



HAL
open science

Computational pragmatics : time, space and discourse

Philippe Muller

► **To cite this version:**

Philippe Muller. Computational pragmatics : time, space and discourse. Computation and Language [cs.CL]. UT3 Paul Sabatier, France, 2014. tel-03284109

HAL Id: tel-03284109

<https://ut3-toulouseinp.hal.science/tel-03284109v1>

Submitted on 12 Jul 2021

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

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

**Computational pragmatics : time,
space and discourse**

PHÉNOMÈNES CONTEXTUELS EN TRAITEMENT
AUTOMATIQUE DES LANGUES :
TEMPS, ESPACE ET DISCOURS

Philippe Muller

Toulouse, 2014

Table des matières

Table des matières	4
Introduction	7
1 Représentation de connaissances temporelles et spatiales : sémantique et ontologie	13
1.1 Raisonnement spatial qualitatif	14
1.2 Ontologie spatiale et temporelle et langage naturel	18
2 Extraction d'information temporelle et spatiale	23
2.1 Repérages de dates et événements	25
2.2 Relations temporelles	28
2.3 Relations spatio-temporelles	32
3 Méthodologies d'évaluation pour structures sémantiques	35
3.1 La recherche de la bonne mesure	38
3.2 Niveaux de représentation pour la prédiction	41
4 Du temps au discours	49
4.1 Analyse de discours : projet empirique	52
4.2 Evaluation de l'interprétation discursive	57
4.3 Perspectives	61
5 Du discours au dialogue	65
5.1 Représentation sémantique structurée du dialogue	66
5.2 Gestion du fonds commun conversationnel	67
5.3 Rôle de l'acquiescement dans l'évolution du fonds conversationnel	70
5.4 Acquiescements, intonation et structure topicale	73
5.5 Perspectives	75
6 Lexique et Discours	77
6.1 Lexique et segmentation discursive	78

6.2	Lexique et relations rhétoriques	81
6.3	Similarités lexicales	85
6.4	Perspectives	90
	Conclusion : représentation des connaissances et traitement des langues	93
	Bibliographie	97

Introduction

Ce rapport couvre une douzaine d'années de travaux au confluent de deux domaines de l'Intelligence Artificielle : la représentation de connaissances et le traitement automatique du langage naturel. Ces deux préoccupations se retrouvent dans les problèmes reliés à l'interprétation de messages en langage naturel, ce qui sera évoqué sous le terme de "sémantique" quand cela concerne un message isolé, dans deux types de contextes différents : des textes écrits, caractérisés par un seul producteur qui en assure l'unité, et des dialogues entre deux agents. La prise en compte des aspects contextuels de l'interprétation des messages sera dénommée par le terme de "pragmatique", et je détaillerai ma contribution à l'étude de la membrane où ont lieu les échanges entre ces deux échelles d'analyse, qu'on a coutume d'appeler l'interface sémantique-pragmatique. Deux types d'information vont jouer un rôle important dans cette étude : les informations temporelles et spatiales, qui situent les événements et leur protagonistes décrits dans des textes ou bien impliqués dans des conversations. Ces informations présentent les caractéristiques idéales pour une étude à l'interface sémantique-pragmatique : de nombreux travaux ont étudié les propriétés formelles des représentations que l'on peut en faire, elles posent des problèmes non triviaux de communication qui impliquent les niveaux sémantique et pragmatique et elles sont cruciales pour l'interprétation du langage naturel en général, et quelques-unes de ses applications en particulier, comme on le verra par la suite. Un signe de leur importance est donnée par l'attention dont elles sont l'objet par le groupe de normalisation "Linguistic annotation framework", ISO TC 37/SC4¹.

La sémantique est généralement considérée à l'échelle de la phrase, et il faut donc montrer comment celle-ci contribue à des phénomènes intrinsèquement contextuels, comme l'information temporelle liée aux événements décrits dans un texte. Au passage, on se posera la question du partage du travail entre l'information qui peut être encodée au niveau du lexique, et celle qui est propre au processus d'interprétation.

Prenons en exemple un extrait légèrement adapté (pour les besoins de la présentation) d'une dépêche AFP du corpus French Timebank constitué par André Bittar, de l'Université Paris-7² :

- (1) Date de publication : [Mercredi 19 Mai 1999](#)_{t₁}
Le président russe Boris Eltsine a **annulé**_{e₁} **hier**_{t₂} une **rencontre**_{e₂} avec le président du gouvernement espagnol Jose Maria Aznar en raison d'une **bronchite**_{e₃}, a **affirmé**_{e₄}

1. <http://www.tc37sc4.org/>

2. <https://gforge.inria.fr/projects/fr-timebank/>

l'ambassade d'Espagne à Moscou. Cependant, le Kremlin a **démenti**_{e₅} que le chef de l'Etat soit **malade**_{e₆}. **Deux jours auparavant**_{t₃}, M. Eltsine avait **montré**_{e₇} des signes évidents de faiblesse dans une **allocution**_{e₈} publique.

Les éléments temporels pertinents sont ici des événements verbaux (rouge/gris) : e_1 , ou nominaux : e_2 et des états : e_6 impliquant les protagonistes et des localisations temporelles, en bleu, soulignées, qui peuvent être explicites (t_1) ou bien sous-spécifiées car faisant implicitement référence à la date de publication (t_2 par rapport à t_1) ou bien une localisation déjà introduite (comme t_3 , qui fait référence ici à t_2). La construction de la situation temporelle correspondant à ce texte intègre des éléments répartis sur plusieurs phrases. Elle met en jeu un lexique propre au temps, des propriétés reliées à certains types de noms, des éléments propres aux groupes verbaux, comme le temps verbal, l'aspect ou la modalité. Plus généralement, un certain nombre de présupposés pragmatiques entrent en jeu ici, liés à la cohérence du texte. Représenter la sémantique de cet extrait impliquera ici d'attribuer des dates aux événements réalisés, de déterminer ceux qui ne le sont pas (comme e_2), d'identifier les acteurs, et d'ordonner les événements réalisés les uns par rapport aux autres dans le temps (par exemple, e_1 a eu lieu avant e_5 et après e_7).

Un autre exemple, illustrant le cas de données spatiales, est tiré du corpus du projet Itipy³, spécifiquement le journal de voyage de J.D. Forbes (Daraux, 2000).

- (2) 13 juillet. Argelès (...). J'ai quitté Argelès après le petit-déjeuner et j'ai remonté la vallée en me promenant jusqu'à Luz, quatre ou cinq lieues.
- (3) 18 juillet. Matinée magnifique. J'ai fait l'ascension du Pic de Bergons. J'ai quitté Luz vers 6 heures, et quoique je sois monté très lentement, je suis arrivé exactement au bout de trois heures, et pas en prenant la voie la plus facile.

On retrouve des éléments analogues au cas temporel : un lexique spécifique (des prépositions exprimant des localisations spatiales, comme *jusqu'à Luz*, exemple 2), des lieux qui servent à positionner des événements et des acteurs dans l'espace (*Luz*, exemple 3), comme les dates servent à positionner ceux-ci dans le temps, et des relations spatiales entre lieux, acteurs et événements (*J'ai quitté Luz vers 6 heures*). Cette dimension s'ajoute à la dimension temporelle : la cohérence du tout est donnée par l'organisation à la fois spatiale et temporelle de la situation décrite. On peut noter aussi que même dans le cas d'un récit explicitement tourné vers la description d'un trajet, le discours ne suit pas exactement l'ordre chronologique et spatial de l'itinéraire : dans l'exemple 3, la dernière phrase élabore la description plus générale de la phrase *J'ai fait l'ascension du Pic de Bergons*. On trouve ici une indication du rôle de la structuration textuelle dans l'interprétation de cet aspect sémantique.

Chronologiquement, j'ai commencé ces investigations en abordant la question des représentations formelles, logiques, et de leurs propriétés. Quelles primitives sont adéquates

3. <http://www.labri.fr/perso/marlet/projects/ITIPY/>

pour représenter des informations spatiales ou temporelles communiquées en langage naturel ? Quelles sont leurs propriétés formelles, quels raisonnements peut-on faire sur ces données ?

Si on reprend l'exemple de la dépêche (1), il y a deux questions à résoudre pour la représentation : d'une part, quels sont les référents pour les dates et les événements, c'est-à-dire quelles sont les primitives de représentation adéquates ? Classiquement, le choix peut être fait entre des instants ou des intervalles de temps, ce qui a des conséquences sur l'expressivité et la manipulation des représentations. D'autre part, quelles sont les relations pertinentes entre entités temporelles ? Il est alors nécessaire d'avoir une théorie formelle explicite fondée sur ces choix, qui permet de caractériser les modèles voulus pour ces représentations, et qui permet aussi de définir ce que l'on peut inférer des représentations construites. Par exemple, une relation d'ordre transitive entre événements implique des relations qui ne sont pas nécessairement exprimées directement dans le texte d'origine : e_7 est avant e_1 qui est avant e_4 donc e_7 est avant e_4 . On verra par la suite que cette base formelle est indispensable notamment pour comparer des situations qui peuvent être construites de façons différentes, par exemple par deux humains à qui l'on demande de décrire les relations temporelles qu'ils comprennent à partir de la lecture du texte, ou encore pour comparer une annotation humaine de ce type avec la production d'un système de compréhension automatique. Les formalismes de représentation du temps sont bien étudiés en Intelligence Artificielle, et depuis longtemps, et il ne s'agit pas ici de contribuer à leur étude formelle, mais simplement à définir l'utilisation de tels modèles pour des approches de sémantique computationnelle.

