM.

[Bond, 2014] Bond, A. H., & Gasser, L. (Eds.). (2014). Readings in distributed artificial intelligence. Morgan Kaufmann.

[Bonk, 2003] Bonk, C., & Dennen, V. (2003). The 3 T's of online assessment: Technology, tools, and (saving) time. In Workshop given at the 19th Annual Conference on Distance Teaching & Learning.

[Bouchon-Meunier , 1995] Bouchon-Meunier, B. (1995). La logique floue et ses applications. Addison-Wesley France.

[Bouron, 1992] Bouron, M. T. (1992). Structures de communication et d'organisation pour la coopération dans un univers multi-agents (Doctoral dissertation, UNIVERSITE PARIS 6).

[Bradshaw, 2004] Bradshaw, J., & Hinton, L. (2004). Benefits of an online discussion list in a traditional distance education course. Turkish Online Journal of Distance Education, 5(3).

[Branon, 2001] Branon, R. F., & Essex, C. (2001). Synchronous and asynchronous communication tools in distance education. TechTrends, 45(1), 36-42.

[Branovic, 2002] Branovic, I., Giorgi, R., & Prete, A. (2002, May). Web-based training on computer architecture: the case for JCachesim. In Proceedings of the 2002 workshop on Computer architecture education: Held in conjunction with the 29th International Symposium on Computer Architecture (p. 10). ACM.

[Bratitsis, 2005] Bratitsis, T., & Dimitrakopoulou, A. (2005). Data Recording and Usage Interaction Analysis in Asynchronous Discussions: The DIAS System. In 12th International Conference on Artificial Intelligence in Education AIED.

[Briot, 2002] Briot, J. P., & Demazeau, Y. (2002). principles and architecture of MAS. Hermès: Paris, 235-266.

[Briot, 2006] Briot, J. P., Guessoum, Z., Aknine, S., Almeida, A. L., Malenfant, J., Marin, O., ... & Lucena, C. (2006, May). Experience and prospects for various control strategies for self-replicating multi-agent systems. In Proceedings of the 2006 international workshop on Self-adaptation and self-managing systems (pp. 37-43). ACM.

[Brooks, 1986] Brooks, R. A. (1986). A robust layered control system for a mobile robot. Robotics and Automation, IEEE Journal of, 2(1), 14-23.

[Calvani, 2010] Calvani, A., Fini, A., Molino, M., & Ranieri, M. (2010). Visualizing and monitoring effective interactions in online collaborative groups. British Journal of Educational Technology, 41(2), 213-226.

[Buisson, 2013] Buisson, J., Galland, S., Gaud, N., Gonçalves, M., & Koukam, A. (2013). Real-time collision avoidance for pedestrian and bicyclist simulation: a smooth and predictive approach. Procedia Computer Science, 19, 815-820.

[Burkett, 2004] Burkett, R. S., Leard, C., & Spector, B. S. (2004). Using an electronic bulletin board in science teacher education: Issues and trade-offs. The Journal of Interactive Online Learning, 3(1), 1-9.

[Campbell, 2002] Campbell, G. (2002). WR Johnson, Lucretius and the Modern World. London: Duckworth, 2000. Pp. x+ 163. ISBN 0-7156-2882-8.£ 9-99. Journal of Roman Studies, 92, 232-232.

[Caropreso, 2000] Caropreso, M. F., Matwin, S., & Sebastiani, F. (2000). Statistical phrases in automated text categorization. Centre National de la Recherche Scientifique, Paris, France.

[Cerratto, 1999] Cerratto, T. I. (1999). Activité collaborative sur réseau: une approche instrumentale de l'écriture en collaboration (Doctoral dissertation). Université Paris 8.

[Chaabi, 2014] Chaabi, Y., Messoussi, R., Hilaire, V., Lekdioui, K., Ruichek, Y., & Touahni, R. (2014, May). An automatic system for the determination of learner's sociological behavior from textual asynchronous conversations analysis in online collaborative learning. In Intelligent Systems: Theories and Applications (SITA-14), 2014 9th International Conference on (pp. 1-7). IEEE.

[Chaabi, 2015] Chaabi, Y., Messoussi, R., Hilaire, V., Ruichek, Y., Lekdioui, K., & Touahni, R. (2015). Design of an Intelligent System to Support Tutors in Learning Communities Using Multi-Agent Systems and Fuzzy Logic. International Review on Computers and Software (IRECOS), 10(8), 845-855.

[Chaib-Draa, 1992] Chaib-Draa, B., Moulin, B., Mandiau, R., & Millot, P. (1992). Trends in distributed artificial intelligence. Artificial Intelligence Review, 6(1), 35-66.

[Champin, 2004] Champin, P. A., Prié, Y., & Mille, A. (2004). MUSETTE: a framework for Knowledge from Experience. In Extraction et gestion des connaissances (EGC'2004)(article court) (pp. 129-134).

[Charaudeau, 2002] Charaudeau, P., & Maingueneau, D. (2002). Dictionnaire d'analyse du discours. Seuil.

[Chen, 2004] Chen, J., Sun, L., Zaiane, O. R., & Goebel, R. (2004, June). Visualizing and discovering web navigational patterns. In Proceedings of the 7th International Workshop on the Web and Databases: colocated with ACM SIGMOD/PODS 2004 (pp. 13-18). ACM.

[Choquet, 2007] Choquet, C., & Iksal, S. (2007). Modeling Tracks for the Model Driven Re-engineering of a TEL System. Journal of Interactive Learning Research, 18(2), 161.

[Clech, 2004] Clech, J. (2004). Contribution méthodologique à la fouille de données complexes (Doctoral dissertation, Lyon 2).

[Cogne, 1998] Cogne, A., David, J. P., & Lacombe, C. (1998). Production d'exercices hypermédias et mise en œuvre pédagogique. In Actes de la conférence Autoformation éducative et enseignement sur mesure, Union des Physiciens et INRP, Montpellier, France.

[Conry, 1991] Conry, S. E., Kuwabara, K., Lesser, V. R., & Meyer, R. A. (1991). Multistage negotiation for distributed constraint satisfaction. Systems, Man and Cybernetics, IEEE Transactions on, 21(6), 1462-1477.

