Analyse discursive et multi-modale des conversations écrites en ligne portées sur la résolution de problèmes

Soufian Antoine Salim

UNIVERSITE BRETAGNE LOIRE

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Soufian Antoine SALIM

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université de Nantes
sous le sceau de l'Université Bretagne Loire

École doctorale : Sciences et technologies de l'information, et mathématiques

Discipline : Informatique, section CNU 27
Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N)

Soutenue le 22 novembre 2017

Analyse discursive et multi-modale des conversations écrites en ligne portées sur la résolution de problèmes

JURY

Présidente :	**M^me^ Béatrice DAILLE**, Professeure des universités, Université de Nantes
Rapporteurs :	**M. Frédéric BÉCHET**, Professeur des universités, Aix Marseille Université
	M. Olivier FERRET, Ingénieur de Recherche, Institut CEA LIST
Examinatrices :	**M^me^ Béatrice DAILLE**, Professeure des universités, Université de Nantes
	M^me^ Géraldine DAMNATI, Docteure Ingénieure, Orange Labs
Directeur de thèse :	**M. Emmanuel MORIN**, Professeur des universités, Université de Nantes
Co-encadrant de thèse :	**M. Nicolas HERNANDEZ**, Maître de conférences, Université de Nantes

Remerciements

Je remercie avant tout mes encadrants Nicolas Hernandez et Emmanuel Morin, pour m'avoir accueilli au laboratoire et pour m'avoir accompagné tout au long de mes travaux. Je les remercie pour leurs conseils, leur disponibilité et leur soutien précieux.

Je remercie également Olivier Ferret et Frédéric Béchet pour avoir accepté d'être rapporteurs pour cette thèse, ainsi que Béatrice Daille et Géraldine Damnati, pour avoir accepté de rejoindre mon jury en tant qu'examinatrices.

Je tiens également à remercier Nathalie Camelin, ainsi que, de nouveau, Olivier Ferret, qui ont ont constitué mon comité de suivi de thèse. Je les remercie d'avoir porté un regard extérieur sur mon travail, et pour leurs remarques stimulantes et encourageantes.

Je suis reconnaissant envers toute l'équipe TALN qui m'a accueilli en son sein ces trois dernières années, ainsi qu'envers tout les membres du LS2N, grâce à qui j'ai pu effectuer cette thèse dans un cadre agréable et convivial.

Enfin, je remercie ma compagne Anaïs, pour ses relectures attentives, mais surtout pour sa présence tout au long de ce travail.

Table des matières

Liste des figures 9

Liste des tableaux 11

Glossaire 13

1 Introduction 17
- 1.1 Contexte général . . . 17
- 1.2 Objectifs de recherche . . . 20
- 1.3 Projet ODISAE . . . 22
- 1.4 Plan de la thèse . . . 23

2 Actes de dialogue 25
- 2.1 Introduction . . . 27
- 2.2 Les actes de discours . . . 27
- 2.3 Des actes de discours aux actes de dialogue . . . 28
  - 2.3.1 Analyse des conversations fonctionnelles . . . 29
  - 2.3.2 Théorie des actes de la conversation . . . 29
  - 2.3.3 Contexte et connaissances communes . . . 30
- 2.4 Schémas d'annotation . . . 31
  - 2.4.1 Concepts pratiques et théoriques . . . 31
  - 2.4.2 DAMSL . . . 33
  - 2.4.3 DIT^{++} . . . 35
  - 2.4.4 ISO 24617-2 . . . 39
- 2.5 Conclusion . . . 41

3 Communication médiée par les réseaux 43
- 3.1 Introduction . . . 45
  - 3.1.1 Les réseaux : un handicap à la communication ? . . . 46
  - 3.1.2 Communication orale, communication écrite . . . 47
  - 3.1.3 Modalités synchrones et asynchrones . . . 48
- 3.2 Spécificités des modalités étudiées . . . 49
  - 3.2.1 Courriels . . . 49
  - 3.2.2 Forums . . . 51
  - 3.2.3 Chats . . . 51

3.3 ISO 24617-2 : étude d'applicabilité 52
3.3.1 Cadre de l'étude 52
3.3.2 Dialogues et polylogues 53
3.3.3 Contraintes techniques 55
3.3.4 Asynchronicité et pseudo-synchronicité 56
3.3.5 Conversations multi-canaux 58
3.4 Conclusion 58

4 Interopérabilité taxonomique : enjeux et proposition 61
4.1 Introduction 63
4.2 Standardisation vs Interopérabilité 64
4.2.1 Le standard ISO 65
4.2.2 Interopérabilité entre taxonomies 65
4.3 Méta-modélisation 66
4.3.1 Intuitions 67
4.3.2 Caractéristiques méta-modélisables d'une taxonomie 67
4.3.3 Le méta-modèle 68
4.3.4 Formalisation des traits 69
4.3.5 Intérêts 71
4.3.6 Extraction des traits primitifs 73
4.4 Cadre expérimental 74
4.4.1 Corpus et taxonomies 74
4.4.2 Méta-modèle expérimental 75
4.5 Expériences 75
4.5.1 Conversion d'annotations 75
4.5.2 Classification inter-taxonomique 77
4.5.3 Méthode 78
4.5.4 Résultats 80
4.6 Autres usages 81
4.7 Conclusion 82

5 Corpus Ubuntu 83
5.1 Introduction 85
5.2 Travaux et ressources similaires 86
5.3 Construction du corpus 87
5.3.1 Modélisation des conversations écrites en ligne 87
5.3.2 Méthode de collecte 89
5.3.3 Pré-traitement des données 89
5.3.4 Statistiques sur le corpus 90
5.4 Annotation d'une partie du corpus en termes d'actes de dialogue et d'informations Opinion-Sentiment-Émotion 91
5.4.1 Taxonomie employée 92

5.4.2 Annotation et accord inter-annotateurs . . . 93
5.4.3 Annotations et statistiques . . . 94
5.5 Conclusion . . . 96

6 Reconnaissance automatique des actes de dialogue 97
6.1 Introduction . . . 99
6.2 Travaux similaires . . . 100
6.3 Méthode . . . 102
6.3.1 Approche et implémentation . . . 102
6.3.2 Traits . . . 103
6.4 Expériences et discussion . . . 103
6.4.1 Comparaison d'un classifieur SVM à l'approche de base . . . 104
6.4.2 Comparaison d'approches et de jeux de traits sur différentes modalités . . . 105
6.5 Comparaison de classifieurs trans-modalités . . . 107
6.6 Conclusion . . . 108

7 Conclusion générale 111
7.1 Synthèse des contributions . . . 111
7.1.1 Modélisation des actes de dialogue . . . 112
7.1.2 Méta-modélisation des taxonomies d'actes de dialogue . . . 112
7.1.3 Corpus Ubuntu-fr . . . 113
7.1.4 Classification automatique des actes de dialogue . . . 113
7.2 Perspectives . . . 114
7.2.1 Réseaux de neurones . . . 114
7.2.2 Interopérabilité inter-taxonomique . . . 115

Annexes 119

A Taxonomie d'actes de dialogue employée pour le corpus Ubuntu-fr 119
A.1 Dimensions sémantiques . . . 119
A.2 Fonctions communicatives . . . 120
A.2.1 Fonctions communicatives génériques . . . 121
A.2.2 Fonctions communicatives spécifiques . . . 121

B Liste des traits primitifs 125
B.1 Composants . . . 125
B.1.1 Participants . . . 125
B.1.2 Objets . . . 125
B.1.3 Verbes . . . 125
B.1.4 Propriétés . . . 126
B.2 Traits primitifs . . . 128

Bibliographie 131

Liste des figures

1.1 Exemple de conversation extraite du forum de la plate-forme ubuntu-fr. 18

2.1 Taxonomie DAMSL. 34
2.2 Aspects des énoncés avancés par Popescu-Belis. 36
2.3 Dimensions retenues par Bunt (2009). 37

3.1 Comparaison entre le caractère arborescent d'une liste de diffusion et la structure linéaire d'un fil de forum. 50

4.1 Exemple de méta-modèle pour trois labels du standard et DAMSL. . . . 68
4.2 Processus d'extraction des traits primitifs pour les fonctions de *feedback*. 72
4.3 Représentation du procédé de conversion d'annotations utilisant le méta-modèle. 76
4.4 Représentation du procédé d'annotation automatique via le méta-modèle par le premier système (A). 79
4.5 Représentation du procédé d'annotation automatique via le méta-modèle par le second système (B). 80

5.1 Modèle de méta-données. 88
5.2 Distribution des labels Opinion, Sentiment et Émotion. 96

Liste des tableaux

2.1 Taxonomies fondatrices en théorie des actes de discours, alignées. . . . 28
2.2 Exemples de fonctions spécifiques. 37
2.3 Fonctions communicatives génériques de la taxonomie DIT^{++}. 38

3.1 Caractéristiques des modalités étudiées. 49

4.1 Comparaison des corpus utilisés. 74
4.2 Statistiques taxonomiques : nombre de labels, nombre de dimensions et nombre de traits requis pour les modéliser. 75
4.3 Scores pour la conversion d'annotations. 77
4.4 Exactitude d'un classifieur classique (label à label) et d'un classifieur inter-taxonomique (label à traits à label). 81

5.1 Un an de données : de janvier à décembre 2014 pour les forums et courriels, d'octobre 2014 à octobre 2015 pour les chats. 91
5.2 Comparaison de la taille des sous-corpus annotés par modalité. 94
5.3 Distribution des dimensions des actes de dialogue. 95
5.4 Distribution des fonctions communicatives des actes de dialogue. . . . 95

6.1 Comparaison des performances d'un classifieur SVM entraîné sur les n-grammes et l'approche de base sur l'ensemble du corpus, toutes modalités confondues, pour la tâche de reconnaissance des dimensions sémantiques et des fonctions communicatives. 104
6.2 Dimensions sémantiques : nombre d'exemples dans la référence et résultats d'un classifieur SVM entraîné avec les n-grammes sur les différentes modalités du corpus. 104
6.3 Fonctions communicatives : nombre d'exemples dans la référence et résultats d'un classifieur SVM entraîné avec les n-grammes sur les différentes modalités du corpus. 105
6.4 Résultats pour la tâche de reconnaissance des dimensions sémantiques sur différentes modalités, différents traits et différentes approches. . . 107
6.5 Expériences pour la tâche de reconnaissance des fonctions communicatives sur différentes modalités, différents traits et différentes approches. 108
6.6 Comparaison des performances de classifieurs entraînés sur des modalités autres que la modalité de test. 109

Glossaire

acte de dialogue L'acte de dialogue correspond à la fonction communicative d'un énoncé : le rôle qu'il a dans une conversation, l'intention du locuteur, et le contenu sémantique qu'il apporte au contexte de la conversation et aux connaissances partagées par les participants. 26

acte de discours Acte social, rhétorique ou pragmatique accompli par le locuteur lorsqu'il produit un énoncé. 25

allocutaire L'allocutaire est celui à qui s'adresse le locuteur. 25

asynchrone Une communication est dite asynchrone quand les participants n'interagissent pas en même temps. Par exemple, les courriels sont une forme de communication asynchrone. 46

canal Un site ou environnement en réseau où des utilisateurs peuvent communiquer. Un forum, un salon de chat ou une liste de diffusion sont tous des canaux. 56

chat Les chats sont une modalité de messagerie instantanée. Les chats permettent l'échange presque instantané de messages textuels entre deux ou plusieurs utilisateurs. Contrairement au courriels et aux forums, les chats permettent une communication presque en temps réel. 49

communauté Une communauté en ligne est une communauté virtuelle dont les membres interagissent entre eux principalement sur internet. Certaines communautés sont focalisées sur l'aide aux utilisateurs, et les canaux de communications qu'elles proposent permettent aux utilisateurs de poser des questions et d'obtenir des réponses. 19

contexte En analyse du dialogue, le contexte contient les informations nécessaires aux participants d'une conversation pour pouvoir interpréter les nouveaux énoncés. Le contexte contient toute les connaissances communes (*common ground*) des participants. 28

conversation Séquence structurée d'énoncés (oraux ou textuels) où différents participants interagissent atour d'un sujet ou d'une tâche, chacun alternant entre le role de locuteur et d'allocutaire sous le contrôle de chaque partie. 15

conversation fonctionnelle Conversation construite dans le but de transmettre des informations et de coordonner des actions pour atteindre un but individuel ou collectif. 27

conversation multi-participant Voir *polylogue*. 51

CRF Les CRF, pour *Conditional Random Fields* ou « champs conditionnels markoviens » en français, sont une classe de classifieurs de type supervisé souvent utilisés pour l'apprentissage statistique de données séquentielles. 98

dialogue Un dialogue est, au sens large, une conversation entre deux ou plusieurs personnes alternant entre les rôles de locuteur et d'allocutaire. Une définition plus restrictive du dialogue implique que la conversation n'implique que deux participants. 25

énoncé Séquence de mots produits par un même locuteur lors d'un tour de parole. Généralement une phrase ou une partie de phrase. Dans le cadre de l'analyse du dialogue, on considère l'énoncé comme l'unité discursive minimale pouvant être porteuse d'un acte de dialogue. 25

fil de discussion Sur les forums, un fil de discussion - parfois appelé *thread* ou *topic* - correspond à une série de messages consécutifs organisés autour d'un même sujet, et constituant une conversation. 49

forum Les forums en ligne sont des sites où les internautes peuvent avoir des conversations sous la formes de messages, ou *posts*. Les messages de forums sont organisés en fils de discussion. Ils peuvent être constitués de plusieurs lignes de texte ainsi que d'autres éléments HTML, comme des images, des balises de mise en forme ou encore des hyperliens. 49

liste de diffusion Les listes de diffusion sont une forme d'utilisation des courriers électroniques qui permettent l'envoi simultané de messages à tous les utilisateurs abonnés. 47

locuteur Le locuteur est le participant d'une conversation qui produit des paroles ou des textes formant un message destiné à un allocutaire. 25

participant Les participants d'une conversation sont les personnes qui contribuent en endossant successivement les rôles de locuteur et d'allocutaire. 27

polylogue Les polylogues sont des conversations faisant intervenir trois participants ou plus, chacun endossant tour à tour les rôles de locuteur et d'allocutaire. 51

SVM Les SVM, pour *Support Vector Machines* ou « machines à vecteurs de support » en français, sont une classe de classifieurs de type supervisé destinés à résoudre des problèmes de discrimination et de régression. 100

synchrone Une communication synchrone est une communication en temps réel où les participants sont disponibles en même temps et où la réception de l'information est simultanée à son émission. Par exemple, les conversations en face à face sont synchrones. 46

Ubuntu Ubuntu est un système d'exploitation libre. C'est l'une des distributions de Linux les plus populaires. Ubuntu est basé sur l'architecture Debian. 21

1 Introduction

1.1 Contexte général

Le développement d'Internet a déclenché des révolutions majeures. Parmi elles, l'avènement du contenu généré par les utilisateurs. Avec le développement du Web 2.0 et l'accroissement des capacités d'interaction entre les internautes, nous avons été témoins d'un phénomène de démultiplication de l'information en ligne. Voilà plus de dix ans que les réseaux sociaux, les blogs, les forums et les autres formes de médias interactifs en ligne sont devenus d'usage courant. Sur ces plate-formes, de nombreux types d'interactions peuvent être identifiés. Nous nous intéressons au cas des conversations orientées vers la résolution de problèmes, *i.e.* celles s'opérant sur des plate-formes dont les utilisateurs sont invités à transmettre leurs demandes d'assistance aux autres internautes, qui en retour tentent d'apporter leur aide. Par exemple, demander comment configurer un routeur, où trouver une pièce pour réparer son véhicule, *etc.*. La figure 1.1 montre un exemple de conversation de ce type, initiée par un utilisateur d'Ubuntu qui demande de l'aide pour résoudre un problème d'installation. Ces conversations se retrouvent, notamment, sur des forums d'entraide (*e.g.* CommentÇaMarche, CNET), des listes de diffusion [1] et des salons de chat (*e.g.* canaux IRC). L'exploitation de cet ensemble massif de données représente un enjeu majeur pour les scientifiques et industriels qui s'intéressent aux problématiques liées à l'assistance aux utilisateurs.

Ce type de conversation se retrouve également dans des environnements privés et plus strictement contrôlés : les canaux d'assistance en ligne mis à disposition directement par les entreprises, pour leurs clients. L'importance de ces canaux n'est pas négligeable : disposer d'un service client en ligne efficace est devenu une part intégrale du succès pour les entreprises opérant sur le net. Les compagnies basées sur le Web savent depuis longtemps que le service client des commerces virtuels est tout aussi important que pour les magasins traditionnels (Bernett, 2000).

Ces demandes d'assistance sont le plus souvent gérées au cas par cas, ce qui mène à un nombre considérable de conversations redondantes pour des problèmes et questions communément soulevées. Dans le cas des forums et des listes de diffusion, les conversations sont généralement perpétuellement sauvegardées, permettant ainsi le partage de support entre utilisateurs. C'est-à-dire qu'elles dotent les utilisateurs de

1. Aussi appelées « listes de discussion ».

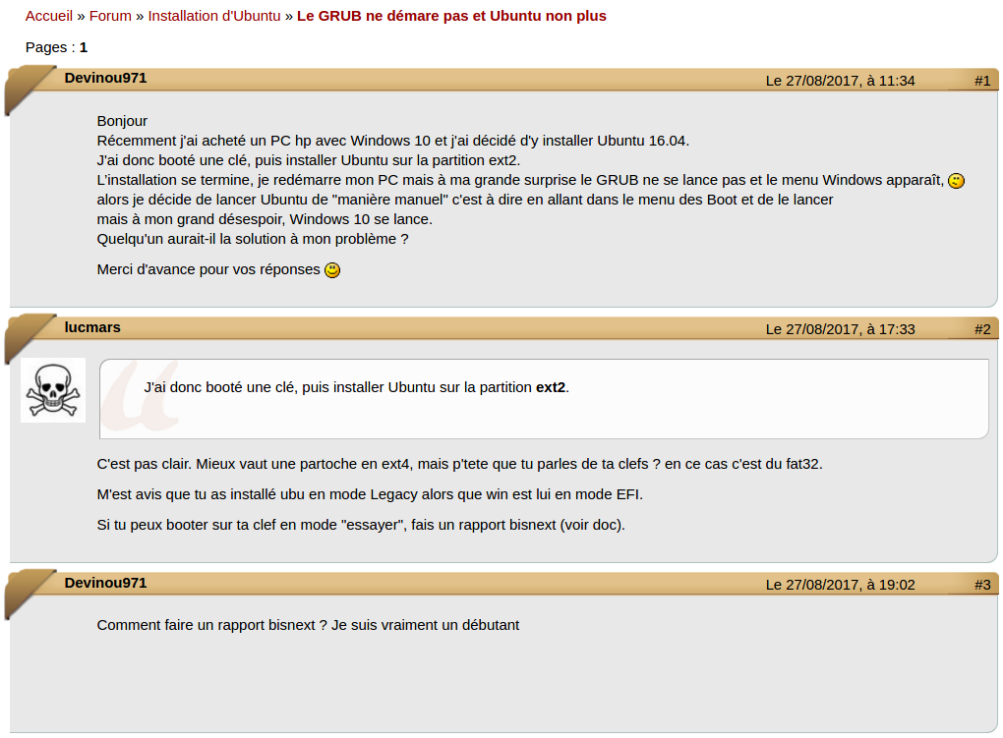

Fig. 1.1.: Exemple de conversation extraite du forum de la plate-forme ubuntu-fr.

la capacité de faire des recherches dans les archives de la plate-forme pour essayer de trouver une solution documentée et directement applicable à leur situation. Bien que moins souvent exploités, les messages transmis dans les salons de chat peuvent aussi faire l'objet d'un archivage automatique. Dans le cas des systèmes de support client, les conversations sont toujours enregistrées et sont généralement manuellement exploitées. Cette exploitation se place, notamment, dans le cadre de l'évaluation des agents, de l'amélioration des techniques de marketing, l'enrichissement des bases de connaissances et des FAQ (Foire Aux Questions).

Un cadre conceptuel et taxonomique bien défini permettant une analyse fine de ce type de conversations représenterait un socle solide sur lequel pourraient reposer différents systèmes liés à l'aide à la résolution des problèmes. Comme, par exemple, des systèmes de recherche d'information (RI) plus performants, pour aider les utilisateurs cherchant des solutions à leurs problèmes dans des archives de conversations. Ou encore, des systèmes automatiques d'enrichissement de bases de connaissances et de FAQ à partir de conversations résolues. Il est aussi possible d'imaginer des systèmes automatisés d'aide à la résolution des problèmes qui proposeraient automatiquement des solutions aux utilisateurs, ou au moins les redirigeraient directement vers un message comportant une solution validée pour un problème identique. Toutes ces applications seraient considérablement enrichies par les masses énormes de données qui sont actuellement librement disponibles sur Internet, mais très peu exploitées.

Dans la littérature, c'est en actes de dialogue que les conversations sont typiquement modélisées. Les actes de dialogue représentent l'intention rhétorique du locuteur, et donc la fonction de l'énoncé dans le dialogue. Par exemple, si un utilisateur se connecte à un salon de chat et envoie le message « salut tout le monde ! », la fonction de l'énoncé est de saluer les participants, et donc de signifier de manière polie que le locuteur est disponible et souhaite dialoguer. Si le locuteur poursuit avec le message « comment faire pour installer la dernière version de Wine ? », la fonction de cet énoncé sera différente : il s'agit d'une question, le locuteur signifie qu'il souhaite que ses interlocuteurs lui fournissent une information. Connaître l'acte de dialogue porté par un énoncé est nécessaire pour comprendre le dialogue, et remplit un rôle fondamentalement différent d'autres problématiques telles que la détection thématique ou l'analyse de sentiment. Savoir qu'un énoncé est porteur d'une émotion positive est important mais secondaire pour un système d'analyse du dialogue, dont la priorité pour être utile est de comprendre la structure de la conversation. De même, comprendre automatiquement que l'énoncé porte sur le logiciel Wine est important mais n'apporte pas de connaissance structurelle, *i.e.* d'informations sur la fonction de l'énoncé où sur ce que le locuteur attend en retour. Un assistant automatique, par exemple, ne peut pas être capable de fournir des réponses aux questions des utilisateurs s'il n'est pas capable de distinguer une question d'une affirmation.

Actuellement, il existe de nombreuses ressources pour la traitement et l'analyse du dialogue. Il existe des corpus de conversations, tels que Enron (Klimt et Yang, 2004), Switchboard (Godfrey et al., 1992), MapTask (Anderson et al., 1991) ou TRAINS (Gross et al., 1993) pour l'anglais. En français, Simuligne (Reffay et al., 2008) ou RATP-Decoda (Bechet et al., 2012) en sont deux exemples. Il existe de nombreux schémas d'annotations pour les actes de dialogue, comme DAMSL (Core et Allen, 1997), COCONUT (Di Eugenio et al., 1998) , LUNA (Raymond et al., 2007) ou encore MapTask (Anderson et al., 1991). Ces dernières années, des efforts de standardisation ont été effectués au travers de la création d'un standard ISO (Bunt et al., 2010) et d'une banque de dialogues annotés (Bunt et al., 2016). En termes de reconnaissance automatique des actes de dialogue, de nombreuses expériences ont été publiées dans la littérature. Généralement basée sur des techniques de classification supervisée, la reconnaissance des actes de dialogue peut être effectuée en utilisant des machines à vecteurs de support (SVM) (Qadir et Riloff, 2011), des champs markoviens conditionnels (CRF) (Kim et Baldwin, 2010), des réseaux de neurones (Khanpour et al., 2016), des classifieurs d'entropie maximale (Lan et al., 2008), des classifieurs naïfs bayésiens (Ivanovic, 2005), *etc*.

Cependant, malgré l'apparente abondance de ressources et de méthodes, les dialogues écrits en ligne et la reconnaissance automatique des actes qui y sont présents sont à la marge du domaine. La majorité des recherches effectuées portent sur les

dialogues oraux (Baron, 1984), dont les caractéristiques sont fondamentalement différentes de l'écrit (Vachek, 1976; Baron, 1984; Biber, 1986; Halliday, 1989; Baron, 1998; David, 2001; Urbanová, 2003). De plus, les conversations articulées autour de la résolution collaborative de problèmes ont leurs propres spécificités (Roschelle et al., 1995). Il n'est pas possible de réutiliser les ressources et outils actuellement utilisés par la communauté tels quels. Pour pouvoir analyser discursivement et automatiquement les conversations écrites en ligne porteuses de demandes d'assistance, il est nécessaire de construire ou au moins d'adapter des modèles, des outils et des ressources. Il s'agit d'une problématique spécifique à la recherche en communication médiée par les réseaux (CMR), et c'est à cette problématique que cette thèse cherche à répondre.

1.2 Objectifs de recherche

Les objectifs du projet de recherche sont l'étude, la modélisation et la reconnaissance automatique des actes de dialogue au sein de conservations en ligne s'étendant sur trois modalités de communication écrites : les forums, les courriels et chats. Ces trois modalités sont sélectionnées parce qu'elles sont les principales modalités de l'écrit en ligne employées pour l'assistance aux utilisateurs, qui constitue notre domaine applicatif. Les langues pour lesquelles ces modalités sont étudiées sont le français et l'anglais. Le modèle que nous souhaitons développer doit avoir les caractéristiques originales suivantes :

1. Il doit permettre l'analyse à **grain fin** du contenu des messages
2. Il doit permettre le traitement **unifié et générique** des modalités synchrones et asynchrones
3. Il doit proposer un fort degré d'**interopérabilité** avec les autres modèles de l'analyse du dialogue

Concernant le premier point, nous considérons que les dialogues sont des phénomènes complexes. De nombreux modèles proposés dans la littérature sont relativement simples, proposant pour annoter les conversations de taxonomies seulement constituées d'une douzaine de labels (Anderson et al., 1991; Qadir et Riloff, 2011; Kim et al., 2010; Kim et Baldwin, 2010). Mais ces taxonomies sont généralement développées dans le but d'accomplir une tâche précise, pas pour soutenir le développement de systèmes d'analyse ultérieurs. Nous souhaitons suivre l'exemple de travaux plus ambitieux tels que DAMSL (Core et Allen, 1997) ou DIT^{++} (Bunt, 2009), dont l'objectif est de soutenir les travaux d'analyse du dialogue plutôt que le développement d'un système spécifique. Cela passe par une taxonomie qui permettre une annotation précise des messages au niveau de l'énoncé.

Pour ce qui est du second point, il apparaît que si les différentes modalités qui nous intéressent présentent des caractéristiques différentes, elles sont toutes fonctionnellement identiques, du moins dans le cadre des conversations orientées autour de la résolution de problèmes. Par exemple, on s'aperçoit que si les conversations présentes sur les listes de diffusion du site ubuntu-fr [2] ont des caractéristiques structurelles différentes des conversations trouvées dans le canal IRC de la même plate-forme, les deux canaux ont le même objectif fonctionnel : permettre à des utilisateurs de poser des questions à la communauté, de discuter, et de trouver ensemble des solutions à des problèmes concernant le système d'exploitation Ubuntu. Par conséquent, il est logique de vouloir analyser au travers du même modèle les conversations présentes sur les forums, les listes de diffusion et les salons de chat.

Enfin, nous justifions le troisième point par une volonté de faciliter la réutilisation de l'existant et l'interopérabilité entre les modèles. Nous constatons avec regret que beaucoup d'efforts effectués dans le domaine, qu'il s'agisse de la construction de corpus ou du développement de nouvelles taxonomies, ne sont pas réutilisables ou exploitables dans un grand nombre de cas. Les chercheurs ayant une tâche spécifique à accomplir préfèrent souvent produire une nouvelle ressource plutôt que s'appuyer sur un existant souvent trop générique, abstrait ou complexe. À l'inverse, les chercheurs travaillant sur des sujets plus fondamentaux n'ont pas beaucoup d'intérêt à utiliser des ressources spécifiques à une tâche ou un domaine particulier. Nous pensons qu'il est donc nécessaire de songer à l'exploitation des ressources et du modèle au moment de leur conception, notamment en rendant les nos contributions « compatibles » avec le standard ISO pour l'annotation des dialogues (Bunt et al., 2010).

Ainsi, dans cette thèse nous nous fixons comme objectif de proposer un modèle pour l'analyse des conversations écrites en ligne porteuses de demandes d'assistance. Nous proposons également une ressource annotée selon ce modèle permettant des applications de traitement automatique de la langue (TAL) et des travaux de reconnaissance automatique. Enfin, nous ajoutons un dernier objectif, celui de confronter ce modèle et cette ressource à une tâche de reconnaissance automatique des actes selon des algorithmes de classification supervisée.

2. `www.ubuntu-fr.org`

1.3 Projet ODISAE

Cette thèse s'inscrit dans le cadre du projet ODISAE [3] [4]. Il s'agit d'un projet de recherche opérationnelle dans le cadre du Fond Unique Interministériel 17 (FUI-17). Le projet a débuté le 15 juillet 2014 et s'est achevé le 15 juillet 2016. Il a réuni sept entreprises et un laboratoire de recherche qui se sont donné pour mission d'explorer les interactions écrites produites dans le cadre de conversations client-agent et d'enrichir les outils logiciels de gestion du support.

Le consortium est réuni autour d'Eptica [5], professionnel de la relation client. Il associait un partenaire universitaire, le Laboratoire d'Informatique de Nantes-Atlantique (LINA), aujourd'hui Laboratoire des Sciences du Numérique de Nantes (LS2N), où cette thèse a été effectuée. Les autres partenaires technologiques sont : La Cantoche Productions (LivingActors [6]), Kwaga [7] (Evercontact [8]), et Jamespot [9]. La plate-forme réalisée a été évaluée par des partenaires utilisateurs coordonnés par le GFII [10] : TokyWoky [11], le centre INSEE Contact et le Comité du Tourisme de l'Aube [12].

L'objectif du projet est de développer des outils d'analyse linguistique pour aider au traitement des échanges avec les centres de support. Ces outils se déclinent en trois groupes : outils d'aide aux agents, outils de gestion pour les administrateurs (*e.g.* optimisation de la FAQ, détection d'experts) et outils de pilotage pour les superviseurs. Le LINA / LS2N a principalement contribué au sous-projet 1 « Moteur d'analyse des discussions », pierre angulaire du projet, en intervenant directement sur trois lots. Il s'agit du lot 1.1 « Modélisation des interactions », du lot 1.2 « Collecte, description et préparation de corpus », et du lot 1.3 « Découpage rhétorique des discussions », qui impliquent des traitements TAL et le développement d'outils de reconnaissance automatique des actes du dialogue.

Les travaux effectués dans le cadre de cette thèse pour ces trois lots constituent la base des contributions présentées respectivement aux chapitres 5 et 6.

3. *Optimizing Digital Interaction with a Social and Automated Environment*
4. `www.odisae.com`
5. `www.eptica.com`
6. `www.livingactor.com`
7. `www.kwaga.com`
8. `www.evercontact.com`
9. `www.fr.jamespot.com`
10. `www.gfii.fr`
11. `www.tokywoky.com`
12. `www.aube-champagne.com/fr/le-comite-departemental-du-tourisme/`

1.4 Plan de la thèse

Nous commençons par nous intéresser aux actes de dialogue. Dans le chapitre 2, nous évoquons en premier lieu la théorie des actes de discours, avant de détailler comment elle a donné naissance à la théorie des actes de dialogue, notamment au travers des notions de contexte et de connaissances communes. Nous terminons ce chapitre en détaillant trois schémas d'annotations importants : DAMSL, DIT^{++} et le standard ISO 24617-2.

Ensuite, nous nous intéresserons aux communications médiées par les réseaux. Nous ouvrons le chapitre 3 en parlant des CMR de manière globale, avant de nous intéresser plus spécifiquement aux trois modalités que nous étudions : les courriels, les forums et les chats. La dernière partie de ce chapitre est consacrée à une étude d'applicabilité du standard ISO, détaillé dans le chapitre précédent, sur des conversations écrites en ligne orientées vers la résolution de problèmes. Nous relevons plusieurs difficultés à utiliser le standard tel quel pour ce type de conversations.

Le chapitre 4 est consacré à la question de l'interopérabilité entre taxonomies. Nous commençons par expliquer la problématique et par défendre une nouvelle proposition, qui consiste à développer un méta-modèle permettant l'exploitation de corpus annotés à l'aide de différentes taxonomies. Nous montrons que l'interopérabilité est une alternative préférable à la standardisation pour résoudre les difficultés observées. Ensuite, nous décrivons les principes du méta-modèle et de sa construction. Nous poursuivons ce chapitre en décrivant des expériences utilisant le méta-modèle pour démontrer qu'il est possible de construire des systèmes de reconnaissance d'actes de dialogue qui ne nécessitent pas de données annotées avec la taxonomie visée mais seulement des données annotées avec une taxonomie capable de capturer suffisamment d'informations pertinentes. Par conséquent, nous démontrons qu'il est possible de travailler avec une taxonomie appropriée pour sa tâche et son domaine tout en bénéficiant des ressources existantes, même dans le cas où ces dernières ne seraient pas totalement pertinentes.

Ensuite, nous consacrons le chapitre 5 à la présentation du corpus Ubuntu-fr que nous avons construit pour soutenir l'analyse discursive des conversations écrites en ligne porteuses de demandes d'assistance, en particulier au travers d'actes de dialogue. Il s'agit d'un corpus de conversations extraites des forums, listes de diffusion et canal IRC de la communauté Ubuntu francophone. Nous détaillons ses principes et le comparons aux corpus existants. Nous détaillons ensuite sa construction, en termes de modélisation, de collecte et de pré-traitements. Enfin, nous présentons les annotations intégrées au corpus : nous détaillons notre méthodologie d'annotation et la taxonomie employée.

Le chapitre 6 est consacré à l'analyse automatique des conversations écrites en ligne en termes d'actes de dialogue. Nous introduisons la problématique et décrivons l'existant, avant de développer notre approche, basée sur des classifieurs SVM et CRF, et notre implémentation. Nous présentons nos expériences effectués sur le corpus Ubuntu-fr et discutons leurs résultats.

Enfin, nous concluons cette thèse en proposant une synthèse des contributions et une perspective des travaux futurs.

Actes de dialogue

2

Table des matières

2.1 Introduction 27

2.2 Les actes de discours 27

2.3 Des actes de discours aux actes de dialogue 28
- 2.3.1 Analyse des conversations fonctionnelles 29
- 2.3.2 Théorie des actes de la conversation 29
- 2.3.3 Contexte et connaissances communes 30

2.4 Schémas d'annotation 31
- 2.4.1 Concepts pratiques et théoriques 31
- 2.4.2 DAMSL 33
- 2.4.3 DIT^{++} 35
- 2.4.4 ISO 24617-2 39

2.5 Conclusion 41

2.1 Introduction

Nous nous intéressons dans cette section aux actes de discours [1], qui sont largement utilisés dans les études sur les phénomènes conversationnels, pour l'annotation de dialogues, et dans la conception d'agents conversationnels [2].