Pour le versant spatial, on retrouve des questions identiques : quelles sont les bonnes primitives de représentation pour les situations spatiales, entre points ou régions de l'espace par exemple, et quelles sont les relations que l'on souhaite exprimer entre les objets, avec la préoccupation de représenter des informations d'origine linguistique. Ce domaine a été exploré en Intelligence artificielle plus récemment que la question du temps, même si des traditions logiques, et bien sûr mathématiques, sont bien plus anciennes. Les vingt dernières années ont donné lieu à de nombreux travaux de représentation de propriétés spatiales "qualitatives", censées correspondre à un niveau plus adéquat de représentation pour le raisonnement de sens commun, notamment à partir de données exprimées en langage naturel, et j'ai contribué à ce domaine en cherchant à le mettre en rapport avec les théories temporelles. L'objectif était alors d'expliquer en même temps sous leur aspect spatial et temporel les situations du type de celles décrites dans le deuxième exemple. Au delà, ces travaux ont un impact plus large dans l'effort de normalisation des connaissances qui est couvert actuellement par le terme d'"ontologie(s)", notamment dans la spécification des couches conceptuelles de haut niveau comme SUMO (Niles et Pease, 2001). Le chapitre 1 détaille plus avant ce premier aspect de mes recherches.

L'arrière-plan linguistique de mes travaux s'est ensuite de plus en plus avancé pour occuper l'essentiel de mes projets, depuis l'étude de la communication d'informations spatio-temporelles entre deux agents dialoguant, jusqu'à l'automatisation de la construction de situations d'un texte décrivant des événements et de leurs rapports dans le temps et l'espace. En partant d'une approche d'extraction d'informations qui recherche et représente les expressions liées au temps, j'ai cherché à construire des modèles qui expliqueraient la cohérence temporelle d'un texte dans des

genres un peu contraints (biographies, textes historiques, dépêches d'agence), cf le chapitre 2. Les questions liées aux inférences permises dans ces situations ont alors posé des problèmes méthodologiques qui sont devenus la matière de plusieurs études sur lesquelles je reviens au chapitre 3. Dans un premier temps, il s'agissait de pouvoir déterminer le référent temporel d'un certain nombre d'expressions et de déterminer les relations que ces référents entretiennent, ce qui nécessite de modéliser le processus d'interprétation, ou du moins de pouvoir approximer ce processus pour rendre compte des données empiriques dont on dispose.

Finalement, cette démarche s'est poursuivie dans l'abstraction en se concentrant sur l'objet discursif lui-même, et la question de la représentation d'un discours comme objet structuré, avec des contraintes liées à sa production par un auteur qui projette des intentions de façon cohérente, au-delà de la cohérence temporelle et spatiale. Schématiquement, l'auteur peut indiquer rhétoriquement l'organisation des situations qu'il décrit, par exemple en structurant son récit en parties, ce qui va induire des sous-événements inclus les uns dans les autres, en marquant explicitement ou implicitement des causalités, ce qui entraîne des précédences de certains événements sur d'autres. Considérons les exemples minimaux de discours suivants :

- (4) Marie **arriva**_{e₉} à la maison. Son mari **faisait**_{e₁₀} la cuisine.
(5) Marie **arriva**_{e₁₁} en retard au cinéma. Elle **attendait**_{e₁₂} son mari à la maison.

La différence entre les exemples (4) et (5) reflète le rôle de cette structure intentionnelle : on ne peut distinguer les deux mini-discours d'un point de vue surfacique, et c'est une interprétation fondée sur ce qui est typiquement plausible dans la situation qui permet de dire que la deuxième phrase de (4) sert d'arrière-plan à la première, tandis que dans l'exemple (5), la deuxième phrase explique la première, ce qui induit des ordres différents pour les événements en question : alors que e_9 a lieu pendant e_{10} , e_{12} a lieu avant e_{11} .

- (6) En 1998_{t₄}, la France **gagna**_{e₁₃} la coupe du monde. Les qualifications pour l'Euro **se passèrent**_{e₁₄} moins bien.
(7) L'année 1998 **fut**_{e₁₅} une année faste. En juin_{t₅}, l'équipe de France **gagna**_{e₁₆} la coupe du monde.

Les exemples 6-7 montrent aussi l'importance de comprendre une organisation du texte au delà de son déroulement linéaire : en 6 les événements décrits sont en séquence, en 7 la première phrase pose le contexte pour la phrase suivante et potentiellement pour une sous-partie du texte. Les ambiguïtés d'interprétation croissent bien sûr exponentiellement avec le nombre de phrases d'un texte.

Le chapitre 4 revient sur le problème de l'interprétation temporelle en contexte discursif et présente les extensions à l'analyse plus générale du discours écrit. La structuration du discours a bien sûr des conséquences bien au-delà des seules dimensions temporelles ou spatiales, pour l'extraction d'informations ou le résumé automatique notamment.

Dans le cas du dialogue, où deux agents interagissent et ont éventuellement des représentations différentes, il s'agit de déterminer la contribution des conventions de langage qui guident les

échanges d'une façon satisfaisante pour les locuteurs, ce qu'on désignera également par le terme de cohérence. A partir de là on peut décider de l'établissement des informations partagées, ce que j'ai étudié dans le cadre de dialogues où cette information est essentiellement spatio-temporelle : l'explication d'un itinéraire par un locuteur à un autre. Considérons un exemple extrait du corpus que nous avons collecté avec Laurent Prévot⁴, où les locuteurs sont notés F (fournisseur de l'explication) et R (récepteur) :

(8) — Dialogue 1.1 —

- | | |
|---|--|
| (F _{7.1}) heum tu vois le rond point de St-Michel ? | <i>(question +
introduction d'un référent)</i> |
| (F _{7.2}) là où il y a le commissariat | <i>(localisation, information)</i> |
| (R _{8.1}) ok ouais | <i>(accusé de réception, réponse)</i> |
| (F _{9.1}) d'accord donc | <i>(accusé de réception)</i> |

L'interaction entre la structure des échanges, les actes de parole effectués par les locuteurs et la tâche poursuivie, ici la construction et l'établissement de situations spatiales dans les connaissances d'un des intervenants, suit des règles d'interprétation qui font intervenir tous les niveaux déjà mentionnés : lexique, sémantique, pragmatique dans la dimension à la fois linguistique (présupposés, anaphores, ...) et sociale, ou socio-linguistique : les conventions portant sur les enchaînements d'actes de parole (indiqués dans le dialogue ci-dessus avec un exemple de *ty-page* pertinent en italiques). Elle présuppose une structuration de l'échange, que l'on a choisi d'articuler autour de *topiques dialogiques*, et de leur cohérence globale.

Parmi ces signes de cohérence, j'ai étudié par exemple certains marqueurs d'actes de dialogue montrant les niveaux d'accord entre locuteurs (acquiescements), cf les tours (R_{8.1}) et (F_{9.1}). On voit alors apparaître des fonctions spécifiques de ces marqueurs en fonction du contexte et des informations spatiales échangées, avec un rôle structurant important. Cela sera exposé chapitre 5.

De façon transversale, on voit réapparaître régulièrement la question du lexique et des informations qui y sont associées, comme unité de départ des représentations sémantiques. Le lexique est vu principalement en sémantique formelle comme un ensemble d'atomes de sens dont les combinaisons, guidées par l'interprétation, forment la représentation. Cette vision est bien sûr très schématique : ces éléments interagissent entre eux, parce que leur usage présente des régularités que le lecteur peut mobiliser pour sa compréhension d'un texte. Certains événements suivent typiquement d'autres événements (cf les exemples (4-7) ci-dessus), et favorisent l'interprétation. Plus généralement, les éléments du lexique sont reliés par associations, et ce tissu sémantique est souvent mis en avant comme étant un élément de la cohérence d'un texte, la "cohésion" lexicale (Halliday et Hasan, 1976). Une partie de mes travaux récents s'est attaché à étudier l'importance de ces liens pour l'interprétation en contexte, et pour expliquer une partie de la cohérence d'un

4. Disponible sur <http://crdo.fr/crdo000742/fr>.

texte. Elle s'est concrétisée notamment par un effort de construction automatisée d'une partie de ces fonctions lexicales, dont la synonymie est un exemple bien représentatif. Le chapitre 6 revient sur ces aspects.