[Cortes, 1995] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

[Cossentino, 2007] Cossentino, M., Gaud, N., Galland, S., Hilaire, V., & Koukam, A. (2007). A holonic metamodel for agent-oriented analysis and design. In Holonic and multi-agent systems for manufacturing (pp. 237-246). Springer Berlin Heidelberg.

[Cox, 1994] Cox, Earl (1994). The fuzzy systems handbook: a practitioner's guide to building, using, maintaining fuzzy systems. Boston: AP Professional. ISBN 0-12-194270-8.

[Cram, 2007] Cram, D., Jouvin, D., & Mille, A. (2007). Visualisation interactive de traces et réflexivité : application à l'EIAH collaboratif synchrone eMédiathèque. Revue Sticef. org, 21(07).

[Damerau, 1964] Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. Communications of the ACM, 7(3), 171-176.

[Dastani, 2005] Dastani, M., & Gomez-Sanz, J. J. (2005). Programming multi-agent systems. The Knowledge Engineering Review, 20(02), 151-164.

[Davis, 1983] Davis, R., & Smith, R. G. (1983). Negotiation as a metaphor for distributed problem solving. Artificial intelligence, 20(1), 63-109.

[De Corte, 1996] De Corte, E. (1996). Changing views of computer-supported learning environments for the acquisition of knowledge and thinking skills. International perspectives on the design of technology-supported learning environments, 129-145.

[De Lièvre, 2006] De Lièvre, B., Depover, C., & Acierno, M. (2006, July). Analyse du soutien fourni aux apprenants par les tuteurs à l'aide d'outils synchrones et asynchrones. In Premières journées communication et apprentissage instrumentés en réseau (pp. 76-99).

[De Lievre, 2011] De Lievre, B., Peraya, D., Quintin, J. J., & Jaillet, A. (2011). Le tutorat en formation à distance. Perspectives en éducation et formation. Bruxelles: Ed. De Boeck.

[Delozanne, 2007] Delozanne, É., Le Calvez, F., Merceron, A., & Labat, J. M. (2007). A structured set of design patterns for learners' assessment. Journal of Interactive Learning Research, 18(2), 309.

[Demazeau, 1991] Demazeau, Y., & Müller, J. P. (Eds.). (1991). Decentralized AI, 2. Elsevier.

[Deneubourg, 1991] Deneubourg, J. L., Goss, S., Franks, N., Sendova-Franks, A., Detrain, C., & Chrétien, L. (1991, February). The dynamics of collective sorting robot-like ants and ant-like robots. In Proceedings of the first international conference on simulation of adaptive behavior on From animals to animats (pp. 356-363).

[Desai, 2008] Desai, D., & Sahu, S. (2008). CRM Change management in an emerging country context: An exploratory study in India. Global Journal of Flexible Systems Management, 9(2/3), 41.

[Desjardins, 2002] Desjardins C. (2002). Vers une méthodologie d'analyse des construits cognitifs collaboratifs produits dans les forums électroniques textuels asynchrones. Thèse en communication, Université de Montréal. 340 pages, 2002.

[Dess, 1984] Dess, G. G., & Davis, P. S. (1984). Porter's (1980) generic strategies as determinants of strategic group membership and organizational performance. Academy of Management journal, 27(3), 467-488.

[Diagne, 2009] Diagne, F. (2009). Instrumentation de la supervision par la réutilisation d'indicateurs: Modèles et Architecture (Doctoral dissertation, Université Joseph-Fourier-Grenoble I).

[Dillenbourg, 1999] Dillenbourg, P. (1999). Collaborative Learning: Cognitive and Computational Approaches. Advances in Learning and Instruction Series. Elsevier Science, Inc., PO Box 945, Madison Square Station, New York, NY 10160-0757.

[Dimitracopoulou, 2006] Dimitracopoulou, A., & Bruillard, E. (2006). Enrichir les interfaces de forums par la visualisation d'analyses automatiques des interactions et du contenu.Revue des Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (STICEF), 13, 40-pages.

[Dimitracopoulou, 2004] Dimitracopoulou, A., Hoppe, U., & Dillenbourg, P. (2004, October). Interaction analysis supporting participants during technology based collaborative activities. In CSCL symposium, Kaleidoscope Noe (pp. 7-9).

[Dimitrova, 2002] Dimitrova, V., Brna, P., & Self, J. (2002, June). The design and implementation of a graphical communication medium for interactive open learner modelling. In Intelligent Tutoring Systems (pp. 432-441). Springer Berlin Heidelberg.

[Domingos, 1997] Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. Machine learning, 29(2-3), 103-130.

[Dreyfus, 2002] Dreyfus, G., Martinez, J. M., Samuelides, M., Gordon, M. B., Badran, F., Thiria, S., & Hérault, L. (2002). Réseaux de neurones-Méthodologie et applications. Réseaux de neurones-Méthodologie et applications.

[Dringus, 2005] Dringus, L. P., & Ellis, T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. Computers & Education, 45(1), 141-160.

[Drogoul, 1993] Drogoul, A. (1993). De la simulation multi-agent à la résolution collective de problèmes (Doctoral dissertation, Université Paris VI).

[Dumais, 1991] Dumais, S. T. (1991). Improving the retrieval of information from external sources. Behavior Research Methods, Instruments, & Computers, 23(2), 229-236.

[Durfee, 1989] Durfee, E. H., & Lesser, V. R. (1989). Negotiating task decomposition and allocation using partial global planning. Distributed artificial intelligence, 2(1), 229-244.

[Dziczkowski, 2008] Dziczkowski, G. (2008). Analyse des sentiments: systeme autonome d'exploration des opinions exprimées dans les critiques cinématographiques (Doctoral dissertation, Thèse de doctorat, Ecole Nationale Supérieurres).

[Erlin, 2009] Erlin, E., Yusof, N., & Rahman, A. A. (2009, April). Students' interactions in online asynchronous discussion forum: A Social Network Analysis. In Education Technology and Computer, 2009. ICETC'09. International Conference on (pp. 25-29). IEEE.