Nous commençons par expliquer ce qu'est la théorie des actes de discours avant de voir comment ces fondements théoriques ont été étendus vers l'analyse des conversations au travers des travaux de Traum et Hinkelman (1992) et Poesio et Traum (1997). Nous poursirons ensuite ce chapitre par la description de trois importants schémas d'annotations : DAMSL, DIT^{++} et le standard ISO 24617-2.

2.2 Les actes de discours

En linguistique et en philosophie du langage, un acte de discours est un énoncé porteur d'une fonction performative. En effet, pour Austin, qui a introduit le terme dans la langue contemporaine, les énonciations doivent être considérées comme des actions effectuées par le locuteur. Apparaît ici l'idée selon laquelle tout acte d'énonciation serait la réalisation d'un acte social. Les verbes qui spécifient ces actions sont appelés verbes performatifs (*i.e.* « Je vous confère le titre de capitaine »). Mais les actes de discours ne sont pas constitués uniquement de ces types de verbes.

Austin (1975) développe une théorie des actes de discours défendant la thèse selon laquelle tout énoncé peut être analysé à trois niveaux. D'abord, au niveau locutoire : il s'agit de sa forme de surface, *i.e.* de la signification de l'énoncé, représenté par ses aspects phonétique, syntaxique et sémantique. Puis au niveau de l'acte illocutoire, porteur de l'intention rhétorique du locuteur. Et enfin, au niveau de l'acte perlocutoire, qui s'intéresse aux conséquences de l'exécution de l'énoncé ou de son interprétation par les allocutaires : son effet pragmatique.

La notion d'acte illocutoire est centrale au concept d'acte de discours. Cet acte permet de décrire les énoncés en termes de fonctions communicatives portées par chacun d'eux (*e.g.* question, réponse, remerciement...). Austin propose cinq classes d'actes de discours : les verdictifs (qui donnent un verdict), les exercitifs (qui exercent un pouvoir), les promissifs (qui engagent le locuteur), les comportatifs (qui expriment l'attitude) et les expositifs (qui exposent de l'information).

Pour Searle (1969), dont la conception des actes de discours diffère légèrement de celle d'Austin, tout acte de discours est illocutoire (sa définition se rapproche ainsi

1. Aussi appelés « actes de langage »
2. Systèmes informatiques conçus pour converser avec des êtres humains.

Austin (1975)	Searle (1976)	Exemples
Expositifs	Assertifs	affirmer, nier, postuler, remarquer...
Exercitifs	Directifs	commander, conseiller, ordonner, pardonner, léguer...
Promissifs	Promissifs	promettre, faire vœu de, garantir, jurer de...
Comportatifs	Expressifs	s'excuser, remercier, féliciter, déplorer, critiquer...
Verdictifs	Déclaratifs	acquitter, condamner, décréter, baptiser...

Tab. 2.1.: Taxonomies fondatrices en théorie des actes de discours, alignées.

de ce que Austin appelle « acte de dialogue »). Il propose cinq classes d'actes : les assertifs (qui affirment un état de fait), les directifs (qui poussent l'interlocuteur à agir), les promissifs (qui engagent le locuteur), les expressifs (qui expriment un état psychologique) et les déclaratifs (qui ont un impact réel, *e.g.* prononcer un jugement) (Searle, 1976). La table 2.1 illustre ces deux taxonomies fondatrices et montre comment elles peuvent être alignées.

Historiquement, cette théorie a rapidement gagné en influence dans un ensemble de disciplines varié. En psychologie, par exemple, il a été suggéré que l'acquisition des actes de discours puisse être un prérequis pour l'acquisition du langage (Bruner, 1975). Des experts littéraires se sont tournés vers Austin pour mettre en lumière des particularités textuelles (Ohmann, 1971). En linguistique, des chercheurs ont trouvé que des notions de la théorie des actes de discours permettaient d'expliquer des problèmes en sémantique (Fillmore, 1971), en syntaxe (Sadock, 1974) et en apprentissage d'une seconde langue (Jakobovits et Gordon, 1974). Même en philosophie, des applications pouvaient être trouvées, par exemple pour déterminer le statut de postulats éthiques (Searle, 1969).

En informatique, les actes de discours sont communément utilisés pour modéliser les conversations dans le cadre d'applications de classification automatique et de recherche d'information (Twitchell et al., 2004). Des modèles pour l'interaction homme-machine ont également été développés en se basant sur ces concepts (Morelli et al., 1991). Ainsi, c'est en termes d'actes de discours que les interactions entre participants d'une conversation sont modélisées par de nombreux travaux liés à la linguistique informatique.

2.3 Des actes de discours aux actes de dialogue

Dans cette section nous nous intéressons aux travaux qui ont été produits autour de la théorie des actes de discours d'Austin et Searle, et qui ont mené au développement du concept d'acte de dialogue tel qu'il est utilisé dans la littérature contemporaine.

2.3.1 Analyse des conversations fonctionnelles

Jusque dans les années 1990, la théorie des actes de discours s'est largement limitée à l'examen d'énoncés isolés, et n'a pas cherché à prendre en charge l'analyse de conversations entières où plusieurs participants peuvent interagir (Vanderveken, 1992). Cependant, les locuteurs accomplissent des actes illocutoires tout au long des conversations qu'ils peuvent avoir avec d'autres participants. Vanderveken souligne que ces derniers répondent et accomplissent à leur tour leurs propres actes de discours, tout en cherchant collectivement à atteindre des objectifs communs. Une application sociale du langage est donc constituée, en général, de séquences ordonnées d'énoncés par différents locuteurs qui cherchent ensemble à poursuivre un même but, comme décider d'une marche à suivre, résoudre un problème, accomplir une action, etc.

Dans le cas de ces deux derniers exemples, on parlerait de « conversation fonctionnelle », *i.e.* d'une conversation construite autour d'une tâche, c'est-à-dire consacrée à la transmission d'information dans le but de réaliser un objectif individuel ou collectif dans le monde réel. C'est à ce type de conversations, qui inclue notamment les conversations porteuses de demandes d'assistance, que s'intéressent la plupart des travaux cherchant à étendre la théorie des actes de discours aux interactions multipartites. Cela s'explique par le fait que les applications informatiques de l'analyse du dialogue sont presque toujours motivées par le besoin de faciliter ou d'automatiser l'exécution d'une tâche par un utilisateur humain.

2.3.2 Théorie des actes de la conversation

Dans cette perspective d'extension de la théorie des actes de discours, Traum et Hinkelman (1992) décrivent une théorie des *actes de la conversation* (*Conversation Act Theory*), qui se veut plus générale. Ils étudient le corpus TRAINS (Gross et al., 1993) [3], tiré du projet éponyme, dont l'objectif est de développer un assistant de planification intelligent qui puisse communiquer en langage naturel avec des opérateurs humains. Le corpus est constitué de dialogues fonctionnels entre un manager devant résoudre des problèmes de planification et une personne jouant le rôle du système, disposant d'informations additionnelles sur la tâche, et chargé d'assister le manager.

Traum et Hinkelman constatent que l'un des traits les plus flagrants des dialogues fonctionnels est la prépondérance des signes d'accord et d'acquittement (*e.g. « There are oranges at Corning, right ? » « Right. »*). C'est l'un des éléments qui les poussent à remettre en question certains postulats généralement implicites dans les travaux antérieurs. Le premier de ces postulats voudrait que les énoncés soient toujours

3. Traum et Hinkelman ont utilisé des données tirées de Gross et al. (1993) avant leur publication, d'où l'apparente incohérence des dates.

entendus et correctement compris par les allocutaires, d'une part, et que les participants ne s'attendent jamais à ce que ce ne soit pas le cas, d'autre part. Mais non seulement les énoncés sont souvent mal compris ou mal perçus, mais en plus Traum et Hinkelman avancent que les conversations sont structurées de manière à prendre en compte ce phénomène : les participants cherchent systématiquement à obtenir des preuves que leur interlocuteur a bien compris ce qu'il voulaient dire. Ces preuves peuvent prendre la forme d'un acquittement explicite (*e.g.* « *Right.* »), d'un acquittement implicite via une réaction pertinente (par exemple en répondant à la question posée), ou encore par des signaux non-verbaux (hochement de tête, *etc.*). Cette quasi-nécessité de l'acquittement les pousse également à remettre en cause l'idée selon laquelle les actes de discours sont des actions réalisées uniquement par le locuteur, et que l'allocutaire n'a qu'une fonction passive face à eux. Les actes de discours ne peuvent être analysés que dans le contexte d'un dialogue multi-agent. Enfin, le troisième postulat que Traum et Hinkelman remettent en cause suite à cette observation, c'est que chaque énoncé n'est porteur que d'un seul acte de discours. En effet, si certains énoncés peuvent non seulement réaliser leur fonction communicative affichée et en plus servent à acquitter un autre énoncé, c'est qu'ils peuvent réaliser deux actes simultanément.

La taxonomie des actes de la conversation qu'ils proposent prend en compte ces trois observations. Elle détaille une catégorisation de ces actes en quatre classes : les actes de prise de parole (*turn-taking acts*), les actes de synchronisation (*grounding acts*), les actes de discours fondamentaux (*core speech acts*), et les actes argumentatifs (*argumentation acts*). En terme d'unité textuelle, les actes de prise de parole se situent à un niveau inférieur à l'énoncé, les actes de synchronisation au niveau de l'énoncé, tandis que les actes de discours fondamentaux (informer, promettre et requérir) se trouvent au niveau de ce qu'ils appellent une « unité de discours ». Cette unité peut contenir un énoncé introductif suivi d'autant d'énoncés de synchronisation que nécessaire pour assurer une bonne communication (*e.g.* « *Because there are oranges in Vermont. Right ? You agree ?* »). Enfin, les actes argumentatifs se situent à un niveau encore supérieur puisqu'ils peuvent contenir un nombre illimité d'unités de discours dont les actes fondamentaux sont utilisés pour former des composés complexes (par exemple le descriptif d'un système, l'exposé d'un problème *etc.*).

2.3.3 Contexte et connaissances communes

Poesio et Traum (1997) s'accordent à dire que les conversations, même fonctionnelles, ont des aspects nettement séparés de la réalisation de la tâche qui en est l'objet, et que l'exercice du langage est une action coordonnée, ce qui impose le développement d'une théorie du *contexte*. Les théories développées à ce sujet se déclinent en deux traditions : d'une part, les approches linguistiques construites autour notamment de

la résolution d'anaphores, et d'autre part les modèles computationnels proposés pour représenter les effets des actes de discours sur les participants d'une conversation, par exemple en termes de croyances, d'obligations et de besoins. C'est cette deuxième approche qui nous intéresse, puisque la première n'a que peu de rapport avec les exercices de planification et de coordination de l'information qui sont propres aux conversations fonctionnelles, et *a fortiori* aux conversations orientées vers la résolution de problèmes. Si la résolution d'anaphores peut évidemment présenter un intérêt pour suivre le fil des conversations, ce problème purement linguistique doit être traité séparement de la question de la synchronisation inter-participants.

Quand Poesio et Traum parlent de contexte, ils font référence à l'information que les participants doivent utiliser pour interpréter les énoncés d'une conversation. Ce contexte est caractérisé notamment par la notion, centrale, de *connaissances communes*, ou « terrain d'entente » (*common ground*) entre les participants. Cette information est cruciale pour pouvoir comprendre à quoi un énoncé fait référence, puisque c'est le contexte qui contient tous les référents disponibles, les référents étant ajoutés aux connaissances communes au travers des nouveaux actes de discours qui sont accomplis. Bien modéliser ces connaissances nécessite donc de bien modéliser les mises à jour du contexte. C'est là que se situe la nuance entre un acte de discours et un acte de dialogue : si l'acte de discours cherche bien à capturer l'intention communicative du locuteur, l'acte de dialogue inscrit cette intention dans un contexte particulier et capture également l'impact que l'énoncé a sur la conversation.

2.4 Schémas d'annotation

Dans cette section, nous présentons trois schémas d'annotations utilisés pour la modélisation des conversations en termes d'actes de dialogue : DAMSL, DIT^{++} et le standard ISO 24617-2. Nous détaillons leurs fondamentaux conceptuels ainsi que leurs taxonomies.

2.4.1 Concepts pratiques et théoriques

Nous décrivons ici les partis pris et postulats conceptuels qui sont partagés par les schémas d'annotations que nous allons présenter.

Mise à jour du contexte :

DIT^{++}, DAMSL et le standard ISO partent du principe que les applications nécessitant une analyse automatique du dialogue doivent prendre en compte les modifications

dynamiques du « terrain d'entente », et pour ce faire proposent d'annoter les *fonctions communicatives* des actes de dialogue. Pour Core et Allen (1997), ces fonctions doivent représenter des manipulations directes du contexte informationnel d'une conversation. La première caractéristique de ces taxonomies est donc qu'elles définissent les actes de dialogue comme des *opérations de mise à jour du contexte*.

Multi-dimensionnalité :

Comme nous l'avons vu, la communication est une activité complexe. Les participants d'une conversation cherchent souvent à accomplir une tâche particulière au travers du dialogue, tout en contrôlant le processus de conversation, mais également en veillant à respecter les conventions sociales et à structurer thématiquement et discursivement la conversation. Puisque les participants accomplissent ces activités variées plus ou moins en même temps, ce n'est pas surprenant que les énoncés soient souvent multi-fonctionnels, et servent plusieurs objectifs à la fois. Par exemple, un énoncé peut répondre à une question, fournir un retour à propos de la compréhension de la question, et passer le tour à l'allocutaire.

Pour répondre à cette problématique, un aspect important des schémas d'annotations que nous allons détailler est leur multi-dimensionnalité. En effet, une des limites de la théorie des actes de discours d'Austin et Searle, qui a été souvent soulignée par les chercheurs, est son incapacité à prendre en compte la pluralité des intentions qu'un locuteur peut chercher à exprimer dans un seul énoncé. Comme préconisé par Traum et Hinkelman (1992), les taxonomies proposées par Core et Allen et Bunt prennent en compte ce problème et autorisent l'application de plusieurs labels à un seul énoncé. On parle alors de dimensions ou de couches (*layers*), chacune permettant d'annoter un aspect différent de l'énoncé.

Généricité :

L'annotation des conversations en termes d'actes de dialogue peut suivre deux approches. La première, ontologique, consiste à proposer une taxonomie spécifique au domaine ou à la tâche étudiée. La seconde, plus ambitieuse, cherche à atteindre une couverture plus générique du dialogue (Leech et Weisser, 2003). C'est le cas de deux schémas d'annotation largement utilisés : DAMSL et DIT^{++}, ainsi que du standard ISO 24617-2, largement basé sur DIT^{++}.

Leur caractère générique est un de leurs attributs les plus importants, et probablement celui qui a le plus contribué à leur popularité. Les annotations proposées sont toutes de suffisamment haut niveau pour pouvoir être appliquées à différents

types de dialogues. Néanmoins, tous se focalisent nettement sur les conversations fonctionnelles, et ils ont d'ailleurs été d'abord développés autour du même corpus TRAINS que les travaux que nous avons détaillés en sous-section 2.3.

2.4.2 DAMSL

DAMSL, dont l'acronyme signifie *Dialogue Act Markup in Several Layers*, ou « balisage d'actes de dialogue en plusieurs couches », est le premier schéma d'annotation à implémenter une approche multidimensionnelle, permettant d'assigner de multiples labels aux énoncés (Core et Allen, 1997).

Catégories de labels :

DAMSL propose quatre catégories distinctes de labels indépendantes les unes des autres : les fonctions prospectives, les fonctions rétrospectives, le niveau d'information et le statut communicatif. Ce choix est justifié par la réalité des conversations, où certains énoncés sont de toute évidence liés entre eux. Par exemple, prenons l'échange suivant :

1. Participant 1 : « *Le ciel commence à se dégager.* »
2. Participant 1 : « *Quelle heure est-il ?* »
3. Participant 2 : « *Il est bientôt midi.* »

Il est immédiatement apparent que l'énoncé 3 fait réponse à l'énoncé 2, et diffère en ce sens de l'énoncé 1, qui n'a pas été sollicité. Pourtant, les deux apportent une information factuelle au contexte, et pourraient être classés comme « informer ». Core et Allen notent que si des travaux avaient déjà tenté de répondre à ce problème en proposant des sous-classes de type « informer-répondre » ou « informer-accepter », ce n'est pas satisfaisant car les actes d'acquitter, de répondre ou d'accepter un énoncé semblent appartenir à un genre d'actes bien distinct de « informer ». Ils disent des fonctions de ces actes qu'elles sont *rétrospectives* (*backward-looking*), puisqu'elles sont orientées vers la partie antérieure de la conversation. Les autres fonctions (*e.g.* affirmer, ordonner, promettre, *etc.*) sont donc dites *prospectives* (*forward-looking*), puisqu'elles en impactent la partie ultérieure. Ces deux groupes de fonctions constituent les deux premières catégories de labels [4] de la taxonomie DAMSL. Ce sont elles qui permettent d'étiqueter les énoncés par leur intention communicative.

Les deux autres catégories définies par DAMSL sont celles des traits énonciatifs (*Utterance Features*). Ces traits ne s'intéressent pas à la fonction communicative

4. Ces super catégories sont appelées « couches » (*layers*) dans la documentation de DAMSL.

de l'énoncé, mais capturent les propriétés de son contenu. Ils indiquent sur quoi l'énoncé porte (s'il porte directement sur la tâche, sur le processus de communication, du processus de résolution de la tâche, ou d'autre chose) : c'est la catégorie *niveau d'information* (*Information Level*). Ils permettent également d'identifier les énoncés qui peuvent être ignorés sans danger (parce qu'incompréhensibles ou interrompus) : c'est la catégorie *statut communicatif* (*Communicative Status*).

Taxonomie :

Les quatre catégories de la taxonomie sont détaillées en figure 2.1 [5] :

Fonctions rétrospectives :

- *Agreement*
  - *Accept*
  - *Accept-Part*
  - *Maybe*
  - *Reject-Part*
  - *Reject*
  - *Hold*
- *Understanding*
  - *Signal-Non-Understanding*
  - *Signal-Understanding*
    - *Acknowledge*
    - *Repeat-Rephrase*
    - *Completion*
  - *Correct-Misspeaking*
- *Answer*
- *Information-Relation*

Fonctions prospectives :

- *Statement*
  - *Assert*
  - *Reassert*
  - *Other-Statement*
- *Influencing Addressee Future Action*
  - *Open-Option*
  - *Directive*
    - *Info-Request*
    - *Action-Directive*
- *Committing Speaker Future Action*
  - *Offer*
  - *Commit*
- *Performative*
- *Other Forward Function*

Niveau d'information :

- *Task*
- *Task Management*
- *Communication Management*
- *Other*

Statut communicatif :

- *Abandoned*
- *Uninterpretable*
- *Self-talk*

Fig. 2.1.: Taxonomie DAMSL.

Les classes situées au premier niveau des listes imbriquées sont appelées « dimensions » par Core et Allen. Les dimensions sont indépendantes les unes des autres. Ainsi par exemple un énoncé peut à la fois acquitter une question (ACKNOWLEDGE) et y répondre (ANSWER). Toutes les dimensions sont optionnelles. Tous les énoncés n'ont pas non plus forcément un label dans chaque catégorie (par exemple, il est possible d'avoir une fonction prospective mais aucune fonction rétrospective), à l'exception du niveau d'information qui doit toujours être indiqué.

L'utilité de la taxonomie de DAMSL est prouvée par le nombre important de travaux s'appuyant dessus, ce qui en fait *de facto* un standard en analyse du dialogue.

5. Nous avons choisi de conserver les noms originaux en anglais pour ne pas dénaturer la taxonomie.

Cependant, comme souligné par Bunt (2006), les dimensions et les « couches » (les quatre catégories de labels) employées dans DAMSL ne sont pas discutées, manquent de signification conceptuelle, et ne s'appuient sur aucun fondement théorique. Comme nous le verrons dans la section suivante, DIT^{++}, lui, tente de proposer un système fondé sur des bases théoriques solides.

2.4.3 DIT^{++}

DIT^{++}, comme DAMSL, cherche à modéliser les actes de dialogue. Sa taxonomie est une extension de celle de la théorie de l'interprétation dynamique (*Dynamic Interpretation Theory*) (Bunt, 1999), originellement basée sur DAMSL (Bunt, 2009). Dans ce schéma, les actes sont interprétés comme des opérations de mise à jour appliquées à l'état informationnel des participants de la conversation. Dans cette perspective, Bunt définit les actes de dialogue comme la conjonction de deux éléments : leur *contenu sémantique* et leur *fonction communicative*. Le contenu sémantique spécifie les objets, propositions et toutes les choses sur lesquelles porte l'acte. La fonction communicative spécifie la manière dont l'acte est supposé impacter l'état informationnel de l'allocutaire. Par exemple, la phrase « *vous avez bientôt fini ?* » peut être interprétée comme une question littérale (le locuteur veut savoir si l'allocutaire est sur le point de finir une tâche), ou comme une expression d'exaspération (le locuteur est gêné par l'activité de l'allocutaire). C'est cette distinction qui doit être capturée par la notion de fonction communicative. Bunt formalise la notion d'acte de dialogue ainsi (Bunt, 2009, p. 13) :

> « *Un acte de dialogue est une unité de description sémantique du comportement communicatif dans le dialogue, précisant comment le comportement est supposé changer l'état informationnel d'un participant qui aurait correctement compris et interprété le comportement. [...] Formellement, un acte de dialogue est un opérateur de mise à jour d'un état informationnel qui s'interprète en appliquant une fonction communicative à un contenu sémantique.* [6] »

Dimensions :

Dans son panorama des taxonomies d'annotation des actes de dialogue, Popescu-Belis (2005) note que les taxonomies multi-dimensionnelles semblent bénéficier d'une justification théorique au vu de la multiplicité des fonctions que les énoncés

6. « *A dialogue act is a unit in the semantic description of communicative behaviour in dialogue, specifying how the behaviour is intended to change the information state of a dialogue participant who understands the behaviour correctly. [...] Formally, a dialogue act is an information-state update operator construed by applying a communicative function to a semantic content.* »

peuvent avoir. Néanmoins, le choix des dimensions à intégrer dans un schéma d'annotation devrait lui-même être justifié théoriquement. Il avance six aspects des énoncés qui devraient être pris en compte pour déterminer ces dimensions. Ces aspects sont détaillés en figure 2.2.

1. Actes de discours : cet aspect correspond à la catégorisation traditionnelle des actes de discours en cinq classes principales, qui bénéficie déjà de solides bases théoriques (Austin, 1975)
2. Tour de parole : les conclusions du champ de l'analyse du dialogue montrent que dans les conversations, des énoncés ont la fonction particulière de gérer les mécanismes de gestion du tour de parole (Shriberg et al., 2004)
3. Paires adjacentes (*adjacency pairs*) : l'analyse de la conversation montre également que des couples d'énoncés sont souvent appareillés relativement à leurs fonctions communicatives, comme par exemple les énoncés de type « réponse » et les énoncés de type « question » (Levinson, 1983; Schegloff et Sacks, 1973)
4. Organisation thématique des conversations : les études en analyse de la conversation ont également démontré que les conversations sont structurées en successions d'épisodes thématiques, au cours desquels les sujets abordés sont amenés à évoluer, et que des énoncés servent à organiser cette évolution (Schegloff et Sacks, 1973)
5. Structure rhétorique : similairement à ce que Thompson et Mann (1987) ont montré pour les discours monologues à travers la théorie RST (*Rhetorical Structure Theory*), des relations discursives rhétoriques peuvent être établies entre les énoncés des conversations (Asher et Lascarides, 2003)
6. Politesse : les fonctions des énoncés en termes de politesse peuvent être formalisées en termes de gestion de la « face », chaque énoncé de ce type pouvant être analysé selon deux axes : d'abord, s'il s'agit de la face du locuteur ou de l'allocutaire, ensuite, si l'interaction vise à *sauver* ou à *menacer* la face (Brown et Levinson, 1983)

Fig. 2.2.: Aspects des énoncés avancés par Popescu-Belis.

Bunt assoit la crédibilité théorique des dimensions qu'il choisit en les basant sur ces six aspects des actes de dialogue. Par ailleurs, il propose de définir précisément ce qu'est un ensemble de dimensions. Dans DIT^{++}, chaque dimension regroupe des fonctions communicatives portant toutes sur un même aspect de ce que peut être la contribution d'un locuteur à la conversation, de manière à ce que : (1) les participants puissent communiquer autour de cet aspect, et (2) cette communication s'opère de façon indépendante des autres aspects, c'est-à-dire qu'un énoncé peut avoir une fonction communicative dans une dimension qui soit totalement indépendante de celles qu'il peut avoir dans d'autres dimensions. Les dimensions retenues sont décrites en figure 2.3. Les énoncés peuvent avoir au plus une fonction par dimension.

1. *Task/Activity*, pour tout ce qui se rapporte à la tâche qui est l'objet de la conversation ;
2. *Auto-Feedback*, pour les actes signifiant le niveau de compréhension et d'interprétation du locuteur ;
3. *Allo-Feedback*, *idem* pour l'allocutaire ;
4. *Turn Management*, pour les actes portant sur la gestion du tour de parole ;
5. *Time Management*, pour les situations où il est nécessaire de signifier que le locuteur a besoin de plus de temps pour contribuer ou qu'il faut faire une pause ;
6. *Contact Management*, pour les actes qui servent à établir et maintenir la communication ;
7. *Own Communication Management*, pour les actes servant à indiquer que le locuteur prépare ou modifie sa contribution au dialogue ;
8. *Partner Communication Management*, pour les actes effectués par un participant endossant le rôle d'allocutaire, servant à assister son partenaire dans la formulation de sa contribution ;
9. *Discourse Structure Management*, pour les actes servant à structurer thématiquement la conversation ;
10. *Social Obligations Management*, pour les actes de gestion sociale du dialogue.

Fig. 2.3.: Dimensions retenues par Bunt (2009).

Fonctions :

Le schéma d'annotation propose deux types de fonctions communicatives : les fonctions génériques (*general-purpose functions*), qui se retrouvent dans toutes les dimensions (*e.g.* PROPOSITIONAL QUESTION, ADDRESS REQUEST), et les fonctions spécifiques (*dimension-specific functions*), qui ne peuvent être appliquées qu'à une dimension particulière (*e.g.* TURN GRABBING, GREETING). La table 2.2 fournit quelques exemples de fonctions spécifiques pour chaque dimension de la taxonomie.

Dimension	Exemples de fonction
Task / Activity	*Open Meeting, Appoint, Hire*
Auto-Feedback	*Perception Negative, Evaluation Positive*
Allo-Feedback	*Interpretation Negative, Evaluation Elicitation*
Turn Management	*Turn Grab, Turn Take, Turn Keep*
Time Management	*Stalling, Pausing*
Contact Management	*Contact Check, Contact Indication*
Own Communication Management	*Self-Correction*
Partner Communication Management	*Completion, Correct Misspeaking*
Discourse Structure Management	*Opening, Topic Introduction*
Social Obligations Management	*Return Greeting, Apology, Thanking*

Tab. 2.2.: Exemples de fonctions spécifiques.

Les fonctions génériques sont elles-mêmes réparties en deux catégories principales : les fonctions de transfert d'information et les fonctions de discussion d'action. La première catégorie comporte les fonctions de sollicitation et de procuration d'information. La seconde comporte les fonctions servant à gérer la planification d'actions, correspondant typiquement aux actes commissifs et directifs. La liste complète est fournie en table 2.3 :

Type	Catégorie	Fonctions
Information	Sollicitatifs	*Propositional Question*, *Set Question*, *Alternatives Question*, *Check Question*, *etc.*, et équivalents indirects (*e.g. Indirect Check Question*)
	Procuratifs	*Inform*, *Agreement*, *Disagreement*, *Correction*, *Propositional Answer*, *Set Answer*, *Confirmation*, *Disconfirmation*, autres variantes dotées de fonctions rhétoriques, comme l'élaboration et la justification, ou de fonctions attitudinales, comme les avertissements
Action	Commissifs	*Offer*, *Promise*, *Address Request*, autres expressifs exprimables via verbes performatifs
	Directifs	*Instruction*, *Address Request*, *Indirect Request*, *Request*, *Suggestion*, autres directifs exprimables via des verbes performatifs (*e.g.* conseils, encouragements)

Tab. 2.3.: Fonctions communicatives génériques de la taxonomie DIT^{++}.

Réception et extension :

La taxonomie DIT^{++} a été utilisée pour un ensemble d'applications variées, notamment dans le cadre d'annotation de conversations, de l'analyse théorique du dialogue, de la modélisation des phénomènes conversationnels, et du développement de systèmes de dialogue. Elle peut être étendue pour prendre en compte plus finement certains phénomènes, notamment au travers de la notion de *qualifieurs de fonctions* et de *relations rhétoriques*.

Les qualifieurs, introduits par Petukhova et Bunt (2010), sont utilisés en conjonction avec les fonctions communicatives pour décrire l'énoncé plus précisément. Ils proposent une représentation fine du comportement des participants selon différents critères : la modalité, qui spécifie la conviction du locuteur ; la conditionnalité, qui représente la capacité du locuteur à effectuer une action ; la partialité, qui limite la portée de l'énoncé à une partie seulement de l'acte auquel il fait réponse ; et le mode, qui est censé capturer l'attitude et l'état émotionnel du locuteur.

De plus, les annotations peuvent être enrichies par l'annotation des relations rhétoriques et des relations de dépendance fonctionnelle entre les actes de dialogue. Par exemple, un locuteur peut répondre à une question, puis, soucieux d'être bien compris, expliquer sa réponse :

1. Participant 1 : « *Si je mets le driver à jour ça réglerait pas le problème ?* »
2. Participant 2 : « *Non.* »
3. Participant 2 : « *C'est un problème matériel, c'est pas dû au driver.* »

Dans cet exemple de dialogue, l'énoncé 2 entretient une relation de dépendance fonctionnelle avec l'énoncé 1 : il s'agit d'une réponse à une question, la réponse n'aurait pas de sens ni de fonction sans la question. En revanche, l'énoncé 3 entretient une relation rhétorique d'explication avec l'énoncé 2. Il n'y a pas dépendance, puisque l'énoncé 3 aurait quand même du sens et une fonction en l'absence de l'énoncé 2, néanmoins ces deux énoncés sont en relation par le fait que le dernier explique la réponse, et cette information peut être annotée avec DIT^{++}.

Ainsi, le schéma DIT^{++}, facilement extensible et appuyé par un vaste ensemble de travaux antérieurs, a été le socle d'un standard international pour l'annotation dialogique : ISO 24617-2 (Bunt et al., 2012).

2.4.4 ISO 24617-2

Le standard ISO 24617-2 (Bunt et al., 2012) a été développé avec la volonté de répondre au besoin croissant d'un schéma d'annotation qui réponde à plusieurs critères :

1. Indépendant du domaine applicatif
2. Théoriquement et empiriquement fondé
3. Compatible avec les dialogues parlés, écrits, et multi-modaux
4. Efficacement utilisable par les annotateurs humains comme les algorithmes automatiques

Le standard est composé de différents composants, tels que le schéma d'annotation, des concepts pour l'annotation de relations rhétoriques, un langage de modélisation du dialogue (DiAML), et des spécifications pour la représentation en XML des annotations. Dans cette section, nous nous intéressons uniquement au schéma d'annotation proposé. Sur ce point, le standard est très largement basé sur DIT^{++}. Les principes fondamentaux restent les mêmes : les actes de dialogue sont similairement définis, et chaque énoncé peut être annoté avec une fonction communicative par

dimension sémantique. Les informations annotées peuvent également être enrichies par des qualifieurs, et les actes de dialogue peuvent être liés entre eux par des relations rhéthoriques ou de dépendance fonctionnelle.

Les taxonomies des dimensions sémantiques et des fonctions communicatives sont simplifiées : le nombre de fonctions a été réduit de 86 à 56, et les dimensions sémantiques sont passées de 10 à 9, avec le retrait de la dimension *Contact Management*. Néanmoins, le standard précise que ces fonctions et dimensions ne sont que les éléments fondamentaux, transverses à toutes les applications et tous les domaines, et peuvent être étendus. Pour adapter le schéma à un domaine ou une tâche particulière, il est possible de l'étendre en ajoutant des fonctions, dimensions, qualifieurs et relations rhétoriques. Cependant, pour rester dans le cadre du standard, ces modifications sont conditionnées au respect des principes fondamentaux du standard ISO, à savoir :

— Le dialogue est multi-fonctionnel, par conséquent le schéma d'annotation doit être multi-dimensionnel.
— Les dimensions sont des types d'activités communicationnelles distinctes, portant sur des aspects distincts de la conversations, et doivent donc capturer des informations catégoriquement distinctes.
— Toutes les dimensions doivent être justifiées théoriquement, observées empiriquement, reconnaissables avec une précision raisonnable par des annotateurs humains et par des systèmes de reconnaissance automatique, et indépendantes des autres dimensions.
— La fonction communicative est définie comme la façon dont le locuteur veut que l'état informationnel de l'allocutaire soit mis à jour.
— Les fonctions communicatives doivent être correctement assignées à des segments fonctionnels, ces derniers étant définis comme les unités de dialogue minimales portant au moins une fonction communicative.
— L'ensemble des fonctions communicatives est divisé en sous-ensembles de fonctions spécifiques pour chaque dimension, ainsi qu'un sous-ensemble de fonctions génériques.
— Pour chaque dimension, les fonctions communicatives spécifiques doivent être connectées sémantiquement, de manière à ce qu'aucun segment fonctionnel ne puisse nécessiter d'être annoté avec plus d'une fonction communicative pour une dimension donnée.

Le standard est donc un schéma d'annotation solide et extensible. Par ailleurs, des efforts récents ont été produits pour démocratiser son usage, notamment au travers de Tilburg DialogBank (Bunt et al., 2016). Ce projet vise à publier des annotations pour plusieurs corpus fréquemment utilisés au travers du standard, tels

que Switchboard, DBOX ou TRAINS. Si la taxonomie et les données proposées ne sont pas spécifiquement conçues pour le domaine qui nous intéresse, il nous est cependant normalement possible de l'étendre et de l'adapter pour l'appliquer aux conversations écrites en ligne porteuses de demandes d'assistance.

2.5 Conclusion

Nous avons montré comment la théorie des actes de dialogue est née de la théorie des actes de discours, et qu'ils représentent un outil pertinent et largement employé par la communauté dans le domaine de l'analyse du dialogue. Nous avons présenté les principaux concepts utilisés par la théorie, comme le contexte discursif et la notion de connaissances communes. Nous avons présenté trois schémas d'annotation importants : DAMSL, DIT^{++} et le standard ISO 24617-2.

Dans ce chapitre, nous nous sommes intéressés à la théorie des actes de dialogue sans nous préoccuper spécifiquement de son application aux conversations qui sont l'objet de ce manuscrit, à savoir les conversations écrites en ligne porteuses de demande d'assistance. Le chapitre suivant portera donc sur les communications médiées par les réseaux, et les spécificités de leur étude en termes d'actes de dialogue.