D'un point de vue plus synthétique, on pourrait résumer la démarche sémantique-pragmatique de la façon suivante : une approche complète de l'interprétation d'un objet linguistique intègre des informations représentées "localement" (au niveau d'un mot, d'un segment discursif, d'une phrase, ...), qui suivent des contraintes "globales" d'origines diverses (conventions d'écriture, sociales, propriétés formelles des informations représentées, liens lexicaux déterminés par l'usage), pour guider l'interprétation dans l'espace des représentations possibles. L'interprétation est donc enrichie par des inférences dont il faut un modèle explicite, et qui doit permettre des représentations sous-spécifiées si nécessaire. Les conséquences pour l'évaluation des approches proposées imposent de considérer des procédures qui prennent en compte le rôle de ces inférences. Dans le cas temporel, les équivalences sémantiques entre certaines descriptions impliquent de gérer explicitement les inférences que l'on estime possibles à partir d'une description pour en retrouver une autre. Le cas le plus simple est lié aux propriétés de certaines relations, la transitivité de la précédence temporelle par exemple. La même question peut être posée à une approche rhétorique du discours. Les conséquences sémantiques des relations d'organisation discursive peuvent-elles conduire à des descriptions équivalentes ? La réponse à cette question a des conséquences notamment sur la comparaison de structures, et l'évaluation de procédures automatiques d'analyse de discours, et sont indissociables de la modélisation.

En résumé, nous pouvons caractériser les problèmes étudiés de la façon suivante : ils concernent tous une question posée à un niveau global du discours ou du dialogue, qui nécessite des résolutions d'abord locales avec des interactions contextuelles ou soumises à des contraintes globales. Cela entraîne des problèmes de méthodologie d'évaluation des approches formelles et automatiques, et fait jouer un rôle au raisonnement sur les représentations, qui est sans doute moins apparent si on reste au niveau "local" d'une phrase, comme par exemple dans le cas de l'extraction d'information mentionné plus haut. Ils posent par ailleurs des problèmes techniques difficiles de prédiction de structures complexes, d'autant plus que les données disponibles sont peu nombreuses et coûteuses à rassembler.

Ces travaux mettent donc naturellement en relation des modèles formels issus de l'Intelligence Artificielle et des techniques plus typiquement issus du Traitement Automatique des Langues sur des phénomènes sémantiques.

Représentation de connaissances temporelles et spatiales : sémantique et ontologie

La représentation de connaissances spatiales, éventuellement dynamiques, a comme horizon ici la spécification de constructions appropriées d'un point de vue ontologique, utilisables pour traiter des expressions en langage naturel. Initialement posée dans une perspective de modélisation du sens commun et de ses formes de raisonnement (limité au temps et à l'espace), cette question a maintenant des ramifications dans les modèles d'ontologie de haut niveau¹. Ces modèles tentent de construire des bases consensuelles de représentation pour des bases de données diverses, en fournissant des primitives bien fondées formellement, et qui permettent d'exprimer la diversité des perspectives voulues par des domaines différents, ce qui peut aller (dans le cas de l'espace et du temps) des besoins des bases de données géographiques jusqu'à la communication homme-machine située. Deux sous-problèmes importants sont à considérer :

- permettre la coordination de formalismes différents pour le temps et l'espace, répondant à des besoins distincts, en fournissant des correspondances formelles ;
- définir des processus de passage entre les sources d'informations, notamment le passage d'information qui proviennent de sources linguistiques vers des représentations de type géographique.

L'origine linguistique des informations à traiter est souvent avancée comme motivation à considérer des théories "qualitatives". [Hernández \(1994\)](#) caractérise les connaissances qualitatives comme le savoir pertinent essentiel par rapport à une tâche donnée, ce qui est généralement traduit par des relations entre entités ou entre grandeurs numériques sous-spécifiées. Hernández recense plusieurs raisons pour lesquelles les données numériques sont parfois inadaptées à la résolution d'un problème, une analyse partagée par un certain nombre d'auteurs (cf. notamment [Cohn, 1996](#)), mais la raison la plus décisive pour nous est l'inadéquation des données numériques pour communiquer naturellement avec l'utilisateur d'un système informatique. Elles obligent à faire des traductions internes malaisées quand il s'agit de données imprécises ou incomplètes comme le sont généralement les données spatiales, ce qui se traduit par l'apparence "vague" de la plupart des expressions spatiales en langage naturel.

Mes contributions dans cette perspective ont exploré deux aspects : d'une part l'expression d'une théorie formelle qui permet de combiner à la fois des propriétés spatiales et temporelles, dans

1. Comme GUM ([Bateman et al., 1995](#)), Dolce ([Masolo et al., 2003](#)), BFO ([Grenon et Smith, 2004](#)), ou encore SUMO ([Niles et Pease, 2001](#))

le but d'obtenir une théorie qualitative du mouvement et de pouvoir raisonner sur des situations spatio-temporelles (section 1.1), et d'autre part l'utilisation de cette théorie logique pour caractériser des classes de référence linguistiques par rapport à leur propriétés temporelles, notamment leur extension et leur structure matérielle (section 1.2).

1.1 Raisonnement spatial qualitatif

Les années 80 ont vu un mouvement de modélisation de connaissances comme support à l'IA, autour de concepts liés à la géométrie, la physique, le temps. Le but commun de ces travaux était la modélisation de savoir de "sens commun", qui caractériserait notre approche intuitive du monde matériel. Le raisonnement temporel (Allen, 1984), la physique qualitative ou "naïve" (Hayes, 1985), le raisonnement géométrique (Forbus, 1983) ont été parmi les premières cibles de ce programme. Si le raisonnement temporel s'est diffusé bien au-delà de l'I.A., les autres travaux cités ont eu moins de répercussions. Ils se sont néanmoins poursuivis en se restreignant un peu plus, et en se concentrant sur les clarifications ontologiques nécessaires à l'utilisation de formalismes utilisables dans des bases de connaissances, ou comme base pour des ontologies de haut niveau comme (Lenat, 1995), et plus récemment dans les efforts de normalisation déjà cités comme Dolce ou GUM. Pour ce qui est de l'espace, les concepts ciblés sont la topologie (la notion d'inclusion spatiale et les frontières de région), l'orientation et la distance entre entités spatiales.

Modèle logique de l'espace-temps Dans ce cadre, on peut se poser les questions aux origines de la physique qualitative : les processus physiques sont spatiaux et temporels, et il est essentiel de résoudre cette combinaison entre théories qualitatives pour représenter ces processus, en gardant un caractère qualitatif et en évitant de trop augmenter la complexité de la théorie résultante. J'ai alors fait une étude des combinaisons entre les théories qualitatives topologiques, qui sont plus ou moins admises comme la fondation des théories de l'espace en attendant de compléter cela avec les notions d'orientation et de distance, et les théories du temps.

Les théories logiques temporelles sont fondées sur des instants ou des intervalles de temps, les spatiales sur des points ou des régions de l'espace. La question des primitives d'une théorie spatio-temporelle est alors celle du choix de la combinaison, ou bien de proposer une base ontologique différente. L'idée de (Clarke, 1981), reprise aussi par (Vieu, 1991), est de considérer des primitives sur un domaine qui correspond à l'espace-temps, qui serait à quatre dimensions si on considère le monde physique euclidien. Une théorie spatiale serait appropriée pour décrire sa topologie, mais il faudrait caractériser la nature de l'ordre temporel dans ce cadre, ce qui n'était pas fait par les travaux ci-dessus.

J'ai alors repris un modèle méréo-topologique conçu comme une théorie spatiale, celui de (Asher et Vieu, 1995), pour le combiner avec un ordre temporel qui permet le même genre de distinctions formelles (inclusion et topologie) plus une notion d'ordre sous-jacent. Un choix pratique a été de prendre un prédicat de connexion analogue à la connexion spatio-temporelle mais réduite à la

dimension temporelle. Cela permet ensuite de définir les correspondances entre propriétés purement temporelles et propriétés de l'espace-temps (Muller, 1998). Formellement j'ai pu spécifier les modèles généraux de la théorie de (Asher et Vieu, 1995) en ajoutant un ordre temporel sous-jacent, construit à partir de classes d'équivalences de points (tels qu'ils sont définis dans l'article précédemment cité), classes d'équivalence par rapport à une relation de contemporanéité, pour en faire un modèle proprement spatio-temporel (Muller, 2002). On peut utiliser des représentations graphiques pour schématiser des trajectoires spatio-temporelles de façon un peu compacte. La figure 1.1 en est un exemple, où le temps est un axe géométrique, et l'espace est réduit à une dimension, l'autre axe. La situation représentée ici peut être décrite comme le mouvement de deux entités d'abord disjointes, l'une vers l'autre jusqu'à ce qu'elles se recouvrent. Les situations initiales et finales sont représentées à gauche.