[Eyssautier, 2004] Eyssautier, C., & Jean-Daubias, S. (2004). A device helping learners to self-assess themselves. In International Conference on Computer Aided Learning in Engineering education (CALIE04) (pp. 6-pages).

[Ferber, 1992] Ferber, J., & Drogoul, A. (1992). Using reactive multi-agent systems in simulation and problem solving. Distributed artificial intelligence: Theory and praxis, 5, 53-80.

[Ferber, 1995] Ferber, J., & Perrot, J. F. (1995). Les systèmes multi-agents: vers une intelligence collective. InterEditions.

[Fiore, 2002] Fiore, A. T., Tiernan, S. L., & Smith, M. A. (2002, April). Observed behavior and perceived value of authors in usenet newsgroups: bridging the gap. InProceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 323-330). ACM.

[Furnkranz, 1998] Furnkranz, J., Mitchell, T., & Riloff, E. (1998). A case study in using linguistic phrases for text categorization on the WWW. In Working Notes of the AAAI/ICML, Workshop on Learning for Text Categorization (pp. 5-12).

[George, 2003] George, S. (2003). Forum contextuel structuré: une étude pour le téléenseignement. In Actes de la 15ème Conférence Francophone sur l'Interaction Homme-Machine, Caen, France,(International Conference Proceedings Series, ACM) (pp. 104-111).

[George, 2004] George, S. (2004). Analyse automatique de conversations textuelles synchrones d'apprenants pour la détermination de comportements sociaux. Revue Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF), 10, p-165.

[George, 2005] George, S. (2005, January). Vers des forums dédiés au contexte éducatif. InSymposium, formation et nouveaux instruments de communication, Amiens..

[Georgeon, 2006] Georgeon, O., Mille, A., & Bellet, T. (2006). Abstract: un outil et une méthodologie pour analyser une activité humaine médiée par un artefact technique complexe. Ingéniérie des Connaissances. Semaine de la connaissance. Nantes.

[Gerosa, 2005] Gerosa, M. A., Pimentel, M. G., Fuks, H., & Lucena, C. J. (2005, May). No need to read messages right now: helping mediators to steer educational forums using statistical and visual information. In Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years! (pp. 160-169). International Society of the Learning Sciences.

[Getoor, 1999] Getoor, L., & Sahami, M. (1999, August). Using probabilistic relational models for collaborative filtering. In Workshop on Web Usage Analysis and User Profiling (WEBKDD'99).

[Gilly, 1988] Gilly, M. (1988). Interaction entre pairs et constructions cognitives: modèles explicatifs. In A.-N. Perret-Clermont & M. Nicolet (Ed.), Interagir et connaître : Enjeux et régulations sociales dans le développement cognitif (pp. 19-28). Cousset (Fribourg) : DelVal.

[Gomez, 1998] Gomez, L. M., Fishman, J. B., & Pea, D. R. (1998). The CoVis Project: Building a Large-Scale Science Education Testbed*. Interactive Learning Environments, 6(1-2), 59-92.

[Gotab, 2009] Gotab, P., Béchet, F., & Damnati, G. (2009, November). Active learning for rule-based and corpus-based Spoken Language Understanding models. In Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on (pp. 444-449). IEEE.

[Gouaïch, 2004] Gouaïch, A., Michel, F., & Guiraud, Y. (2004). MIC*: a deployment environment for autonomous agents. In Environments for Multi-Agent Systems (pp. 109-126). Springer Berlin Heidelberg.

[Grimsey, 2004] Grimsey, D., & Lewis, M. K. (2004). Public Private Partnerships: The Worldwide Revolution of. Infrastructure Provision and Project Finance.

[Guessoum, 1996] Guessoum, Z. (1996). Un environnement opérationnel de conception et de réalisation de SMA (Doctoral dissertation, thèse LAFORIA, Paris VI).

[Guessoum, 2005] Guessoum, Z., Faci, N., & Briot, J. P. (2005). Adaptive replication of large-scale multi-agent systems–towards a fault-tolerant multi-agent platform. In Software Engineering for Multi-Agent Systems IV (pp. 238-253). Springer Berlin Heidelberg.

[Gunawardena, 1997] Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. Journal of educational computing research, 17(4), 397-431.

[Guzdial, 2002] Guzdial, M., & Carroll, K. (2002, January). Exploring the lack of dialogue in computer-supported collaborative learning. In Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community (pp. 418-424). International Society of the Learning Sciences.

[Gwenegan, 2005] Gwenegan, R. (2005). Structuration et analyse de traces hybrides issues de situation d'apprentissage. Rapport de Master, Laboratoire CLIPS-IMAG.

[Hartley, 2002] Hartley, D., & Mitrovic, A. (2002, June). Supporting learning by opening the student model. In Intelligent Tutoring Systems (pp. 453-462). Springer Berlin Heidelberg.

[Heller, 2004] Heller, J., Levene, M., Keenoy, K., Hockemeyer, C., & Albert, D. (2004). An e-Learning Perspective of Cognitive and Pedagogical Aspects of Trails. Kaleidoscope project deliverable, 22(1).

[Henri, 1992] Henri, F. (1992). Computer conferencing and content analysis (pp. 117-136). Springer Berlin Heidelberg.

[Henri, 2005] Henri, F., & Charlier, B. (2005, January). L'analyse des forums de discussion: pour sortir de l'impasse. In Symposium Symfonic.

[Henri, 1996] Henri, F., & Rigault, C. R. (1996). Collaborative distance learning and computer conferencing. In Advanced educational technology: Research issues and future potential (pp. 45-76). Springer Berlin Heidelberg.

[Heraud, 2004] Heraud, J. M., France, L., & Mille, A. (2004, September). Pixed: An ITS that guides students with the help of learners' interaction log. In International conference on intelligent tutoring systems, workshop analyzing student tutor interaction logs to improve educational outcomes. Maceio, Brazil (pp. 57-64).

[Herring, 1999] Herring, S. (1999). Interactional coherence in CMC. Journal of Computer Mediated Communication, 4(4), 0-0, Special issue on Persistent Conversation, Thomas Erickson (ed.).

[Hilaire, 2000] Hilaire, V. (2000). Vers une approche de spécification, de prototypage et de vérification de Systemes Multi-Agents (Doctoral dissertation).