Le travail présenté dans ce chapitre a fait l'objet d'une publication à TALN-RÉCITAL 2015 (Salim, 2015).

3 Communication médiée par les réseaux

Table des matières

3.1 Introduction 45
3.1.1 Les réseaux : un handicap à la communication ? . . 46
3.1.2 Communication orale, communication écrite 47
3.1.3 Modalités synchrones et asynchrones 48

3.2 Spécificités des modalités étudiées 49
3.2.1 Courriels 49
3.2.2 Forums 51
3.2.3 Chats 51

3.3 ISO 24617-2 : étude d'applicabilité 52
3.3.1 Cadre de l'étude 52
3.3.2 Dialogues et polylogues 53
3.3.3 Contraintes techniques 55
3.3.4 Asynchronicité et pseudo-synchronicité 56
3.3.5 Conversations multi-canaux 58

3.4 Conclusion 58

3.1 Introduction

L'intérêt des institutions académiques et industrielles pour les communications médiées par les réseaux (CMR) [1] s'est développé rapidement après que ces modes de communication se soient imposés sur le web. Dans le domaine du Traitement Automatique des langues (TAL), de nombreux concepts et outils ont été imaginés et développés autour de l'étude de conversations orales ou médiées par d'autres systèmes (*e.g.* conversations téléphoniques), et les appliquer à des conversations écrites extraites des CMR se révèle parfois fastidieux. Les aspects prosodiques des conversations orales n'ont pas d'équivalents directs, la forte variance orthographique des mots qu'on y trouve met en échec les systèmes les plus performants (on parle alors - peut-être à tort - de « bruit »), les métadonnées nouvellement accessibles ne sont pas exploitées, etc. Or, il existe un intérêt très fort pour les conversations écrites médiées par les réseaux, qui constituent une part extrêmement importante des conversations liées à l'assistance aux utilisateurs, et notamment les conversations gérées par les services de gestion de la relation client (GRC) des entreprises implantées sur le web. Pour traiter correctement ces conversations, qui représentent un enjeu majeur pour l'industrie, il faut se doter d'outils appropriés prenant en compte leurs spécificités.

Baron (1984) note que la plupart des travaux d'analyse du dialogue portent sur les dialogues oraux. Le domaine ne faisant pas exception, la plupart des travaux ayant développé le concept d'acte de dialogue se sont fondés sur l'étude de conversations orales. Dans nombre de travaux fondateurs, il s'agit même du même corpus, TRAINS (Traum et Hinkelman, 1992; Poesio et Traum, 1997; Core et Allen, 1997). Ce corpus est constitué de dialogues entre un agent preneur de décision et un assistant disposant d'une carte ; leur objectif étant de trouver le parcours optimal pour un train de marchandises. La popularité de ce corpus s'explique par le fait que ces dialogues sont proches de ce que l'on attendrait d'une communication entre un agent humain et un assistant virtuel chargé de communiquer le contenu d'une base de connaissances. Cependant, deux aspects de ces conversations, et donc des travaux qui en sont issus, les éloignent de nos objectifs : d'abord, parce que tout ce qui a trait à la communication orale n'est pas transposable aux conversations écrites, et ensuite parce que tout ce qui a trait aux dialogues ne s'applique pas forcément aux polylogues, où un nombre indéfini de locuteurs peut intervenir dans n'importe quel ordre. Or, ces deux aspects sont communs aux modalités qui servent souvent de supports aux conversations orientées vers la résolution de problème dans les communautés internet, à savoir les forums, les listes de diffusion et les services de messagerie instantanée (ou « chats »), et elles s'inscrivent directement dans le champ des communications médiées par les réseaux.

1. *Computer-Mediated Communication (CMC)*

Dans ce chapitre nous nous intéressons à ces communications. Nous présentons les CMR et leurs spécificités dans un contexte général avant de nous intéresser plus précisément aux modalités qui font l'objet de cette thèse : les forums, les courriels et les chats. Enfin, nous confrontons le modèle proposé par le standard ISO (Bunt et al., 2012) (voir section 2.4.4) aux problématiques propres aux CMR au travers d'une étude d'applicabilité.

3.1.1 Les réseaux : un handicap à la communication ?

Une partie significative des travaux initiaux en CMR est focalisée sur la nature de ces réseaux, et les implications que ces caractéristiques ont sur la communication. Les CMR sont textuelles, et donc la communication non verbale est pour une grande partie éliminée. Les CMR, quand utilisées dans un format asynchrone (comme les courriels ou les forums), ne permettent pas de fournir de retours immédiats, ce qui limite la capacité du locuteur à corriger un message si l'interprétation que l'allocutaire en fait est inexacte. La théorie de la richesse médiatique (*Media Richness Theory*) affirme que les CMR sont des environnements de communication plus pauvres que les conversations en face à face (Daft et Lengel, 1986). Quand les retours ne sont pas immédiats et que les participants ne peuvent pas se baser sur des indices non-verbaux, l'ambiguïté croit, et les risques d'échec de communication croissent également. Il s'agit de l'argument de l'infériorité des CMR, qui seraient trop éloignés du mode de communication humain « naturel » pour être aussi efficaces.

Ce jugement n'est pas énoncé de manière égal pour toutes les modalités inclues dans les CMR. Ces modalités varient en termes de degré de synchronicité : ainsi, les services de messagerie instantanées sont « plus synchrones »[2] que les courriels ou les forums. Et s'il est parfois affirmé que les canaux de communications asynchrones ont le bénéfice de permettre aux utilisateurs de réfléchir plus posément à leurs messages, la majorité des travaux théoriques et les observations empiriques accordent leur crédit au postulat que le manque de synchronicité constitue un obstacle à la communication, et que la satisfaction des participants n'est jamais aussi élevée que lorsque la communication s'opère en face à face (Simon, 2006). Comme nous l'avons vu, la théorie des actes de dialogue elle-même a été bâtie sur les notions de contexte et de connaissances communes, et sur l'idée que les actes de synchronisation qui sont accomplis par les participants au fur et à mesure d'une conversation sont les clefs de leur construction et de leur entretien (voir section 2.3.2). Le caractère pseudo-synchrone ou asynchrone des conversations écrites en ligne représenterait donc un premier obstacle à l'utilisation de schémas classiques pour leur modélisation et leur reconnaissance.

2. On dit parfois qu'ils sont « pseudo-synchrones ».

Cependant, les travaux plus récents ont montré que l'argument de l'infériorité des CMR n'est pas toujours pertinent, et que les premiers travaux n'ont pas pris la peine de prendre en compte le fait que les utilisateurs observés étaient contraints d'utiliser un moyen de communication avec lequel ils n'étaient pas familiers, et que l'habitude et l'entraînement permettaient aux utilisateurs d'apprendre à s'adapter à ces nouvelles modalités (Spitzberg, 2006). Ce faisant, ils développent de nouvelles techniques de communication et améliorent à la fois leur propre satisfaction et le degré d'efficacité de la communication. La modélisation des actes de dialogue doit prendre cette adaptation en compte : puisque de nombreux aspects de la conversation disparaissent ou sont fortement réduits en passant d'un dialogue verbal à un dialogue passant par les CMR, les moyens et méthodes de modélisation du dialogue doivent eux aussi être adaptés pour être en mesure de capturer la conversation telle qu'elle se produit. Par exemple, s'il n'est pas possible dans un environnement virtuel de se baser sur le contact visuel, l'orientation du corps ou la gestuelle pour déterminer à qui un message est destiné au sein d'une conversation de groupe, il est possible pour le faire en se basant soit sur les méta-données disponibles (comme le champ « *to* » d'un courriel), ou encore de se baser sur les codes sociaux du Web acquis par les participants. Par exemple, dans les salons de chat il est commun de préfixer un message du pseudonyme de la personne à qui l'on s'adresse spécifiquement.

3.1.2 Communication orale, communication écrite

La différence entre l'oral et l'écrit a reçu une attention considérable dans la littérature (Vachek, 1976; Baron, 1984; Biber, 1986; Halliday, 1989; Baron, 1998; David, 2001; Urbanová, 2003). La distinction fondamentale entre les deux réside dans la notion de proximité temporelle et spatiale entre les participants : à l'oral les participants de la conversation sont proches, à l'écrit ils sont distants. La séparation entre les deux catégories de communication a été très nette, historiquement parlant : les conversations écrites représentaient des communications plus formelles, plus rigoureusement construites, à positionner dans un registre de langage soutenu et propices à de longs échanges réfléchis. À l'inverse, les conversations orales étaient plus dynamiques, plus familières, plus expressives et plus adaptées pour des échanges porteurs d'émotions et de sentiments.

Cependant, le développement des communications médiées par les réseaux ces dernières décennies a contribué à rendre cette distinction moins nette. Bien qu'exprimé majoritairement sous forme de texte, l'univers des CMR manifeste des caractéristiques autrefois attribuées aux communications orales, ce qui leur a valu d'être parfois qualifiées de « formes d'oralité tertiaire ». Il n'existe pas de consensus dans la communauté sur la place des CMR, certains linguistes les positionnant comme

une forme spécifique de communication écrite (Ko, 1996), d'autres préférant parler d'« oral écrit » ou d'« écrit parlé » (David, 2001).

Mais si les CMR ne sont pas exactement similaires aux autres formes de communication écrites, plusieurs de leurs caractéristiques les séparent clairement des conversations orales : l'absence de retour (*feedback*) instantané, l'absence de communication concurrente (*i.e.* deux énoncés ne peuvent pas être produits en même temps), et l'absence de traits prosodiques, paralinguistiques [3] et kinésiques.

3.1.3 Modalités synchrones et asynchrones

Une communication synchrone est une communication en temps réel où les participants sont connectés et disponibles en même temps et où la réception de l'information est simultanée à son émission. Une communication asynchrone, quant à elle, est une communication qui autorise les participants à ne pas être connectés en même temps. De fait la conversation peut s'étaler dans le temps et l'historique de celle-ci reste disponible (Oztok et al., 2013).

Ainsi, toutes les formes de communication n'ont pas le même degré de synchronicité. Les conversations orales sont parfaitement synchrones : les interlocuteurs peuvent réagir en temps réel. Les courriels et les forums sont des modalités asynchrones : les participants écrivent leurs messages en relative isolation les uns des autres, puis les transmettent dans leur intégralité, en une seule fois. Il n'est pas possible pour les destinataires de percevoir le message en temps réel au fur et à mesure qu'il est composé, et il n'est pas possible pour l'émetteur de savoir quand et si le message a été lu. L'immédiateté est perdue, mais le caractère asynchrone apporte également des avantages : notamment, le fait que les contributions individuelles soient sauvegardées et qu'elles puissent être consultées des jours, des mois, voire des années plus tard.

Les chats occupent une place un peu particulière en terme de synchronicité. Comme pour les courriels et les forums, les messages ne sont pas transmis au fur et à mesure mais uniquement une fois la transmission déclenchée. Cependant, les messages sont courts et rapides, et le concept de « salon de chat » où les participants doivent se connecter pour recevoir les messages implique que les communications sont perçues quasi-instantanément. De plus, les messages sont sauvegardés mais uniquement à court terme : un utilisateur peut consulter les messages précédents, mais s'il quitte le salon et se reconnecte quelques heures plus tard, tout sera perdu. Chaque participant intervient quand il le souhaite (*e.g.* « *désolé je fumais une clope* ») et peut consulter à

3. Ce point est sujet à débat : selon le canal de communication, la plate-forme technique de la conversation peut permettre aux participants de « personnaliser » leurs contributions, par exemple en modifiant la forme visuelle de leurs messages (gras, italique, couleurs...).

sa guise l'historique de la conversation courante. Par conséquent, on qualifie cette modalité hybride de *pseudo-synchrone*.

3.2 Spécificités des modalités étudiées

Il existe un large nombre de modalités de CMR, qui peuvent inclure des sources telles que Twitter et Wikipedia, cependant toutes ne sont pas massivement utilisées pour l'aide à la résolution de problèmes. Nous nous intéressons particulièrement aux chats, aux forums et aux courriels, qui sont les canaux favoris des plate-formes communautaires d'entraide en ligne. Les courriels, et dans une moindre mesure les chats, sont également des modalités fréquemment utilisées par les systèmes de gestion de la relation client. Toutes ces modalités ont en commun le fait d'être des modalités écrites, ainsi que le fait d'être potentiellement multi-participants. Nous décrivons ici leurs spécificités. La table 3.1 compare les principales caractéristiques de ces modalités.

	Courriels	Forums	Chats
Synchronicité	Asynchrone	Asynchrone	Pseudo-synchrone
Longueur des messages	Longs	Longs	Courts
Structure	Arborescente	Linéaire	Linéaire
Zones de citation	Oui	Oui	Non
Multi-participants	Variable	Oui	Variable

Tab. 3.1.: Caractéristiques des modalités étudiées.

3.2.1 Courriels

Les courriels sont d'abord utilisés pour la communication privative d'un ordinateur à un autre, entre deux parties, mais le système peut être utilisé pour des communications de groupe via une liste de diffusion. Le dispositif technique des listes de diffusion permet de distribuer un courriel à un ensemble défini de destinataires, et permet également souvent l'archivage des messages et leur accès public. Les messages des forums comme ceux des listes de diffusion sont donc consultables en ligne, cependant dans le cas des secondes les message sont reçus directement dans la boîte mail de l'usager. Cete particularité cause une implication plus forte du participant (Akrich, 2012).

Un bref examen de n'importe quel corpus de courriels permet de constater que les messages présentent plusieurs caractéristiques qui les rendent très différents à la fois des transcriptions tirées de conversations parlées et des autres modalités étudiées. D'abord, le fait que les messages ne soient pas entièrement constitués de contenu « neuf ». Pour Lampert et al. (2009), les courriels peuvent être découpés en

trois zones : les zones de locution (*sender zones*), qui contiennent le texte écrit par l'expéditeur ; les zones de contenu cité (*quoted conversation zones*), qui contiennent à la fois le contenu retransmis d'autres conversations et le contenu cité du message auquel l'auteur répond ; et les zones d'encadrement (*boilerplate zones*) qui contiennent le contenu réutilisé sans modification dans plusieurs messages, comme la signature ou les coordonnées de l'auteur. Si l'analyse de la conversation peut profiter de l'extraction d'informations tirées des zones d'encadrement, et si les zones de citations peuvent aider des systèmes à lier les messages entre eux, ce sont surtout les zones de locution qui nous intéressent si l'on cherche à analyser les conversations en termes d'actes de dialogue.

En termes de contenu, les courriels sont typiquement constitués de plusieurs énoncés, et sont accompagnés de nombreuses méta-données, comme le nom et l'adresse de l'expéditeur, de l'adresse du destinataire, d'une signature automatique, *etc*. Les courriels sont généralement constitués uniquement de texte mais ils peuvent comprendre du HTML, des images, et être édités pour que le style du texte soit profondément modifié (couleur, gras, *etc*.). La structure conversationnelle des listes de diffusion est intéressante puisqu'elle n'est pas nécessairement linéaire : chaque courriel faisant réponse, non pas à la liste de diffusion elle-même, mais à un message en particulier. Cette structure est illustrée par la figure 3.1.

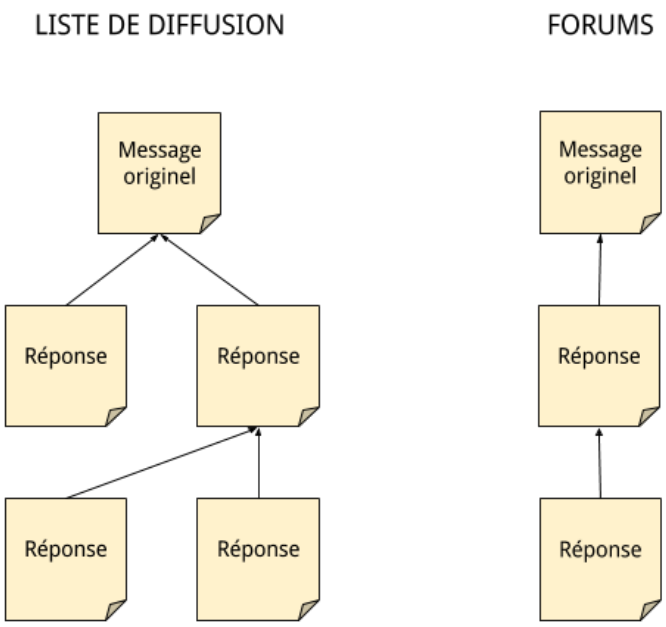

Fig. 3.1.: Comparaison entre le caractère arborescent d'une liste de diffusion et la structure linéaire d'un fil de forum.

3.2.2 Forums

Les forums sont similaires aux listes de diffusion par plusieurs aspects, mais leur structure conversationnelle diffère. S'il est possible de répondre à n'importe quel courriel, créant ainsi une structure de discussion arborescente, il n'est pas possible de répondre à n'importe quel message d'un forum : les messages sont répartis en fils de discussion, où ils se suivent linéairement. Typiquement, chaque fil de discussion correspond à un sujet donné, spécifié par l'auteur du premier message du fil. On peut donc distinguer deux types de messages : les messages initiaux, qui cadrent le sujet, et leurs réponses. Au sein de ces messages, Qadir et Riloff (2011) font la distinction entre actes de discours et texte expositif (qui apporte de l'information factuelle), et considèrent que les messages de forums diffèrent des documents monologues en ce qu'ils contiennent un mélange des deux, formant ainsi un genre « hybride ». Mais s'il est en effet courant de constater la présence de « messages-documents » sur certains forums, où l'auteur rédige du texte expositif sans caractère dialogique, ce n'est pas le cas dans les forums consacrés à l'entraide entre utilisateurs, où les fils correspondent systématiquement à des conversations.

Si des sites comme Stack Exchange [4] ou Reddit [5] sont généralement classés comme des forums, leur structure non-linéaire les rapproche structurellement des listes de diffusion. En effet, chaque nouveau message peut se rattacher à n'importe quel message antérieur, de manière à aboutir à une structure de conversation arborescente. Stack Exchange a également la particularité de traiter les messages différemment selon qu'il s'agisse d'une « réponse » ou d'un « commentaire ». Dans cette thèse nous adoptons une description restrictive des forums, qui exclut ces types de sites, ainsi que les *bulletin boards* et *image boards* qui ont également leurs propres particularités.

3.2.3 Chats

Les chats représentent également un canal fréquemment utilisé pour communiquer autour de la résolution de problèmes, et se distinguent des courriels et forums par leur caractère pseudo-synchrone. La plupart des chats sont organisés autour de « salons » où plusieurs participants peuvent se connecter et participer simultanément. Il existe aussi des systèmes de messagerie instantanée qui ne font intervenir que deux participants. Par exemple, les services après-vente proposent parfois des canaux de messagerie instantanée où un client peut converser instantanément avec un agent de l'entreprise. Dans le domaine de l'assistance, Stede et Schlangen (2004) notent qu'il existe des différences fondamentales entre les chats de nature exploratoire et axés vers la recherche d'information (*information-seeking chats*), comme par exemple ceux

4. `stackexchange.com`
5. `reddit.com`

où un client s'adresse à un agent pour obtenir plus d'informations sur un produit, et ceux orientés autour de l'accomplissement d'une tâche. Ils observent notamment que si les conversations tournées vers l'obtention d'information sont articulées autour d'une série de topiques et sous-topiques liés au domaine, les autres sont plutôt mues par un ensemble de sous-objectifs.

La forme de surface des messages de chat diffère de celle des forums et courriels par plusieurs aspects :

— Les messages sont souvent constitué d'un unique énoncé
— Grammaticalement, les énoncés sont souvent simples ou incorrects
— La ponctuation est fréquemment omise

Globalement, les chats sont la modalité la plus proche des conversations parlées. Greenfield et Subrahmanyam (2003) notent que les participants des salons de chats font souvent appel à des stratégies conventionnement propres à la communication en face à face. Comme, par exemple, utiliser le nom de l'interlocuteur pour signaler qu'un message lui est destiné, ou encore répéter des portions d'énoncés pour signifier à quoi un message répond. Mais les participants font aussi appel à des nouvelles stratégies propres à la modalité, comme l'utilisation de couleurs pour rendre l'auteur des messages plus simples à identifier, ou l'utilisation de codes conventionnels, comme « asl[6] » ou « brb[7] ».

3.3 ISO 24617-2 : étude d'applicabilité

En traitement automatique, les actes de dialogue sont largement utilisés dans les études sur les phénomènes conversationnels (Traum, 2000), ainsi que dans les travaux d'annotation de dialogues (Bunt et al., 2012) et le développement d'agents virtuels interactifs (Kopp et al., 2005; Swartout et al., 2006). Cependant, si les actes de dialogue sont régulièrement invoqués dans des travaux portant sur des données issues des communications médiées par les réseaux, les spécificités propres à ce type de conversations y sont peu explorées et la question de savoir si les outils d'analyse disponibles sont bien adaptés à ce type de conversations est rarement posée.

3.3.1 Cadre de l'étude

Notre objectif est la modélisation des actes de dialogue dans les conversations écrites médiées par les réseaux. Dans ce contexte, le schéma du standard ISO (Bunt

6. « *age sex location* »
7. « *be right back* »

et al., 2012) constitue une référence : fondé sur les mêmes principes que le schéma DAMSL (Core et Allen, 1997), il permet l'annotation précise des actes de dialogue (voir section 2.4.4). Dans ce chapitre, nous proposons une critique du standard ISO vis-à-vis de la prise en compte de quatre aspects des communications médiées par les réseaux : la gestion du multi-participant, les contraintes techniques, le caractère pseudo-synchrone et asynchrone de certaines modalités et la communication multi-canaux. Nous discutons ces spécificités par rapport au corpus multi-canaux (chat, courriel et forum) Ubuntu-fr (Hernandez et al., 2016), construit à partir de conversations orientées vers l'assistance aux utilisateurs, auquel nous consacrerons le chapitre 5.

Nous acceptons sans réserve le cadre conceptuel du standard. Notre critique porte avant tout sur la pertinence des dimensions et des fonctions proposées pour modéliser des communications médiées par les réseaux. Le schéma d'annotation du standard a donc été confronté à des données réelles pour permettre une analyse de son applicabilité. Les exemples utilisés dans cette section sont tous extraits du corpus Ubuntu-fr, cependant leur orthographe a été corrigée pour le confort du lecteur.

3.3.2 Dialogues et polylogues

Le standard a été pensé pour modéliser des dialogues entre deux participants, et n'est pas forcément approprié pour l'annotation de polylogues. Ce caractère transparaît dans la définition des dimensions sémantiques et de certaines fonctions communicatives.

Les dimensions portant sur le traitement cognitif d'un énoncé antérieur (FEEDBACK) et sur la production linguistique d'un message (COMMUNICATION MANAGEMENT) sont au nombre de quatre et se distinguent par le « *sujet* » ayant opéré le traitement ou produit le message : AUTO FEEDBACK pour désigner le traitement du locuteur (*e.g.* « *et donc ?* », « *Ah ouais d'accord* »), ALLO FEEDBACK pour désigner le traitement de l'interlocuteur (*e.g.* « *Tu vois ce que je veux dire ?* »), OWN COMMUNICATION MANAGEMENT pour désigner la communication du locuteur (*e.g.* « *(désolé pour les fautes)* ») et PARTNER COMMUNICATION MANAGEMENT pour désigner celle de son interlocuteur (*e.g.* « *Petit conseil, lorsqu'on ouvre une discussion le sujet ne se met JAMAIS en majuscule* »). Le caractère dialogique est aussi intégré au niveau de fonctions communicatives. Ainsi les commissives impliquent une action du locuteur (*e.g.* « *bah, je vais essayer* ») et les directives impliquent une action de l'interlocuteur (*e.g.* « *Pose ta question* »).

Dans une conversation multi-participant, le sujet sémantique d'un acte de dialogue peut prendre des valeurs autres que soi-même ou autre (l'interlocuteur). Nous observons ainsi des cas de locuteurs pluriels (*e.g.* ici dans une fonction directive :

« *demande-nous conseil à chaque fois que tu as une incertitude* », là dans une fonction commissive : « *si cela ne marche pas, on t'aidera à mettre à jour ton BIOS* »). Il en va de même pour le cas des interlocuteurs pluriels, qui se produisent souvent dans le cadre de demandes d'informations relatives à la tâche (*e.g.* « *avez-vous une solution à me proposer ?* ») ou de remerciements dans la dimension SOCIAL OBLIGATIONS MANAGEMENT (« *merci pour votre aide* »). L'auto-désignation plurielle est majoritairement utilisée par un participant pour désigner le groupe des participants susceptibles de venir en aide à son interlocuteur, dont il fait partie (*e.g.* « *on est là pour ça* »). L'allocutaire pluriel est généralement utilisé par le demandeur d'assistance pour désigner ce même groupe, dont il ne fait pas partie (*e.g.* « *vous avez une idée de la source de mon problème ?* »). Nous notons aussi des cas d'auto-désignation (*i.e.* « *nous* ») qui comprennent à la fois le locuteur et l'interlocuteur (*e.g.* « *ne nous enflammons pas :D* », « *comme ça on verra la même chose (j'ai également cet outil)* »). Le schéma d'interaction qui oppose un locuteur (singulier ou pluriel) à un interlocuteur (singulier ou pluriel) est le schéma privilégié puisque la notion d'acte de dialogue implique une dualité locuteur/allocutaire, et ce même en contexte multi-participants.

Nous observons aussi quelques cas de désignation d'un tiers (*e.g.* ici une suggestion relative à la tâche « *et puis, elle peut aussi essayer le pilote proprio* », là une information de fin de contact « *il est parti :)* »). Ce type de phénomène est très rare dans notre corpus (moins de cinq occurrences observées dans notre corpus en recherchant par les pronoms et adjectifs personnels, démonstratifs et possessifs). Parfois ce tiers est indéfini (*e.g.* « *si quelqu'un peut t'aider il se manifestera ;)* »). Nous observons aussi des cas de désignation d'interlocuteur indéfini (*e.g.* « *merci d'avance à tous ceux qui pourront m'aider* » ou « *je tiens à remercier pour leur dévouement les bénévoles de Ubuntu* »). Les fonctions communicatives concernées sont essentiellement relatives aux dimensions SOCIAL OBLIGATIONS MANAGEMENT (*e.g.* THANKING) et minoritairement à la dimension TASK. Il s'agit alors principalement des actes qui décrivent l'activité à un niveau méta-discursif[8], *i.e.* d'actes de demande d'aide ou d'engagement à aider.

Au vu de ces observations, il apparaît que plusieurs « *sujets* » peuvent être impliqués pour désigner celui qui accomplit l'acte ou qui est concerné par son exécution : le locuteur singulier, le locuteur pluriel distinct de l'interlocuteur, le locuteur pluriel comprenant l'interlocuteur, l'interlocuteur singulier ou pluriel et le tiers singulier ou pluriel. Dans la taxonomie proposée par le standard, les sujets sont directement intégrés aux définitions des fonctions et des dimensions. Il serait pertinent d'étendre le schéma de manière à pouvoir capturer les différents types de sujets des énoncés.

8. C'est-à-dire d'énoncés qui portent sur la discussion elle-même, comme par exemple « *répondez s'il vous plaît* ».

3.3.3 Contraintes techniques

Dans le cadre des communications écrites médiées par les réseaux, les logiciels employés imposent une unité logique minimale d'information qui transite. Il s'agit du message pour le chat, du post pour le forum et du courriel pour l'échange de courriers sur Internet. Nous parlerons globalement de message.

Dans le standard, des dimensions sont dédiées à la gestion du transfert de ces messages. C'est le cas des dimensions OWN COMMUNICATION MANAGEMENT et PARTNER COMMUNICATION MANAGEMENT, à travers lesquelles le locuteur signale une erreur de forme dans le message communiqué, et, éventuellement, la corrige (que l'erreur se trouve dans sa propre communication ou dans celle de son interlocuteur). Nous observons que les énoncés relevant de cette dimension sont quasiment inexistants dans les forums et les courriels du corpus. Enfin, c'est également le cas des dimensions AUTO FEEDBACK et ALLO FEEDBACK, via lesquelles le locuteur confirme sa bonne compréhension des messages reçus et s'assure de celle de son interlocuteur.

La problématique de gestion de transmission des données est centrale dans les sciences des réseaux et des télécommunications. Lorsque l'on observe le protocole de communication de référence, Transmission Control Protocol (RFC 793 [9]), mis en place dans les architectures réseaux pour assurer la qualité du service communication, on peut en retirer quelques fonctions communicatives spécialisées manquantes au standard. Outre la demande de disponibilité pour émettre ou recevoir un message et la signalisation de cette disponibilité, il devrait exister des fonctions spécifiques pour indiquer l'indisponibilité, demander la confirmation de l'émission ou de la réception d'un message, signaler le succès ou l'échec d'émission ou de réception d'un message (*e.g.* « *je viens de poster les infos concernant le CPU* », « *mince je t' avais posté une réponse toute à l'heure et j'ai eu moi aussi un bad gateway, je pensais que le message était passé quand même* », « *merci pour les trois courriels de réponse* »), signifier le début (*e.g.* « *en ce qui concerne les questions techniques afférentes à mon paquet je posterai sous peu* »), le retard (*e.g.* « *pas eu le temps de répondre* »), l'interruption, la reprise ou la fin d'une émission de messages à différents délais. À ces fonctions devraient s'ajouter des fonctions de gestion des réémissions ou de réception de doublons (*e.g.* « *heu pourquoi as-tu posté six messages identiques ?* »). Les exemples observés dans notre corpus confirment ce manque. Par ailleurs, il peut être utile de distinguer le cas où le locuteur indique sa disponibilité (cas prévu dans le standard) du cas où le locuteur s'exprime sur la disponibilité d'un autre participant (*e.g.* « *il est parti :)* » ou « *tu es absent ?* »).

Les dimensions FEEDBACK du standard informent et questionnent le traitement d'un énoncé antérieur par le locuteur ou l'interlocuteur. Différents niveaux de traitement

9. https://www.ietf.org/rfc/rfc793.txt

cognitif sont considérés : l'attention, la perception, l'interprétation sémantique, l'évaluation pragmatique, et l'exécution effective de l'acte portée par l'énoncé. Nous observons que les *feedbacks* concernant l'attention et la perception sont très peu présents par rapport aux traitements du niveau évaluation ou exécution. L'accès à l'historique de la conversation et l'aspect asynchrone, ou tout moins alterné de la communication, rendent quasiment inutile l'usage de ces niveaux. Le participant peut à sa guise prendre le temps de relire un message qu'il aurait de prime abord mal perçu.

Concernant la dimension dédiée à la gestion de la communication, les fonctions spécifiques du standard et les fonctions générales nous semblent suffisantes pour discuter la production matérielle (*e.g.* la réalisation linguistique) des messages. Nous relevons des signalements d'erreurs concernant la communication du locuteur (*e.g.* « *euh je voulais écrire :* ») ou de celle de l'interlocuteur (*e.g.* « *t'as un problème toi avec ton logiciel de chat* »), des directives (*e.g.* « *n'écris pas en SMS* »), des demandes d'information (« « *peut* » *ou* « *peu* » *?* »). Nous notons aussi des cas de signalement où le sujet est un tiers (*e.g.* « *vu son français très aléatoire... ;-)* ») ce qui renforce notre conviction qu'il faut représenter spécifiquement ce type d'informations.

Nous notons que certains systèmes offrent la possibilité d'éditer les messages antérieurs. L'acte de correction peut donc être davantage qu'une mise à jour de la pensée du locuteur à un temps donné. Il peut aussi modifier littéralement des propos antérieurs produits (*e.g.* « *et j'ai mis à jour le post avec le contenu de mon fichier sites-available/default* »). La persistance de la conversation n'est donc pas garantie. Cela peut par ailleurs conduire à des risques d'incohérence (état de non vérité) de messages faisant référence à d'autres qui ne sont plus ce qu'ils ont été.

3.3.4 Asynchronicité et pseudo-synchronicité

Dans le standard, au moins deux dimensions sont concernées par la notion de synchronicité : la dimension TIME MANAGEMENT (gestion du temps) et la dimension DISCOURSE MANAGEMENT (gestion de la structure du discours).

On retrouve dans notre corpus les deux fonctions spécifiques définies dans la dimension TIME MANAGEMENT, à savoir : une fonction qui vise à gagner un peu de temps (*e.g.* « *mhhhh* ») et une autre qui marque explicitement une pause (*e.g.* « *je te ping quand je suis dispo* »). Les fonctions générales du standard nous permettent de décrire les demandes d'information (*e.g.* « *tout à l'heure, c'est quand ?* »), les directives (*e.g.* « *le repas est prêt, donne moi vingt minutes* »), les réponses (*e.g.* « *dans une ou deux heures* ») et les acceptations (*e.g.* « *ok ok, même trente si tu veux* »). Cependant, il n'existe pas de fonction spécifique qui permettrait de gérer les situations de reprise après interruption, présentes dans le corpus (*e.g.* « *désolé pour le retard*

(pas d'alerte) », « *re* », « *désolé de réagir aussi tard* »). D'une manière générale, les fonctions de gestion du temps sont très rares dans notre modalité pseudo-synchrone, et presque complètement absentes des modalités asynchrones.

Concernant la dimension DISCOURSE MANAGEMENT, le standard définit des fonctions spécifiques permettant d'introduire ou de changer de sujet, ou d'annoncer des événements tels que la fin d'une conversation. Les utilisateurs de forums ont la possibilité de spécifier un sujet (ou « topique ») à la création du fil de conversation, et les utilisateurs de listes de diffusion peuvent le faire à l'envoi de chaque courriel. Pour ce qui est des salons de chats, les conversations se produisent en flux continu, sans indication formelle du début ni de la fin de conversation. Différents sujets peuvent être discutés en même temps sur les salons de chats, ce qui produit des conversations « emmêlées » les unes avec les autres (Riou et al., 2015). Pour notre corpus, nous avons retenu comme information contribuant à la définition d'un sujet : l'expression d'une motivation (*e.g.* « *j'aurai besoin d'un coup de pouce pour une installation Ubuntu sur une vieille tour* »), l'exposé d'un problème (*e.g.* « *je pense avoir un problème de pilote de carte graphique* ») et les demandes d'information et d'action.

L'aspect asynchrone des listes de diffusion et des forums conduit à un étalement dans le temps des conversations. Il serait donc pertinent de disposer de fonctions spécifiques à la dimension DISCOURSE MANAGEMENT pour assurer le suivi d'un sujet. Ces fonctions, par un marquage explicite, permettraient de cadrer un propos dans un sujet non clôturé (*e.g.* « *Pour un cours, moi je suggère de regarder dans Youtube* ») et la réintroduction d'un sujet clôturé (*e.g.* « *tu te souviens de mon problème de connexion réseau en statique, hier ?* »). Malgré l'existence du champ « sujet » pour les courriels et les forums, nous observons un usage des fonctions d'introduction de sujets dans tous les canaux. Les fonctions de suivi et de réintroduction de sujet sont néanmoins principalement observées dans les courriels et les forums. La fonction de suivi s'observe aussi dans les chats lorsqu'un locuteur émet plusieurs demandes par exemple.