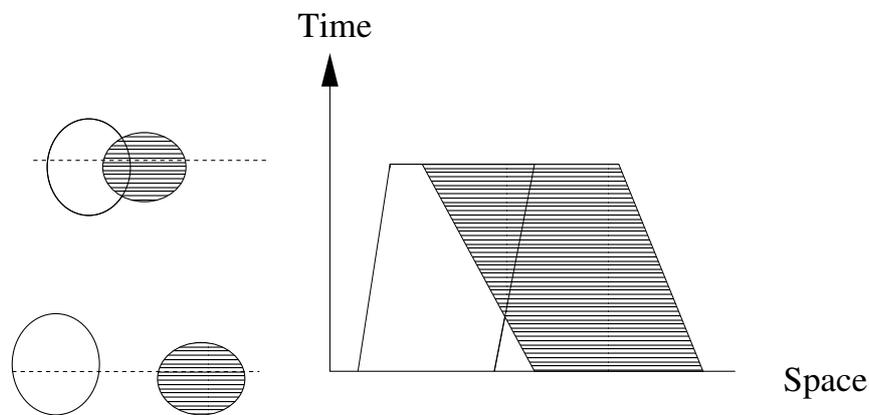


FIGURE 1.1 – Diagramme schématisé d'une trajectoire spatio-temporelle

L'évaluation de ce genre d'approche, au-delà de la preuve de cohérence interne du système, se situe généralement dans la productivité des axiomes : propriétés, théorèmes, et aussi dans les notions que l'on peut définir à l'intérieur du cadre formel. La suite en détaille quelques exemples.

Continuité qualitative L'expressivité d'un tel modèle peut en effet s'apprécier dans les définitions qu'il permet. Dans le but de modéliser qualitativement le raisonnement sur le mouvement, au moins d'un point de vue topologique, certains travaux avaient tenté de préciser une notion de continuité du changement spatial (Galton, 1993; Cui et al., 1992), affaiblie dans le cadre qualitatif, en plongeant une méréotopologie dans une théorie de l'action, et en spécifiant alors toutes les contraintes entre changements de situations spatiales qui respectaient la continuité du mouvement. Par exemple, un changement continu entre 1) une situation où un objet est inclus dans une région de l'espace et 2) une situation où l'objet est complètement déconnecté de cette région implique que l'objet soit passé par des situations intermédiaires où il recouvre partiellement la région puis est en contact externe avec lui avant d'atteindre l'état final. À partir d'une théorie qui porte directement sur des régions de l'espace-temps on a pu montrer qu'un principe général de continuité qualitative suffit à retrouver l'essentiel de ces règles de transition. On peut

par ailleurs affaiblir cette notion de continuité pour définir des types de changement qualitatifs correspondant par exemple à des degrés de transformation d'une entité (Muller, 2002), reprenant notamment les distinctions proposées dans (Galton, 1997) en les formalisant dans un cadre plus simple.

Raisonnement spatio-temporel et algèbres de relations La propriété de continuité mentionnée auparavant est un des cas de raisonnement qu'on peut vouloir faire sur des représentations spatio-temporelles. La théorie axiomatique définie plus haut étant assez expressive, il est vain de pouvoir espérer raisonner directement à partir des axiomes, et de façon analogue à la pratique en raisonnement temporel ou spatial, on s'est efforcé de définir des cadres restreints de raisonnement sur des sous-langages de contraintes, suivant les exemples des algèbres relationnelles, temporelle de Allen (Allen, 1984) ou topologique de (Renz et Nebel, 1997).

On peut définir une algèbre de relations qui définit un cadre de contraintes dès qu'on dispose d'un ensemble de relations mutuellement exclusives et qui décrivent exhaustivement les possibilités logiques. Si on considère l'algèbre de Allen par exemple, qui définit 13 relations pouvant être définies entre deux intervalles de temps, n'importe quelle situation temporelle entre deux intervalles peut être partiellement décrite par une disjonction (un sous-ensemble) de ces 13 relations. Les propriétés logiques de ces relations permettent d'inférer une disjonction entre deux entités I et J, si l'on sait par ailleurs quelque chose sur I et K et K et J, K étant un troisième objet. Cette opération de composition, avec l'union et l'intersection de relations forme une algèbre qui permet de définir un réseau de contraintes. En effet, on peut définir la composition générale de disjonctions : avec $R_1 = \{r_1, r_2, \dots, r_k\}$ et $R_2 = \{s_1, s_2, \dots, s_m\}$, r_i, s_i étant des relations de base

$$R_1 \circ R_2 = \{r_1, r_2, \dots, r_k\} \circ \{s_1, s_2, \dots, s_m\} = \bigcup_{i,j} (r_i \circ s_j)$$

Etant donné un ensemble de relations entre entités (intervalles de temps, régions de l'espace) on peut alors parler de "clôture" d'un ensemble de contraintes quand toutes les inférences de cette sorte sont calculées. Cette procédure permet de tester la cohérence d'un ensemble de contraintes, d'expliciter un certain nombre d'informations avec une complexité faible, et dans le cas de situations dont l'origine est une description en langage naturel, elle permet la comparaison d'annotations d'une même situation. Dans le cas de l'espace, les relations² DC, EC, PO, TPP, NTPP, TPP⁻¹, NTPP⁻¹ et l'égalité, aussi appelées RCC8, forment un ensemble de relations exhaustives et exclusives, définissant là aussi une algèbre de relations, et le problème de la satisfaction d'un ensemble de contraintes est bien étudié, certaines classes tractables sont connues (Renz et Nebel, 1997).

On peut étendre ce cadre à une théorie spatio-temporelle, dans la mesure où RCC8 peut être interprété comme un calcul sur les régions de l'espace-temps. L'équivalent de réseaux de contraintes doit aller plus loin, en permettant de raisonner sur des combinaisons de mouvements

2. Respectivement : non-connexion, connexion externe, recouvrement partiel, partie propre tangentielle ou non et leurs inverses.

et de données temporelles. La continuité est une forme de raisonnement possible, par interpolation. Une autre forme que nous avons définie dans notre cadre correspond au schéma général suivant : étant donné un mouvement d'un objet ou une région u par rapport à un autre v pendant y , et une relation temporelle R_1 entre x et y , que peut-on dire des trajectoires u et v pendant x (cf figure 1.2, qui reprend la façon d'illustrer ces situations introduites plus haut) ?

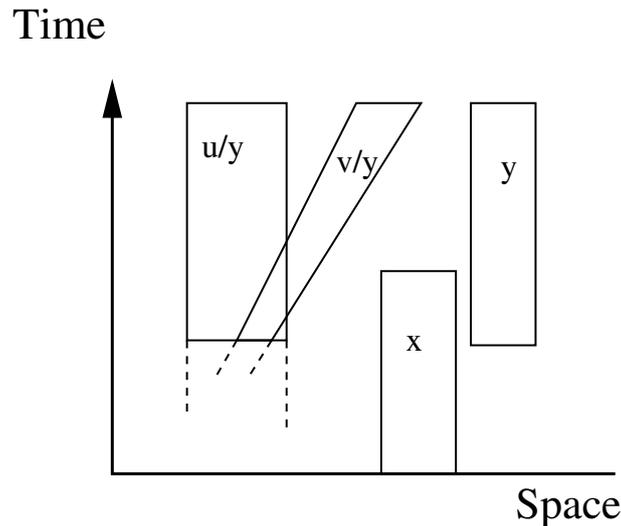


FIGURE 1.2 – Exemple de combinaison de contraintes spatio-temporelles. La notation u/y désigne la partie de l’histoire de u pendant l’événement y .

Par exemple, si u est “en dehors” (relation DC) de v pendant un événement y , et si x est un sous-événement inclus dans y , on doit pouvoir déduire que u est en dehors de v pendant x . Ce cas simple est bien sûr un seul parmi toutes les situations à considérer.

La définition de catégories de mouvements pertinents est alors essentielle, et nous avons idéalisé une étude linguistique sur l’expression du déplacement (cf section 1.2) pour définir des types de mouvement topologiques qui semblaient utiles dans cette perspective. Toutes les combinaisons possibles avec des relations temporelles entre mouvements décrits ont alors été prouvées à l’intérieur de la théorie (Muller, 1998, 2002). Cela permet d’avoir une fondation formelle à un ensemble de raisonnements qualitatifs, l’alternative étant encore une fois de lister les cas possibles sous certaines hypothèses extérieures à la théorie, par exemple comme dans (Gerevini et Nebel, 2002), dans le cas de la cohérence de situations spatio-temporelles (“scénarios”).

Extrapolation spatiale Une autre dimension du raisonnement spatial s’est révélée au contact des travaux de Jérôme Lang, spécialiste de théorie de la décision, et de ses travaux sur la notion de “persistance” temporelle. Celle-ci concerne l’extrapolation de connaissances à travers le temps : sachant que certaines propriétés sont vraies à certains instants, peut-on en déduire leur valeur de vérité entre ces instants, sous la contrainte d’hypothèses raisonnables, par exemple celle de

changement minimal, associé à des probabilités variables (de Saint-Cyr et Lang, 1997). L'extrapolation spatiale peut-être vue comme une extension de ce cadre : étant donné une propriété vraie à certains endroits (par exemple : on trouve de l'eau en sous-sol), peut-on en déduire de façon raisonnée la situation à d'autres endroits, dans le voisinage des points connus ? L'hypothèse de persistance correspond ici à une dépendance spatiale, à savoir que la propriété spatiale étudiée présente des régularités. Le but peut être par exemple de prendre une décision pour savoir où il serait le plus "informatif" de tester la vérité de la propriété en question. Dans notre article (Lang et Muller, 2001), nous proposons de modéliser ce problème en utilisant des fonctions de croyance (Shafer, 1976), qui permettent de gérer la différence entre l'ignorance à un endroit due à un manque d'indicateurs voisins, et la contradiction entre sources d'informations contradictoires. Par ailleurs la gestion de la persistance de propriétés spatiales était associée d'une mesure de la dépendance des sources pour tempérer leur poids quand elles concordent.