[Hilali, 2009] Hilali, H. (2009). Application de la classification textuelle pour l'extraction des règles d'association maximales. Université du Québec à Trois-Rivières.

[Hiltz, 1990] Hiltz, S. R. (1990). Evaluating the virtual classroom. Online education: Perspectives on a new environment, 133-183.

[Ho, 2002] Ho, S. (2002). Evaluating students' participation in on-line discussions. In The Eighth Australian World Wide Web Conference (AusWeb 2002). Sunshine Coast, Queensland.

[Huang, 2003] Huang, J., & Morgan, G. (2003). A functional approach to evaluating content knowledge and language development in ESL students' science classification texts. International Journal of Applied Linguistics, 13(2), 234-262.

[Huberman, 1991] Huberman, A. M., Miles, M. B., & De Backer, C. (1991). Analyse des données qualitatives: recueil de nouvelles méthodes (pp. 21-24). Bruxelles: De Boeck Université.

[Hung, 2007] Hung, H., Jayagopi, D. B., Yeo, C., Friedland, G., Ba, S. O., Odobez, J. M., ... & Gatica-Perez, D. (2007). Using audio and video features to classify the most dominant person in a group meeting. In "" (No. LIDIAP-CONF-2007-016).

[Ignat, 2006] Ignat, C., & Rousselot, F. (2006). Representation de textes a l'aide de la d'etiquettes semantiques dans le cadre dela classification automatique. Revue roumaine de linguistique, 51(3-4), 421-439.

[Jaillet, 2004] Jaillet, S. (2004). Catégorisation automatique de documents. Actes de la douzième session des journées des doctorants DOCTISS'04. LIRMM UMR, Montpellier – France.

[Jalam, 2003] Jalam, R. (2003). Apprentissage automatique et catégorisation de textes multilingues (Doctoral dissertation, Université Lumière-Lyon 2).

[Jean-Daubias, 2005] Jean-Daubias, S., & Eyssautier, C. (2005). An environment helping teachers to track students' competencies. In Workshop LeMore (Learner Modelling for Reflection, to Support Learner Control, Metacognition and Improved Communication between Teachers and Learners) at AIED 2005 (12th International Conference on Artificial Intelligence in Education) (pp. 9-pages).

[Jégou, 2010] Jégou, H., Douze, M., & Schmid, C. (2010). Représentation compacte des sacs de mots pour l'indexation d'images. In RFIA 2010-Reconnaissance des Formes et Intelligence Artificielle.

[Jennings, 1998] Jennings, N. R., Sycara, K., & Wooldridge, M. (1998). A roadmap of agent research and development. Autonomous agents and multi-agent systems, 1(1), 7-38.

[Jermann, 2004] Jermann, P. R. (2004). Computer support for interaction regulation in collaborative problem-solving (Doctoral dissertation, Université de Genève).

[Kay, 2006] Kay, R. (2006). Using asynchronous online discussion to learn introductory programming: An exploratory analysis. Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie, 32(1).

[Kay, 2006] Kay, R. H. (2006). Developing a comprehensive metric for assessing discussion board effectiveness. British Journal of Educational Technology,37(5), 761-783.

[Kazmierska, 2008] Kazmierska, J., & Malicki, J. (2008). Application of the Naïve Bayesian Classifier to optimize treatment decisions. Radiotherapy and Oncology, 86(2), 211-216.

[Keenoy, 2004] Keenoy, K., de Freitas, S., Levene, M., Jones, A., Brasher, A., Waycott, J., ... & Montandon, L. (2004). Personalised trails and learner profiling within e-Learning environments. Personalised and Collaborative Trails of Digital and Non-Digital Learning Objects, 22(4.1), 1-43.

[Klaas, 2005] Klaas, M. (2005). Toward indicative discussion fora summarization. UBC CS TR-2005, 4.

[Kligyte, 2001] Kligyte, G. (2001). I think I know what is good for you: User interface design for a CSCL sytem (Doctoral dissertation, Thesis from the Master of Arts at the Media Lab, the University of Art and Design Helsinki).

[Klisc, 2009] Klisc, C., McGill, T. J., & Hobbs, V. J. (2009). The effect of assessment on the outcomes of asynchronous online discussion as perceived by instructors. Australasian Journal of Educational Technology, 25(5), 666-682.

[Kolp, 2006] Kolp, M., Giorgini, P., & Mylopoulos, J. (2006). Multi-agent architectures as organizational structures. Autonomous Agents and Multi-Agent Systems, 13(1), 3-25.

[Kosba, 2005] Kosba, E., Dimitrova, V., & Boyle, R. (2005). Using student and group models to support teachers in web-based distance education. In User Modeling 2005 (pp. 124-133). Springer Berlin Heidelberg.

[Kukich, 1992] Kukich, K. (1992). Techniques for automatically correcting words in text. ACM Computing Surveys (CSUR), 24(4), 377-439.

[Zadeh, 1968] L.A. Zadeh. (1968)."Fuzzy algorithms," Info. & Ctl., Vol. 12, 1968, pp. 94-102.

[Lance, 1967] Lance, G. N., & Williams, W. T. (1967). A general theory of classificatory sorting strategies II. Clustering systems. The computer journal, 10(3), 271-277.

[Lauirillard, 2002] Lauirillard, D. (2002). Rethinking university teaching: A framework for the effective use of learning technologies. (2nd Ed.). London and New York: Routledge.

[Leh, 2002] Leh, A. S. (2002). Action research on hybrid courses and their online communities. Educational Media International, 39(1), 31-38.

[Lemieux, 2013] Lemieux, F., Desmarais, M. C., & Robillard, P. N. (2013). Analyse chronologique des traces journalisées d'un guide d'étude pour apprentissage autonome. STICEF

[Levan, 2004] Levan, S. K., & Vickoff, J. P. (2004). Travail collaboratif sur Internet: concepts, méthodes et pratiques des plateaux projet. Vuibert.

[Levenshtein, 1966] Levenshtein, V. I. (1966, February). Binary codes capable of correcting deletions, insertions, and reversals. In Soviet physics doklady (Vol. 10, No. 8, pp. 707-710).