La dimension SOCIAL OBLIGATIONS MANAGEMENT évolue aussi pour appréhender le traitement différé des messages. La fonction THANKING ne fait pas la distinction entre les remerciements opérés en réaction à une contribution, par exemple un « *merci !* » qui suivrait une réponse utile à une question, et les remerciements par anticipation qui concluent souvent les demandes d'assistance dans les conversations écrites en ligne. Une fonction spécifique serait nécessaire pour distinguer ces deux types d'énoncés qui sont porteurs d'une intention sensiblement différente.

3.3.5 Conversations multi-canaux

Les réseaux permettent l'interconnexion de services, notamment d'échanges de contenus, et rendent ainsi possibles une communication sur plusieurs canaux de manière simultanée ou différée. Ainsi, via une URL, un participant peut mettre à disposition des autres participants une image, un texte, une vidéo, une page web, *etc*. Ce contenu peut compléter son propos, voire constituer son propos (*e.g.* « *tu peux pastebiner ?* », « *Je te copy/paste le contenu* », « *j'ai ce message suivant :* <LIEN VERS UN FICHIER TEXTE> », « *chez moi ça fonctionne :* <LIEN VERS CAPTURE D'ÉCRAN> »).

Les systèmes de communication de type chat permettent généralement de jongler d'un canal chat multi-participants à un canal chat privé (*e.g.* « *si tu as besoin de conseils là-dessus je peux te donner quelques infos en private* », « *je t'écris en MP*[10] »). De manière différée, on peut observer des conversations qui se poursuivent sur plusieurs canaux : d'une conversation dans un forum à une conversation par courriel (*e.g.* « *je traîne le sujet qui suit depuis plusieurs semaines, voir lien :* <LIEN VERS UN AUTRE FIL DU FORUM> ») ou du chat vers la documentation (*e.g.* « *j'ai repris ton problème dans la doc, ça va aider les gens avec un problème similaire* »). Le sens de communication s'élargit. Les technologies de communication en ligne donnent un pouvoir d'ubiquité en permettant d'être sur plusieurs canaux en même temps (*e.g.* dans une situation où deux participants se trouvent sur le canal chat, l'un dit : « *continue donc sur # ubuntu-fr-doc* », ce à quoi le second répond : « *j'y suis maintenant* »).

Ces observations mènent à penser que pour traiter correctement les communications médiées par les réseaux, une nouvelle dimension est nécessaire pour encadrer les actes dédiés à la gestion des canaux de communication. Cette dimension serait constituée d'actes spécifiques, comme le signalement d'un changement de canal ou la présentation d'un hyperlien.

3.4 Conclusion

Dans ce chapitre, nous avons montré que si l'étude des communications médiées par les réseaux était un domaine très prolifique, la recherche en actes de dialogue est restée majoritairement concentrée sur la communication orale. Néanmoins, nous avons vu que les réseaux ne constituent pas nécessairement un handicap à la communication. La modélisation de ces conversations doit en revanche prendre en compte les stratégies de communication spécifiques qui ont été développées autour des différentes modalités des CMR, ainsi que les méta-données fournies par leurs

10. Message privé.

supports techniques, pour capturer correctement les actes de dialogue qui y sont accomplis par les utilisateurs.

Nous avons procédé à une étude d'applicabilité du standard ISO 24617-2 pour l'annotation des actes de dialogue. Nous avons montré que le standard était plus adapté à l'annotation de dialogues que de polylogues. Or, la plupart des conversations se produisant dans les listes de diffusion, les salons de chat et les forums sont des conversations multi-participant. Pour mieux modéliser les conversations de ces modalités des CMR, le schéma devrait être adapté pour prendre en compte des situations où de multiples participants qui interagissent entre eux peuvent faire référence à des tiers ou à des locuteurs ou allocutaires pluriels.

De même, nous avons montré que l'aspect technique des systèmes de communication en ligne limite fortement la pertinence de certaines dimensions sémantiques du standard, telles que OWN COMMUNICATION MANAGEMENT et PARTNER COMMUNICATION MANAGEMENT, qui sont pensées pour des dialogues oraux où les participants contrôlent mutuellement leur communication en temps réel. De plus, ce même aspect technique des CMR ouvre la porte à de nouveaux besoins, comme ceux d'actes de dialogue signifiant l'échec de l'émission d'un message ou signalant un problème technique avec le canal de communication.

Nous avons aussi montré que l'aspect asynchrone ou pseudo-synchrone des modalités étudiées avait un impact sur la pertinence de certains éléments du schéma. Ainsi, les dimensions et fonctions de gestion du temps sont quasiment inutiles lorsque la conversation se produit dans un contexte asynchrone. De même, le caractère asynchrone des forums et des listes de diffusion pousse les participants à accomplir des actes propres à ce contexte, comme par exemple remercier son interlocuteur par avance pour sa future réponse.

Enfin, nous avons montré que les réseaux ont ouvert la voie à des comportements nouveaux et sans équivalent en face-à-face, comme le fait de pouvoir communiquer simultanément sur plusieurs canaux à la fois. Les actes qui sont propres à ces comportements, comme le fait d'envoyer un lien vers une ressource externe (vidéo, texte, autre conversation), n'existent pas dans le standard. Il faudrait y remédier pour pouvoir modéliser correctement et exhaustivement les conversations des modalités de l'écrit en ligne.

Toutes ces réflexions mènent à la conclusion que pour correctement modéliser les conversations qui sont l'objet de cette thèse, une nouvelle taxonomie est nécessaire. Nous nous basons donc sur les éléments de ce chapitre pour développer une taxonomie des actes de dialogue dans les conversations écrites en ligne porteuses de demandes d'assistance. Cependant, la communauté regorge déjà de taxonomies en

tous genres, et en proposer une énième peut se révéler futile si aucun effort n'est fait pour la rendre « compatible » avec les autres, et notamment avec le standard ISO. Le chapitre qui suit porte donc sur la question de l'interopérabilité entre taxonomies.

4 Interopérabilité taxonomique : enjeux et proposition

Table des matières

4.1 Introduction 63

4.2 Standardisation vs Interopérabilité 64
4.2.1 Le standard ISO 65
4.2.2 Interopérabilité entre taxonomies 65

4.3 Méta-modélisation 66
4.3.1 Intuitions 67
4.3.2 Caractéristiques méta-modélisables d'une taxonomie 67
4.3.3 Le méta-modèle 68
4.3.4 Formalisation des traits 69
4.3.5 Intérêts 71
4.3.6 Extraction des traits primitifs 73

4.4 Cadre expérimental 74
4.4.1 Corpus et taxonomies 74
4.4.2 Méta-modèle expérimental 75

4.5 Expériences 75
4.5.1 Conversion d'annotations 75
4.5.2 Classification inter-taxonomique 77
4.5.3 Méthode 78
4.5.4 Résultats 80

4.6 Autres usages 81

4.7 Conclusion 82

4.1 Introduction

Les actes de dialogue constituent un aspect fondamental du domaine de l'analyse du dialogue, et la disponibilité des annotations qui leur sont associées est essentielle pour les composants d'apprentissage statistique de nombreuses applications, telles que les agents conversationnels, les outils de *e-learning* ou les systèmes de gestion de relation client. Cependant, en fonction des objectifs applicatifs et de recherche, il peut être difficile d'obtenir des annotations pertinentes. Dans ce chapitre, nous détaillons une nouvelle méthode destinée à limiter les coûts de construction de systèmes de classification des actes de dialogue basés sur l'apprentissage machine. Les méthodes de classification supervisées sont la norme pour ce type de tâche (Tavafi et al., 2013), et comme l'annotation de nouvelles données est souvent une entreprise coûteuse et complexe, rendre les données existantes réutilisables autant que possible serait un avantage certain pour de nombreux chercheurs.

Plusieurs corpus annotés en termes d'actes de dialogue sont disponibles pour les chercheurs, comme Switchboard [1] (Godfrey et al., 1992), MapTask [2] (Anderson et al., 1991), MRDA [3] (Shriberg et al., 2004), BC3 [4] (Ulrich et al., 2008) *etc*. La plupart de ces corpus sont annotés avec des taxonomies plus ou moins similaires entre elles. Par exemple, le corpus Switchboard est annoté en utilisant une variante du schéma DAMSL (Core et Allen, 1997), MapTask utilise sa propre taxonomie, et BC3 utilise celle du MRDA (Shriberg et al., 2004), elle-même basée sur DAMSL. Intuitivement, il paraît normal que différents chercheurs utilisent différentes taxonomies puisque les informations capturées par tel ou tel schéma d'annotation ne sont pas toutes pertinentes dans le cadre de n'importe quelle tâche, domaine ou modalité. De même, les taxonomies généralistes peuvent ignorer des informations qui seraient cruciales pour une tâche donnée, ou spécifique à un domaine particulier. C'est aussi la raison pour laquelle de nombreux chercheurs développent leurs propres taxonomies, ou utilisent alternativement une variante ou version simplifiée d'une taxonomie existante. Par exemple, Cohen et al. (2004), développent leur propre taxonomie d'« actes de discours de courriels », Kim et al. (2010) font de même pour les forums du site CNET [5], Qadir et Riloff (2011) utilisent une taxonomie dérivée de celle de Searle (1976), tandis que Ivanovic (2005) et Raymond et al. (2007) utilisent une variante de DAMSL. Parfois, les nouvelles taxonomies sont utilisées pour l'annotation de données qui finissent par n'être utilisées que le temps de quelques expériences, et les données elles-mêmes ne sont pas mises à disposition de la communauté.

1. www.isip.piconepress.com/projects/switchboard/
2. groups.inf.ed.ac.uk/maptask/
3. www1.icsi.berkeley.edu/~ees/dadb/
4. cs.ubc.ca/cs-research/lci/research-groups/natural-language-processing/bc3.html
5. www.cnet.com/forums/

Tout cela représente un gaspillage considérable de ressources et de temps, et est à l'origine d'une problématique de taille. Annoter des données en termes d'actes de dialogue est coûteux et prend du temps, pourtant la plupart des annotations qui résultent de ces efforts ne sont pas utilisées autant qu'elles le pourraient parce que chacun utilise sa propre taxonomie, ou est investi dans son propre domaine. Il y a un réel besoin de corpus divers mais partageant les mêmes types d'annotations, comme le démontrent les efforts significatifs qui ont été récemment investis dans le développement de la Tilburg DialogBank (Bunt et al., 2016). Ce projet a pour but de publier des annotations pour plusieurs corpus connus en utilisant le langage d'annotation DiAML du standard ISO 24617-2 (Bunt et al., 2012). Mais bien que ce projet soit louable et utile, il est important de garder à l'esprit que le standard n'est pas la réponse à toutes les tâches et à tous les problèmes qui peuvent se présenter. Il y a trop d'informations potentielles à annoter dans les dialogues pour pouvoir raisonnablement espérer un jour aboutir à un schéma d'annotation parfaitement exhaustif et indépendant du domaine. Bien qu'il s'agisse d'un standard, aucun standard ne sera jamais suffisant pour couvrir toutes les situations de dialogue possibles, et aucun standard ne peut être utilisé pour toutes les tâches imaginables en analyse du dialogue. Bien que le standard propose des règles et des consignes pour étendre le schéma d'annotation, les appliquer revient à créer une nouvelle taxonomie, plus ou moins similaire à l'originale.

Par conséquent, plutôt que tenter de résoudre le problème de la réutilisabilité des corpus en proposant un standard plus complet ou plus efficace, ce qui serait au-delà de nos capacités, nous proposons une approche différente : l'adoption d'un méta-modèle pour l'abstraction des taxonomies d'actes de dialogue. Le méta-modèle que nous détaillons est construit en décomposant les labels des actes de dialogue en traits fonctionnels primitifs, que nous définissons comme des aspects des actes qui peuvent être capturés par différents labels dans différentes taxonomies. Dans ce chapitre, nous démontrons qu'il est possible d'utiliser un méta-modèle de taxonomies pour convertir des annotations, mais aussi qu'un tel modèle peut être utilisé pour entraîner des systèmes de classification sur un corpus annoté à partir d'une taxonomie différente de celle dont les annotations qu'il produit sont tirées. C'est-à-dire qu'il permet théoriquement de produire automatiquement des annotations pour une taxonomie pour laquelle il n'existe aucune annotation.

4.2 Standardisation vs Interopérabilité

Dans cette section, nous présentons les efforts de standardisation effectués ces dernières années, et en montrons leurs limites. Nous évoquons les avantages présentés par une approche plus axée vers l'interopérabilité entre taxonomies, et montrons qu'un méta-modèle serait un outil pertinent pour progresser dans cette voie.

4.2.1 Le standard ISO

Comme nous l'avons mentionné, une approche possible pour répondre au manque d'interopérabilité des taxonomies d'actes de dialogue est le développement de nouveaux standards et de leurs ressources associées. Dans cette perspective, la Tilburg DialogBank (Bunt et al., 2016) constitue le plus récent effort pour proposer des données annotées génériques et fiables à la communauté. Il s'agit essentiellement de ressources langagières contenant des dialogues de diverses sources re-segmentées et ré-annotées en conformité avec le standard ISO 24617-2. Les dialogues proposés proviennent de divers corpus, tels que MapTask ou Switchboard.

Les efforts de ses auteurs sont basés sur leur conviction que le standard est à la fois complet et indépendant du domaine et de l'application. Cependant, nous sommes convaincus que les efforts de standardisation pourraient bénéficier d'outils efficaces pour améliorer l'interopérabilité des annotations existantes qui ne sont pas conformes aux spécifications du standard. D'abord, parce que même s'il est vrai que le standard est plus complet sémantiquement et moins dépendant du domaine et de l'application que les autres schémas d'annotation existants, comme démontré par Chowdhury et al. (2016), il n'est pas universel. Par exemple, un chercheur travaillant comme nous le faisons sur des conversations écrites extraites de forums en ligne passerait à côté d'importants aspects du discours en ligne en utilisant le standard, tels que la présence de liens hypertextes ou la possibilité pour les participants de changer dynamiquement de canal de communication. Dans le même temps, il ou elle serait encombré par un nombre significatif de dimensions et de fonctions communicatives qui sont presque absentes de ses données, comme les fonctions des dimensions de gestion du temps ou du tour de parole. De plus, nous sommes convaincus que les dialogues sont si complexes et si riches que l'on ne peut pas espérer qu'un seul schéma d'annotation soit capable de capturer toutes les informations qui peuvent être pertinentes pour n'importe quel système d'analyse. Il y aura toujours des informations manquantes qui auraient pu être utiles pour quelque chose, et la poursuite de l'exhaustivité à tout prix peut parfois mener au développement d'outils lourds et peu pratiques. De telles ambitions peuvent mener au phénomène connu sous le nom de *feature creep*, qui correspond à l'ajout continuel de nouveaux composants qui ne sont utiles que pour des cas d'utilisation très spécifiques et bien au-delà des objectifs initiaux de l'outil, ce qui peut avoir pour conséquence de produire des outils compliqués et opaques plutôt que simples et efficaces.

4.2.2 Interopérabilité entre taxonomies

Il est donc peut-être préférable de construire différentes taxonomies pour différentes situations, et de concentrer les efforts déployés vers plus d'interopérabilité. Petu-

khova et al. (2014) proposent un bon exemple de tels efforts en fournissant une méthode pour effectuer des requêtes sur les corpus MapTask et AMI au travers d'annotations du standard. Ils rapportent notamment que l'aspect multi-dimensionnel du schéma rend leur méthode plus précise : *i.e.* , encoder les actes de dialogue avec plusieurs fonctions communicatives est une bonne façon de rendre la taxonomie plus interopérable. En effet, comme nous l'avons établi auparavant, le fait que les énoncés puissent généralement avoir plusieurs fonctions est bien établi dans la littérature. Traum (2000) note qu'il y a deux façons de capturer cette multiplicité dans une taxonomie : soit en annotant chaque fonction séparément, ce qui demande que chaque énoncé puisse porter plusieurs annotations simultanément, soit en regroupant ces fonctions en ensembles cohérents et en codant les énoncés avec les labels complexes qui en résultent.

La première option est celle préférée par le standard, puisqu'elle a l'avantage de réduire la taille de l'ensemble considéré pour chaque décision d'annotation, et parce qu'elle capture plus efficacement la multi-fonctionnalité des énoncés. L'idée étant qu'il est préférable d'utiliser plusieurs ensembles d'annotations mutuellement exclusifs qu'un ensemble au sein duquel les annotations partagent souvent différents traits fonctionnels. Par exemple, considérons le dialogue suivant [6] :

1. Participant 1 : « *Now take a left* »
2. Participant 1 : « *And then uuh* »
3. Participant 2 : « *Turn right ?* »
4. Participant 1 : « *Yeah* »

Avec le standard, il serait possible d'annoter le second énoncé avec à la fois les labels INSTRUCT et STALLING. Cependant, dans le schéma d'annotation du HCRC (utilisé pour MapTask), le label INSTRUCT est séparé de UNCODABLE, et par conséquent l'énoncé ne peut être annoté qu'avec l'un ou l'autre de ces labels. Concrètement, les taxonomies multi-dimensionnelles séparent les traits fonctionnels des énoncés pour résoudre ce type de problème. Mais cette séparation n'a pour but que de faciliter l'annotation d'énoncés au sein d'une même taxonomie : pour pouvoir rendre un schéma d'annotation compatible avec les autres, il faut identifier les traits fonctionnels au sein des labels d'une même dimension.

4.3 Méta-modélisation

Dans cette section nous présentons l'idée de construire un méta-modèle des taxonomies de fonctions communicatives. Un tel modèle a pour ambition de permettre la

6. Cet exemple est tiré du corpus MapTask.

modélisation des taxonomies elles-mêmes, au travers de la caractérisation de leurs labels en traits primitifs, ce qui ouvrirait la porte à plusieurs possibilités ambitieuses qui sont détaillées ici.

4.3.1 Intuitions

La première intuition, qui constitue la base de ce travail, est que les labels des taxonomies de fonctions communicatives - comme celles de DIT++, de DAMSL, de MapTask *etc.* - peuvent être caractérisées avec des traits primitifs. Par exemple, on peut dire du label ANSWER de DAMSL qu'il a pour caractéristique de devoir avoir été élicité par un allocutaire. On propose alors le trait primitif « élicité » qui sera alors l'un des constituants de la définition du label ANSWER.

La deuxième intuition, c'est que les taxonomies ont une cohérence interne en termes de traits. Ainsi, si la définition du label ANSWER de DAMSL inclut le trait « élicité », il y a de fortes chances pour que l'on retrouve ce même trait dans d'autres labels, ou son opposé, dans le cas où un label est en partie défini par le fait que les énoncés qu'il capture ne sont *pas* élicités.

Enfin, la dernière intuition est que ces traits primitifs sont souvent partagés entre différentes taxonomies de différents auteurs, même si ce n'est pas toujours explicite et que différentes taxonomies n'emploient pas forcément le même vocabulaire dans leurs définitions. Compiler tous les traits d'une taxonomie permet ainsi de comparer ses labels avec ceux d'autres taxonomies pour lequel le même travail a été fait.

4.3.2 Caractéristiques méta-modélisables d'une taxonomie

Une taxonomie peut être caractérisée à différents niveaux :

1. Les définitions des concepts de base
2. Le choix de couverture des différents aspects de la conversation
3. Les définitions des différents labels
4. Le niveau de granularité de la taxonomie
5. Les relations de spécialisation entre les labels
6. Les relations d'exclusivité et de compatibilité entre les différents labels

En qualifiant les labels des différentes taxonomies en termes de traits primitifs, on peut modéliser les schémas selon un méta-modèle qui modélise les niveaux 3, 4, 5 et 6. Le niveau 1 ne peut pas être résolu par un méta-modèle et le niveau 2

dépend du choix des dimensions sémantiques, dans le cas d'une taxonomie multi-dimensionnelle.

Par exemple, la figure 4.1 représente trois labels de DAMSL et les labels les plus proches de la taxonomie du standard. Elle montre la présence de traits primitifs dans les énoncés en fonction de leurs labels. En caractérisant ces labels par des traits primitifs, on obtient une grille de lecture rapide des taxonomies, et les différences entre les labels apparaissent plus simplement. Le niveau 3 est représenté par chaque ligne, qui correspond à la définition du label. Le niveau 4 est représenté par le choix du niveau de précision des traits primitifs. Le niveau 5 peut être inféré des définitions : par exemple, on voit que le label STATEMENT de DAMSL est parent du label ASSERT parce qu'il est presque identique à l'exception d'une règle de trait plus restrictive : les énoncés annotés en ASSERT doivent être persuasifs, c'est-à-dire que le locuteur doit chercher à remplacer une croyance de l'allocutaire, alors que ce n'est pas une nécessité pour les énoncés annotés en STATEMENT. Pour ce qui est du niveau 6, lui aussi apparaît au travers de cette représentation : si une case verte sur une ligne est rouge sur une autre, alors les deux labels sont mutuellement exclusifs.

Fig. 4.1.: Exemple de méta-modèle pour trois labels du standard (haut) et DAMSL (bas). Les traits indiqués en vert, « A » (pour *Always*), sont toujours présents dans les énoncés porteurs du label (la définition inclut le trait), ceux en rouge, « N » (pour *Never*), ne sont jamais présents dans les énoncés porteurs du label (la définition inclut la négation du trait), et ceux en bleu peuvent être présents dans les énoncés (la définition n'inclut pas le trait). La dénomination des traits utilise plusieurs raccourcis : `S` signifie « Speaker » (locuteur), `A` signifie « Addressee » (allocutaire), `p` signifie « proposition énoncée » et ¬ symbolise la négation. Par conséquent, `S.believes(p)` peut être exprimé par « le locuteur pense que la proposition énoncée est vraie ».

4.3.3 Le méta-modèle

La manière dont les actes de dialogue doivent être définis a été longuement discutée dans la littérature. Traum (2000) soulève plusieurs questions à propos de différentes perspectives qui devraient être considérées quand on définit des actes de dialogue, telles que « est-ce que les taxonomies utilisées pour annoter des corpus de dialogues

doivent être dotées d'une sémantique formelle ? » ou « est-ce que la même taxonomie peut être utilisée pour différentes catégories d'agents ? ». L'objectif de ce chapitre n'est pas de promouvoir ou déprécier une approche plutôt qu'une autre, mais de suggérer une manière de les faire se rencontrer.

Nous postulons que la plupart des taxonomies d'actes de dialogue peuvent être généralisées en faisant usage de traits primitifs comme attributs définitoires de leurs labels. Par exemple, une réponse (ANSWER) dans le standard ne peut pas avoir un aspect porté sur la discussion d'action [7], mais une réponse dans DAMSL le peut. Dans les deux cas, le label ne peut être appliqué qu'à un énoncé élicité par l'allocutaire. On peut donc identifier quelques traits propres à ces labels et s'en servir pour définir ANSWER pour chacune des taxonomies. Le fait que la réponse doive être élicitée par l'allocutaire constitue un trait commun, et le fait que la réponse ne puisse pas participer à la planification d'action constitue un trait distinctif.

On définit un méta-modèle comme l'ensemble de tous les traits qui peuvent être utilisés pour définir tous les labels d'un ensemble donné de taxonomies. Quelques avantages d'un tel outil sont illustrés par la figure 4.1. Pour reprendre notre exemple précédent, on voit que les labels pour ANSWER sont faciles à comparer quand ils sont définies comme des ensembles de traits, et que le faire ne demande aucun discernement humain : dans les colonnes « `S.believes(p)` » et « `p.isInformation` », les cellules sont vertes pour le standard mais bleues pour DAMSL. Cela signifie qu'une réponse doit être honnête dans le standard, mais les réponses mensongères sont acceptées dans DAMSL. De plus, dans le standard une réponse doit être informationnelle - *i.e.* elle ne peut pas être dédiée à la discussion d'une action, ni constituer un acte déclaratif - ce qui n'est pas le cas dans le standard. Par exemple, répondre à une demande d'action peut être une réponse au sens de DAMSL mais pas au sens du standard. Un ordinateur pourrait les comparer, ce qui serait impossible en se basant simplement sur leurs définitions écrites. On peut observer dans la figure 4.1 que quand deux labels partagent tous leurs codes-couleur, ils sont essentiellement identiques, quand à certains endroits une cellule qui est bleue pour l'une est rouge ou verte pour l'autre, le deuxième label est une spécialisation (ou « fille ») de la première, et quand une cellule est verte pour l'une mais rouge pour l'autre, ils sont mutuellement exclusifs.

4.3.4 Formalisation des traits

Nous avons choisi de formaliser les traits en utilisant quelques composants basiques qui peuvent être liés entre eux : les *participants* ((`S`)peaker, (`A`)ddressee) utilisent des *verbes* (*e.g.* `provides()`, `wants()`, `believes()`) pour manipuler des

7. *i.e.* Elle ne peut pas porter sur la planification d'une action, comme dans l'énoncé « d'accord je vais redémarrer mon ordinateur »

objets (*e.g.* (p)roposition, (f)eedback, (a)ction), et ces objets ont des *propriétés* (*e.g.* `isPositive`).

L'exemple suivant liste les traits du label AUTO NEGATIVE FEEDBACK du standard, destiné à marquer les énoncés fournissant des retours négatifs, comme « Je ne comprends pas » par exemple :

$$\texttt{S.provides(f)} \wedge \texttt{f.isAuto} \wedge \neg\ \texttt{f.isPositive}$$

Les traits sont séparés par des symboles de conjonction. Le premier signifie « le locuteur fournit un retour », le second « le retour concerne la compréhension d'un énoncé par le locuteur », et le troisième « le retour est négatif ».

Cette manière de formaliser les traits offre de multiples avantages. Notamment, elle permet d'éviter la redondance des traits, et elle permet l'utilisation d'opérateurs logiques (*e.g.* $\neg$, $\vee$, $\wedge$). De plus, utiliser un tel format rend possible le fait de décomposer les traits en sous-composants qui pourront ensuite être utilisés pour entraîner un classifieur (par exemple, la présence d'un objet (a)ction dans le trait). Nous avons aussi choisi de représenter les traits d'une manière réminiscente des prédicats logiques pour qu'ils puisent être analysés et évalués : bien que le fait de représenter le dialogue au sein d'un cadre logique est une idée qui a déjà été explorée dans la littérature (Sadek, 1991), nous n'avons pas rencontré de travaux tentant d'utiliser les représentations individuelles de labels d'actes de dialogue pour des tâches de reconnaissance automatique. Cependant, cet aspect de notre recherche - analyser des énoncés pour tenter de trouver des correspondances avec des définitions logiques - est hors de la portée de ce travail. Pour le moment, chaque trait est traité comme un booléen par les algorithmes développés et la convention de nommage n'impacte pas les expériences menées, *i.e.* « $\texttt{S.provides(f)} \wedge \texttt{f.isAuto} \wedge \neg\ \texttt{f.isPositive}$ » est équivalent à « $trait_a = vrai, trait_b = vrai, trait_c = vrai$ ».

Cependant, le principal avantage de cette formulation est qu'elle nous permet d'utiliser des concepts tels que « *belief* » ou « *feedback* » au travers de multiples traits, et d'implémenter des notions théoriquement fondées dans les briques de base du méta-modèle. Ces éléments reflètent les fondations conceptuelles des taxonomies comprises dans le méta-modèle. Dans celui utilisé dans ce travail, les traits primitifs utilisés indiquent que les chercheurs derrière DAMSL, le standard ISO et HCRC ont souscrit à une certaine vision de la structure du dialogue. En effet, les traits sont majoritairement construits autour des notions de croyance, de désir et d'intention (Grosz et Sidner, 1986; Georgeff et al., 1998), ainsi qu'autour de la notion linguistique de *connaissances communes* (Traum et Hinkelman, 1992). Cependant il est important de noter que le méta-modèle lui-même n'est pas motivé linguistiquement, et qu'il peut incorporer des éléments de n'importe quelle théorie. Par exemple,

si le méta-modèle devait intégrer les modes de réponse verbale (*Verbal Response Modes*) (Stiles, 1978), ses traits devraient nécessairement capturer des notions telles que le cadre de référence ou la source d'expérience. En pratique, les traits primitifs peuvent décrire les opérations sur les connaissances, les intentions et les croyances du locuteur et de l'allocutaire, ainsi que les caractéristiques des mécanismes de discussion d'action et de synchronisation dialogique.

4.3.5 Intérêts

Les intérêts d'un méta-modèle de taxonomies de fonctions communicatives sont multiples et s'orientent sur quatre axes : faciliter la compréhension des taxonomies existantes, faciliter la construction de nouvelles taxonomies, faciliter l'annotation (et surtout la ré-annotation) de données, et surtout un tel modèle peut permettre d'exploiter un corpus annoté pour une tâche de reconnaissance automatique des actes dans une taxonomie différente de celle du corpus.

Pour les utilisateurs de taxonomies, le modèle permettrait des représentations visuelles qui permettent d'appréhender plus facilement une nouvelle taxonomie s'ils sont déjà familiers avec une autre, en faisant apparaître immédiatement leurs différences. On peut également faire apparaître facilement les différences entre deux labels proches, et aider les utilisateurs à choisir quelle taxonomie est plus appropriée à son domaine en le laissant choisir les traits primitifs qui lui semblent importants.

Pour les constructeurs de taxonomie, une représentation des labels en termes de traits permet d'automatiser une grande partie du travail passé la sélection initiale des traits et leur répartition au sein de différents labels : les relations d'exclusivité et de hiérarchie peuvent être calculées automatiquement, et il est possible de détecter automatiquement si la taxonomie présente des incohérences internes : les cas d'ambiguïté entre labels sont éliminés facilement, et il est même possible de détecter simplement si la taxonomie ne couvre pas tous les énoncés qu'un annotateur peut rencontrer. Par exemple, admettons une taxonomie basée sur deux traits, x et y, et deux labels, A et B. Si le label A est défini par x et le label B par x et y, il est possible qu'un annotateur rencontre un énoncé qui corresponde au trait y mais pas au trait x, auquel cas aucun label ne correspond. Ce genre de situations apparaît fréquemment quand on annote un corpus en utilisant une taxonomie sophistiquée, et les décisions « au jugé » des annotateurs contribuent à réduire les accords inter-annotateurs et l'exploitabilité des annotations.

Pour les annotateurs, il est possible de réduire leur charge de travail, souvent importante, dans deux situations. La première, c'est celle où un annotateur cherche à annoter un corpus avec une taxonomie A alors qu'i a déjà été annoté avec une taxonomie B : même si les labels ne correspondent pas exactement, se concentrer

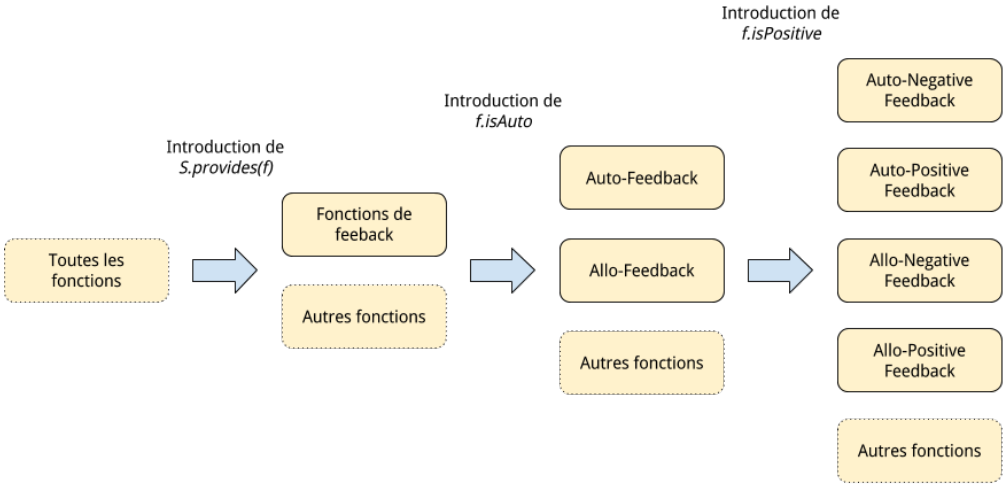

Fig. 4.2.: Processus d'extraction des traits primitifs pour les fonctions de *feedback*.

sur l'annotation de la seule information manquante - un certain nombre de traits absents des labels utilisés sur le corpus - est bien moins coûteux en temps que tout réannoter, même en s'inspirant des annotations existantes. Typiquement, lorsque ce type de tâche est accompli, les annotateurs définissent des ponts entre certains labels des deux taxonomies et effectuent une conversion automatique, et se contentent de réannoter les parties qui n'ont pas été couvertes. Mais opérer en termes de traits devrait permettre de convertir autant d'information, sinon plus, sans se fier à des ponts entre différents labels de différentes taxonomies qui introduisent souvent des erreurs lorsque leurs définitions ne sont pas exactement les mêmes (ce qui est courant, comme nous l'avons vu pour la définition de ANSWER entre DIT^{++} et DAMSL). La seconde situation pertinente, c'est celle où la définition d'un label est modifiée en cours de route, ce qui peut se produire lorsqu'un corpus est annoté avec une nouvelle taxonomie, et que les deux se construisent en parallèle. Au lieu de devoir repasser sur toutes les instances du label pour en choisir un nouveau, seule l'information manquante, en termes de traits primitifs, doit être ajoutée.

Enfin, la méta-modélisation des taxonomies devrait rendre possible l'annotation automatique par l'apprentissage machine de données annotées dans une taxonomie différente de la taxonomie de destination. En entraînant des classifieurs sur les traits des labels plutôt que sur les labels eux-mêmes, il est théoriquement possible d'apprendre des informations à partir d'un corpus que l'on peut projeter sur un autre corpus avant de réassembler ces informations en labels différents des labels d'origine. Et si la taxonomie d'origine ne couvre pas entièrement la taxonomie de destination en termes de labels, il devrait même être possible d'apprendre ces informations à partir de plusieurs taxonomies différentes.

4.3.6 Extraction des traits primitifs

Nous avons basé nos traits sur les définitions écrites exactes de leurs labels, telles que publiées dans la littérature par leurs auteurs. Nous avons cherché à obtenir le nombre minimal de traits primitifs pour permettre la distinction catégorique des actes de dialogue. Par exemple, pour le label AUTO-NEGATIVE FEEDBACK utilisé dans notre exemple précédent, la définition exacte telle que fournie dans les consignes du standard ISO 24617-2 est la suivante :

> *« Fonction communicative d'un acte de dialogue accompli par le locuteur, L, de manière à informer l'allocutaire, A, que le traitement de l'énoncé précédent par S a rencontré un problème. »* [8]

Théoriquement, n'importe quel nombre de traits peut être extrait d'une telle définition. Peut-être un trait signifiant qu'un énoncé porte une information, un autre pour signaler qu'il ne s'agit pas d'une information portée sur la tâche, une autre pour marquer l'énoncé comme potentiellement non-verbal *etc*. Notre formalisation de ce label est « `S.provides(f)` $\wedge$ `f.isAuto` $\wedge$ $\neg$ `f.isPositive` ». Pour atteindre ce résultat à partir de la définition écrite, nous avons utilisé un principe simple : de nouveaux traits ne devraient être introduits que pour permettre de distinguer le label de ses parents et pairs [9]. Le processus est représenté par la figure 4.2.