Il apparaît nettement depuis ces premiers travaux que le problème considéré est un cas particulier d'apprentissage de classes dans un espace métrique avec une hypothèse de régularité sur les régions de décision. Notre approche revenait alors à un apprentissage par plus proches voisins, avec comme aspects non-classiques une mesure de la dépendance des voisins en fonction de leur éloignement, et une mesure en fonctions de croyance au lieu d'avoir une probabilité. Nous sommes là un peu au-delà de la représentation d'informations spatiales uniquement, et ces travaux sont restés en suspens.

1.2 Ontologie spatiale et temporelle et langage naturel

L'exploration formelle de propriétés spatio-temporelles peut faire oublier un moment l'objectif initial : spécifier des primitives de représentation utilisables pour des données exprimées en langage naturel, et fournir des catégories pertinentes pour ces données. Nous revenons ici sur l'exploitation des formalismes introduits plus haut, d'une part sur l'utilisation sémantique du formalisme et d'autre part sur la façon dont on peut l'exploiter pour caractériser certaines catégories d'entités pertinentes dans une analyse linguistique.

L'utilisation du modèle théorique de la section précédente se décline de deux façons : il fournit d'abord une sémantique à des classes de mouvement qui ont une pertinence linguistique et offre une alternative pour la référence en sémantique formelle, ce que l'on peut apprécier sur la question des références nominales temporalisées.

Classes de mouvements La spécification de modèles formels qualitatifs vise essentiellement l'adéquation pour la représentation d'expressions en langage naturel. Parmi celles-ci, nous nous sommes attachés aux descriptions de mouvement, sous certaines contraintes : sont considérés les événements qui décrivent le déplacement d'une *cible* relativement à un *site*, pour reprendre la terminologie de (Vandeloise, 1986), appelés trajector/landmark ou figure/ground dans la littérature anglophone (Langacker, 1987; Talmy, 1983). Les mouvements sont plus naturellement exprimés par des constructions verbales, mais ce qui suit s'appliquerait aussi à la représentation d'expres-

sions nominales équivalentes. Parmi les catégories de verbes, on peut reprendre les distinctions faites par (Asher et Sablayrolles, 1995) pour distinguer les mouvements qui correspondent à un changement de lieu de la cible du mouvement (*partir de la ville*), un changement de position de la cible (*courir en tous sens*) ou un changement de posture (*se redresser*). Les constructions sont soit transitives, soit intransitives avec un éventuel syntagme prépositionnel qui introduit un site. Ces situations n’impliquent pas nécessairement des sites explicites, et nous avons restreint l’étude dans un premier temps à un cas totalement spécifié qui est celui de constructions transitives de la forme “cible déplacement site”, comme *La voiture a traversé la ville*. Ce travail, mené en collaboration avec Laure Sarda, a consisté à formaliser une partie de la sémantique de ce groupe de verbes, surtout dans leur dimension spatiale topologique, en utilisant le formalisme défini plus haut (Muller et Sarda, 1999, 1998). Pour un ensemble de classes sémantiques attestées dans le lexique, on a pu alors définir une sémantique spatio-temporelle, en laissant la place pour les autres informations de nature géométriques, ou bien les aspects fonctionnels propres à chaque verbe.

Cette catégorisation, plutôt ancienne, trouve alors son application dans les efforts récents de normalisation d’annotations sémantiques sur des données linguistiques, puisqu’elle est actuellement considérée dans la définition de la norme d’annotation consacrée aux informations spatiales, ISO-Space³, cf (Pustejovsky et al., 2010), combinée à celle de (Asher et Sablayrolles, 1995) pour fournir les catégories de mouvement annotés.

Référence temporelle nominale Le modèle ontologique présenté section 1.1 peut être utilisé aussi en sémantique formelle pour prendre en compte les aspects temporels de la référence nominale (Denis et Muller, 2004). Certains problèmes liés à la quantification existentielle des syntagmes nominaux peuvent en effet être considérés comme des problèmes temporels.

L’interprétation temporelle est majoritairement considérée comme un problème verbal en sémantique formelle. Interpréter temporellement une phrase revient à indiquer le prédicat verbal, et positionner éventuellement l’indice (le temps, ou l’état ou l’événement correspondant) par rapport au contexte d’énonciation, ou bien une perspective déjà établie. Les autres prédicats sont donc implicitement atemporels. De nombreux prédicats nominaux dénotent pourtant des propriétés temporaires de leurs arguments sans pour autant dénoter des états, et il se pose alors la question du lien entre le moment où ces propriétés sont vérifiées et les indices verbaux, et savoir si cette information influence l’interprétation de la phrase, comme pour le prédicat *étudiant* dans l’exemple 1.1 ou bien le prédicat *otage* dans l’exemple 1.2 :

(1.1) Tout étudiant en art a visité le Louvre.

(1.2) Le président recevra les otages à la Maison Blanche.

On retrouve cette distinction dans les propriétés individuelles/temporaires (ou “individual”/“stage”) de (Carlson, 1980) et ce problème a été notamment étudié par Musan (1999) et Tonhauser (2002).

3. <http://sites.google.com/site/isospaceorg/home>

Une façon simple de regrouper les prédicats existentiels est de considérer la référence comme étant intrinsèquement matérielle : toute variable désigne un individu dans un espace-temps qu'il occupe de façon limitée, et on peut relier ces individus par des relations temporelles.

Le travail réalisé avec Pascal Denis ([Denis et Muller, 2004](#)) montre comment définir une interprétation compositionnelle qui laisse la place à des indices temporels pour les noms et les verbes, quitte à laisser les relations indéterminées ou dépendantes d'une interprétation contextuelle. La sémantique des quantificateurs universel et existentiel est adaptée pour prendre ce changement en compte. La question de la référence elle-même peut aussi profiter du changement de cadre ontologique pour redéfinir les catégories de référence, comme la distinction stage/individual, ou la distinction entre objet et substance par exemple

Catégories de référence Ces questions de fondements ontologiques pour la référence en langage naturel ont été poursuivies dans ([Gambiarotto et Muller, 2003](#); [Muller, 2007](#)). Essentiellement, il s'agit d'étudier les propriétés temporelles impliquées dans la référence à des objets matériels, à travers les distinctions linguistiques suivantes : un objet par rapport à la substance qui le constitue, un objet singulier par rapport à un collectif regroupant des objets du même type, un prédicat comptable par rapport à un terme de masse, un prédicat substantiel par rapport à un prédicat temporaire, et enfin la distinction entre un objet et un événement qui l'implique, notamment quand il s'agit de les quantifier. Il existe une tradition ontologique qui considère les objets matériels comme des processus occupant une région de l'espace-temps⁴ et dérivent d'événements primitifs. De façon intermédiaire, certains auteurs admettent une dimension temporelle aux objets⁵, et nous avons étudié les conséquences de ces propriétés sur certaines catégories d'objets exprimées en langage naturel. Un problème ontologique classique en sémantique est par exemple la possibilité de différencier un objet et la matière qui le constitue. Suivant ([Link, 1983](#)), pour pouvoir appliquer des propriétés différentes à ces deux objets, on peut introduire une fonction de constitution reliant un objet à sa matière (ainsi une bague peut être neuve et l'or qui la constitue être très vieux). Un problème soulevé par ([Bach, 1986](#)) est alors qu'on peut quand même avoir des niveaux de constitution différents : un bonhomme de neige est fait de neige, faite avec de l'eau, faite d'hydrogène et d'oxygène, etc, tous ayant des propriétés distinctes. Une ontologie spatio-temporelle résout le problème en distinguant les trajectoires des entités, et posant qu'un objet est un épisode spatio-temporel de ce qui le constitue, qui peut également être un épisode de composants temporellement plus étendus, chacun étant distinct mais partageant des épisodes commun avec les autres. La même approche permet de garder une approche structurelle des entités plurielles, de définir les propriétés matérielles d'entités massiques et comptables ([Muller, 2007](#)).