[Lewis, 1992] Lewis, D. D. (1992, June). An evaluation of phrasal and clustered representations on a text categorization task. In Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 37-50). ACM.

[Li, 1992] Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. Information Theory, IEEE Transactions on, 38(6), 1842-1845.

[Liu, 2007] Liu, S. (2007). Assessing online asynchronous discussion in online courses: An empirical study. In 12th Annual TCC Worldwide Online Conference (pp. 24-32).

[M.E.N._EducEval, 2008] M.E.N._EducEval, Ministère de l'éducation nationale : évaluations diagnostiques, http://educ-eval.education.fr/evaldiag.htm (2008).

[Martınez, 2003] Martınez, A., Dimitriadis, Y., Rubia, B., Gómez, E., & De la Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. Computers & Education, 41(4), 353-368.

[Mason, 1991] Mason, R. (1991). Methodologies for evaluating applications of computer conferencing. Collaborative Learning Through Computer Conferencing, Berlin: Springer Verlag.

[May, 2007] May, M., George, S., & Prévôt, P. (2007, March). Tracking, analyzing, and visualizing learners' activities on discussion forums. In International Conference on Web-based Education (WBE 2007) (pp. 649-656). IASTED, ACTA Press.

[Mazza, 2003] Mazza, R., & Dimitrova, V. (2003). CourseVis: Externalising student information to facilitate instructors in distance learning. In Proceedings of the International conference in Artificial Intelligence in Education (pp. 117-129). Sydney, Australia.

[McKlin, 2001] McKlin, T., Harmon, S. W., Evans, W., & Jones, M. G. (2001). Cognitive Presence in Web-Based Learning: A Content Analysis of Students' Online Discussions.

[Meena, 2009] Meena, M. J., & Chandran, K. R. (2009, December). Naive Bayes text classification with positive features selected by statistical method. In Advanced Computing, 2009. ICAC 2009. First International Conference on (pp. 28-33). IEEE.

[Meyer, 2003] Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychological bulletin, 129(5), 674.

[Michelet, 2007] Michelet, S., Adam, J. M., & Luengo, V. (2007). Adaptive learning scenarios for detection of misconceptions about electricity and remediation. International Journal of Emerging Technologies in Learning, 2(1), 2007.

[Miller, 1999] Miller, D. R., Leek, T., & Schwartz, R. M. (1999, August). A hidden Markov model information retrieval system. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 214-221). ACM.

[Mitton, 2009] Mitton, R. (2009). Ordering the suggestions of a spellchecker without using context. Natural Language Engineering, 15(02), 173-192.

[Mochizuki, 2005] Mochizuki, T., Kato, H., Yaegashi, K., Nagata, T., Nishimori, T., Hisamatsu, S. I., ... & Suzuki, M. (2005, May). Promotion of self-assessment for learners in online discussion using the visualization software. In Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years! (pp. 440-449). International Society of the Learning Sciences.

[Mohamadally, 2006] Mohamadally, H., & Boris, F. (2006). SVM: Machines à Vecteurs de Support ou Séparateurs à Vastes Marges. Report, January.

[Moulin, 1996] Moulin, B., & Chaib-Draa, B. (1996). An overview of distributed artificial intelligence. Foundations of distributed artificial intelligence, 1, 3-55, John Wiley & Sons, New-York.

[Müller, 1996] Müller, J. P. (1996). The design of intelligent agents: a layered approach (Vol. 1177). Springer Science & Business Media.

[Nonnecke, 2001] Nonnecke, B., & Preece, J. (2001). Why lurkers lurk. AMCIS 2001 Proceedings, 294.

[Oflazer, 1996] Oflazer, K. (1996). Error-tolerant finite-state recognition with applications to morphological analysis and spelling correction. Computational Linguistics, 22(1), 73-89.

[Oumaira, 2011] Oumaira.I, al (2011). Instrumentation des activités des tuteurs à l'aide d'un système multi-agents d'analyse automatique des interactions. A R I M A – Volume 14 pp. 125-147 – 2011.

[Oussalah, 2005] Oussalah, M. (2005). Ingénierie des composants: concepts, techniques et outils. Vuibert informatique.

[Pernin, 2005] Pernin, J. P. (2005). CSE, un modèle de traitement de traces. Research repport, CLIPS-IMAG, 65-75.

[Pessiot, 2004] Pessiot, J. F., Caillet, M., Amini, M. R., & Gallinari, P. (2004, February). Apprentissage non-supervisé pour la segmentation automatique de textes. In CORIA (pp. 213-228).

[Pléty, 1998] Pléty, R. (1998). Comment apprendre et se former en groupe. Retz pédagogie.

[Pollock, 1984] Pollock, J. J., & Zamora, A. (1984). Automatic spelling correction in scientific and scholarly text. Communications of the ACM, 27(4), 358-368.

[Pothin, 2007] Pothin, J. B., & Richard, C. (2007). Apprentissage de métrique appliqué à la classification de textes par méthodes à noyaux. In 21° Colloque GRETSI, Troyes, FRA, 11-14 septembre 2007. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images.

[Qiang, 2010] Qiang, G. (2010, May). An effective algorithm for improving the performance of naïve bayes for text classification. In 2010 Second International Conference on Computer Research and Development.

[O'Leary, 2007] R. O'Leary, (2007).Virtual Learning Environments.Adapted, with permission,from the LTSN starter guide 'Virtual Learning Environments'. (Date Accessed : 12 June 2007)

[Rao, 1991] Rao, A. S., & Georgeff, M. P. (1991). Modeling rational agents within a BDI-architecture. KR, 91, 473-484.

[Reffay, 2003] Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. In Designing for change in networked learning environments (pp. 343-352). Springer Netherlands.

[Reffay, 2006] Reffay, C., & Lancieri, L. (2006). Quand l'analyse quantitative fait parler les forums de discussion. Revue des Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (STICEF), 13, 24-pages.

[Reimann, 2003] Reimann, P. (2003). How to support groups in learning: More than problem solving. Artificial Intelligence in Education (AIED 2003). Supplementary Proceedings, 3-16.

[Reyes, 2003] Reyes, P., & Tchounikine, P. (2003). Supporting emergence of threaded learning conversations through augmenting interactional and sequential coherence. In Designing for Change in Networked Learning Environments (pp. 83-92). Springer Netherlands.