Chacun de ces trois traits est donc utilisé pour distinguer AUTO-NEGATIVE FEEDBACK des autres labels. « `S.provides(f)` » signifie que l'énoncé porte sur le traitement d'un énoncé précédent, et ainsi sert à distinguer les fonctions de *feedback* des fonctions générales [10]. « `f.isAuto` » signifie que le *feedback* porte sur le locuteur lui-même, et est utilisé pour distinguer le label de ALLO-NEGATIVE FEEDBACK qui porte sur le traitement d'un énoncé par l'allocutaire. « $\neg$ `f.isPositive` » signifie que le *feedback* signale un problème ; ce trait est utilisé pour faire la distinction avec AUTO-POSITIVE FEEDBACK. Aucun trait supplémentaire n'est nécessaire pour distinguer efficacement chacun de ces labels. Cette méthode a pour but de produire un méta-modèle approprié pour comparer des labels, et non pas de capturer toutes les informations représentées par une annotation.

8. *« Communicative function of a dialogue act performed by the sender, S, in order to inform the addressee, A that S's processing of the previous utterance(s) encountered a problem. »*
9. Si la taxonomie est « plate », *i.e.* non hiérarchique, tous les labels sont traités comme des pairs.
10. Bien que ce ne soit pas spécifié dans les consignes du schéma d'annotation, INFORM et dans certains cas ANSWER peuvent être considérés comme des parents de toutes les labels de *feedback*.

4.4 Cadre expérimental

Les expériences détaillées dans ce manuscrit portent sur la conversion et la reconnaissance d'actes de dialogue au travers de plusieurs taxonomies. Nous commençons par présenter les corpus utilisés pour nos expériences, et ensuite notre implémentation du méta-modèle.

4.4.1 Corpus et taxonomies

Corpus	Taxonomie	Annot.	Annot. ISO	Conversations	Polylogues
Switchboard	SWBD-DAMSL	200 000	750	Libre	Non
MapTask	HCRC	2 700	200	Tâche	Non

Tab. 4.1.: Comparaison des corpus utilisés.

Deux corpus semblent particulièrement pertinents pour notre tâche : le corpus Switchboard (Godfrey et al., 1992) [11] et le corpus Maptask (Anderson et al., 1991) [12]. La table 4.1 compare les deux corpus.

Switchboard est un très large corpus (plus de 200 000 énoncés), constitué d'une collection de plus de 2 400 dialogues téléphoniques entre 543 participants. Les dialogues ont tous impliqué deux participants auxquels ont été soumis un sujet de conversation aléatoire. Switchboard a été annoté avec le schéma SWBD-DAMSL (Jurafsky et al., 1997). DAMSL est le premier schéma d'annotation à implémenter une approche multi-dimensionnelle (*i.e.* qui permet d'appliquer plusieurs annotations à un seul énoncé) et est *de facto* un standard en analyse du dialogue. SWBD-DAMSL est une variante de DAMSL destinée à réduire la multi-dimensionnalité de ce dernier (Core et Allen, 1997). Une partie du corpus Swichboard, soit à peu près 750 énoncés, a également été annotée avec le standard ISO 24617-2 (Bunt et al., 2016).

Le corpus MapTask est également un corpus relativement large (plus de 2 700 annotations). Il fait intervenir deux participants autour d'une tâche de lecture et d'interprétation de carte : l'un d'entre eux, l'*instruction follower*, doit suivre les instructions de son interlocuteur, l'*instruction giver*, pour reproduire sur une carte la route qui lui est indiquée. Il a été annoté avec le système de codage du HCRC (Carletta et al., 1996), qui comporte douze labels. Une partie de ce corpus, un peu plus de 200 énoncés, a également été annotée en utilisant le schéma du standard ISO, ce qui en fait un candidat idéal pour notre première tâche, la conversion d'annotations d'une taxonomie à une autre.

11. `https://catalog.ldc.upenn.edu/ldc97s62`
12. `http://groups.inf.ed.ac.uk/maptask/`

4.4.2 Méta-modèle expérimental

Nous avons construit un méta-modèle pour les labels des taxonomies du standard, de SWBD-DAMSL et de HCRC de la manière décrite en section 4.3.6. Il contient 119 traits différents construits à partir de 3 types de participants, 19 verbes, 6 types d'objets et 39 propriétés d'objets. La table 4.2 compare la complexité des taxonomies. Comme le système du HCRC ne comporte que douze labels et est mono-dimensionnel, il se distingue par sa simplicité. La liste complète des traits du méta-modèle expérimental est disponible en annexe B.

Taxonomie	Labels	Dimensions	Traits
SWBD-DAMSL	42	1	41
Standard ISO	56	9	83
HCRC	12	1	18

Tab. 4.2.: Statistiques taxonomiques : nombre de labels, nombre de dimensions et nombre de traits requis pour les modéliser.

4.5 Expériences

D'abord, nous expérimentons avec la conversion d'annotations au sein du même corpus pour démontrer la capacité du méta-modèle à agir comme un pont entre taxonomies. Ensuite, nous présentons nos résultats avec des classifieurs inter-taxonomiques, entraînés sur des données annotées avec une taxonomie différente de celle dont les labels qu'ils produisent sont issus.

4.5.1 Conversion d'annotations

Dans le contexte de la construction de la Tilburg DialogBank, des efforts importants ont été consacrés à la réannotation de corpus avec le standard, comme cela a été le cas pour le corpus Switchboard (Fang et al., 2012). De telles entreprises ont rencontré certaines difficultés (Bunt et al., 2013). Un certain degré d'automatisation a été mis en œuvre, sous la forme d'alignements manuellement définis entre labels ayant des correspondances multivoques (*many-to-one*) ou monovoques (*one-to-one*). Notre expérience explore une nouvelle méthode automatique de conversion d'annotations, passant par un méta-modèle.

Pour cette expérience nous n'appliquons pas d'algorithme supervisé pour la classification des actes de dialogues. Nous nous contentons d'employer le méta-modèle à convertir des annotations d'une taxonomie à une autre sur les mêmes données. Puisque des énoncés du corpus Switchboard sont annotés à la fois avec des labels SWBD-DAMSL et du standard, nous l'utilisons dans cette expérience. Nous utilisons

également les énoncés du corpus MapTask qui ont été annotés avec à la fois la taxonomie HCRC et le schéma d'annotations du standard ISO 24617-2. Il est à noter cependant que si le texte annoté est le même, la segmentation des phrases en énoncés porteurs de dialogue n'est pas symétrique. Par conséquent, nous avons utilisé l'algorithme de Knuth-Morris-Pratt (*KNP search*) pour aligner les énoncés d'un corpus vers l'autre.

Les annotations de la taxonomie source sont d'abord converties en traits primitifs (l'ensemble des traits de tous les labels portés par l'énoncé), puis sont réassemblées en nouvelles annotations de la taxonomie cible (y compris l'annotation NONE). On cherche d'abord à effectuer la deuxième étape en calculant la similarité cosinus entre les traits du label original et les traits des labels de la taxonomie cible. Le système se charge ensuite de choisir le label ayant l'ensemble de trait le plus similaire à celui du label d'origine. Nous répétons ensuite l'expérience en utilisant un algorithme naïf de Bayes. Le système classifie les ensembles de traits en labels cibles. Le procédé est détaillé dans la figure 4.3.

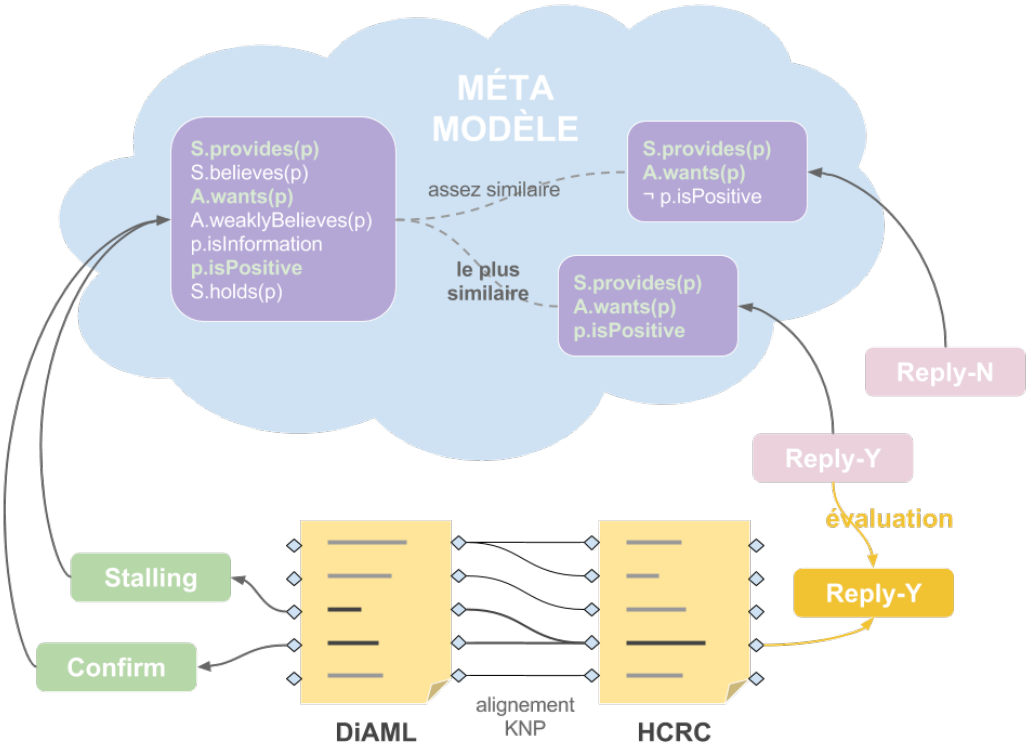

Fig. 4.3.: Représentation du procédé de conversion d'annotations utilisant le méta-modèle.

Ce système a ensuite été évalué en validation croisée, sur dix plis. Les résultats pour chacune des méthodes sont rapportées dans la table 4.3.

Nous comparons nos résultats à une approche de base simple, que nous appelons l'approche par conversion directe. Elle consiste à utiliser un classifieur naïf de Bayes entraîné sur des combinaisons de labels provenant des taxonomies sources et cibles. Ce classifieur ne fait pas du tout usage du méta-modèle.

Les résultats ont été évalués sur un échantillon de 746 actes de dialogue du corpus Switchboard (SWB) et 675 actes du corpus MapTask. Ils sont rapportés en table 4.3.

Corpus	Source → Cible	Exactitude
Approche de base : conversion directe		
MapTask	ISO → HCRC	0,60
MapTask	HCRC → ISO	0,70
SWBD	ISO → SWBD-DAMSL	0,60
SWBD	SWBD-DAMSL → ISO	0,78
Annotations converties via un algorithme basé sur une métrique de similarité		
MapTask	ISO → HCRC	0,60
MapTask	HCRC → ISO	0,76
SWBD	ISO → SWBD-DAMSL	**0,65**
SWBD	SWBD-DAMSL → ISO	0,87
Annotations converties via un algorithme basé sur un classifieur NaiveBayes		
MapTask	ISO → HCRC	**0,71**
MapTask	HCRC → ISO	**0,82**
SWBD	ISO → SWBD-DAMSL	0,64
SWBD	SWBD-DAMSL → ISO	**0,93**

Tab. 4.3.: Scores pour la conversion d'annotations.

Nous observons que les deux méthodes sont plus performantes que l'approche de base. Nous observons aussi qu'un simple classifieur entraîné sur très peu de données peut avoir de meilleures performances pour la tâche de conversion d'annotations que l'usage d'une simple métrique de similarité, à l'exception de l'expérience pour ISO vers SWBD-DAMSL, pour laquelle les résultats sont presque identiques. Cette expérience montre que le méta-modèle a un intérêt pour la tâche de conversion automatique d'annotations.

4.5.2 Classification inter-taxonomique

Trois ensembles de résultats sont rapportés pour cette expérience. Le premier est notre approche de base : il comprend les résultats pour une tâche classique de reconnaissance des actes de dialogue : sur dix itérations, un modèle est entraîné sur neuf dixièmes des données et évalué sur le reste. Cette méthode requiert des données annotées avec la taxonomie cible pour fonctionner. Les deux ensembles de résultats suivants sont ceux de systèmes qui tentent d'atteindre des niveaux d'exactitude

similaires, mais cette fois en utilisant des données provenant d'annotations d'une taxonomie différente de celle des annotations de sortie.

Le premier de ces systèmes, le système A, fonctionne comme suit :

1. Un modèle est entraîné sur les annotations correctes du corpus source annoté avec sa taxonomie source
2. Les annotations de la taxonomie source sont projetées sur les données du corpus cible
3. Les annotations projetées sont converties en annotations de la taxonomie cible en suivant la méthode décrite en sous-section 4.5.1

Le fonctionnement de ce premier système est détaillé en figure 4.4.

Le second système, le système B, tente d'apprendre directement les traits primitifs plutôt que les labels :

1. Un modèle est entraîné sur les *traits primitifs* corrects du corpus source annoté suivant la taxonomie source
2. Le corpus cible est automatiquement annoté en termes de traits primitifs, 3)
3. Les annotations de la taxonomie cible sont reconnues à partir des traits primitifs suivant la méthode décrite en sous-section 4.5.1

Le fonctionnement du système est détaillé en figure 4.5.

4.5.3 Méthode

Pour la classification, nous avons utilisé un SVM comme algorithme et les tokens, lemmes et étiquettes morpho-syntaxiques comme traits. Chaque type de trait est utilisé pour construire un modèle de type « sac-de-n-grammes ». Le classifier SVM a été implémenté via la librairie de classification et d'analyse de texte $liblinear$ (Fan *et al*, 2008). Nous avons utilisé un modèle de bigrammes sans en retirer les mots outils. Nous avons utilisé une heuristique basée sur *WordNet* (Miller, 1995) pour la lemmatisation et le *Stanford Toolkit* (Manning et al., 2014) pour effectuer l'étiquetage morpho-syntaxique.

Comme l'une de nos taxonomies est multi-dimensionnelle, permettant à chaque énoncé d'être étiqueté séparément (et optionnellement) dans différentes dimensions (*i.e.* catégories), un système qui chercherait à choisir un label parmi tous ceux de la taxonomie ne serait pas approprié. Plutôt qu'utiliser un SVM multi-classes

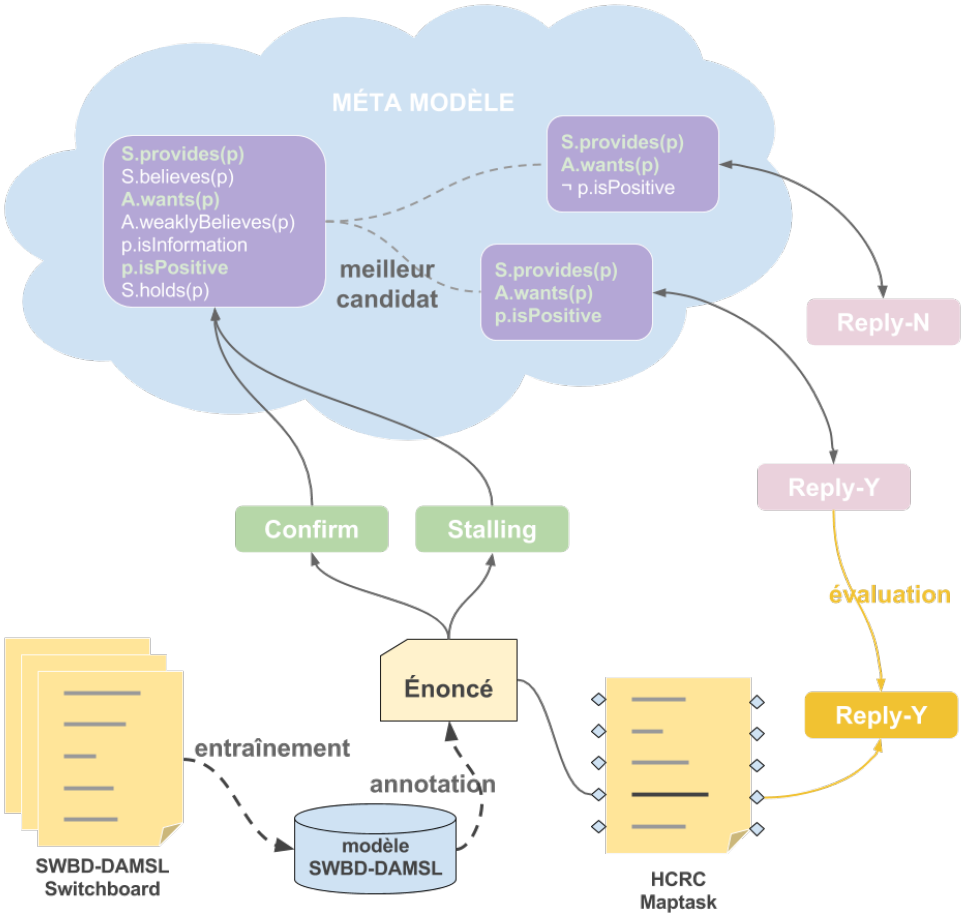

Fig. 4.4.: Représentation du procédé d'annotation automatique via le méta-modèle par le premier système (A).

sur l'ensemble des labels, ce qui ne serait pas non plus approprié puisque dans le standard seul un label par dimension peut être appliqué à un énoncé, nous choisissons de les séparer en ensembles dimensionnels. Nous rajoutons ensuite un label « nul » (NONE) à chaque ensemble de labels pour capturer les énoncés qui ne devraient recevoir aucune annotation dans une dimension particulière. Par conséquent, pour le standard les résultats fournis sont la moyenne des résultats obtenus pour chaque dimension. Si certains résultats semblent anormalement élevés, c'est parce que certaines dimensions - telles que ALLO FEEDBACK par exemple - seront principalement annotées avec le label NONE. Celà ne constitue pas un problème pour notre évaluation puisque le système utilisé comme approche de base bénéficie également de ce biais.

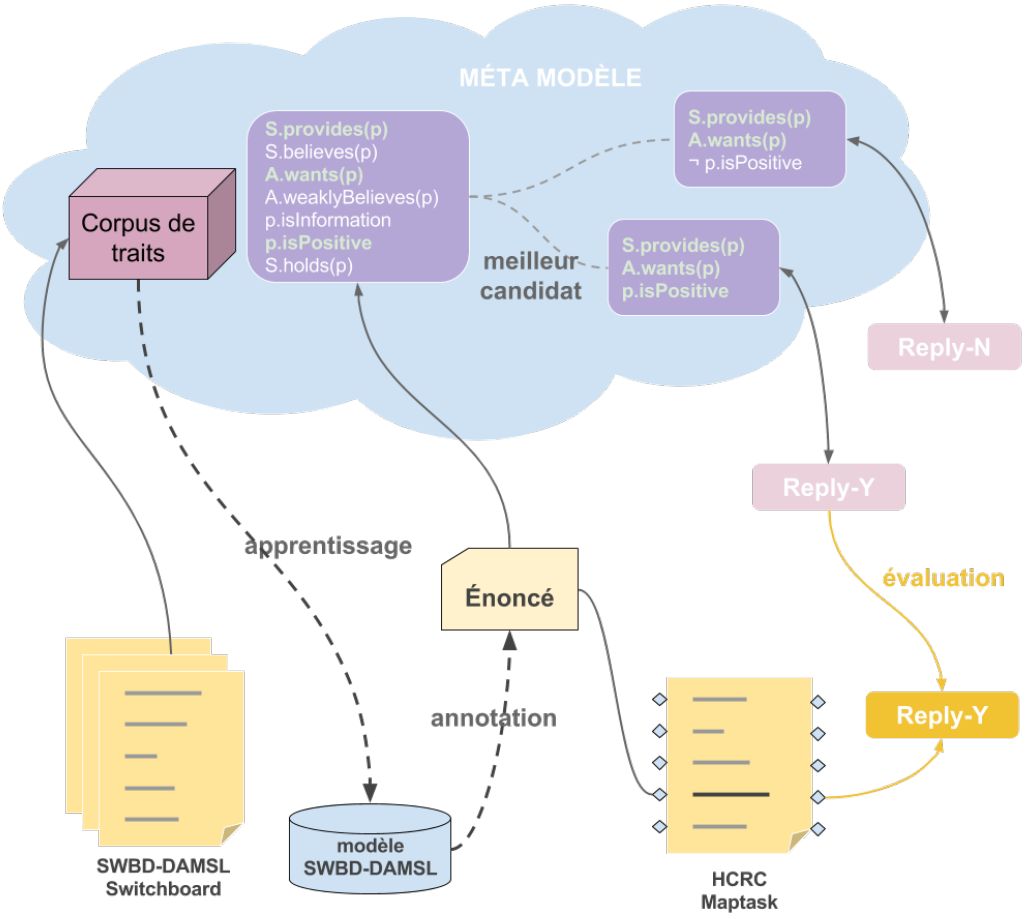

Fig. 4.5.: Représentation du procédé d'annotation automatique via le méta-modèle par le second système (B).

4.5.4 Résultats

Les résultats sont fournis en table 4.4. Nous observons que le système B a des performances bien plus faibles que le système A. Son exactitude est 22 et 13 points derrière celle du classifieur direct, pour le standard ISO et HCRC, respectivement. Le système A, en revanche, est seulement battu de 9 et 8 points. Cela suggère que de nombreux traits sont difficiles à apprendre, comparés aux labels d'actes de dialogue. Cela s'explique par le fait que les mêmes ensembles de caractéristiques aient été utilisés indistinctement pour apprendre à identifier les différents traits primitifs. Cependant il parait évident que cette approche est naïve puisque les traits capturent des aspects très différents des énoncés. Certains sont explicitement syntaxiques (*e.g.* « $p.hasSubjectInversion$ »), d'autres sémantiques (*e.g.* « $p.isPositive$ »), discursifs (*e.g.* « $p.isSummary$ »), lexicaux (*e.g.* « $p.isTagQuestion$ »), *etc*. Une approche plus spécialisée doit être explorée pour améliorer leur reconnaissance automatique.

Nous pouvons constater que bien que le système B ait de faibles performances, le système A est plutôt efficace, moins de dix points derrière les résultats d'un classifieur direct. L'augmentation du taux d'erreur peut être attribuée à deux facteurs : (1) les taux d'erreur de la phase de conversion d'annotation, et (2) un taux d'erreur accru du classifieur d'actes de dialogues dû aux différences structurelles et linguistiques entre les corpus utilisés dans cette expérience.

Source	Cible	Exactitude
Approche de base : classification directe		
SWBD (ISO)	SWBD (ISO)	0,83
MapTask (HCRC)	MapTask (HCRC)	0,59
A : reconnaissance, décomposition puis recomposition		
SWBD (DAMSL)	SWBD (ISO)	0,74 (-0,09)
SWBD (DAMSL)	MapTask (HCRC)	0,51 (-0,08)
B : décomposition, reconnaissance puis recomposition		
SWBD (DAMSL)	SWBD (ISO)	0,61 (-0,22)
SWBD (DAMSL)	MapTask (HCRC)	0,46 (-0,13)

Tab. 4.4.: Exactitude d'un classifieur classique (label à label) et d'un classifieur inter-taxonomique (label à traits à label).

4.6 Autres usages

Bien que l'objectif principal de ce travail est de rendre possible l'exploitation d'un corpus annoté pour entraîner des modèles capables de produire des annotations basées sur sur une taxonomie différente, il existe d'autres utilisations pratiques de ce méta-modèle.

Comme indiqué dans figure 4.1, un tel outil peut faciliter la compréhension des taxonomies. De nouveaux utilisateurs peuvent facilement appréhender les différences entre une nouvelle taxonomie et une qu'ils connaissent déjà, étant donné que les définitions listées et les représentations de manière graphique facilitent la compréhension de légères nuances entre labels similaires. Et, plus important encore, le choix d'une taxonomie plutôt qu'une autre pour une tâche donnée peut être motivé par la présence ou l'absence de ces traits pertinents dans leurs labels.

On peut également utiliser un méta-modèle pour construire ou peaufiner une taxonomie. En lieu et place de la méthode traditionnelle, pour laquelle les chercheurs travaillent en se basant sur les labels, ils commenceraient par choisir les traits qu'ils souhaitent saisir dans les énoncés (*e.g.* S.WANTS(X), S.OFFER(X), X.ISINFOTRANSFER, X.ISACTION), et ensuite les regrouperaient en labels pertinents (*e.g.* QUESTION = S.WANTS(X) + X.ISINFOTRANSFER, REQUEST = S.WANTS(X) + X.ISACTION, COMMIT

= S.OFFER(X) + X.ISACTION). Une telle méthode permettrait également une vérification programmatique de l'intégrité d'une taxonomie, en s'assurant que tous les énoncés possibles peuvent être annotés et que deux labels ne peuvent pas s'appliquer pour un même énoncé. Par exemple, la taxonomie présentée dans ce paragraphe ne permettrait pas de catégoriser un énoncé pour lequel le locuteur fournit une information.

4.7 Conclusion

Dans ce chapitre, nous avons présenté un méta-modèle pour l'abstraction des taxonomies d'actes de dialogue. Nous pensons que les méta-modèles ont de nombreuses applications utiles pour l'analyse du dialogue et la recherche taxonomique. Le principal but de ce travail est de fournir une méthode pour construire des systèmes de reconnaissance d'actes de dialogue qui ne nécessitent pas de données annotées avec la taxonomie visée mais seulement des données annotées avec une taxonomie capable de capturer suffisamment d'informations pertinentes. Nous avons montré qu'un classifieur entraîné sur les annotations SWBD-DAMSL pouvait produire des annotations pour le schéma d'annotation ISO ou HCRC avec une exactitude assez proche d'un classifieur standard.

Étant donné que nous avons montré que deux taxonomies peuvent être rendues interopérables par l'utilisation de traits primitifs, nous avons fait le choix de réutiliser au maximum les traits présents dans le standard ISO pour le développement de notre propre taxonomie. Cette dernière, dédiée à l'analyse des conversations écrites en ligne porteuses de demandes d'assistance, est utilisée pour l'annotation de notre corpus, Ubuntu-fr, que nous présentons dans le chapitre suivant.

Le travail présenté dans ce chapitre a fait l'objet d'une publication à Cicling 2017 (Salim, Hernandez, et Morin, 2017).

Corpus Ubuntu

5

Table des matières

5.1 Introduction 85

5.2 Travaux et ressources similaires 86

5.3 Construction du corpus 87
- 5.3.1 Modélisation des conversations écrites en ligne . . 87
- 5.3.2 Méthode de collecte 89
- 5.3.3 Pré-traitement des données 89
- 5.3.4 Statistiques sur le corpus 90

5.4 Annotation d'une partie du corpus en termes d'actes de dialogue et d'informations Opinion-Sentiment-Émotion . . 91
- 5.4.1 Taxonomie employée 92
- 5.4.2 Annotation et accord inter-annotateurs 93
- 5.4.3 Annotations et statistiques 94

5.5 Conclusion 96

5.1 Introduction

Dans ce chapitre nous présentons le résultat de nos efforts pour la construction et l'annotation d'un large corpus libre en français compilant à la fois des conversations écrites en ligne sur des modalités synchrone (chat) et asynchrones (forum et courriel) et la documentation associée (au format HTML) sur une même période, et ce, autour d'un même type de situation discursive : l'assistance à la résolution de problèmes. Les objectifs scientifiques qui ont motivé la construction de ce corpus sont multiples :

1. Améliorer le traitement linguistique des différentes modalités en exploitant leur caractère comparable (*e.g.* en travaillant sur la portabilité d'analyseurs construits sur une modalité pour en traiter une autre)
2. Maîtriser l'alignement voire la « traduction » d'un contenu porté par une modalité textuelle (texte explicatif ou communication écrite en ligne) vers une autre
3. Mieux comprendre la structure et le fonctionnement de ce type de conversation ainsi que leurs interactions
4. Permettre des activités de Traitement Automatique des Langues (TAL) (*e.g.* construction de modèles statistiques dédiés à la reconnaissance de structures discursives)
5. Permettre la diffusion de nos données dans une perspective de reproductibilité et de réutilisation (« *open science* » (Nielsen, 2011))

Nous voulons soutenir des applications de recherche d'information inter-modalités textuelles (*e.g.* recherche d'une solution dans une modalité distincte de la demande), de gestion automatique de la documentation (*e.g.* détection de l'absence ou de l'obsolescence d'une solution) ou d'assistance à la production et au diagnostic (*e.g.* aide à la formulation de problème, évaluation de la complétude de messages réponses, suivi d'un sujet sur plusieurs fils de discussion et modalités). Pour pouvoir soutenir de telles applications, nous envisageons l'analyse des conversations en termes d'organisation en actes de dialogue, et en fonction des objets conceptuels manipulés, à savoir les problèmes et leurs possibles solutions.

Nos objets d'étude sont plus spécifiques que les communications médiées par les réseaux (CMR). Nous avons travaillé sur la proposition d'un modèle de conversations écrites en ligne offrant un cadre de traitement générique prenant en compte les spécificités des modalités manipulées. La définition de ce modèle passe par l'adaptation de la taxonomie des actes de dialogues et des relations discursives définies dans le standard ISO (Bunt et al., 2012).

Un des aspects les plus importants de ce corpus est sa partie annotée : pour permettre l'application de techniques d'apprentissage supervisé, près de 6 000 énoncés issus des trois modalités ont été annotés en termes d'actes de dialogue, d'opinion, de sentiment et d'émotion (OSE).

5.2 Travaux et ressources similaires

La communauté TAL dispose de peu de ressources portant sur les médias sociaux (Seddah et al., 2012; Falaise, 2014; Yun et Chanier, 2014). Seul le corpus Simuligne du projet LETEC [1] [2] couvre plusieurs modalités (chats, forums, courriels, support pédagogique) en français (Chanier et al., 2009). Mais leur corpus diffère du nôtre en termes d'objet d'étude, d'échelle et de volume : le leur contient 11 506 messages, 600 348 unités lexicales et 67 participants. Falaise (2014) propose un corpus de chat générique comprenant cinq millions de tours de paroles. Pour étudier l'analyse automatique de contenu généré par les utilisateurs, Seddah et al. (2012) ont diffusé un *treebank* contenant 1 700 phrases issues de services de microblogging et de forums web. Notre initiative fait suite à celle de Uthus et Aha (2013) et de Lowe et al. (2015) qui ont proposé d'exploiter la communauté Ubuntu pour construire un large corpus libre de conversations de chat en anglais. Il n'existe pas à notre connaissance de corpus français annoté en actes de dialogue portant sur les modalités des conversations écrites en ligne.

Le projet CoMeRe [3] [4] tente de combler le manque de ressources en CMR en publiant tous les efforts accomplis dans le domaine via le nœud français de l'infrastructure CLARIN [5] [6] (Chanier et al., 2014). Le projet a pour vocation de fournir un noyau de corpus de communication médiée par les réseaux en OpenData, c'est-à-dire librement consultables, diffusables et transformables. Ces corpus portent sur une variété de modalités : blogs, tweets, SMS, courriels, forums, conférences en ligne, *etc*. Les formats des corpus suivent les recommandations de la *Text Encoding Initiative* (TEI), plus spécifiquement de la TEI-CMC [7], et sont diffusés via le réseau Ortolang [8]. Le développement de ce projet témoigne d'un intérêt grandissant pour la mise à disposition de ressources francophones pour l'étude des communications en ligne.

Bien que notre initiative ne se situe pas sous l'égide du projet CoMeRe (Chanier et al., 2014), nous adhérons à leurs principes. D'une part, nous considérons qu'il est néces-

1. *Learning and Teaching Corpus*
2. `lrl-diffusion.univ-bpclermont.fr/mulce2/`
3. `http://corpuscomere.wordpress.com/apropos`
4. Pour « Communications Médiées par les Réseaux ».
5. *Common Longuage Resources and Technology Infrastructure*
6. `https://www.clarin.eu/`
7. `corpuscomere.wordpress.com/tei/`
8. `www.ortolang.fr/api/content/comere/latest/comere.html`

saire de standardiser les descriptions des données et des méta-données, et d'autre part nous pensons qu'il est nécessaire de réfléchir aux possibilités d'exploitation et de diffusion d'un corpus en amont de sa construction.

5.3 Construction du corpus

Les débats autour de l'encodage des interactions écrites multi-modales en sont encore à un stade très précoce. Un groupe d'intérêt spécifique [9] (*Special Interest Group* - SIG) de la TEI considère les propositions de CLARIN et CoMeRe pour adapter la TEI aux différents genres des CMR. Ces genres incluent les tweets, les discussions sur les wikis, les blog, les SMS, les canaux audio, *etc*. Notre travail se concentre sur les objets conversationnels, plus précisément les conversations se déroulant sur les forums, les listes de diffusions et les salons de chat. Bien que nous sommes en accord avec les modèles de données recommandés par le SIG, nous considérons qu'une approche qui ne fait pas de distinction claire entre le modèle et son implémentation fait fausse route. En effet, la recommandation spécifie comment le contenu doit être formaté et structuré, ignorant ainsi l'importance du principe d'annotation en « stand-off » (Thompson et McKelvie, 1997), qui veut que les annotations sont clairement séparées des données, pour permettre d'éviter de les modifier et d'altérer ou perdre de l'information qui peut se révéler importante. Mais bien que ne retenons pas la TEI ou son extension aux CMR pour représenter notre corpus, nous adhérons aux modèles de données qu'ils sous-tendent.

5.3.1 Modélisation des conversations écrites en ligne

Nous définissons un modèle générique pour décrire les méta-données conversationnelles. Le modèle résulte de la généralisation des structures conversationnelles des modalités observées. Il prend en compte les évolutions récentes des formats de messages en ligne [10] ainsi que les recommandations actuelles de la TEI en termes de méta-données. Il intègre les attributs communs et spécifiques pour inclure une catégorisation thématique des conversations, du nombre de vues, des rôles des participants (*e.g.* ambassadeur, expert, client), *etc*. Les structures de Conversation sont accessibles au travers de certains traits de Message : *daytime* et *inReplyToMessageIds*. La figure 5.1 montre les relations entre les principaux objets conversationnels.