4. ([Whitehead, 1929](#); [Quine, 1960](#); [Russell, 1914](#))

5. ([Noonan, 1976](#); [Heller, 1990](#); [Goodman, 1977](#); [Sider, 2001](#))

Bilan

Les derniers points mentionnés peuvent paraître mineurs dans une perspective de formalisation de connaissances. Au delà des problèmes linguistiques qu'ils soulèvent, les exemples introduits posent pourtant des questions au coeur de la constitution d'ontologies dites "de haut niveau", comme SUMO, DOLCE ou GUM dont nous avons déjà parlé. Si les auteurs de telles ontologies prennent garde de rester agnostiques sur certains choix en permettant par exemple l'existence des différentes alternatives (points/régions, instants/intervalles), il reste qu'il faut pouvoir expliquer les correspondances quand il y en a. Cela peut être fait dans la sémantique des théories, et il faut alors utiliser un formalisme pour représenter les connaissances à un moment donné, si possible celui qui est le plus approprié. C'est-à-dire celui qui permet une bonne couverture de concepts, n'introduit pas de paradoxes, et permet de définir des phénomènes différents. Sans prétendre apporter des réponses définitives à ces questions qui mobilisent des communautés significatives, l'entreprise formelle que constituent les travaux présentés ici contribue à la bonne définition de ces questions, y compris quand elle s'intéresse à son application à des expressions linguistiques, et au mapping langage naturel/ontologies. Pour reprendre un passage de (Masolo et al., 2009) :

In particular, from the viewpoint of applications, the "alliance" between ontologies and lexica can improve the infrastructures of the emerging Semantic Web, supplying lexical coverage to formalized conceptual distinctions.

Le raisonnement spatial qualitatif a d'autres applications, dans le cadre des systèmes d'information géographiques, ou des bases de connaissances manipulant des données visuelles, mais il est toujours considéré en relation avec un niveau "intuitif" d'information spatiale, qui correspond aux besoins de communication entre une base de données et un utilisateur, ou bien un automate et un opérateur humain, cf (Bateman, 2010b,a). Là encore, mes travaux participent à un effort de normalisation important, particulièrement à travers le groupe de réflexion ISO-space, auquel je suis associé.

Le chapitre suivant abordera les aspects plus empiriques reliés au traitement des informations temporelles et spatiales en langage naturel.

Extraction d'information temporelle et spatiale

Le chapitre précédent insistait sur des questions de choix de représentation de l'information spatiale, éventuellement dynamique, avec comme objectif de fonder une approche ontologique formelle, et comme application la sémantique de certaines expressions en langage naturel, décrivant principalement des événements de mouvement. J'ai étendu la portée de ces préoccupations en les considérant dans le contexte plus large d'un texte, où l'interprétation d'une partie est dépendante de l'ensemble du texte et de sa cohérence. C'est ici que les dimensions sémantique et pragmatique se rejoignent, et je détaillerai le cas de l'interprétation d'informations temporalisées et d'informations spatiales, et leur interaction avec la structure d'un texte, dans des études cette fois plus empiriques.

La sémantique du temps possède une longue tradition, qu'il est d'usage de faire remonter à un passage de l'oeuvre du logicien [Reichenbach \(1947\)](#). Les approches de l'interprétation temporelle en linguistique formelle ont connu une activité importante fin des années 1980-début des années 1990¹, mais leurs pendants empiriques étaient plus limités, et peu de systèmes de traitement du langage étaient évalués au-delà de quelques exemples². L'intérêt de la communauté TAL a été cependant ravivé avec l'apparition de données plus volumineuses, notamment le corpus anglais TimeBank à la suite de la production de la norme TimeML ([Pustejovsky et al., 2002](#)), puis l'organisation des campagnes TempEval³. Le développement de systèmes automatiques s'est affirmé, à la fois sur la détection d'informations temporelles et leur mise en relation à partir des travaux de [Mani et Wilson \(2000\)](#). On peut en effet décomposer cet effort au niveau TAL en trois étapes, en oubliant une approche de compréhension exhaustive pour se restreindre à une approche superficielle focalisée sur les aspects temporels :

1. le repérage des expressions pertinentes (événements réalisés, adverbiaux de localisation temporelle),
2. la détermination des attributs nécessaires au calcul de la référence temporelle (dates, durées, ...) des événements et des adverbiaux,
3. la mise en relation des temps et des événements inter-clauses ou inter-phrases. Les correspondants en sémantique formelle pour 2- sont l'interprétation d'adverbiaux temporels

1. Avec notamment ([Webber, 1988](#); [Kamp et Reyle, 1993](#); [Asher et Lascarides, 1993](#))

2. ([Grover et al., 1995](#); [Kameyama et al., 1993](#); [Passonneau, 1988](#); [Song et Cohen, 1991](#))

3. Cf. ([Verhagen et al., 2007](#)), ([Verhagen et al., 2010](#)) et récemment ([UzZaman et al., 2013](#)).

(par exemple [Asher et al., 1995](#)), et 3- l'interprétation temporelle dynamique à la ([Webber, 1988](#); [Kamp et Reyle, 1993](#)).

Reprenons l'exemple de l'introduction :

(2.1) Date de publication : Mercredi 19 Mai 1999_{t₁}

Le président russe Boris Eltsine a **annulé**_{e₁} hier_{t₂} une **rencontre**_{e₂} avec le président du gouvernement espagnol Jose Maria Aznar en raison d'une **bronchite**_{e₃}, a **affirmé**_{e₄} l'ambassade d'Espagne à Moscou. Cependant, le Kremlin a **démenti**_{e₅} que le chef de l'Etat soit **malade**_{e₆}. Deux jours auparavant_{t₃}, M. Eltsine avait **montré**_{e₇} des signes évidents de faiblesse dans une **allocution**_{e₈} publique.

Les expressions mises en évidence sont les éléments à repérer pour la sous-tâche 1. Pour la sous-tâche 2, on veut par exemple attribuer une date explicite à l'expression *hier*, à savoir le mardi 18 mai 1999, déterminer le temps et l'aspect du procès *e₁* et le relier à cette date. A ce stade, une normalisation des informations ciblées est nécessaire, de même qu'un choix de représentation pour les relations reliant les entités temporelles. Les phrases suivantes sont interprétées dans la même perspective, en ajoutant éventuellement des liens entre les procès introduits.

Depuis quelques années, la normalisation des représentations sémantiques que l'on peut désirer extraire de ces expressions a progressé, majoritairement grâce au groupe ISO-TimeML, à l'intérieur du groupe LAF-ISO, cf section 2.1.

J'ai contribué à toutes les étapes de ce problème, en partant des limites des premières approches empiriques, notamment sur les relations entre événements. Le programme que je m'étais fixé était initialement de transposer les hypothèses faites en sémantique formelle sur l'interprétation incrémentale de textes narratifs, notamment issus de ([Kamp et Reyle, 1993](#)) et ses successeurs. Afin d'évaluer ces modèles, la méthodologie suivait les étapes suivantes :

- une collecte de données, avec une normalisation explicite (TimeML dans ses évolutions successives) ;
- l'extraction d'expressions temporelles combinée à un modèle sémantique partiel ;
- des modèles pragmatiques d'interprétation ;
- un cadre évaluatif adapté à ce problème de nature discursive.

Cet effort a débuté par les DEA de Xaver Tannier et Axel Reymonet, avec qui nous avons commencé la collecte de données, ciblée sur des dépêches d'agence et quelques textes dans des genres comportant par nature des informations temporalisées (biographies, textes historiques). La mise en place de la méthodologie complète a correspondu à peu près à la période de stabilisation de la norme TimeML, et elle a permis de constituer un bon extracteur superficiel d'expressions temporelles ([Muller et Tannier, 2004](#)), étendu dans ([Parent et al., 2008](#)), et a permis de commencer à tester certains modèles pragmatiques d'interprétation, avec des résultats assez limités ([Muller et Tannier, 2004](#); [Muller et Reymonet, 2005](#)).

Il est apparu assez vite, et cela s'est confirmé dans les années qui suivirent avec de nombreux travaux d'équipes différentes, que les modèles d'interprétation de la littérature linguistique supposaient soit un genre narratif idéalisé que l'on ne rencontrait jamais tel quel dans les textes étudiés (ou alors de façon très locale et qui ne permettait pas de rendre compte des phénomènes

généraux) soit un modèle plus général de l'organisation textuelle, dont la structure temporelle est une partie spécifique ; nous y avons donc été naturellement mené, cf chapitre 4. La tâche de détermination des relations temporelles entre événements dans un texte n'est possible à l'heure actuelle qu'avec de fortes restrictions⁴, et c'est un problème encore très actif. Je reviens sur mes contributions récentes à ce problème à la section 2.2.

Par ailleurs, ces travaux ont eu un prolongement méthodologique important, essentiellement sur la partie évaluative, et j'y consacre en grande partie le chapitre 3. Le domaine est actuellement mouvant, et la dernière campagne d'évaluation a montré de façon cruciale que le cadre méthodologique avait besoin d'évoluer⁵.

Je reviens maintenant sur chacune des étapes du traitement des informations temporelles dans des textes en langage naturel, les aspects liés à l'extraction d'entités temporelles (section 2.1), la détection de relations temporelles (section 2.2), avant d'aborder les prolongements aux informations spatiales à la section 2.3.