[Reyes, 2005] Reyes, P., & Tchounikine, P. (2005, May). Mining learning groups' activities in Forum-type tools. In Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years! (pp. 509-513). International Society of the Learning Sciences.

[Rish, 2001] Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46). IBM New York.

[Ritschard, 2009] Ritschard, G., Marcellin, S., & Zighed, D. A. (2009). Arbre de décision pour données déséquilibrées: sur la complémentarité de l'intentisé d'implication et de l'entropie décentrée.

[Rodet, 2003] Rodet, J. (2003). Le clavardage (chat), média de support à l'apprentissage?. Distances et savoirs, 1(3), 399-406.

[Rodriguez, 2005] Rodriguez, S., Hilaire, V., & Koukam, A. (2005). Formal specification of holonic multi-agent systems framework. In Computational Science–ICCS 2005 (pp. 719-726). Springer Berlin Heidelberg.

[Rodriguez, 2011] Rodriguez, S., Hilaire, V., Gaud, N., Galland, S., and Koukam, A. (2011). Holonic Multi-Agent Systems, chapter 11, pages 238–263. Self-Organising Software From Natural to Artificial Adaptation - Natural Computing. Springer, first edition.

[Rourke, 2001] Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. International Journal of Artificial Intelligence in Education (IJAIED), 12, 8-22.

[Rovai, 2002] Rovai, A. P. (2002). Building sense of community at a distance. The International Review of Research in Open and Distributed Learning, 3(1).

[Rozé, 2005] Rozé, G. (2005). Structuration et analyse de traces hybrides issues de situation d'apprentissage. Rapport de Master 2ème année. Grenoble, France. Université Joseph Fournier. Recuperado el, 9.

[Sahami, 1998] Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998, July). A Bayesian approach to filtering junk e-mail. In Learning for Text Categorization: Papers from the 1998 workshop (Vol. 62, pp. 98-105).

[Saldarriaga, 2010] Saldarriaga, S. P. (2010). Approches textuelles pour la catégorisation et la recherche de documents manuscrits en-ligne (Doctoral dissertation, Université de Nantes).

[Salton, 1971] Salton, G. (1971). The SMART retrieval system—experiments in automatic document processing. Chapter 14, Relevance Feedback in Information Retrieval, p. 313–323. Gerard Salton (editor), Prentice-Hall Inc. : New Jersey.

[Salton, 1986] Salton, G., & McGill, M. J. (1986). Introduction to modern information retrieval. New York: McGraw-Hill

[Savary, 2003] Savary, A., & Jacquemin, C. (2003). Reducing information variation in text. In Text-and Speech-Triggered Information Access (pp. 145-181). Springer Berlin Heidelberg.

[Scapin, 1986] Scapin, D. L. (1986). Guide ergonomique de conception des interfaces homme-machine (p. 92). Rocquencourt: INRIA.

[Scardamalia, 1994] Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. The journal of the learning sciences, 3(3), 265-283.

[Schmid, 1994] Schmid, H. (1994). Probabilistic pos tagging using decision trees. In Proceedings of International Conference on New methods in Language Processing.

[Schummer, 2005] Schummer, T., Strijbos, J. W., & Berkel, T. (2005). A new direction for log-files analysis in CSCL: Examples with a spatio-temporal metric. In Proceedings of the 2005 Conference on Computer Support for Collaborative Iearning, CSCL2005 (pp. 567-576).

[Searle, 1992] Searle, S. R., Casella, G., & McCulloch, C. E. (1992). Variance components John Wiley and Sons. New York, New York, USA.

[Sebastiani, 2002] Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47.

[Sébastien, 2006] Sébastien George, I. N. S. A. (2006). Bridging the Gap between Human Communications and Distance-Learning Activities. Cognitively Informed Systems: Utilizing Practical Approaches to Enrich Information Presentation and Transfer: Utilizing Practical Approaches to Enrich Information Presentation and Transfer, 102.

[Seidita, 2010] Seidita, V., Cossentino, M., Hilaire, V., Gaud, N., Galland, S., Koukam, A., and Gaglio, S. (2010). The metamodel : a starting point for design processes construction. International Journal of Software Engineering and Knowledge Engineering (IJSEKE), 20(4) :575–608

[Seo, 2007] Seo, K. K. (2007). Utilizing peer moderating in online discussions: Addressing the controversy between teacher moderation and nonmoderation. The American Journal of Distance Education, 21(1), 21-36.

[Settouti, 2006] Settouti, L. S., Prié, Y., Mille, A., & Marty, J. C. (2006). Systèmes à base de trace pour l'apprentissage humain. In Colloque international TICE.

[Shannon, 1948] Shannon, C. E. (1948). BA mathematical theory of communication, Bell System Tech journal.

[Siebra, 2005] Siebra, S., Salgado, A. C., Brézillon, P., & Tedesco, P. (2005). A learning interaction memory using contextual information. In Proceedings of the CONTEXT-05 Workshop on Context and Groupware, Paris, France.

[Silva, 2004] Silva, J. (2004). El rol moderador del tutor en la conferencia mediada por computador. Edutec: Revista Electrónica de Tecnología Educativa, 17.

[Soller, 2001] Soller, A., Monés, A. M., Jermann, P., & Muehlenbrock, M. (2001). From mirroring to guiding: a review of state of the art technology for supporting collaborative learning. In Proceedings of the First European Conference on Computer-Supported Collaborative Learning.

[Souque, 2000] SOUQUE, A. Approche critique des produits IdL: analyse comparative des correcteurs orthographiques de Word 2000 et OpenOffice 2.

[Sparck, 1972] Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of documentation, 28(1), 11-21.

[Stroppa, 2007] Stroppa, N., & Yvon, F. (2007). Formal models of analogical proportions.

[Suchman, 2015] Suchman, L. (2015). Technologies of accountability: Of lizards and airplanes.

[Sycara, 1989] Sycara, K. (1989). Multi-agent compromise via negotiation. Distributed artificial intelligence, 2(1), 119-139.

[Szyperski, 1997] Szyperski, C. (1997). Component Software: Beyond Object-Oriented Programming (Acm Press).