Les Forums structurent thématiquement les Conversations qui sont constituées de Messages. Un Forum est une structure récursive. Il correspond au concept de Salon pour les chats. Les Forums, Conversations, Messages et Participants sont sont uniquement identifiés. Un Forum, une Conversation, et un Message peuvent

9. `http://wiki.tei-c.org/index.php/SIG:Computer-Mediated_Communication`
10. `http://6tools.ietf.org/html/rfc6854`

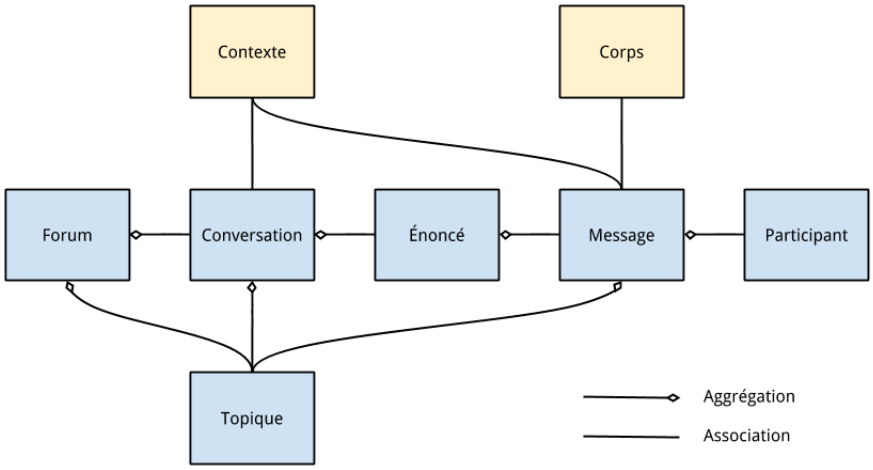

Fig. 5.1.: Modèle de méta-données.

avoir des sous-objets (*e.g.* « Problème de carte SD ») et peuvent être catégorisés par Topiques (*e.g.* « Hardware »). Un Contexte contient les informations relatives au type de médium, au statut « privé »/« public », au statut « résolu », au nombre de « likes », de vues, *etc*. Un Message connecte les Participants via les relations de type FROM, TO, CC et BCC. Il est constitué d'Énoncés, chacun d'entre eux pouvant être associé à un label Opinion-Sentiment-Émotion (OSE) et à un acte de dialogue. Le Corps du message spécifie son type MIME et l'encodage de son contenu. Un Participant est caractérisé par un nom d'utilisateur, une adresse, un rôle (*e.g.* client), *etc*.

Le modèle est implémenté en schémas W3C XML et en un système de types UIMA [11]. Le XML sert à stocker et à échanger les données brutes après extraction depuis leur format source. Le second est utilisé dès que les données entrent dans les chaînes de traitement TAL. Le typage est basé sur la plate-forme UIMA de Apache (Ferrucci et Lally, 2004) puisqu'elle fournit le degré d'abstraction nécessaire pour la manipulation de données annotées. Nos données annotées sont sérialisées sous la forme de fichiers XML Metadata Interchange (XMI), une pratique standard pour l'échange de métadonnées UML. Ces choix nous écartent de la TEI mais pas nécessairement de son modèle conceptuel.

Les principales fonctions du format sont le stockage et l'échange de données. Leur principal objectif est de permettre à n'importe quel utilisateur de modifier les annotations et d'en ajouter de nouvelles aux données dans leurs formes originales sans les altérer. Ceci implique l'adoption de certains principes (*e.g.* l'annotation en « *stand off* ») et l'usage de certains outils, tels que UIMA.

11. `http://docs.oasis-open.org/uima/v1.0/os/uima-spec-os.html#_Toc205201040`

5.3.2 Méthode de collecte

Nos données ont été collectées depuis un certain nombre de ressources publiques hébergées par la communauté Ubuntu francophone [12]. En plus d'une documentation en ligne [13], la communauté dispose de plusieurs dispositifs de communication : des forums [14], des listes de diffusion [15], et des canaux IRC [16]. Leurs contenus sont accessibles publiquement en ligne.

La documentation est fournie sous licence libre CC BY-SA v3.0 [17]. Pour les autres ressources, les utilisateurs délèguent l'usage de leurs messages à l'éditeur, Ubuntu-fr. Cela implique que si un utilisateur exprime son refus de participer au corpus, nous devrons en retirer les messages dont il ou elle est l'auteur.

La documentation n'est pas statique et doit être versionnée. Elle ne peut être récupérée que via des techniques de *web scraping*, c'est-à-dire en téléchargeant les pages HTML et en extrayant le contenu qui nous intéresse. Nous avons récupéré quotidiennement une version de la documentation depuis novembre 2014. Les courriels sont archivés incrémentalement et sont disponibles publiquement. Les forums croissent incrémentalement également mais aucune archive publique n'est disponible, et nous avons également dû utiliser des techniques de *web scraping* pour collecter leurs données. Les messages chat ne sont pas du tout sauvegardés, cependant nous avons déployé un bot d'archivage en novembre 2014 qui collecte les messages envoyés depuis cette date.

Nous disposons donc de tous les messages de forums et de tous les courriels depuis la création de la plate-forme en 2004, ainsi que toutes les données de la documentation et des canaux de chat depuis novembre 2014. Les données ont été collectées jusqu'à fin octobre 2015. Ces sous-corpus sont représentatifs du style moderne d'expression en ligne et témoignent de l'évolution des communications en ligne sur une période de dix ans. De plus, les données peuvent être facilement étendues avec des données en langue anglaise déjà disponibles (Lowe et al., 2015).

5.3.3 Pré-traitement des données

Pour préparer les données à des tâches d'annotation et de traitements TAL, nous avons préparé les données en effectuant les traitements suivants :

12. `http://ubuntu-fr.org`
13. `http://doc.ubuntu-fr.org`
14. `http://forum.ubuntu-fr.org`
15. `http://lists.ubuntu.com/mailman/listinfo/ubuntu-fr`
16. `http://irc.freenode.net/ubuntu-fr`
17. `https://creativecommons.org/licenses/by-sa/3.0/`

1. Extraction de méta-données. Nous avons développé des analyseurs *ad hoc* pour instancier le modèle présenté en sous-section 5.3.1 pour chaque modalité. Pour les chats, nous avons développé un moteur d'analyse spécifique pour identifier le destinataire de chaque message.
2. Extraction de texte. Nous avons résolu les problèmes d'encodage et de type MIME pour chaque message. Les contenus des courriels ont également été analysés pour reconnaître les sections citées.
3. Identification et désentremêlement de conversations. La structure des forums permet d'identifier immédiatement le fil des messages. Les structures de conversations dans les listes de diffusion peuvent être identifiées en se basant sur le champ *inReplyTo* des méta-données ainsi que sur la similarité entre les sujets des courriels (en terme de distance d'édition) envoyés dans la même période temporelle. Cependant, pour les chats, de multiples conversations peuvent se produire simultanément sur le même canal. Dans (Riou et al., 2015), nous étudions la portabilité pour le Français d'une méthode état-de-l'art de désentremêlement de conversations (Elsner et Charniak, 2010) ainsi que la pertinence d'utiliser les informations discursives disponibles dans ce but.
4. La segmentation de texte en unités lexicales et en phrases. Dans ce but, nous avons développé une approche à base de règles exploitant la ponctuation, la typographie et le disposition du texte. Les segmenteurs ont été conçus pour traiter du HTML et des phénomènes propres aux CMR.

En ce qui concerne l'anonymisation des données, nous considérons plusieurs alternatives : nous pouvons l'éviter entièrement en supprimant les messages à la demande, nous pouvons la restreindre aux méta-données, ou nous pouvons nous contenter de fournir nos outils de collecte et de traitement, sans données.

5.3.4 Statistiques sur le corpus

La table 5.1 détaille le contenu d'un échantillon d'un an de données. Bien qu'elle ne détaille que des échantillons récents pour une meilleure comparaison, dans le cas des courriels la taille d'un tel échantillon aurait été bien plus importante dans les premières années de la plate-forme. Par exemple, l'échantillon de courriels de 2005 contient 7 034 messages. Ceci reflète les profondes évolutions du comportement des utilisateurs ces dix dernières années. Ce type de données peut être utilisé pour effectuer des études diachroniques. Cependant ce type d'analyse est hors du périmètre de ce chapitre et ne sera pas détaillé ici.

Les statistiques fournies en table 5.1 montrent des différences significatives entre les modalités disponibles. Les conversations ont deux fois plus de chances d'obtenir une réponse sur les forums que sur les listes de diffusion ou les canaux de chat, et elles

	courriels	forums	chats	doc.
pages (total)	-	-	-	4 600
conversations (total)	100	23K	4.4K‡	-
fils sans réponse (pourcentage)	23 %	12 %	21 %‡	-
messages (total)	448	189K	114K	-
messages par conversation (médian)	3	5	4‡	-
mots (total)	40K†	25M	1M	4M
mots par message (médian)	59†	46	7	-
participants (total)	75	12K	2.4K	-
réf. à la documentation (par conv.)	2.71†	0.78	0.02	-
réf. aux forums (par conv.)	0.44†	2.13	0.11	-
réf. aux listes de diffusion (par conv.)	0.17†	0.00	0.00	-
réf. à des ressources externes (par conv.)	5.08†	3.82	2.58	-
durée d'une conversation (médian)	4h33	5h51	-	-

Tab. 5.1.: Un an de données : de janvier à décembre 2014 pour les forums et courriels, d'octobre 2014 à octobre 2015 pour les chats. Le symbole † signifie que le texte cité n'est pas pris en compte. Le symbole ‡ signifie qu'il s'agit d'une estimation basée sur les résultats d'une expérience de désentremêlement de chats sur un échantillon de 1 229 messages (Riou et al., 2015).

y restent actives plus longtemps. Les courriels sont plus longs, et ont bien plus de chances de référencer des ressources externes et particulièrement la documentation Ubuntu. Les forums référencent principalement les autres forums, tandis que presque personne ne redirige d'autres utilisateurs vers la liste de diffusion.

5.4 Annotation d'une partie du corpus en termes d'actes de dialogue et d'informations Opinion-Sentiment-Émotion

Pour étudier les interactions entre utilisateurs sur différentes modalités, nous avons annoté une portion du corpus en termes d'actes de dialogue et d'information OSE (Opinion, Sentiment et Émotion). À partir de novembre 2014, 29 conversations sur les forums, 45 sur les listes de diffusion et 6 jours d'activité IRC ont été annotés. Pour permettre des approches d'apprentissage automatique, nous avons décidé d'annoter au moins 1 200 énoncés pour chaque modalité. Les messages ont été segmentés en énoncés selon une approche semi-automatique : les messages ont d'abord été automatiquement segmentés en phrases, puis dans un second temps cette segmentation a été modifiée au fur et à mesure de l'annotation par les annotateurs.

5.4.1 Taxonomie employée

Nous avons basé notre système d'annotation sur la taxonomie du standard ISO. Mais bien que nous acceptons sans réserve le cadre conceptuel du standard, nous avons dû adapter la taxonomie pour notre corpus pour pallier aux spécificités décrites en section 3.3. Le détail complet de la taxonomie est disponible en annexe A.

Pour ce faire, pour chaque dimension, fonction communicative et qualifieur, nous nous sommes posé deux questions : « est-ce utile ? » et « est-ce présent dans nos données ? ». Nous avons cherché la réponse à la première question en confrontant la taxonomie à différents cas d'utilisation, la seconde a requis une approche dirigée par l'étude des données. Et pour chaque énoncé, nous avons également considéré si nous avions une fonction appropriée dans la taxonomie. Par conséquent, trois annotateurs, chacun ayant une formation en TAL, ont effectué plusieurs sessions d'exploration des données à partir de novembre 2014.

Ces expériences ont mené à l'élaboration de plusieurs modifications du schéma d'annotation. Nous avons ajouté la dimension EXTRA DISCOURSE pour capturer tout le texte non-discursif que les participants peuvent inclure dans leurs messages (tels que les portions de texte cité, ou les caractères de formatage). Nous avons également ajouté la dimension PSYCHOLOGICAL STATE pour capturer les énoncés portant sur l'état psychologique des participants à la conversation (*e.g.* « Ça commence à m'énerver », « :D »). À l'inverse, certaines dimensions du standard ne semblent pas pertinentes pour les modalités étudiées. Ainsi, la dimension TIME MANAGEMENT n'est pas utile dans le contexte de conversations écrites asynchrones ou pseudo-synchrones. D'autres, comme COMMUNICATION MANAGEMENT ont une présence très marginale dans les données. La taxonomie a donc été simplifiée pour la rendre utilisable dans le cadre de tâches d'apprentissage automatique des actes de dialogue. Les dimensions ALLO FEEDBACK et AUTO FEEDBACK ont été réunies au sein de la dimension FEEDBACK, et nous avons fait le choix d'utiliser un qualifieur ALLO ou AUTO pour en indiquer le sujet. De même, nous avons fusionné OWN COMMUNICATION MANAGEMENT et PARTNER COMMUNICATION MANAGEMENT puisque la distinction est capturée par le qualifieur nouvellement introduit.

Les fonctions communicatives ont été simplifiées de plusieurs manières, tout en préservant leur structure globale. Par exemple, la taxonomie ne distingue pas les différents types de questions, d'actes commissifs et d'actes directifs ; à la place nous utilisons REQUEST FOR INFORMATION, REQUEST FOR ACTION et REQUEST FOR DIRECTIVES. Les résultats d'annotations préliminaires ont montré que les fonctions spécifiques relatives aux dimensions peu présentes (*e.g.* STALL, SHIFT TOPIC) sont quasiment absentes du corpus.

Les tables 5.3 et 5.4 détaillent les dimensions et fonctions retenues pour la tâche d'annotation.

L'analyse de sentiments serait utile pour évaluer la satisfaction des utilisateurs ainsi que les performances des supports client. Nous pensons également qu'il serait intéressant de pouvoir capturer les retours de satisfaction indépendamment de l'évaluation effective de la réponse (*e.g.* « merci pour la réponse rapide, cependant cette solution ne fonctionne pas pour moi »). C'est pourquoi nous avons choisi une forme détaillée d'annotation des sentiments et avons basé nos annotations OSE sur les classes sémantiques à grain fin définies dans le projet Ucomp [18] (Fraisse et Paroubek, 2014). La notion d'opinion est couverte par les labels DEPRECIATE et PROMOTE, la notion de sentiment est couverte par les labels SATISFIED et UNSATISFIED, et la notion d'émotion est couverte par sept labels : HAPPINESS, UNHAPPINESS, FEAR, DISGUST, LOVE, SURPRISE et ANGER. Les catégories émotionnelles ont été généralisées depuis les labels originaux pour prévenir certaines ambiguïtés. Les énoncés ne portant aucun sentiment seraient catégorisés sous le label générique INFORMATION. Cependant, après examen des résultats d'annotation nous nous sommes aperçus que seule une faible proportion des énoncés, environ 10 % d'entre eux, était porteur d'un label OSE. Malgré tout, nous nous sommes aperçus que la pertinence d'une réponse peut être évaluée par d'autres moyens, notamment par la présence de fonctions des dimensions FEEDBACK et SOCIAL OBLIGATIONS MANAGEMENT.

Pour ce qui est des autres qualifieurs présents dans le standard, nous avons choisi de tous les conserver. Le qualifieur de partialité peut être utile pour identifier les réponses incomplètes, bien que leur détection automatique puisse être difficile. La certitude est également utile pour évaluer la confiance d'un participant en ses propres réponses. La conditionnalité est également utile puisqu'elle peut souvent indiquer qu'un utilisateur est sur le point d'abandonner ou qu'un client s'engage dans une dynamique de menace (*e.g.* « Si je ne suis pas remboursé, je n'achèterai plus jamais rien chez vous »). Malheureusement, les résultats d'annotation sont en deçà de ce qu'on aurait pu espérer : moins d'un énoncé sur vingt est porteur d'un marqueur clair de partialité, de certitude ou de conditionnalité.

5.4.2 Annotation et accord inter-annotateurs

L'annotation s'est déroulée selon un processus itératif et incrémental. À chaque itération, de nouvelles données ont été annotées principalement par un seul annotateur, une post-doctorante avec de solides connaissances en linguistique et en TAL. Les nouveaux phénomènes rencontrés et les cas ambigües ont fait l'objet d'intenses discussions entre trois chercheurs travaillant sur le sujet. Les définitions en furent

18. www.ucomp.eu

par conséquent renforcées et les données précédemment annotées furent révisées, jusqu'à ce que les objectifs quantitatifs et qualitatifs aient été atteints.

Pour mesurer l'homogénéité des données annotées, deux annotateurs ont annoté trois nouvelles conversations (soit environ 110 énoncés, totalisant 1 200 mots). Chaque annotateur a annoté 110 énoncés. Puisque les annotateurs n'ont pas toujours segmenté les phrases de la même manière, nous avons décidé de calculer le coefficient d'accord au niveau des tokens. Nous avons obtenu les valeurs de Kappa suivantes : 0,69 et 0,70 pour les dimensions et fonctions communicatives, respectivement. Ces valeurs indiquent un accord substantiel.

5.4.3 Annotations et statistiques

L'annotation couvre plus de 4 700 actes de dialogue. La table 5.2 compare la taille des sous-corpus annotés pour chaque modalité en termes de conversations, messages, tokens et actes de dialogue. Les sous-corpus forums, courriels et chats sont de taille suffisamment similaires pour pouvoir être comparés.

	# conversations	# messages	# tokens	# actes
chats		2 320	17 448	1 989
forums	29	258	25 205	1 338
courriels	45	200	19 798	1 382

Tab. 5.2.: Comparaison de la taille des sous-corpus annotés par modalité.

La table 5.3 montre la variation entre les comportements conversationnels des participants en fonction de la modalité. Les actes de gestion sociale du dialogue sont nettement plus présents dans les courriels. Dans les chats, la gestion du discours (*e.g.* « voici ma question ») est rare, contrairement aux forums et en particulier aux courriels. En revanche, les classes EVALUATION (« OK fais voir », ATTENTION PERCEPTION INTERPRETATION (« J'ai compris ») et PSYCHOLOGICAL STATE (« Ah ça va mieux ») sont bien plus prévalentes, ce qui montre que la synchronisation informationnelle et émotionnelle des participants est plus importante dans les conversations synchrones que dans les conversations asynchrones. Les énoncés qui portent sur la gestion du temps (*e.g.* STALLING), du contact (« Il y a quelqu'un ? ») et de la communication (« Euh je voulais dire Unix ») ne sont présents que dans les chats, et absentes des modalités asynchrones.

La table 5.4 confirme que dans les courriels, les salutations, les adieux, les remerciements et autres actes sociaux semblent attendus et protocolaires pour encadrer un message. Nous observons également que les conversations chats sont plus directes et orientées vers l'action : les participants IRC sont deux fois plus sujets à accomplir un

Dimensions	Chats	Forums	Courriels
Domain/Activities	82,35	80,1	67
Social Obligation Management	9,25	12,85	30,35
Discourse Management	0,85	4,8	2
Evaluation	3,95	1,65	0,15
Psychological State	1,45	0,45	0,35
Attention Perception Interpretation	0,6	0,15	0,05
Communication Management	1,35	0	0
Contact Management	0,9	0	0
Time Management	0,2	0	0

Tab. 5.3.: Distribution des dimensions des actes de dialogue.

acte commissif et consacrent plus d'énoncés à réclamer des actions, des instructions ou de l'information.

Fonctions	Chats	Forums	Courriels
Inform	26,95	31,3	33,35
Answer	17,65	20,05	11,2
Request for Information	16,35	9,2	6,95
Answer Positively	9,1	5,45	3,05
Request for Action	8,85	8,45	6,15
Greetings	5,65	5,1	6,8
Correct	4,75	3	1,3
Answer Negatively	3,4	2,85	1,9
Thanking	2,05	2,85	3,4
Commit	1,95	1,05	1
Request for Directive	0,85	0,9	0,35
Valediction	0,5	1,35	6,15
Apologizing	0,45	0,65	0,6
Anticipate Thanking	0,4	1,95	2,95
Final Self Introduction	0	0,9	9,2
Announce	0,35	0,8	1,25

Tab. 5.4.: Distribution des fonctions communicatives des actes de dialogue (les classes trop rares pour permettre une comparaison ne sont pas représentées).

La figure 5.2 montre la distribution de labels OSE. Bien que le nombre d'énoncés porteurs d'un marqueur clair d'opinion, de sentiment ou d'émotion soit relativement faible - moins de 10 % des énoncés - nous pouvons tout de même identifier une tendance. Les chats semblent plus émotionnels, tandis que les forums sont plus portés sur l'expression d'opinions.

Globalement, il semble que les forums représentent une forme intermédiaire de conversation en ligne, à placer entre le style plus formel des courriels et les chats, plus directs et informels par nature.

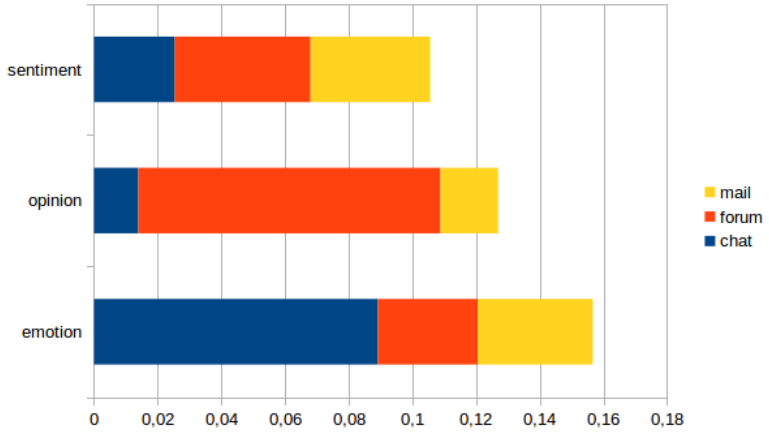

Fig. 5.2.: Distribution des labels Opinion, Sentiment et Émotion.

5.5 Conclusion

Nous avons construit un large corpus francophone de conversations écrites en ligne extraites de forums, courriels et chats collectés sur la plate-forme Ubuntu. Le corpus est destiné à permettre l'étude des modalités de conversations écrites en ligne orientées vers l'assistance aux utilisateurs.

Le corpus est construit autour d'un modèle robuste de méta-données basé sur des principes forts, tels que le principe d'annotation en « *stand off* ». Nous avons montré comment les données ont été collectées et traitées, et nous avons détaillé leurs contenus. Une portion du corpus a été annotée en termes d'actes de dialogue et d'opinion, de sentiment et d'émotion. Nous avons adapté le schéma d'annotation du standard ISO pour développer la taxonomie employée pour l'annotation du corpus.

Les données annotées sont prêtes à être exploitées pour des tâches de reconnaissance automatique des actes de dialogue. Le chapitre suivant est consacré aux travaux que nous avons effectués dans cette perspecive.

Le travail présenté dans ce chapitre a fait l'objet d'une communication à LREC 2016 (Hernandez, Salim, et Loginova-Clouet, 2016).

6 Reconnaissance automatique des actes de dialogue

Table des matières

6.1 Introduction 99

6.2 Travaux similaires 100

6.3 Méthode 102
6.3.1 Approche et implémentation 102
6.3.2 Traits 103

6.4 Expériences et discussion 103
6.4.1 Comparaison d'un classifieur SVM à l'approche de base 104
6.4.2 Comparaison d'approches et de jeux de traits sur différentes modalités 105

6.5 Comparaison de classifieurs trans-modalités 107

6.6 Conclusion 108

6.1 Introduction

Les capacités d'interaction entre internautes se sont spectaculairement accrues au cours des dernières décennies, et avec elles les plate-formes d'échange participatives. Ces plate-formes, qui se déclinent en différentes modalités - salons de chat, listes de diffusion, forums de discussions - sont souvent utilisées par les particuliers comme par les professionnels pour demander et offrir de l'aide. Ce type d'échanges, les conversations écrites en ligne orientées vers la résolution de problèmes, représentent un champ de recherche encore peu exploré par la communauté scientifique. Cette forme de communication peut néanmoins être exploitée, et est très similaire aux conversations entre agents et utilisateurs des services de gestion de la relation client (GRC) des entreprises qui assurent une présence en ligne. C'est pourquoi les conversations porteuses de demandes d'assistance intéressent déjà les industriels, qui s'efforcent de développer des systèmes adaptés à la gestion de ce type d'échanges.

Développer un système capable de les analyser automatiquement représente ainsi non seulement un enjeu industriel important, mais permettrait également d'améliorer les plate-formes d'échanges collaboratives qui sont quotidiennement sollicitées par des millions d'utilisateurs. Cependant, les techniques d'identification de la structure des conversations n'ont pas été développées autour des conversations écrites en ligne. Il serait intéressant de déterminer l'efficacité de ces techniques sur des données extraites de conversations orientées vers la résolution de problèmes en fonction de leur modalité. Cette problématique nous confronte à des problématiques propres aux communications médiées par les réseaux (CMR) : données en français de différents registres et à forte variation orthographique, faibles performances des outils linguistiques, absence d'informations prosodiques, structures de conversations non-linéaires, *etc*.

Dans ce chapitre, nous nous focalisons sur la reconnaissance des actes de dialogue présents dans les énoncés de notre corpus. L'approche la plus commune et efficace pour la tâche de classification d'énoncés en termes d'actes de dialogue est d'employer des techniques supervisées d'apprentissage machine (Tavafi et al., 2013). C'est également l'approche que nous retenons. Les données ne manquent pas : dans le cas des forums et des listes de diffusion, les conversations sont généralement perpétuellement sauvegardées pour permettre aux utilisateurs de faire des recherches dans les archives. Ce n'est pas toujours le cas pour les messages transmis dans les salons de chat, mais ils peuvent eux aussi faire l'objet d'un archivage automatique et de nombreux outils existent à cette fin.

Ce chapitre a pour objectif de comparer les approches communes pour la classification des actes de dialogue dans un corpus de conversations écrites en ligne orientées vers la résolution de problèmes décliné en différentes modalités, à savoir le corpus

Ubuntu-fr que nous avons présenté au chapitre 5. Dans un premier temps, nous examinerons l'état de l'art et évoquerons les travaux similaires. Ensuite, présentons l'approche que nous adoptons pour notre tâche de reconnaissance automatique, avant d'exposer nos méthodes d'évaluation. Enfin, nous présenterons les expériences effectuées et discuterons des résultats obtenus.

6.2 Travaux similaires

Comme nous l'avons vu, la plupart des travaux existant dans le domaine de la détection des actes de dialogue s'intéressent à l'étude de conversations orales. C'est notamment le cas de nombreux travaux fondateurs en matière d'analyse du dialogue, qui reposent souvent sur le même corpus, TRAINS (Traum et Hinkelman, 1992; Poesio et Traum, 1997; Core et Allen, 1997; Bunt, 2009). C'est dû à l'objectif de ces chercheurs, à savoir développer des systèmes de dialogues dans lesquels le système prend la place d'un agent humain pour assister l'opérateur dans sa tâche. Le caractère textuel des conversations médiées par les réseaux nous prive ainsi d'importantes informations prosodiques, qui ont été démontrées utiles pour la classification d'actes de dialogue (Ang et al., 2005; Fernandez et Picard, 2002). Les conversations qu'ils étudient et modélisent sont également souvent bi-participants, tandis que nous nous intéressons principalement à des conversations multi-participants, ce qui accroît la complexité de la tâche. En effet, on ne peut pas systématiquement supposer que le second participant est le destinataire du message. Enfin, le rôle des utilisateurs est souvent pris en compte par le système pour effectuer ses choix, or cette information n'est souvent pas disponible ou non-applicable à nos conversations (*e.g.* , un système de *e-learning* bénéficie grandement de savoir si l'énoncé vient d'un élève ou d'un professeur puisque l'un est supposé poser des questions tandis que l'autre est chargé d'y répondre et de donner des instructions). Enfin, l'écrasante majorité des travaux portent sur la reconnaissance des actes de dialogue dans des corpus en anglais, mais très rarement en français.

Plusieurs travaux ont déjà porté sur des tâches de classification d'actes de dialogue dans des corpus de conversations en ligne. C'est l'approche supervisée qui est employée dans l'immense majorité des cas. Tavafi et al. (2013) étudient les travaux effectués dans le domaine et testent un panorama de techniques de classification employées dans la littérature. Le travail soutient l'idée que les actes de dialogue doivent être étiquetés séquentiellement pour une meilleure performance. Leur conclusion est que le modèle SMV-HMM (machine à vecteurs de support et *Hidden Markov Model*) est plus performant que les autres techniques testées, à savoir l'utilisation de champs conditionnels markoviens (CRF) et l'utilisation d'un SVM multi-classes. Cependant, l'approche séquentielle ne peut pas nécessairement être appliquée aux chats puisque plusieurs conversations peuvent se produire simultanément sur le

même canal, produisant des conversations entremêlées dont les messages ne peuvent pas être immédiatement identifiés. Ils testent notamment leurs techniques sur des corpus de réunions (MRDA) et de conversations téléphoniques (SWBD), mais aussi de courriels (BC3) et de forums (CNET). Leur travail n'étudie pas le cas des chats. Seul le corpus CNET correspond bien à notre tâche, puisqu'il contient des échanges visant à résoudre des problèmes utilisateurs. L'expérience présentée a cependant l'inconvénient de ne pas rapporter des résultats très satisfaisants (58 % micro-précision, 17 % macro-précision). En outre les labels employés, qui sont assez restreints, sont principalement composés de déclinaisons de « question » et de « réponse ». Ils reposent sur une définition assez limitée de ce qu'est un acte de dialogue pour pouvoir être appliqués à leur ensemble varié de corpus. Leur travail présente l'intérêt de comparer la classification des actes de dialogue au travers différentes modalités, mais le fait d'utiliser une taxonomie différente pour chacune d'entre elles limite fortement les conclusions que l'on peut tirer des résultats rapportés.

Cohen et al. (2004) s'intéressent à la classification de courriels en termes d'actes de discours. Leur taxonomie est constituée d'actes composés d'un nom et d'un verbe (*e.g.* REQUEST MEETING), et n'est pas basée sur la théorie des actes de dialogue. Par ailleurs, ils ne cherchent pas à classifier les énoncés mais les messages entiers, ce qui éloigne également leur travail de notre tâche. Lampert et al. (2006) en revanche s'intéressent à la classification des énoncés dans les courriels, et basent leur taxonomie sur les VRM (*Verbal Response Mode*), qui trouvent leur source dans la théorie des actes de langage d'Austin. Leurs résultats montrent que leur approche basée sur un classifieur SVM est crédible. Cependant les labels de la taxonomie VRM (DISCLOSURE, EDIFICATION, ADVISEMENT, CONFIRMATION, QUESTION, ACKNOWLEDGEMENT, INTERPRETATION et REFLECTION) ne sont pas suffisantes pour permettre une analyse fine des conversations. Elles ne permettent par exemple pas de faire la différence entre les fonctions commissives et les fonctions expressives (*e.g.* « je vais me coucher » et « j'aime l'opéra ») ni de distinguer une demande d'action d'une demande d'information, ni de capturer les énoncés de politesse, *etc*.

En ce qui concerne les forums, Qadir et Riloff (2011) cherchent à classifier les actes de discours dans les messages. Ils distinguent le texte contenant ces actes de ce qu'ils nomment « texte expositif », qu'ils définissent comme le discours comportant uniquement de l'information factuelle. Leur taxonomie est composée de quatre labels de Searle (1976) : les commissifs, les directifs, les expressifs et les représentatifs. Ils ignorent les déclaratifs, trop rares dans leur corpus, et les représentatifs, qu'ils ne semblent pas considérer comme des actes de langage, contrairement à la théorie qui leur attribue un caractère illocutoire (à savoir, l'intention de faire connaître leur contenu sémantique). Malgré cette taxonomie assez limitée, leur travail rapporte des résultats encourageants, d'autant plus que les traits utilisés pour leur classifieur -

un SVM - ne demandent pas d'analyse linguistique et se prêtent bien aux phrases peu grammaticales que l'on peut rencontrer dans les corpus de CMR.

Les chats sont également fréquemment utilisés pour la résolution collaborative de problèmes, et se distinguent des autres modalités étudiées par leur caractère synchrone. Ha et al. (2013) s'y intéressent dans le cadre du développement de systèmes de dialogue : ils cherchent à la fois à identifier le type d'acte accompli par l'utilisateur et à choisir le type d'acte que le système doit produire. Leur approche utilise un classifieur d'entropie maximale (*MaxEnt*). Leur corpus est constitué de situations de tutorat plus que de situations d'aide à la résolution de problèmes, ce qui est différent même si les deux ont des similarités. Cette distinction se ressent dans leur taxonomie, qui inclut des labels particulièrement présents dans les dialogues de tutorat (*e.g.* HINT, POSITIVE FEEDBACK). Leurs résultats dépassent l'état de l'art pour la tâche de classification, et ils parviennent également à prédire le *timing* des interventions du tuteur avec une bonne précision. Ivanovic (2005) cherche à classifier les énoncés des messages chats et se base sur une taxonomie dérivée du schéma d'annotation DAMSL (Core et Allen, 1997). Il parvient à une précision de 80 % en combinant un classifieur naïf de Bayes et un modèle de n-grammes. Kim et al. utilisent la même taxonomie qu'Ivanovic et l'appliquent à un corpus de conversations biparticipants (Kim et Baldwin, 2010) et multiparticipants (Kim et al., 2012). Un classifieur basé sur les CRF obtient des précisions très élevées (plus de 97 %) pour les deux tâches en combinant des traits lexicaux, structurels et relationnels.

6.3 Méthode

Nous appréhendons le problème de la modélisation des conversations comme une tâche de classification des énoncés en termes d'actes de dialogue basée sur un apprentissage statistique supervisé.

Cette section détaille l'implémentation de notre approche, les traits choisis pour caractériser les énoncés, notre méthode d'évaluation, et enfin le corpus utilisé.

6.3.1 Approche et implémentation

Nous avons choisi de développer un système basé sur un classifieur SVM multiclasses. La première raison qui a motivé notre choix, c'est qu'un CRF ou un HMM-SVM ne permettrait pas de comparer nos trois modalités, puisque les conversations chats ne peuvent pas facilement être reconstituées pour permettre leur étiquetage séquentiel, et nous cherchons à effectuer une comparaison homogène sur différentes modalités. De plus, étant donné que les labels de notre taxonomie sont fortement déséquilibrés, nous voulions éviter une approche qui soit dépendante de l'observa-

tion *a priori* des annotations. Enfin, l'efficacité des SVMs a été démontrée pour la reconnaissance d'actes de dialogue (Lampert et al., 2006; Qadir et Riloff, 2011). Sauf mention contraire, ce classifieur est utilisé dans la plupart des expériences rapportées. L'implémentation utilise la librairie *liblinear* [1] (Fan et al., 2008).

6.3.2 Traits

Nous cherchons à tester l'efficacité de traits classiquement utilisés en reconnaissance des actes de dialogue sur un corpus de modalités des CMR. L'efficacité des traits lexicaux tels que les *n*-grammes, et en particulier les unigrammes, a été établie dans le cadre de la tâche de classification des actes de dialogue (Kim et Baldwin, 2010; Sun et Morency, 2012; Ferschke et al., 2012; Ravi et Kim, 2007; Carvalho et Cohen, 2005). Nous avons donc choisi de faire des unigrammes et des bigrammes les traits principaux du classifieur. Différentes expériences ont été effectuées avec des combinaisons d'unigrammes, de bigrammes, de trigrammes et de quadrigrammes, mais c'est la combinaison des deux premiers qui a produit les meilleurs résultats. Les autres traits que nous avons testés sont communément utilisés pour des tâches similaires, et incluent les racines des mots, leurs lemmes, leurs étiquettes morpho-syntaxiques (ou *parts-of-speech*), ainsi que des informations contextuelles (*e.g.* la position de l'énoncé (Ferschke et al., 2012), l'auteur de l'énoncé (Sun et Morency, 2012), la taille de l'énoncé (Ferschke et al., 2012; Lampert et al., 2006)...).