2.1 Repérages de dates et événements

La détection d'expressions temporelles et d'événements a suscité un certain nombre de travaux, pour l'essentiel restreints à l'anglais dans le cadre des campagnes MUC ou ACE, avec récemment l'introduction de la tâche sur le chinois et l'espagnol⁶, et à l'intérieur du groupe de travail préparant la norme TimeML⁷ et les campagnes TempEval. La norme TimeML est le fruit d'un groupe de travail dirigé par James Pustejovsky (Pustejovsky et al., 2005) qui vise à standardiser les annotations sémantiques reliées à la temporalité dans un texte en langage naturel. Le guide d'annotation est tourné vers l'anglais, mais les éléments du standard sont a priori indépendants de la langue considérée, même si quelques ajustements ont été proposés pour certaines langues, comme l'italien (Caselli et al., 2011) ou le français (Bittar, 2010) par exemple. La norme traite de trois sortes d'entités annotables : les adverbiaux, comme les dates (objet TIMEX), les éventualités (EVENT, qui encodent en fait aussi des états), et les signaux (SIGNAL), surtout les prépositions temporelles ; elle introduit aussi plusieurs sortes de relations entre ces entités : liens d'ordonnement (TLINK), de modalité (SLINK), ou bien aspectuels (ALINK), couvrant ainsi l'essentiel des informations temporelles que l'on peut associer à un texte. Les TIMEX ont plusieurs attributs, comme le type (date, heure, durée), la valeur absolue correspondant à la localisation temporelle, le temps d'ancrage, pour les localisations anaphoriques, la fréquence. Les éventualités ont pour attribut une classe parmi : occurrence (la plupart des événements), état, action ou état intensionnel, action d'énonciation ou de perception (dire, révéler, entendre, voir, etc.) et finalement les événements aspectuels (commencer, finir de, continuer, etc.). Les autres attributs sont principalement : le temps grammatical, l'aspect progressif ou non (s'il est déterminable), la

4. (Bramsen et al., 2006; Chambers et al., 2007)

5. Voir notamment le constat fait en conclusion de (Chambers, 2013).

6. <http://www.nist.gov/speech/tests/ace/ace07/>.

7. <http://www.timeml.org>

forme (nom, verbe, adjectif). Les signaux identifient les items, généralement des prépositions, marquant une relation temporelle entre événements et localisations temporelles.

Si en français on peut citer les travaux de (Bernard et Maurel, 1993; Molines, 1989) sur l'interprétation des adverbiaux temporels, il n'existait pas de travaux ayant la couverture de ceux sur l'anglais ou avec des validations empiriques comparables. Le travail que j'ai initié avec Xavier Tannier dans (Muller et Tannier, 2004) avait l'ambition de se placer dans un cadre complet similaire à TimeML, mais la quantité de données concernées était alors assez faible. Il a été cependant repris en collaboration avec Michel Gagnon et Gabriel Parent, restreint au repérage des adverbiaux temporels (et au calcul de leur référence selon la norme TimeML) et aux événements. Nous avons ajouté quelques attributs pertinents pour les tâches en aval de cette reconnaissance (calcul des référents des dates et liens avec les événements), qui permettent notamment d'introduire la notion de perspective temporelle, nous inspirant des travaux de (Molines, 1989). Nous faisons donc la distinction entre une perspective absolue, déictique car relative au temps d'énonciation (TE), ou anaphorique car relative à un point de perspective temporel ou focus temporel (TF), introduit plus tôt dans le texte. Les règles et le lexique ayant été développés sur un petit corpus de développement (une vingtaine de dépêches AFP), nous avons testé leur portée sur un corpus de test, obtenu par mise à jour du corpus de (Baldwin, 2002), qui utilisait une norme précédente (Timex2, issu de la campagne ACE). Le F-score sur les expressions partiellement reconnues était de 91%, 81 sur les expressions exactes. Un grand nombre d'adverbiaux sont des dates assez simples à repérer et dont la valeur est facile à calculer, aussi est-il intéressant de regarder le cas des "dates incomplètes" (*lundi, hier, deux jours avant,...*) qui nécessitent une interprétation contextuelle, soit en les rapportant à la date de parution de la dépêche ou à un point de perspective introduit dans le texte. La proportion de ces adverbiaux anaphoriques correctement reconnue est de 68%, et se décompose en 76% et 52% sur les deux sous-catégories, ce qui montre l'effort restant sur ce premier étage d'extraction.

Depuis, les travaux d'André Bittar (Bittar, 2010) sont arrivés indépendamment à des résultats similaires, à 1-2% près, en suivant une méthode assez proche. On peut mentionner aussi (Teissède et al., 2010) qui développe un système analogue sur une sous-partie des expressions couvertes par TimeML, avec leur propre normalisation.

A l'intérieur des travaux similaires sur l'anglais on peut distinguer deux types d'approches : des approches à base de règles de pattern-matching de complexité variable, écrites "manuellement" (Mani et Wilson, 2000; Mazur et Dale, 2007) ou bien des approches à base d'apprentissage automatique, qui intègrent presque toujours une analyse syntaxique partielle (du simple étiquetage morpho-syntaxique à du chunking plus ou moins élaboré), cf notamment (Boguraev et Ando, 2005). Le calcul des références temporelles est alors fait par des règles lexicales ad hoc, à l'exception de (Baldwin, 2002) qui tente de généraliser aussi l'interprétation à base d'apprentissage, avec des résultats plus mitigés.

La reconnaissance d'adverbiaux temporels obtient généralement des scores comparables avec les deux méthodes, avec des F-scores voisins de 80-85% pour la détection et le calcul de référence, selon la manière de compter (voir plus bas) et les normes (Timex2 pour Ace, Timex3 pour TimeML). En particulier, les travaux de (Negri et Marseglia, 2004), (Mani et Wilson, 2000) et

(Mazur et Dale, 2007) montrent bien qu'on peut espérer de bons résultats avec un système basé sur un ensemble de règles, et c'est la voie que nous avons suivie pour le français. Une approche inductive dans ce cas précis n'a pas grand sens quand on considère les nombreuses variations de formes des adverbiaux temporels, et le lexique mis en jeu, qui nécessiterait d'avoir tous les cas possibles dans le corpus d'entraînement. On peut juger aussi que pour ce problème, un système de règles est plus facile à maintenir et à généraliser à des corpus différents, avec moins de données de départ. Quoi qu'il en soit, à l'heure actuelle les systèmes à base de règles obtiennent de meilleures performances sur cette tâche que les quelques approches à base d'apprentissage automatique (Ahn et al., 2007). Contrairement aux adverbiaux, une analyse syntaxique complète de la phrase est utilisée pour l'élaboration de règles d'annotation des événements. Le lexique Verbaction⁸ fruit du travail d'une équipe de l'INALF, enrichi par la suite à l'ERSS est utilisé comme ressource supplémentaire. Des stratégies différentes ont été établies pour chaque catégorie (verbes, noms et adjectifs) dénotant des événements tout en s'appuyant sur le même principe de règles de réécriture sur l'arbre syntaxique. Nous obtenons 70% de f-score sur le corpus de test (83% sur le développement). Il faut noter que la reconnaissance d'événements dans la norme TimeML a un accord inter-annotateurs peu élevé pour une tâche de ce type, aux alentours de 81% sur l'anglais (Pustejovsky et al., 2006), et qui peut être pris comme une borne maximale de la tâche.

En segmentant les résultats en événements nominaux, verbaux, il apparaît que le problème principal de notre approche est notre faible capacité à extraire efficacement les nominalisations et les (rares) adjectifs dénotant des événements. La tâche d'annotation des événements nominaux semble aussi avoir le moins bon accord inter-annotateur et sa spécification nécessite donc un approfondissement. Le guide d'annotation TimeML pour l'anglais ne fournit pas suffisamment de détails sur l'annotation des noms, mais le travail récent d'André Bittar aborde en détail cet aspect de l'annotation temporelle (Bittar, 2010). Tous les scores ont été d'ailleurs légèrement surpassés par son analyseur, qui réalise 76% de f-score indifféremment sur les corpus de développement et de test.

Par comparaison, sur l'anglais, le système EVITA (Saurí et al., 2005), qui utilise une analyse partielle des phrases, un système de patrons lexicaux-syntaxiques utilisant certaines informations de la base lexicale Wordnet (Fellbaum, 1998), obtient un taux de précision de 74% pour le classement des événements et un rappel de 87%, mais il est à noter que ces résultats ne sont pas obtenus avec un corpus d'évaluation indépendant du corpus d'entraînement. Il est intéressant de comparer ce travail avec le système STEP (Bethard et Martin, 2006), entièrement basé sur de l'apprentissage automatique, utilisant le même corpus d'entraînement et des traits syntaxiques incluant ceux d'EVITA. En termes de précision et de rappel, les résultats obtenus avec l'ensemble total de caractéristiques est légèrement supérieur à ceux obtenus en n'utilisant que le sous-ensemble des traits pris par EVITA. Dans la littérature, on retrouve d'autres approches qui visent à identifier les événements cités dans un texte, mais aucune de celles-ci n'aborde le problème de manière aussi complète que les deux systèmes que nous venons de citer.