[Teixeira, 1998] Teixeira De Oliveira, R. F. (1998). Gestion des reseaux avec connaissance des besoins: utilisation des agents logiciel (Doctoral dissertation).

[Torzek, 2004] TORZEK, N. (2004). Contribution à l'étude des messages électroniques francophones Quelques résultats et leurs conséquences pour le TALN. Journée de l'ATALA.

[Traverso, 2003] Traverso, V. (2003). Les genres de l'oral: le cas de la conversation. In Les genres de l'oral. Apr 2003, Lyon, France. halshs-01001725

[Upton, 2009] Upton, K., & Kay, J. (2009). Narcissus: group and individual models to support small group work. In User modeling, adaptation, and personalization (pp. 54-65). Springer Berlin Heidelberg.

[Vassileva, 2005] Vassileva, J., Cheng, R., Sun, L., & Han, W. (2005). Designing mechanisms to stimulate contributions in collaborative systems for sharing course-related materials. Designing Computational Models of Collaborative Learning Interaction, Workshop at ITS 2004, pp. 59-64, Maceió, Brazil

[Vassileva, 2003] Vassileva, J., McCalla, G., & Greer, J. (2003). Multi-agent multi-user modeling in I-Help. User Modeling and User-Adapted Interaction, 13(1-2), 179-210.

[Veronis, 2004] Veronis, J., & Guimier de Neef, E. (2004). Le traitement automatique des nouvelles formes de communication écrite (e-mails, forums, chats, SMS, etc.). Journée de l'ATALA, Paris.

[Véronis, 2006] Véronis, J., & Guimier de Neef, E. (2006). Le traitement des nouvelles formes de communication écrite. Compréhension automatique des langues et interaction, 227-248.

[Villanova, 2002] Villanova-Oliver, M. (2002). Adaptabilité dans les systèmes d'information sur le Web: modélisation et mise en oeuvre de l'accès progressif (Doctoral dissertation, Institut National Polytechnique de Grenoble-INPG, Grenoble, France).

[Vinot, 2003] Vinot, R., Grabar, N., Valette, M., & STIM-DIAM, A. P. H. P. (2003). Application d'algorithmes de classification automatique pour la détection des contenus racistes sur l'internet. Actes de TALN, 2003, 257-284.

[Wiener, 1993] Wiener, E, Kanki. B, Helmreich. R, eds. (1993). Crew coordination and training in the advanced-technology cockpit. Crew Resource Management. (Pp 199-230). San Diego, Calif.: Academic Press.

[Wiener, 1995] Wiener, E. D. (1995). A neural network approach to topic spotting in text (Doctoral dissertation, University of Colorado).

[Wooldridge, 1994] Wooldridge, M., & Jennings, N. R. (1994). Towards a theory of cooperative problem solving. In Distributed Software Agents and Applications (pp. 40-53). Springer Berlin Heidelberg.

[Wooldridge, 1999] Wooldridge, M., & Weiss, G. (1999). Intelligent agents. Multiagent Systems. A modern approach to Distributed Artificial Inteligence, 27-77.

[Wu, 2013] Wu, Y. (2013). Conversation Analysis-A Discourse Approach to Teaching Oral English Skills. International Education Studies, 6(5), 87.

[Yang, 2002] Yang, S. P. (2002). Problem-based learning on the World Wide Web in an undergraduate kinesiology class: An integrative approach to education (pp. 1-182).

[Yang, 1999] Yang, Y., & Liu, X. (1999, August). A re-examination of text categorization methods. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 42-49). ACM.

[Yang, 2002] Yang, Y., Slattery, S., & Ghani, R. (2002). A study of approaches to hypertext categorization. Journal of Intelligent Information Systems, 18(2-3), 219-241.

[Zadeh, 1973] Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. Systems, Man and Cybernetics, IEEE Transactions on, (1), 28-44.

[Zadeh, 1984] Zadeh, L. A. (1984). Making computers think like people: the termfuzzy thinking is pejorative when applied to humans, but fuzzy logic is an asset to machines in applications from expert systems to process control. Spectrum, IEEE, 21(8), 26-32.

[Zadeh, 1996] Zadeh, L. A. (1996). Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh (Vol. 6). World Scientific.

[Zadeh, 1965] Zadeh, L. s(1965). Fuzzy sets. Information and control, 8(3), 338-353.

[Zaïane, 2001] Zaïane, O. R., & Luo, J. (2001). Towards evaluating learners' behaviour in a web-based distance learning environment. In Advanced Learning Technologies, 2001. Proceedings. IEEE International Conference on (pp. 357-360). IEEE.

[Zapata-Rivera, 2001] Zapata-Rivera, J. D., & Greer, J. (2001). Externalising learner modelling representations. In Proceedings of Workshop on External Representations of AIED: Multiple Forms and Multiple Roles (pp. 71-76).

Table des figures

Figure 2-1 : Diagramme d'une architecture BDI [Wooldridge, 1999] 52
Figure 2-2 : Agent réactif [Drogoul, 1993] 53
Figure 2-3 : Aperçu du processus ASPECS 59
Figure 2-4 : Diagramme de classes UML de Métamodèle CRIO relatif à la Plateforme Janus 0,5 [Cossentino, 2007] 61
Figure 2-5 : Diagramme UML simplifié de la plate-forme Janus et du domaine de la solution (PSM) du métamodèle CRIO 62
Figure 2-6 : L'architecture complète de la plate-forme Janus 64
Figure 3-1 : Architecture générale d'un réseau de neurones artificiels 82
Figure 3-2 : Classification des personnes par âge en trois sous-ensembles [Zadeh, 1984]. 85
Figure 3-3 : Structure d'un système flou [Zadeh, 1984]. 86
Figure 3-4 : Variables linguistiques (V,T(V),X) pour décrire la température [Zadeh, 1996]. 88
Figure 3-5 : La défuzzification par la méthode du centre de maximum [Zadeh, 1973, 1968] 91
Figure 3-6 : La défuzzification par la méthode du centre de gravité 91
Figure 4-1 : Exemple de messages pendant une discussion 97
Figure 5-1 : Architecture de système 113
Figure 5-2 : Architecture du Lemmatiseur 122
Figure 5-3 : Les étapes de classification 126
Figure 5-4 : Arborescence de discussion correspondent à un fil de discussion 137
Figure 5-5 : Système flou 140
Figure 5-6 : Fonctions d'appartenance Actes catégorie Animateur, Vérificateur, Quêteur et Indépendant sont définis de manière similaire 141
Figure 5-7 : Fonctions d'appartenance Réaction Directe et indirecte 142
Figure 5-8 : Fonctions d'appartenance Volume d'intervention 142
Figure 5-9 : Fonction d'appartenance Résultat 143
Figure 5-10 : Centres de gravité profil animateur, vérificateur, quêteur et indépendant 144
Figure 6-1 : Evolution du profil de l'apprenant (1) sur toute la durée du projet. 164
Figure 6-2 : Evolution du profil de l'apprenant (2) sur toute la durée du projet. 165