6.4 Expériences et discussion

Dans cette section, nous décrivons les différentes expériences, rapportons leurs résultats et les discutons. Les deux tâches sont : la classification des énoncés en termes de dimensions sémantiques, et en termes de fonctions communicatives. Tous les scores sont calculés par validation croisée ($k = 10$). La partition s'est faite en prenant en compte les conversations, c'est-à-dire qu'une même conversation ne peut pas se retrouver à la fois dans le corps d'apprentissage et de test. Les métriques utilisées sont les suivantes : l'exactitude (nombre d'énoncés correctement classifiés sur l'ensemble des énoncés), la précision (nombre de vrais positifs sur l'ensemble des vrais et faux positifs) et le rappel (nombre de vrais positifs sur l'ensemble de vrais positifs et faux négatifs). La macro-moyenne (moyenne des moyennes) et la micro-moyenne (moyenne globale) sont rapportées pour la précision et le rappel. Pour évaluer nos résultats, nous avons choisi le maximum de vraisemblance comme approche de base. La segmentation des messages en énoncés ne fait pas partie des tâches évaluées dans cette section. Pour toutes ces expériences, nous utilisons la segmentation semi-automatique effectuée lors de la construction du corpus Ubuntu.

1. *L2-regularized L2-loss support vector classification (dual)*, $\epsilon = 0.1$, $C = 1$.

Nous commençons par rapporter les résultats de notre classifieur SVM sur l'ensemble du corpus multimodal avant de comparer les différentes modalités au travers de plusieurs expériences.

6.4.1 Comparaison d'un classifieur SVM à l'approche de base

Les premières expériences consistent à entraîner un classifieur SVM multi-classes sur l'ensemble des données (courriels, forums et chats) en utilisant les traits lexicaux. Les résultats de ces expériences sont rapportés en table 6.1. Pour la classification des dimensions sémantiques, l'exactitude est assez haute, avec 92 % des instances correctement classifiées. Les fonctions communicatives, quant à elles, sont correctement classifiées dans 63 % des cas. Les macro-moyennes, plus faibles, indiquent que le classifieur est moins performant sur certains labels moins représentés. Le classifieur SVM multi-classes ne peut pas tenir compte de la possibilité qu'un énoncé puisse contenir plusieurs actes de dialogue, cependant dans les faits une infime minorité des énoncés contenant des annotations dans différentes dimensions, cette limitation n'a qu'un impact minime sur les résultats. Dans tous les cas, le classifieur obtient de meilleures performances que l'approche de base.

Tâche	Exactitude	Macro P.	Macro R.	Micro P.	Micro R.
dimensions : n-grammes	**0,92**	**0,56**	**0,37**	**0,81**	**0,91**
dimensions : max. vrais.	0,60	0,09	0,09	0,55	0,55
fonctions : n-grammes	**0,63**	**0,42**	**0,36**	**0,52**	**0,58**
fonctions : max. vrais.	0,23	0,03	0,03	0,14	0,14

Tab. 6.1.: Comparaison des performances d'un classifieur SVM entraîné sur les n-grammes et l'approche de base sur l'ensemble du corpus, toutes modalités confondues, pour la tâche de reconnaissance des dimensions sémantiques et des fonctions communicatives.

Label	Courriels			Forums			Chats		
	Nb.	P.	R.	Nb.	P.	R.	Nb.	P.	R.
Domain Activities	894	0,86	0,95	1040	0,70	0,96	1638	0,90	0,99
Social Obligations M.	405	0,70	0,97	161	0,86	0,83	184	0,99	0,74
Discourse M.	27	0,67	0,07	64	0,28	0,17	17	0,00	0,00
Evaluation	2	0,00	0,00	22	0,79	0,50	79	0,65	0,41
Extra Discourse	48	1,00	0,85	2	0,00	0,00	0	-	-
Psychological State	5	0,00	0,00	6	0,00	0,00	29	0,22	0,07
Communication M.	0	-	-	0	-	-	27	0,93	0,48
Contact M.	0	-	-	0	-	-	18	0,50	0,06

Tab. 6.2.: Dimensions sémantiques : nombre d'exemples dans la référence et résultats d'un classifieur SVM entraîné avec les n-grammes sur les différentes modalités du corpus (les deux labels non représentés sont absents ou ont une précision et un rappel de 0,00 sur toutes les modalités).

6.4.2 Comparaison d'approches et de jeux de traits sur différentes modalités

Label	Courriels			Forums			Chats		
	Nb.	P.	R.	Nb.	P.	R.	Nb.	P.	R.
Inform	461	0,56	0,81	415	0,33	0,66	541	0,46	0,70
Answer	155	0,29	0,22	255	0,30	0,38	354	0,37	0,34
Request for Information	96	0,74	0,62	118	0,67	0,59	328	0,85	0,80
Request for Action	85	0,41	0,27	113	0,42	0,32	178	0,52	0,36
Answer Positively	42	0,47	0,33	73	0,52	0,47	183	0,68	0,61
Greetings	94	0,86	0,95	62	0,82	0,97	113	0,98	0,82
Correct	18	0,00	0,00	34	0,00	0,00	95	0,52	0,18
Final Self Introduction	127	0,45	0,98	10	0,67	0,20	0	-	-
Answer Negatively	26	0,25	0,08	37	0,65	0,35	68	0,20	0,12
Thanking	47	0,64	0,62	38	0,76	0,66	41	0,83	0,73
Valediction	85	0,77	0,87	17	0,69	0,53	10	0,67	0,40
Anticipate Thanking	41	0,67	0,78	24	0,55	0,50	8	0,33	0,12
Commit	14	0,33	0,07	14	0,33	0,07	39	0,48	0,38
Report Speech	4	0,00	0,00	49	0,26	0,24	0	-	-
Boilerplate	48	0,98	0,88	0	-	-	0	-	-
Announce	17	0,50	0,12	11	0,00	0,00	7	0,00	0,00
Request for Directive	5	0,50	0,60	11	0,50	0,09	17	0,42	0,29
Apologizing	8	1,00	0,38	9	0,86	0,67	9	1,00	0,33

Tab. 6.3.: Fonctions communicatives : nombre d'exemples dans la référence et résultats d'un classifieur SVM entraîné avec les n-grammes sur les différentes modalités du corpus (les deux labels non représentés sont absents ou ont une précision et un rappel de 0,00 sur toutes les modalités).

Dans cette section nous présentons les résultats obtenus avec des modèles propres aux différentes modalités (courriels, forums et chats). Les multiples ensembles de traits (racines, lemmes, formes morpho-syntaxiques et traits contextuels) ont été appliqués pour cette tâche. Les tables 6.4 et 6.5 rapportent les résultats pour la tâche de classification des dimensions sémantiques et des fonctions communicatives, respectivement. Il apparaît que le jeu de traits n'a que peu d'incidence sur les résultats, et que les approches basées sur les n-grammes sont quasiment toujours optimales, avec une légère amélioration lorsqu'ils sont couplés aux étiquettes morpho-syntaxiques. On peut en conclure qu'un système basé sur la classification multi-classes des énoncés atteint rapidement ses limites : certains labels sont caractérisés par des mots très discriminants (*e.g.* « OK » pour ANSWER POSITIVELY, « bonjour » pour GREETINGS), d'autres ne le sont pas et nécessitent une approche différente pour être correctement reconnus. Par exemple, le label CORRECT indique que la fonction communicative d'un énoncé est de corriger un énoncé précédent. Typiquement ce type d'énoncés apparaît de la façon suivante :

1. « *oui c'est ce que je voulais rire* »
2. « *voulais dire* »

Dans une situation comme celle-ci, l'examen de l'énoncé (2) pris en isolation ne comporte aucune indication qu'il s'agit d'une correction, même pour un humain. On ne peut deviner que sa fonction est de corriger l'énoncé (1) que en ayant connaissance de l'énoncé précédent : en effet, l'énoncé (2) est très proche de la fin de l'énoncé (1), qu'il corrige. Les traits utilisés dans nos expériences ne sont pas capables de capturer cette information, et les classifieurs obtiennent un rappel et une précision de 0 sur les forums et les courriels. Les résultats sont meilleurs pour les chats, la convention y étant d'utiliser le symbole « * » en fin ou début d'énoncé pour indiquer qu'il s'agit d'une correction. Ce symbole constituant une unité lexicale, les traits utilisés suffisent parfois à correctement identifier les actes de correction.

La table 6.2 présente les scores pour les différentes dimensions sémantiques et montre une forte variation entre les labels. On constate qu'elles sont très déséquilibrées : la plus importante, DOMAIN ACTIVITIES , regroupe 75 % des énoncés. La seconde, SOCIAL OBLIGATIONS MANAGEMENT, en regroupe plus de 15 %. Les huit autres se partagent donc seulement 10 % des énoncés restants, certaines ne contenant qu'une poignée d'énoncés. La table 6.3 présente les scores des fonctions communicatives. On s'aperçoit que les labels représentant les fonctions communicatives sont un peu mieux équilibrés que celles représentant les dimensions, cependant on note tout de même que plusieurs d'entre eux ne sont pratiquement pas représentés dans le corpus.

Les tables 6.4 et 6.5 incluent également les résultats d'une expérience effectuée en remplaçant le classifieur SVM par un CRF. Nous avons utilisé la librairie Mallet (McCallum, 2002) pour notre implémentation. Le corpus de chats a été exclu de l'expérience puisqu'il ne peut pas être analysé séquentiellement avant que les conversations qu'il contient ne soient « démêlées ». Les modèles sont entraînés sur les n-grammes, auxquels sont ajoutés un trait indiquant l'auteur du message. On constate que l'approche basée sur les CRF obtient de très bons résultats sur les forums, où elle bat le classifieur SVM. Elle obtient de moins bons résultats sur les courriels, en particulier dans le cas des fonctions communicatives pour lesquelles les labels sous représentés sont nettement moins bien annotés, comme l'indique les macro-moyennes presque deux fois moins élevées qu'avec un SVM. Ce résultat est contre intuitif, puisque les courriels étant plus formellement construits et tendant plus à respecter un schéma standard, on pourrait s'attendre à ce qu'il soit pertinent de les étiqueter séquentiellement. En réalité, cet apprentissage structurel opéré par le CRF cause beaucoup d'erreurs dès qu'un courriel sort du schéma typique.

Corpus	Traits	Exact.	Macro-P	Macro-R	Micro-P	Micro-R
Courriels	max. de vraisemblance	0,40	0,13	0,13	0,40	0,40
	n-grammes	**0,94**	0,42	0,40	0,81	0,93
	n-grammes + P.O.S.	**0,94**	0,44	**0,41**	0,81	**0,94**
	lemmes	**0,94**	0,46	**0,41**	0,81	**0,94**
	n-grammes + lemmes	**0,94**	0,46	**0,41**	0,81	0,93
	racines	0,92	0,38	0,36	0,71	0,89
	n-grammes + racines	**0,94**	0,46	**0,41**	0,81	0,93
	n-grammes + contexte	0,93	0,46	0,40	0,79	0,91
	n-grammes (CRF)	0,91	**0,56**	0,37	**0,90**	0,91
Forums	max. de vraisemblance	0,47	0,12	0,12	0,46	0,46
	n-grammes	**0,92**	0,38	0,36	**0,71**	0,89
	n-grammes + P.O.S.	**0,92**	0,37	0,37	**0,71**	**0,90**
	lemmes	**0,92**	0,37	0,36	**0,71**	**0,90**
	n-grammes + lemmes	0,91	0,37	0,36	**0,71**	0,89
	racines	**0,92**	0,38	0,36	**0,71**	0,89
	n-grammes + racines	0,91	0,37	0,36	**0,71**	0,89
	n-grammes + contexte	0,89	0,29	0,38	0,68	0,86
	n-grammes (CRF)	0,91	**0,69**	**0,46**	**0,71**	0,89
Chats	max. de vraisemblance	0,68	0,10	0,10	0,68	0,68
	n-grammes	0,89	0,42	0,30	0,89	0,89
	n-grammes + P.O.S.	**0,90**	**0,54**	**0,32**	**0,90**	**0,90**
	lemmes	0,89	0,42	0,29	0,89	0,89
	n-grammes + lemmes	0,89	0,46	0,31	0,89	0,89
	racines	**0,90**	0,47	0,30	**0,90**	**0,90**
	n-grammes + racines	**0,90**	0,48	0,31	**0,90**	**0,90**
	n-grammes + contexte	**0,90**	0,47	0,30	**0,90**	**0,90**

Tab. 6.4.: Résultats pour la tâche de reconnaissance des dimensions sémantiques sur différentes modalités, différents traits et différentes approches.

6.5 Comparaison de classifieurs trans-modalités

Dans cette section, nous présentons les résultats d'expériences effectuées avec des classifieurs trans-modalités, c'est-à-dire entraînés sur des données issues de modalités autres que la modalité de test. Tous les classifieurs sont de type SVM et ont été entraînés sur des n-grammes. La table 6.6 détaille les résultats des expériences. Chaque classifieur a été entraîné sur deux des trois modalités, la troisième servant d'ensemble de test. Par exemple, les résultats obtenus sur la modalité « courriels » sont ceux d'un classifieur entraîné sur les données de chats et de forums.

Sans surprise, dans la plupart des cas effectuer l'apprentissage sur des données issues d'une modalité différente de celle pour laquelle on cherche à classifier des énoncés conduit à une perte de performance. Les chats sont la modalité la plus affectée par cette perte de performance : on constate une baisse de 11 points d'exactitude pour la tâche de classification des dimensions sémantiques et de 9 points pour celle de classification des fonctions communicatives. On peut en conclure que les champs

Corpus	Traits	Exact.	Macro-P	Macro-R	Micro-P	Micro-R
Courriels	max. de vraisemblance	0,13	0,05	0,05	0,13	0,13
	n-grammes	**0,72**	**0,45**	**0,41**	0,59	0,68
	n-grammes + P.O.S.	0,71	0,40	0,40	0,58	0,66
	lemmes	0,71	0,43	0,40	0,58	0,66
	n-grammes + lemmes	0,71	**0,45**	0,40	0,58	0,67
	racines	0,70	0,43	0,40	0,57	0,66
	n-grammes + racines	0,70	0,43	0,40	0,57	0,65
	n-grammes + contexte	0,70	0,43	0,40	0,57	0,65
	n-grammes (CRF)	0,69	0,23	0,22	**0,68**	**0,69**
Forums	max. de vraisemblance	0,15	0,04	0,04	0,15	0,15
	n-grammes	0,61	0,40	0,28	0,41	0,51
	n-grammes + P.O.S.	0,60	0,33	0,29	0,39	0,50
	lemmes	0,61	0,37	0,29	0,41	0,51
	n-grammes + lemmes	0,62	0,35	0,29	**0,47**	0,44
	racines	0,61	0,36	0,28	0,40	0,50
	n-grammes + racines	0,61	0,33	0,28	0,40	0,51
	n-grammes + contexte	0,57	0,29	0,28	0,37	0,46
	n-grammes (CRF)	**0,68**	**0,66**	**0,46**	**0,47**	**0,59**
Chats	max. de vraisemblance	0,15	0,04	0,04	0,15	0,15
	n-grammes	**0,56**	0,35	**0,27**	**0,56**	**0,56**
	n-grammes + P.O.S.	0,54	0,33	0,26	0,54	0,54
	lemmes	0,55	0,35	0,26	0,55	0,55
	n-grammes + lemmes	**0,56**	0,35	**0,27**	**0,56**	**0,56**
	racines	0,55	0,35	**0,27**	0,55	0,55
	n-grammes + racines	**0,56**	**0,36**	**0,27**	**0,56**	**0,56**
	n-grammes + contexte	**0,56**	0,35	**0,27**	**0,56**	**0,56**

Tab. 6.5.: Expériences pour la tâche de reconnaissance des fonctions communicatives sur différentes modalités, différents traits et différentes approches.

lexicaux correspondant aux différentes fonctions et dimensions varient en fonction de la modalité. C'est-à-dire que pour accomplir un même acte, les unités lexicales employées seront différentes d'une modalité à l'autre.

6.6 Conclusion

Nous avons présenté nos travaux en classification automatique des énoncés en termes d'actes de dialogue, dans un corpus de conversations écrites en ligne à modalités multiples portant sur la résolution collaborative de problèmes. Nous avons rapporté les résultats de nombreuses expériences visant à confronter les approches traditionnelles de classification des actes de dialogue à des données extraites de différentes modalités CMR. Il s'agit à notre connaissance du premier travail qui examine des données tirées de différentes modalités avec la même taxonomie et les mêmes approches. Nous avons rapporté les variations observées entre les modalités, et nous avons montré que des résultats intéressants peuvent être atteints même en se limitant à l'utilisation de traits purement lexicaux.

Modalité	Exactitude	Macro P.	Macro R.	Micro P.	Micro R.
			Dimensions sémantiques		
Forums	0,90 (-0,02)	0,39 (-0,01)	0,30 (-0,06)	0,69 (-0,02)	0,88 (-0,01)
Chats	0,78 (-0,11)	0,28 (-0,14)	0,23 (-0,07)	0,78 (-0,11)	0,78 (-0,11)
Courriels	0,82 (-0,06)	0,27 (-0,15)	0,29 (-0,11)	0,69 (-0,12)	0,79 (-0,14)
			Fonctions communicatives		
Forums	0,61 (+0,00)	0,36 (-0,04)	0,27 (-0,01)	0,40 (-0,01)	0,50 (-0,01)
Chats	0,47 (-0,09)	0,33 (-0,02)	0,28 (+0,01)	0,57 (-0,01)	0,47 (-0,09)
Courriels	0,55 (-0,06)	0,44 (+0,04)	0,30 (+0,02)	0,41 (+0,00)	0,48 (-0,03)

Tab. 6.6.: Comparaison des performances de classifieurs entraînés sur des modalités autres que la modalité de test.

Le travail présenté dans ce chapitre a fait l'objet de deux publications à TALN 2016 (Salim, Hernandez, et Morin, 2016; de Mazancourt, Salim, et Recourcé, 2016).

7 Conclusion générale

Cette thèse a pour but de fournir un cadre pour l'analyse discursive des conversations écrites en ligne orientées vers la résolution collaborative de problèmes. Nous avons articulé nos travaux autour du concept d'acte de dialogue, fondamental pour la modélisation des conversations en termes de fonctions communicatives des énoncés. Ce travail, qui s'inscrit dans le cadre de la recherche sur les communications médiées par les réseaux, s'est focalisé sur la nature des conversations écrites en ligne porteuses de demandes d'assistance, sur la modélisation taxonomique des actes de dialogue et sur l'interopérabilité entre taxonomies de fonctions communicatives. Ces travaux ont donné lieu à la création d'un nouveau corpus libre et francophone pour l'étude de ce type de conversations, qui nous a servi à étudier l'application d'algorithmes d'apprentissage supervisé pour la reconnaissance des actes de dialogue dans ce domaine.

7.1 Synthèse des contributions

Cette thèse a donné lieu à cinq publications nationales et internationales, aux conférences TALN-RECITAL, GSCL (NLP4CMC), LREC, JEP-TALN et CICLING.

Ces publications incluent :

— Un état de l'art sur l'analyse automatique des conversations écrites en ligne porteuses de demandes d'assistance (Salim, 2015)

— Une publication sur la pertinence des informations discursives pour la tâche de désentremelement des conversations chat (Riou et al., 2015)

— Une publication sur le corpus Ubuntu-fr (Hernandez et Salim, 2015)

— Une publication sur la reconnaissance automatique des actes de dialogue (Salim et al., 2016)

— Une publication sur le méta-modèle pour l'interopérabilité entre taxonomies d'actes de dialogues (Salim et al., 2017)

— Une démonstration des outils produits dans le cadre du projet ODISAE a également été présentée à TALN (de Mazancourt et al., 2016)

7.1.1 Modélisation des actes de dialogue

Nous avons montré que les actes de dialogue sont un outil pertinent et largement employé par la communauté scientifique dans le domaine de l'analyse du dialogue. Cependant, nous avons relevé qu'étant donné que l'existant dans le domaine est basé de façon prédominante sur les conversations orales, et étant donné que les modalités écrites des conversations en ligne leur sont très différentes, il serait difficile d'appliquer tels quels les ressources et outils existants. Pour le démontrer, nous avons effectué une étude d'applicabilité du standard ISO 24617-2 pour l'annotation de conversations écrites en ligne en nous basant sur l'examen de messages extraits de la plate-forme francophone ubuntu-fr. Nous avons montré que plusieurs aspects des conversations écrites en ligne limitent la pertinence du standard pour leur annotation : le fait qu'elles soient multi-participants, leur caractère asynchrone ou pseudo-synchrone, leur caractère multi-canal, ainsi que les contraintes techniques qui leurs sont propres.

Nous avons par ailleurs proposé une nouvelle taxonomie, dérivée méthodiquement de la taxonomie du standard, dont la fonction est de répondre aux problèmes soulevés dans notre étude d'applicabilité. Cette taxonomie respecte le cadre conceptuel du standard ainsi que ses consignes d'adaptation.

7.1.2 Méta-modélisation des taxonomies d'actes de dialogue

Pour répondre au problème du manque d'interopérabilité des taxonomies d'actes de dialogue, nous avons proposé une solution originale : l'emploi d'un méta-modèle pour l'abstraction des taxonomies d'actes de dialogue. Le méta-modèle est construit en décomposant les labels des actes de dialogue en traits fonctionnels primitifs, définis comme des aspects des actes qui peuvent être capturés par différents labels dans différentes taxonomies.

Nous avons construit un méta-modèle expérimental pour les taxonomies de SWBD-DAMSL, du standard ISO et de MapTask. Au travers d'une première expérience, nous avons montré qu'il était possible d'identifier le label correct d'un énoncé dans une taxonomie donnée en ayant seulement connaissance de son label dans une autre taxonomie couverte par le méta-modèle. Une seconde expérience a montré qu'il était possible d'utiliser ce principe pour entraîner un classifieur SVM à produire des annotations pour une taxonomie différente de la taxonomie utilisée pour annoter les données d'apprentissage. Nous avons donc montré qu'il était possible de construire des systèmes de reconnaissance d'actes de dialogue qui ne nécessitent pas de données annotées avec la taxonomie visée mais seulement des données annotées avec une taxonomie capable de capturer suffisamment d'informations pertinentes.

7.1.3 Corpus Ubuntu-fr

Nous avons construit un large corpus libre et francophone de conversations écrites en ligne porteuses de demande d'assistance, extraites de la plate-forme Ubuntu. Le corpus se décline en trois sous-corpus, un pour chaque modalité étudiée dans cette thèse : forums, chats et courriels. Les trois sous-corpus contiennent des conversations de même nature, de la même période et portant sur le même domaine, pour permettre leur comparaison.

Le corpus a été construit autour d'un modèle robuste de méta-données basé sur des principes forts, tels que le principe d'annotation en « *stand off* ». Les données ont été pré-traitées pour les préparer à des tâches d'annotation et à des traitements TAL : les méta-données (*e.g.* auteur, nombre de vues, destinataires, *etc*.) ont été extraites des conversations, les encodages ont été unifiés en UTF-8, la structure des conversations a été identifiée et marquée, et les documents ont été segmentés en énoncés et en unités lexicales. De plus, plus de 4 700 énoncés contenus dans les trois modalités ont été annotés en termes d'actes de dialogue, de qualifieurs et d'opinion, de sentiment et d'émotion. Ces annotations ont été prévues pour permettre l'application d'algorithmes d'apprentissage statistique sur le corpus.

7.1.4 Classification automatique des actes de dialogue

Pour ouvrir la voie à des systèmes automatiques d'analyse des conversations, nous avons effectué un travail expérimental sur la tâche de reconnaissance automatique des actes de dialogue. Nous avons utilisé le corpus Ubuntu-fr pour effectuer nos expériences. Ces expériences portent sur deux sous-tâches : la reconnaissance des dimensions sémantiques des actes de dialogue, et la reconnaissance de leurs fonctions communicatives.

Nous avons rapporté les résultats de nombreuses expériences effectuées sur des conversations extraites d'IRC, de forums et de listes de diffusion. Nous avons employé des SVM et des CRF entraînés avec différents jeux de traits typiquement utilisés pour ce type de tâche : des traits lexicaux, morpho-syntaxiques et contextuels. Les classifieurs ont systématiquement obtenu des résultats supérieurs à l'approche de base. Nous avons rapporté les variations observées entre modalités et montré l'importance des informations lexicales ainsi que l'importance d'avoir une approche séquentielle pour l'analyse des conversations.

Ces travaux en reconnaissance automatique ont été exploités dans le cadre du projet ODISAE, pour lequel nous avons contribué un moteur d'analyse de conversations issues des CMR en termes d'actes de dialogue.

7.2 Perspectives

L'objectif de cette thèse a été de fournir un cadre pratique pour l'analyse des conversations écrites en ligne porteuses de demandes d'assistance en termes d'actes de dialogue. Nous avons effectué plusieurs avancées dans cette direction, cependant il reste encore des pistes à explorer.

7.2.1 Réseaux de neurones

Les récentes avancées de la recherche en apprentissage profond ont permis d'importants gains de performances dans plusieurs domaines du traitement automatique de la langue. Ces progrès rapides ont mené Manning et al. (2014) à comparer le phénomène à un « tsunami » qui aurait frappé toutes les conférences majeures en linguistique computationnelle. Le domaine de l'analyse du dialogue n'est pas différent, et plusieurs travaux concernant l'application de réseaux de neurones à la tâche de reconnaissance des actes de dialogue ont été publiés (Khanpour et al., 2016; Lee et Dernoncourt, 2016; Kalchbrenner et Blunsom, 2013), obtenant des résultats supérieurs à l'état de l'art sur plusieurs corpus.

Il serait intéressant de confronter ces systèmes basés sur l'apprentissage profond à des corpus de conversations écrites en ligne. Les dialogues étant des phénomènes complexes, l'une des principales difficultés en construisant un système de reconnaissance des actes de dialogue se trouve dans le choix de traits pertinents à utiliser pour permettre une classification efficace. L'avantage des réseaux de neurones est qu'ils ne dépendent pas de leur concepteur pour identifier les traits, et collectent automatiquement les informations statistiquement pertinentes à partir de l'observation de mots, de phrases et de séquences.

Cerisara et al. (2017) emploient les *word embeddings* produits par l'algorithme *Word2Vec* (Mikolov et al., 2013) pour l'améliorer les performances d'un réseau de neurones dans le cadre d'une tâche de reconnaissance des actes de dialogue. Lilleberg et al. (2015) explorent l'utilisation des *embeddings* pour ajouter de l'information sémantique aux traits employés par un classifieur SVM. En se basant sur leurs travaux, nous avons déjà effectué des expériences préliminaires pour essayer de reconnaître des actes de dialogue en utilisant un SVM ayant été entraîné sur des *embeddings*. À ce stade, les résultats sont très encourageants, et nous pensons qu'il serait intéressant d'explorer cette option.

7.2.2 Interopérabilité inter-taxonomique

Nous avons présenté nos travaux sur le développement d'un méta-modèle destiné à rendre les taxonomies plus inter-opérables. À ce stade, nos expériences représentent une importante preuve de concept : il est effectivement possible de produire des annotations pour une taxonomie différente de celle utilisée pour annoter les données d'entraînement. Cependant, il devrait être possible d'aller plus loin dans cette direction s'il on souhaite aboutir à un système véritablement performant.

Les classifieurs que nous avons utilisés dans cette expérience sont simples, et nous avons montré qu'ils étaient peu efficaces lorsqu'il est question de reconnaître directement les trait fonctionnels. Le meilleur système que nous ayons présenté consiste à reconnaître un label, puis à le convertir vers le label le plus pertinent dans la taxonomie de destination. Mais il serait intéressant de travailler sur l'alternative, à savoir apprendre et reconnaître directement les traits fonctionnels : en effet, cette démarche est nettement plus flexible, puisqu'elle permettrait d'apprendre à partir de données extraites de différents corpus, chacun pouvant être annoté avec sa propre taxonomie. À ce stade des expérimentations, les mêmes ensembles de caractéristiques ont été utilisés pour la reconnaissance des différents traits primitifs, mais une approche différente pour chaque type de trait devrait grandement améliorer le système, et permettre la reconnaissance directe des traits primitifs.

Cette solution a deux avantages. Le premier, bien sûr, est qu'elle permet d'utiliser plus de corpus, donc plus de données. Le second, tout aussi important, est qu'elle permet d'utiliser différents corpus pour apprendre tous les traits fonctionnels présents dans la taxonomie de destination, ce qui n'est pas possible autrement si la taxonomie de destination comporte des traits absents des annotations du corpus d'entraînement. Par exemple, admettons que la taxonomie pour laquelle nous souhaitons produire des annotations contienne les traits A, B et C. Pour notre apprentissage, nous disposons du corpus X, dont la taxonomie comprend les traits A et B, et le corpus Y, dont la taxonomie comprend les traits A et C. Si nous pouvons apprendre directement les traits, nous pouvons nous permettre de fusionner les corpus X et Y pour apprendre les trois traits, A, B et C.

Employer des techniques sophistiquées pour obtenir des résultats comparables à ceux d'un classifieur état-de-l'art en se basant sur les traits fonctionnels pourrait grandement réduire le besoin de produire de nouvelles annotations pour pouvoir utiliser une nouvelle taxonomie dans le contexte d'une tâche de reconnaissance des actes de dialogue. Il est extrêmement complexe et coûteux d'acquérir de nouvelles données annotées dans le domaine. Un tel système, s'il est suffisamment performant, pourrait grandement simplifier l'accomplissement de travaux futurs dans le domaine. On peut même imaginer l'extension du principe à d'autres tâches : le manque de

données annotées représente un des principaux obstacles au développement de nouveaux outils linguistiques basés sur des techniques d'apprentissage supervisé. Permettre la réutilisation de corpus annotés à l'aide de taxonomies différentes peut constituer un moyen intéressant de pallier à ce manque.

Annexes

A Taxonomie d'actes de dialogue employée pour le corpus Ubuntu-fr

La taxonomie présentée dans cette annexe se conforme aux principes conceptuelles du standard ISO 24617-2. Les actes de dialogue sont considérés comme la conjonction d'une *dimension sémantique* et d'une *fonction communicative*.

Les dimensions sémantiques représentent l'aspect de la conversation sur lequel l'énoncé porte, tandis que les fonctions communicatives représentent l'intention rhétorique du locuteur, *i.e.* ce qu'il cherche à accomplir en produisant l'énoncé.

A.1 Dimensions sémantiques

DOMAIN / ACTIVITIES :

Concerne ce qui se rapporte strictement à la tâche, au domaine et à l'activité autour de laquelle la conversation est structurée, comme les instructions et les questions sur le sujet de la discussion (*e.g.* « ça fait longtemps que tu es sous Ubuntu ? »).

SOCIAL OBLIGATION MANAGEMENT :

Concerne les énoncés porteurs d'actes de gestion sociale du dialogue et de politesse, comme les remerciements ou les salutations.

DISCOURSE MANAGEMENT :

Concerne les actes qui ont pour but de façonner la structure du discours, comme les annonces (*e.g.* « laissez-moi vous expliquer ») ou les demandes de changement de sujet (*e.g.* « parlons plutôt du problème de driver »).

EXTRA DISCOURSE :

Concerne les actes qui ont une fonction communicative mais qui ne sont pas discursifs, propres aux modalités de l'écrit en ligne, comme l'envoi d'un lien, d'un log ou le copié/collé d'un message d'erreur.

Evaluation :

Concerne les actes de *feedback* qui portent sur le résultat de l'évaluation du contenu d'un énoncé précédent, *i.e.* de la comparaison de l'information nouvelle avec les connaissances antérieures du locuteur (*e.g.* « hein ? mais juste avant tu m'as dit de faire le contraire ? »).

Attention Perception Interpretation :

Concerne les actes de *feedback* qui rapportent la bonne ou mauvaise perception (*e.g.* j'ai bien reçu ton sms »), compréhension (*e.g.* j'ai rien compris »), ou sur le manque d'attention du locuteur ou d'un allocutaire (*e.g.* tu m'écoutes ? »).

Psychological State :

Concerne les actes qui portent sur l'état psychologique et mental du locuteur (*e.g.* ça me fait marrer tout ça ! »).

Contact Management :

Concerne les énoncés porteurs d'actes de gestion et de contrôle de la transmission, que ce soit au niveau de son établissement, de son maintien, de sa reprise ou de son interruption (*e.g.* « je déco dans 5 minutes »).

Communication Management :

Concerne les actes de gestion des retours et des corrections portant sur la forme, le rendu, l'encodage ou la réalisation linguistique des énoncés générés (*e.g.* « pourquoi on voit pas les accents quand j'écris ? »).

Time Management :

Concerne les énoncés porteurs d'actes de gestion du temps, comme ceux signifiant que le locuteur a besoin de plus de temps pour contribuer ou qu'une pause dans le dialogue est nécessaire (*e.g.* « euuuuh... hmmm... »).

A.2 Fonctions communicatives

Comme dans le standard ISO, nous distinguons les fonctions génériques des fonctions spécifiques.

A.2.1 Fonctions communicatives génériques

Les fonctions communicatives génériques peuvent être utilisées conjointement avec n'importe quelle dimension.

INFORM :

Le locuteur fournit une information à l'allocutaire.

ANSWER :

Le locuteur fournit une information précédemment demandée par l'allocutaire.

ANSWER POSITIVELY :

Le locuteur répond positivement à une requête de l'allocutaire.

ANSWER NEGATIVELY :

Le locuteur répond négativement à une requête de l'allocutaire.

COMMIT :

Le locuteur annonce qu'il a l'intention d'une action, telle que décrite dans l'énoncé.

REQUEST FOR ACTION :

Le locuteur cherche à obtenir de l'allocutaire qu'il accomplisse une action, telle que décrire dans l'énoncé.

REQUEST FOR INFORMATION :

Le locuteur désire obtenir une information, qu'il suppose que l'allocutaire détient.

REQUEST FOR DIRECTIVE :

Le locuteur désire obtenir des instructions pour obtenir un résultat décrit dans l'énoncé.

A.2.2 Fonctions communicatives spécifiques

Les fonctions communicatives spécifiques ne peuvent être employées que pour les énoncés d'une dimension particulière.

Comunication Management

CORRECT :

Le locuteur corrige une erreur de communication qui a été commise par lui-même ou pas l'allocutaire.

Time Management

STALL :

Le locuteur a besoin d'un peu plus de temps pour formuler un énoncé, ou a besoin d'effectuer une pause dans le dialogue.