Liste des tableaux

Tableau 2-1 : Le choix des données à observer dépend de l'environnement d'interaction, et des objectifs de l'observation [Gwenegan, 2005]. 28
Tableau 2-2 : Classification des indicateurs 31
Tableau 2-3 : Les 20 dimensions des profils d'apprenants [Cogne et al., 1998; Eyssautier et al., 2004]. 47
Tableau 4.1 : Caractéristiques des profils comportementaux d'apprenants travaillant en groupe – Adaptation à partir des travaux de [Pléty, 1998] 100
Tableau 4.2 : Exemple des actes de langage proposés. 102
Tableau 5.1 : Exemple d'une liste des mots vides 117
Tableau 5.2 : Les étiquettes pour la langue française 118
Tableau 5.3 : Exemple des actes de langage proposés. 124
Tableau 6-1 : Résultat d'analyse intuitive pour le groupe 1 phase rédaction de charge des charges .. 155
Tableau 6-2 : Résultat d'analyse intuitive pour le groupe 1 phase analyse et conception 155
Tableau6-3 : Indicateurs calculés par le système 157
Tableau 6-4 : Indicateurs calculés par le système pour le groupe 1 phase rédaction cahier des charges 158
Tableau 6-5: Indicateurs calculés par le système pour le groupe 1 phase analyse et conception..... 158
Tableau 6-6 : Indicateurs calculés par le système pour le groupe 1 phase rédaction cahier des charges 159
Tableau 6-7 : Indicateurs calculés par le système pour le groupe 1 phase analyse et conception..... 159
Tableau 6-8 : illustrent les résultats de l'analyse intuitive de tuteur et de l'analyse automatique, pour les premières phases du projet. 161
Tableau 6-9 : Le calcul des écarts entre l'analyse intuitive des tuteurs et les résultats de système de la méthode de classification « Actes de langage » 162
Tableau 6-10 : Le calcul des écarts entre l'analyse intuitive des tuteurs et les résultats de système de la méthode de classification « Naive bayes » 163

Résumé :

La place importante de tutorat dans la réussite d'un dispositif de formation en ligne a ouvert un nouvel axe de recherche dans le domaine des EIAH (Environnements Informatiques pour l'Apprentissage Humain). Nos travaux se situent plus particulièrement dans le champ de recherches des ACAO. Dans un contexte collaboratif, le tutorat et les outils d'awareness constituent des solutions admises pour faire face à l'isolement qui très souvent, mène à l'abandon de l'apprenant.Ainsi, du fait des difficultés rencontré par le tuteur pour assurerun encadrement et un suivi appropriés à partir des traces de communication (en quantités conséquentes) laissées par les apprenants. Nous proposons une approche multi-agents pour analyser les conversations textuelles asynchrones entre apprenants, ces interactions étant révélatrices de comportements sociaux –animateur, indépendant, etc. –qu'il nous paraîtimportant de pouvoir repérer lors d'une pédagogie de projet, pour permettre aux apprenants de situer leurs travaux par rapport aux autres apprenants et de situer leur groupe par rapport aux autres groupes d'une part, et d'autre part permettre au tuteur d'accompagner les apprenants dans leur processus d'apprentissage, repérer et soutenir les individus en difficulté pour leur éviter l'abandon. Ces indicateurs seront déduits à partir des grands volumes d'échanges textuels entre apprenants. L'approche a été ensuite testée sur une situation réelle, qui a montré une parfaite concordance entre les résultats observés par des tuteurs humains et ceux déterminés automatiquement par notre système.

Mots-clés : Apprentissage collectif, Analyse de texte, acte langage, Système Multi-Agents, naïve bayésienne, profils sociaux, logique floue, CSCL, Pédagogie de projet

Abstract:

The growing importance of online training has put emphasis on the role of remote tutoring. A whole new area of research, dedicated to environment for human learning (EHL), is emerging. We are concerned with this field. More specifically, we will focus on the monitoring of learners. The instrumentation and observation of learners activities by exploiting interaction traces in the EHL and the development of indicators can help tutors to monitor activities of learners and support them in their collaborative learning process. Indeed, in a learning situation, the teacher needs to observe the behavior of learners in order to build an idea about their involvement, preferences and learning styles so that he can adapt the proposed activities. As part of the automatic analysis of collaborative learner's activities, we describe a multi agent approach for supporting learning activities in a Virtual Learning Environment context. In order to assist teachers who monitor learning processes, viewed as a specific type of collaboration, the proposed system estimates a behavioral (sociological) profile for each student. This estimation is based on automatic analysis of students' textual asynchronous conversations. The determined profiles are proposed to the teacher and may provide assistance to teacher during tutoring tasks. The system was experimented with students of the master "software quality" of the Ibn Tofail University. The results obtained show that the proposed approach is effective and gives satisfactory results.

Keywords: Collaborative Learning, Conversation Analysis, Fuzzy Logic, Multi-Agent System, Naïve Bayes, Social Profiles

SPIM

■ École doctorale SPIM - Université de Technologie Belfort-Montbéliard
F - 90010 Belfort Cedex ■ tél. +33 (0)3 84 58 31 39
■ ed-spim@univ-fcomte.fr ■ www.ed-spim.univ-fcomte.fr

utbm
université de technologie
Belfort-Montbéliard