Extra Discourse

BOILERPLATE :

Le locuteur réutilise des énoncés communs à plusieurs messages : signature automatique d'un courriel, information de contact, *disclaimer etc*. Ces énoncés sont généralement automatiquement ajoutés à un message.

OTHER EXTRA :

Le locuteur choisit de conserver un artefact de formattage pour guider la lecture d'un message par les allocutaires (par exemple, dans un forum : « <cite>Shadok a écrit :</cite> »).

Discourse Management

REPORT SPEECH :

Le locuteur rapporte des énoncés produits par une autre personne, qui n'est pas nécessairement un participant à la conversation.

INTRODUCE TOPIC :

Le locuteur veut introduire le topique décrit dans l'énoncé à la conversation.

REINTRODUCE TOPIC :

Le locuteur veut ré-introduire le topique décrit dans l'énoncé à la conversation.

SHIFT TOPIC :

Le locuteur exprime son désir de changer de sujet de conversation.

CLOSE TOPIC :

Le locuteur exprime son désir de clore le sujet courant de la conversation.

ANNOUNCE :

Le locuteur annonce le contenu des prochains énoncés qu'il a l'intention de produire.

SUMMARIZE :

Le locuteur résume le contenu des énoncés précédents.

CONCLUDE :

Le locuteur conclut une séquence d'énoncés.

Social Obligations Management

GREETING :

Le locuteur salue l'allocutaire.

APOLOGIZING :

Le locuteur exprime des excuses dans l'énoncé.

VALEDICTION :

Le locuteur annonce que l'énoncé sera sa dernière contribution à la conversation.

SELF INTRODUCTION :

Le locuteur se présente aux autres participants de la conversation.

FINAL SELF INTRODUCTION :

Le locuteur se présente à la fin d'une série d'énoncés ; il s'agit généralement d'une signature en fin de message.

THANKING :

Le locuteur exprime sa gratitude envers l'allocutaire pour ses actes ou ses messages.

ANTICIPATE THANKING :

Le locuteur remercie par avance les allocutaires, pour leurs actes ou paroles futures, contrairement à THANKING qui porte sur des actes antérieurs à l'énoncé.

B Liste des traits primitifs

B.1 Composants

B.1.1 Participants

A	*Addressee*. L'allocutaire.
S	*Speaker*. Le locuteur.
T	*Third-party*. Un tiers.

B.1.2 Objets

tu	*Turn*. Un tour de parole.
f	*Feedback*. Un retour sur les énoncés précédents.
d	*Dialogue*. Un dialogue (ou polylogue).
p	*Proposition*. Une proposition.
a	*Action*. Une action.
to	*Topic*. Un sujet de conversation.

B.1.3 Verbes

isInDialogue()	Le participant est prêt à participer au dialogue ***d***, c'est-à-dire peut envoyer et recevoir des messages.
pause()	Le participant souhaite suspendre le dialogue ***d*** un court instant.
conditions()	Le participant conditionne un engagement par une proposition ***p***.
retracts()	Le participant retire un énoncé prononcé pendant son tour de parole.
commits()	Le participant s'engage à accomplir le contenu de la proposition ***p***.
holds()	Le participant souhaite conserver le tour de parole ***tu*** le temps de formuler un nouvel énoncé.

signalsError()	Le participant signifie qu'il a commis une erreur lors d'un énoncé prédédent.
introduces()	Le participant présente un participant ***S*** ou ***T***.
has()	Le participant a le tour de parole ***tu***.
complete()	Le participant complète un énoncé entamé par un de ses allocutaires.
provides()	Le participant fournir une information ***p*** ou un retour ***f***.
options()	Le participant indique une action possible ***a*** sans demander d'engagement de la part de son allocutaire.
knows()	Le participant connaît l'information décrite par la proposition ***p***.
believes()	Le participant croit en la véracité de la proposition ***p***.
weaklyBelieves()	Le participant souhaite qu'une de ses croyances décrite par la proposition ***p*** soit confirmée par un de ses allocutaires.
able()	Le participant est capable d'effectuer l'action ***a***.
wants()	Le participant désire obtenir une information ***p***, l'accomplissement d'une action ***a*** ou la discussion d'un sujet ***to***.
corrects()	Le participant corrige une erreur et la remplace par la proposition ***p***.
performs()	Le participant accomplit un acte performatif ***p***.

B.1.4 Propriétés

isExpansion	L'énoncé s'inscrit dans la continuité d'un ou plusieurs énoncés précédents.
inSet	L'énoncé est porteur de plusieurs alternatives (comme dans une question à choix multiples, par exemple).
inXORSet	L'énoncé et porteur de plusieurs alternatives mutuellement exclusives.
isPositive	La polarité de l'énoncé est positive.
isNegative	La polarité de l'énoncé est négative
isThirdPartyTalk	L'énoncé rapporte les propos d'un tiers.
isSummary	L'énoncé résume plusieurs autres énoncés.

hasOrClause	L'énoncé contient une clause conjonctive « ou ».
hasSubjectInversion	Le sujet et le verbe sont inversés (en anglais, est un marqueur fort d'une forme interrogative).
isAttention	Le *feedback* porte sur l'attention portée à un énoncé.
isEvaluation	Le *feedback* porte sur la compréhension de l'énoncé.
isInterpretation	Le *feedback* porte sur l'interprétation de l'énoncé.
isPerception	Le *feedback* porte sur la perception de l'énoncé.
isExecution	Le *feedback* porte sur l'exécution de l'énoncé.
isEcho	L'énoncé est une répétition ou une reformulation d'un énoncé précédent.
isHarmful	L'action discutée aurait des effets négatifs pour l'allocutaire.
isBeneficial	L'action discutée aurait des effets positifs pour l'allocutaire.
isSelfTalk	L'énoncé n'est pas destiné à un allocutaire, le locuteur « parle tout seul ».
isQuotation	L'énoncé est une citation.
isOpen	L'énoncé est porteur d'une question ouverte.
isDeclarative	L'énoncé est porteur d'une question déclarative (*e.g.* « alors comme ça tu entres à l'université ? »).
inHistory	La proposition a déjà été introduite par un énoncé précédent.
isAllo	L'allocutaire est le sujet de la proposition porté par l'énoncé.
isYesNo	L'énoncé attend une réponse de type « oui » ou « non ».
isApology	Le locuteur présente des excuses au travers de l'énoncé.
isRegret	L'énoncé exprime le regret du locuteur.
isThanks	L'énoncé exprime la gratitude du locuteur.
isAppreciative	L'énoncé exprime l'appréciation du locuteur.
isExclamation	L'énoncé est une exclamation.
expandsOnAnswer	L'énoncé élabore une réponse précédente.
isNonVerbal	L'énoncé n'est pas constitué de mots.
isIndecisive	L'affirmation portée par l'énoncée est indécise.
isPropositional	L'énoncé peut-être réduit à un « oui » ou un « non ».
isOther	L'énoncé ne constitue ni un « oui » ni un « non ».

isTagQuestion	L'énoncé est une question s'achevant par un « tag », *e.g.* « n'est ce pas ? ».
isPartial	L'énoncé est explicitement partiel et doit être complété par un autre énoncé.
isRhetorical	L'énoncé est une question rhétorique.
isOpinion	L'énoncé est présenté comme véhiculants l'opinion personnelle du locuteur plutôt qu'un fait avéré.
isInformation	L'énoncé est strictement informationnelle et ne comporte pas d'élément de discussion d'action.

B.2 Traits primitifs

Les traits primitifs suivant constituent le méta-modèle agrégeant les taxonomies de DiAML, SWBD-DAMSL et MapTask.

— *!A.believes(p)*
— *!A.has(tu)*
— *!A.wants(S.has(tu))*
— *!A.wants(p)*
— *!S.commits(S.performs(a))*
— *!S.conditions(A.wants(S.performs(a)))*
— *!S.weaklyBelieves(p)*
— *!f.isAllo*
— *!f.isEcho*
— *!f.isPositive*
— *!p.isNegative*
— *!p.isPartial*
— *!p.isPositive*
— *!p.isPropositional*
— *!p.isYesNo*
— *A.able(a)*
— *A.believes(!p)*
— *A.believes(S.able(a))*
— *A.believes(a.isBeneficial)*
— *A.commits(A.performs(a))*
— *A.conditions(S.wants(A.performs(a)))*
— *A.has(tu)*
— *A.isInDialogue(d)*
— *A.knows(p)*
— *A.options(A.performs(a))*
— *A.options(S.performs(a))*
— *A.performs(p)*
— *A.provides(f)*
— *A.provides(p)*
— *A.wants(S.has(tu))*
— *A.wants(S.performs(a))*
— *A.wants(f)*
— *A.wants(p)*
— *A.weaklyBelieves(!p)*
— *A.weaklyBelieves(p)*
— *S.able(a)*
— *S.believes(!A.knows(p))*
— *S.believes(A.believes(p))*
— *S.believes(p)*
— *S.commits(S.performs(a))*
— *S.complete(A)*
— *S.conditions(A.wants(A.performs(a)))*
— *S.conditions(A.wants(S.performs(a)))*
— *S.conditions(p)*
— *S.corrects(A)*
— *S.corrects(S)*
— *S.has(tu)*
— *S.holds(p)*
— *S.introduces(S)*
— *S.isInDialogue(d)*
— *S.options(A.performs(a))*
— *S.pause()*
— *S.performs(p)*
— *S.provide(S.isInDialogue(d))*
— *S.provides(S.isInDialogue(d))*
— *S.provides(f)*
— *S.provides(p)*
— *S.retracts(S)*
— *S.signalsError(S)*
— *S.wants(!A.performs(a))*
— *S.wants(!S.has(tu))*
— *S.wants(!S.isInDialogue(d))*
— *S.wants(!S.performs(a))*
— *S.wants(!d)*
— *S.wants(!to)*
— *S.wants(A.believes(p))*
— *S.wants(A.has(tu))*
— *S.wants(A.performs(a))*

— *S.wants(A.provides(A.isInDialogue(d))*
— *S.wants(A.provides(f))*
— *S.wants(A.provides(p == True))*
— *S.wants(A.provides(p))*
— *S.wants(S.has(tu))*
— *S.wants(S.knows(p))*
— *S.wants(S.performs(a))*
— *S.wants(d)*
— *S.wants(!d)*
— *S.wants(f)*
— *S.wants(to)*
— *S.weaklyBelieves(p)*
— *a.isBeneficial*
— *a.isHarmful*
— *f.isAllo*
— *f.isAppreciative*
— *f.isAttention*
— *f.isEcho*
— *f.isEvaluation*
— *f.isExecution*
— *f.isInterpretation*
— *f.isPerception*
— *f.isPositive*
— *f.isSummarize*
— *p.expandsOnAnswer*
— *p.hasOrClause*
— *p.hasSubjectInversion*
— *p.inHistory*
— *p.inSet*
— *p.inXORSet*
— *p.isApology*
— *p.isDeclarative*
— *p.isExclamation*
— *p.isExpansion*
— *p.isIndecisive*
— *p.isInformation*
— *p.isNonVerbal*
— *p.isOpen*
— *p.isOpinion*
— *p.isOther*
— *p.isPartial*
— *p.isPositive*
— *p.isPropositional*
— *p.isQuotation*
— *p.isRegret*
— *p.isRhetorical*
— *p.isSelfTalk*
— *p.isTagQuestion*
— *p.isThanks*
— *p.isThirdPartyTalk*
— *p.isYesNo*

Bibliographie

Madeleine Akrich. Les listes de discussion comme communautés en ligne : outils de description et méthodes d'analyse. Rapport technique, 2012.

Anne H Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. The HCRC map task corpus. *Language and speech*, 34(4) :351–366, 1991.

Jeremy Ang, Yang Liu, et Elizabeth Shriberg. Automatic dialog act segmentation and classification in multiparty meetings. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 1, pages I–1061. IEEE, 2005.

Nicholas Asher et Alex Lascarides. *Logics of Conversation*. Cambridge University Press, 2003.

John Langshaw Austin. *How to Do Things With Words*. Oxford University Press, second edition, 1975.

Naomi S Baron. Computer mediated communication as a force in language change. *Visible language*, 18(2) :118, 1984.

Naomi S Baron. Letters by phone or speech by other means : The linguistics of email. *Language & Communication*, 18(2) :133–170, 1998.

Frederic Bechet, Benjamin Maza, Nicolas Bigouroux, Thierry Bazillon, Marc El-Beze, Renato De Mori, et Eric Arbillot. Decoda : a call-centre human-human spoken conversation corpus. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, et Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, 2012. European Language Resources Association (ELRA).

Howard Bernett. E-Commerce, Customer Service, and the Web-Enabled Call Center, 2000.

Douglas Biber. Spoken and written textual dimensions in english : Resolving the contradictory findings. *Language*, pages 384–414, 1986.

Penelope Brown et Stephen C Levinson. *Politeness : Some Universals in Language Use*. Cambridge University Press, 1983.

Jerome S Bruner. From Communication to Language - A Psychological Perspective. *Cognition*, 3(3) :255–287, 1975.

Harry Bunt. Dynamic interpretation and dialogue theory. *The structure of multimodal dialogue*, 2 :1–8, 1999.

Harry Bunt. Dimensions in Dialogue Act Annotation. In *Proceedings of the 2006 International Conference on Language Resources and Evaluation (LREC 2006)*, pages 919–924, Genoa, Italy, 2006.

Harry Bunt. The DIT++ taxonomy for functional dialogue markup. In *Proceedings of the AAMAS 2009 Workshop "Towards a Standard Markup Language for Embodied Dialogue Acts" (EDAML 2009)*, pages 13–24, Budapest, Hungary, 2009.

Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, et David Traum. Towards an iso standard for dialogue act annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 2010.

Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, et David R Traum. ISO 24617-2 : A semantically-based standard for dialogue annotation. In *Proceedings of the 2012 International Conference on Language Resources and Evaluation (LREC 2012)*, pages 430–437, Istanbul, Turkey, 2012.

Harry Bunt, Alex C Fang, Xiaoyue Liu, Jing Cao, et Volha Petukhova. Issues in the addition of ISO standard annotations to the Switchboard corpus. In *Proceedings of the 9th Joint ISO ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-9)*, pages 59–70, Potsdam, Germany, 2013.

Harry Bunt, Volha Petukhova, Andrei Malchanau, Alex Fang, et Kars Wijnhoven. The DialogBank. In *Proceedings of the 2016 International Conference on Language*

Resources and Evaluation (LREC 2016), pages 3151–3158, Portorož, Slovenia, 2016.

Jean Carletta, Amy Isard, Jacqueline Kowtko, et G Doherty-Sneddon. HCRC dialogue structure coding manual. Rapport technique, 1996.

Vitor R Carvalho et William W Cohen. On the collective classification of email speech acts. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 345–352. ACM, 2005.

Christophe Cerisara, Pavel Kral, et Ladislav Lenc. On the effects of using word2vec representations in neural networks for dialogue act recognition. *Computer Speech & Language*, 2017.

Thierry Chanier, Marie-Noelle Lamy, Christophe Reffay, Marie-Laure Betbeder, et Maud Ciekanski. Letec (learning and teaching corpus) simuligne. 2009.

Thierry Chanier, Céline Poudat, Benoit Sagot, Georges Antoniadis, Ciara R Wigham, Linda Hriba, Julien Longhi, et Djamé Seddah. The comere corpus for french : structuring and annotating heterogeneous cmc genres. *JLCL-Journal for Language Technology and Computational Linguistics*, 29(2) :1–30, 2014.

Shammur Absar Chowdhury, Evgeny A. Stepanov, et Giuseppe Riccardi. Transfer of Corpus-Specific Dialogue Act Annotation to ISO Standard : Is it worth it ? In *Proceedings of the 2016 International Conference on Language Resources and Evaluation (LREC 2016)*, pages 132–135, Portorož, Slovenia, 2016.

William W Cohen, Vitor R Carvalho, et Tom M Mitchell. Learning to Classify Email into “Speech Acts”. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pages 309–316, Barcelona, Spain, 2004.

Mark G Core et James Allen. Coding dialogs with the damsl annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, volume 56, Boston, MA, 1997.

Richard L Daft et Robert H Lengel. Organizational information requirements, media richness and structural design. *Management science*, 32(5) :554–571, 1986.

Crystal David. *Language and the Internet*. 2001.

Hugues de Mazancourt, Soufian Salim, et Gaëlle Recourcé. Un analyseur de conversations pour la relation client. In *Actes de la conférence conjointe JEP-TALN-RECITAL*

2016, volume 5 : Démos, volume 5, pages 3–5, Paris, France, 2016.

Barbara Di Eugenio, Pamela W Jordan, et Liina Pylkkänen. The coconut project : dialogue annotation manual. Rapport technique, University of Pittsburgh, 1998.

Micha Elsner et Eugene Charniak. Disentangling chat. *Computational Linguistics*, 36 (3) :389–409, 2010.

A. Falaise. Corpus de français tchaté getalp_org. In Thierry Chanier, editor, *Banque de corpus CoMeRe*. Ortolang.fr, Nancy, 2014.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, et Chih-Jen Lin. Liblinear : A library for large linear classification. *The Journal of Machine Learning Research*, 9 :1871–1874, 2008.

Alex C Fang, Jing Cao, Harry Bunt, et Xiaoyue Liu. The annotation of the Switchboard corpus with the new ISO standard for dialogue act analysis. In *Proceedings of the 8th joint ISO-ACL Sigsem workshop on interoperable semantic annotation*, page 13, Pisa, Italy, 2012.

Raul Fernandez et Rosalind W Picard. Dialog act classification from prosodic features using support vector machines. In *Speech Prosody 2002, International Conference*, 2002.

David Ferrucci et Adam Lally. Uima : an architectural approach to unstructured information processing in the corporate research environment. *Natural Language Engineering*, 10(3-4) :327–348, 2004.

Oliver Ferschke, Iryna Gurevych, et Yevgen Chebotar. Behind the article : Recognizing dialog acts in wikipedia talk pages. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 777–786. Association for Computational Linguistics, 2012.

Charles J Fillmore. Some Problems for Case Grammar. *Monograph Series on Languages and Linguistics*, 24 :35–56, 1971.

Amel Fraisse et Patrick Paroubek. Toward a unifying model for opinion, sentiment and emotion information extraction. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, et Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3881–3886, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA).

Michael Georgeff, Barney Pell, Martha Pollack, Milind Tambe, et Michael Wooldridge. The belief-desire-intention model of agency. In *Proceedings of the International Workshop on Agent Theories, Architectures, and Languages (LNAI 1555)*, pages 1–10, Berlin, Germany, 1998. Springer.

John J Godfrey, Edward C Holliman, et Jane McDaniel. Switchboard : Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, volume 1, pages 517–520. IEEE, 1992.

Patricia Marks Greenfield et Kaveri Subrahmanyam. Online discourse in a teen chatroom : New codes and new modes of coherence in a visual medium. *Journal of Applied Developmental Psychology*, 24(6) :713–738, 2003.

Derek Gross, James Allen, et David Traum. The TRAINS 91 Dialogues. Rapport technique, University of Rochester, 1993.

Barbara J Grosz et Candace L Sidner. Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3) :175–204, 1986.

Eun Young Ha, Christopher M Mitchell, Kristy Elizabeth Boyer, et James C Lester. Learning Dialogue Management Models for Task-Oriented Dialogue with Parallel Dialogue and Task Streams. In *Proceedings of the 14th SIGdial Workshop on Discourse and Dialogue (SIGdial 2013)*, pages 204–2013, Metz, France, 2013.

Michael A Halliday. *Spoken and written language*. Oxford University Press, 1989.

Nicolas Hernandez et Soufian Salim. Construction d'un large corpus libre de conversations écrites en ligne synchrones et asynchrones en français à partir de ubuntu-fr. In *International Research Days Social Media and CMC Corpora for the eHumanities*, Rennes, France, October 2015.

Nicolas Hernandez, Soufian Salim, et Elizaveta Loginova-Clouet. Ubuntu-fr : a Large and Open Corpus for Supporting Multi-Modality and Online Written Conversation Studies. In *Proceedings of the 2016 International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Slovenia, 2016.

Edward Ivanovic. Dialogue act tagging for instant messaging chat sessions. In *Proceedings of the ACL Student Research Workshop (SRW 2005)*, pages 79–84, Ann Arbor, MI, USA, 2005.

Leon A Jakobovits et Barbara Gordon. *The Context of Foreign Language Teaching*. Newbury House Publishers, 1974.

Dan Jurafsky, Elizabeth Shriberg, et Debra Biasca. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. Rapport technique, 1997.

Nal Kalchbrenner et Phil Blunsom. Recurrent convolutional neural networks for discourse compositionality. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, Sofia, Bulgaria, 2013.

Hamed Khanpour, Nishitha Guntakandla, et Rodney Nielsen. Dialogue act classification in domain-independent conversations using a deep recurrent neural network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics : Technical Papers*, pages 2012–2021, Osaka, Japan, 2016. The COLING 2016 Organizing Committee.

Lawrence Kim, Nam Suand Cavedon et Timothy Baldwin. Classifying dialogue acts in one-on-one live chats. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010)*, pages 862–871, Cambridge, MA, USA, 2010.

Nam Su Kim, Lawrence Cavedon, et Timothy Baldwin. Classifying dialogue acts in multi-party live chats. In *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation (PACLIC 2012)*, pages 463–472, Bali, Indonesia, 2012.

Su Nam Kim, Li Wang, et Timothy Baldwin. Tagging and Linking Web Forum Posts. In *Proceedings of the 14th Conference on Computational Natural Language Learning (CoNLL 2010)*, pages 192–202, Stroudsburg, PA, USA, 2010.

Bryan Klimt et Yiming Yang. The Enron Corpus : A New Dataset for Email Classification Research. In Jean-François Boulicaut, Floriana Esposito, Fosca Giannotti, et Dino Pedreschi, editors, *Machine learning : ECML 2004*, volume 3201, pages 217–226. Springer, 2004.

Kwang-Kyu Ko. Structural characteristics of computer-mediated language : A comparative analysis of interchange discourse. *Electronic Journal of Communication/La revue électronique de communication*, 6(3) :n3, 1996.

Stefan Kopp, Lars Gesellensetter, Nicole C Krämer, et Ipke Wachsmuth. A conversational agent as museum guide–design and evaluation of a real-world application. In *International Workshop on Intelligent Virtual Agents*, pages 329–343. Springer, 2005.

Andrew Lampert, Robert Dale, et Cécile Paris. Classifying Speech Acts Using Verbal Response Modes. In *Proceedings of the 2006 Australasian Language Technology*

Workshop (ALTW 2006), pages 34–41, Sydney, Australia, 2006.

Andrew Lampert, Robert Dale, et Cécile Paris. Segmenting Email Message Text into Zones. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)*, pages 919–928, Singapore, 2009.

Kwok Cheung Lan, Kei Shiu Ho, Pong Luk, Robert Wing, et Hong Va Leong. Dialogue act recognition using maximum entropy. *Journal of the Association for Information Science and Technology*, 59(6) :859–874, 2008.

Ji Young Lee et Franck Dernoncourt. Sequential short-text classification with recurrent and convolutional neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, pages 515–520, San Diego, California, June 2016. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/N16-1062`.

Geoffrey Leech et Martin Weisser. Generic speech act annotation for task-oriented dialogues. In *Proceedings of the 2003 Corpus Linguistics Conference (CL 2003)*, pages 441–446, Lancaster, UK, 2003.

Stephen C Levinson. *Pragmatics*. Cambridge University Press, 1983.

Joseph Lilleberg, Yun Zhu, et Yanqing Zhang. Support vector machines and word2vec for text classification with semantic features. In *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2015 IEEE 14th International Conference on*, pages 136–140. IEEE, 2015.

Ryan Lowe, Nissan Pow, Iulian Serban, et Joelle Pineau. The ubuntu dialogue corpus : A large dataset for research in unstructured multi-turn dialogue systems. In *Proceedings of the Meeting of the Special Interest Group on Dialogue and Discourse (SIGDIAL)*, volume abs/1506.08909, pages 285–294, Prague, Czech Republic, 2015.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, et David McClosky. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics : System Demonstrations*, pages 55–60, Baltimore, MD, USA, 2014.

Andrew Kachites McCallum. Mallet : A machine learning for language toolkit. 2002.

Tomas Mikolov, Kai Chen, Greg Corrado, et Jeffrey Dean. Efficient estimation of word representations in vector space. In *ICLR Workshop Papers*, 2013.

George A Miller. WordNet : a lexical database for English. *Communications of the ACM*, 38(11) :39–41, 1995.

RA Morelli, JD Bronzino, et JW Goethe. A computational speech-act model of human-computer conversations. In *Proceedings of the 17th IEEE Northeast Bioengineering Conference*, pages 263–264, Hartford, CT, USA, 1991.

Michael Nielsen. *Reinventing Discovery : The New Era of Networked Science*. Princeton, N.J. Princeton University Press, 2011.

Richard Ohmann. Speech Acts and the Definition of Literature. *Philosophy & Rhetoric*, 4(1) :1–19, 1971.

Murat Oztok, Daniel Zingaro, Clare Brett, et Jim Hewitt. Exploring asynchronous and synchronous tool use in online courses. *Computers & Education*, 60(1) :87–94, 2013.

Volha Petukhova et Harry Bunt. Introducing communicative function qualifiers. In *Proceedings of the 2nd International Conference on Global Interoperability for Language Resources (ICGL 2010)*, pages 123–133, Hong Kong, China, 2010.

Volha Petukhova, Andrei Malchanau, et Harry Bunt. Interoperability of Dialogue Corpora through ISO24617 24617-2-based Querying. In *Proceedings of the 2014 International Conference on Language Resources and Evaluation (LREC 2014)*, pages 4407–4414, Reykjavik, Iceland, 2014.

Massimo Poesio et David R Traum. Conversational Actions and Discourse Situations. *Computational Intelligence*, 13(3) :309–347, 1997.

Andrei Popescu-Belis. Dialogue Acts : One or More Dimensions ? *ISSCO Working Papers*, (62), 2005.

Ashequl Qadir et Ellen Riloff. Classifying Sentences As Speech Acts in Message Board Posts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, pages 748–758, Edinburgh, UK, 2011.

Sujith Ravi et Jihie Kim. Profiling student interactions in threaded discussions with speech act classifiers. *Frontiers in Artificial Intelligence and Applications*, 158 :357, 2007.

Christian Raymond, Giuseppe Riccardi, KJ Rodrigez, et Joanna Wisniewska. The luna corpus : an annotation scheme for a multi-domain multi-lingual dialogue

corpus. In *Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 11)*, pages 185–186, 2007.

Christophe Reffay, Thierry Chanier, Muriel Noras, et Marie-Laure Betbeder. Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche. *Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (Sticef)*, 15 :xx, 2008.

Mathieu Riou, Nicolas Hernandez, et Soufian Salim. Using discursive information to disentangle french language chat. In *Natural Language processing for Computer-Mediated Communication, Social Media workshop at GSCL Conference*, pages 23–27, Essen, Germany, 2015.

Jeremy Roschelle, Stephanie D Teasley, et al. The construction of shared knowledge in collaborative problem solving. In *Computer-supported collaborative learning*, volume 128, pages 69–197, 1995.

M David Sadek. Dialogue acts are rational plans. In *Proceedings of the ESCA/ETRW Workshop on The structure of multimodal dialogue*" *(VENACO II)*, pages 19–48, Maratea, Italy, 1991.

Jerrold M Sadock. *Toward a Linguistic Theory of Speech Acts*. Academic Press, 1974.

Soufian Salim. État de l'art : analyse des conversations écrites en ligne porteuses de demandes d'assistance en termes d'actes de dialogue. In *Actes des 17e Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues*, pages 60–71, Caen, France, 2015. Association pour le Traitement Automatique des Langues.

Soufian Salim, Nicolas Hernandez, et Emmanuel Morin. Comparaison d'approches de classification automatique des actes de dialogue dans un corpus de conversations écrites en ligne sur différentes modalités. In *Actes de la 23ème Conférence sur le Traitement Automatique des Langues Naturelles*, pages 29–42, 2016.

Soufian Salim, Nicolas Hernandez, et Emmanuel Morin. Dialogue act taxonomy interoperability using a meta-model. In *Proceedings of the 18th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2017)*, 2017.

Emanuel A Schegloff et Harvey Sacks. Opening Up Closings. *Semiotica*, 8(4) : 289–327, 1973.

John R Searle. *Speech Acts : An Essay in the Philosophy of Language*. Cambridge University Press, 1969.

John R Searle. *A Taxonomy of Illocutionary Acts*. Linguistic Agency University of Trier, 1976.

Djamé Seddah, Benoit Sagot, Marie Candito, Virginie Mouilleron, et Vanessa Combet. The French Social Media Bank : a treebank of noisy user generated content. In *Proceedings of COLING 2012*, pages 2441–2458, Mumbai, India, 2012. The COLING 2012 Organizing Committee.

Elizabeth Shriberg, Raj Dhillon, Sonali Bhagat, Jeremy Ang, et Hannah Carvey. The icsi meeting recorder dialog act (mrda) corpus. Rapport technique, DTIC Document, 2004.

Andrew F Simon. Computer-mediated communication : Task performance and satisfaction. *The Journal of Social Psychology*, 146(3) :349–379, 2006.

Brian H Spitzberg. Preliminary development of a model and measure of computer-mediated communication (cmc) competence. *Journal of Computer-Mediated Communication*, 11(2) :629–666, 2006.

Manfred Stede et David Schlangen. Information-Seeking Chat : Dialogues Driven by Topic Structure. In *Proceedings of the 8th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2004)*, pages 117–124, Barcelona, Spain, 2004.

William B Stiles. Verbal response modes and dimensions of interpersonal roles : A method of discourse analysis. *Journal of Personality and Social Psychology*, 36(7) : 693, 1978.

Congkai Sun et Louis-Philippe Morency. Dialogue act recognition using reweighted speaker adaptation. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 118–125. Association for Computational Linguistics, 2012.

William R Swartout, Jonathan Gratch, Randall W Hill Jr, Eduard Hovy, Stacy Marsella, Jeff Rickel, David Traum, et al. Toward virtual humans. *AI Magazine*, 27(2) :96, 2006.

Maryam Tavafi, Yashar Mehdad, Shafiq Joty, Giuseppe Carenini, et Raymond Ng. Dialogue Act Recognition in Synchronous and Asynchronous Conversations. Master's thesis, University of British Columbia, 2013.

Henry S. Thompson et David McKelvie. Hyperlink semantics for standoff markup of read-only documents. In *Proceedings of SGML Europe '97 : the next decade – pushing the envelope*, pages 227–229, 1997.

Sandra A Thompson et William C Mann. Rhetorical Structure Theory : A Framework for the Analysis of Texts. *IPrA Papers in Pragmatics*, 1(1) :79–105, 1987.

David R Traum. 20 Questions on Dialogue Act Taxonomies. *Journal of Semantics*, (1) :7–30, 2000.

David R Traum et Elizabeth A Hinkelman. Conversation acts in task-oriented spoken dialogue. *Computational intelligence*, 8(3) :575–599, 1992.

Douglas P Twitchell, Mark Adkins, Jay F Nunamaker, et Judee K Burgoon. Using Speech Act Theory to Model Conversations for Automated Classification and Retrieval. In *Proceedings of the 9th International Working Conference on the Language-Action Perspective on Communication Modelling (LAP 2004)*, pages 121–129, New Brunswick, NJ, USA, 2004.

J. Ulrich, G. Murray, et G. Carenini. A Publicly Available Annotated Corpus for Supervised Email Summarization. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence EMAIL Workshop (AAAI 2008 EMAIL Workshop)*, pages 77–82, Chicago, IL, USA, 2008.

Ludmila Urbanová. On expressing meaning in english conversation. *Semantic Indeterminacy. Brno : Masaryk University*, 2003.

David C Uthus et David W Aha. Multiparticipant chat analysis : A survey. *Artificial Intelligence*, 199 :106–121, 2013.

Josef Vachek. *Selected writings in English and general linguistics*, volume 92. Walter de Gruyter GmbH & Co KG, 1976.

Daniel Vanderveken. La théorie des actes de discours et l'analyse de la conversation. *Cahiers de Linguistique Française*, pages 9–62, 1992.

H. Yun et T. Chanier. Corpus d'apprentissage favi (français académique virtuel international). In Thierry Chanier, editor, *Banque de corpus CoMeRe*. Ortolang.fr, Nancy, 2014.

UNIVERSITE BRETAGNE LOIRE

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Soufian Antoine SALIM

Analyse discursive et multi-modale des conversations écrites en ligne portées sur la résolution de problèmes

Multi-modal discursive analysis of problem-solving written online conversations

Résumé
Nous nous intéressons aux conversations écrites en ligne orientées vers la résolution de problèmes. Dans la littérature, les interactions entre humains sont typiquement modélisées en termes d'actes de dialogue, qui désignent les types de fonctions remplies par les énoncés dans un discours. Nous cherchons à utiliser ces actes pour analyser les conversations écrites en ligne. Un cadre et des méthodes bien définies permettant une analyse fine de ce type de conversations en termes d'actes de dialogue représenteraient un socle solide sur lequel pourraient reposer différents systèmes liés à l'aide à la résolution des problèmes et à l'analyse des conversations écrites en ligne. De tels systèmes représentent non seulement un enjeu important pour l'industrie, mais permettraient également d'améliorer les plate-formes d'échanges collaboratives qui sont quotidiennement sollicitées par des millions d'utilisateurs. Cependant, les techniques d'identification de la structure des conversations n'ont pas été développées autour des conversations écrites en ligne. Il est nécessaire d'adapter les ressources existantes pour ces conversations. Cet obstacle est à placer dans le cadre de la recherche en communication médiée par les réseaux (CMR), et nous confronte à ses problématiques propres. Notre objectif est de modéliser les conversations écrites en ligne orientées vers la résolution de problèmes en termes d'actes de dialogue, et de proposer des outils pour la reconnaissance automatique de ces actes.

Mots clés
actes de dialogue, conversation, discours, dialogue, communication médiée par les réseaux.

Abstract
We are interested in problem-solving online written conversations. These conversations may be found on online channels such as forums, mailing lists or chat rooms. In the literature, human interactions are usually modelled in terms of dialogue acts. Dialogue acts are typically used to represent the discursive functions of utterances in dialogue. We want to use dialogue acts for the analysis of online written conversations. Well-defined methods and models allowing for the fine-grained analysis of these conversations would represent a solid framework to support different user-assistance and dialogue analysis systems. This would represent an important stake for the customer support industry, but could also be used to improve collaborative assistance platforms that are accessed daily by millions of users. However, current conversations analysis techniques were not developed with written online conversations in mind. It is necessary to adapt existing resources for these conversations. This effort is related to the field of research in computer-mediated conversations (CMC). Our goal is to build a dialogue act model for problem-solving online written conversations, and to offer tools for the automatic recognition of these acts.

Key Words
dialogue acts, conversation, discourse, dialogue, computer-mediated communication.