8. Cf. <http://w3.univ-tlse2.fr/erss/ressources/verbaction/main.html>

A partir de là nous avons étendu le corpus de référence en intégrant d'autres genres : biographies (une vingtaine pour l'instant), textes historiques, juridiques.

Une partie de ces données a été réutilisée et nettoyée par André Bittar pour la constitution d'un corpus plus complet, mieux équilibré et validé, le FrenchTimeBank (Bittar, 2009).

2.2 Relations temporelles

Processus d'interprétation Mon objectif initial était d'évaluer les principes d'interprétations temporelles que l'on pouvait trouver dans la littérature en sémantique formelle⁹ puis dans les premières approches computationnelles des années 1990¹⁰, qui n'étaient évaluées que sur quelques instances.

La plupart de ces approches reposaient sur des règles d'interprétation incrémentales équivalentes à des automates simples avec des mémoires encodant les points de référence temporel pertinents, suivant le paradigme de Reichenbach. Les contraintes de Reichenbach étaient plus ou moins modifiées pour avoir des règles d'enchaînement des temps moins rigides, ou bien remplacées par des contraintes inspirées par exemple de la théorie du centrage (Kameyama et al., 1993).

Nous avons donc monté une chaîne expérimentale pour tenter de reproduire un processus d'interprétation selon ces lignes et évaluer sa correspondance avec des données écologiques.

La méthodologie suivie dans (Muller et Tannier, 2004) et (Muller et Reymonet, 2005) a été de comparer un petit nombre de stratégies utilisant tout ou partie d'une chaîne incluant les extracteurs de date et d'événements et ajoutant des règles d'interprétation sémantique dérivées des règles des travaux déjà mentionnés.

Ce que l'on cherche à annoter est en fait une approximation du modèle que se fait le lecteur de l'information temporelle contenue dans un texte et cela fait intervenir une part de raisonnement. Comme on l'a déjà souligné, il est difficile d'accorder des annotateurs humains entre eux, car ils peuvent se contenter d'exprimer quelques relations entre événements (par exemple qu'un événement e_1 a lieu pendant un autre e_2 , et que e_2 a lieu avant e_3 , étant implicite qu'alors e_1 a lieu avant e_3). Une possibilité serait de demander explicitement d'établir un lien pour chaque paire d'événements, mais cela conduirait à une tâche très lourde pour l'humain, puisqu'il y a $O(n^2)$ paires pour n événements reconnus dans un texte.

Pour comparer deux annotations, (Setzer, 2001) mais aussi de façon simplifiée (Katz et Arosio, 2001) ont proposé de prendre les fermetures transitives des annotations par rapport à quelques règles d'inférence intuitives (comme celle que nous venons de prendre en exemple). On peut ensuite mesurer, pour chaque relation proposée aux annotateurs, que les événements reliés sont bien les mêmes avec les mesures classiques de précision et de rappel. Les relations proposées dans les travaux de Setzer à l'époque formaient un ensemble *ad hoc*, les travaux équivalents

9. (Kamp et Reyle, 1993; Asher et Lascarides, 1993; Steedman, 1995; Webber, 1988)

10. (Hwang et Schubert, 1992; Singh et Singh, 1995; Grover et al., 1995; Kameyama et al., 1993; Passonneau, 1988; Song et Cohen, 1991)

alors n'étant pas normalisés, ce que l'on peut constater dans l'assemblage hétéroclite de l'atelier ACL sur le sujet (Harper et al., 2001).

S'il est difficile d'estimer dans quelle mesure un système de raisonnement se rapproche des capacités humaines, il existe néanmoins des systèmes formels plus précis pour modéliser l'information temporelle comme l'algèbre de relations d'Allen, cf chapitre 1. La norme TimeML a fini par choisir un ensemble de relations équivalentes à celles de Allen (même si les noms sont différents), à l'exception des relations de recouvrement, cf figure 2.1.

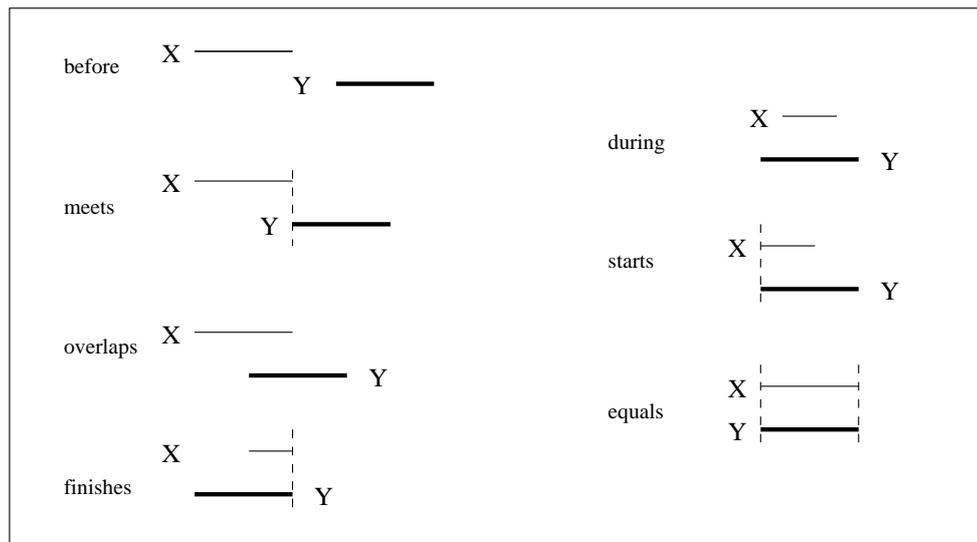


FIGURE 2.1 – Relations de Allen entre intervalles temporels (sans les relations inverses)

Dans (Muller et Tannier, 2004) nous prônions d'appliquer le calcul de relations dans l'algèbre de Allen pour pleinement saturer la représentation, et pas seulement un petit ensemble de règles restreint aux cas simples. Le résultat est alors que toute paire d'événements est reliée par une relation ou une disjonction de relations (éventuellement non informative), et que l'on peut décider ou non d'évaluer tous ces arcs ou seulement les relations "simples" (non disjonctives). Un problème se pose alors quand on utilise le jeu de relations de Allen : celui-ci consiste en 13 relations qui sont sans doute un peu trop précises par rapport aux distinctions que l'on peut vouloir faire d'un point de vue linguistique. Si les annotations doivent être plus sous-spécifiées, un annotateur doit pouvoir choisir de le faire, mais sans avoir à naviguer dans toutes les disjonctions possibles. Dans l'article précité, nous proposons un sous ensemble de relations moins précises pour l'annotation, qui correspondait à des disjonctions de relations de Allen, que nous pensions intuitives, et qui permettaient d'utiliser le calcul de Allen de façon sous-jacente. Nous ignorions alors que ces relations correspondaient en fait à un ensemble proposé par (Bruce, 1972) pour le raisonnement temporel, et plus ou moins oublié depuis. La question du choix de représentation est importante, même si l'évaluation en TAL suppose de fixer des normes partagées pour comparer les travaux similaires. Nous revenons cependant au chapitre 3 sur des moyens expérimentaux de choisir un

formalisme adapté à certains des critères présentés ici : utilisable par des annotateurs, avec une sémantique bien fondée, et pratique pour l'automatisation de la tâche.

Il reste aussi la question de l'estimation de la réussite de la tâche. La mesure la plus simple est de considérer comme référent l'ensemble des relations simples que l'on peut déduire de l'annotation manuelle. Dans (Muller et Tannier, 2004), nous proposons aussi des mesures d'accord des relations disjonctives, essentiellement par recouvrement entre les disjonctions du système ou de la référence, donnant des mesures de précision et rappel relâchées, appelées alors finesse et cohérence. Une mesure similaire a été proposée indépendamment dans le cadre de la campagne TempEval 2007 sur un jeu de relations simplifié.

Je suis plus tard revenu plus précisément sur les critères qui permettent d'évaluer de telles informations, car il est ensuite apparu que ces mesures et celles utilisées par ailleurs dans la littérature avaient un certain nombre de biais gênants. Là encore, je reviens sur ces questions méthodologiques au chapitre 3.

Les expériences de (Muller et Tannier, 2004) avaient donc été faites avec cette méthodologie, sur un ensemble limité d'une vingtaine de dépêches d'agence AFP, contenant de 10 à 40 événements, en considérant la tâche de détection d'adverbiaux et d'événements réalisée pour séparer les problèmes. Nous avons deux stratégies de base pour estimer la difficulté de la tâche : une annotation au hasard de chaque paire successive d'événements dans le texte parmi les relations "à gros grain" ; une deuxième stratégie basique consiste à supposer que les événements sont introduits dans le texte dans l'ordre chronologique : la seule relation prédite est donc "before", qui est par ailleurs la relation la plus courante dans le corpus. Ceci est fait tant que le tout reste cohérent avec les informations données par les dates. Cette stratégie est appelée "natural reading order" (NRO) dans (Bramsen et al., 2006). Ensuite, nous avons testé deux sortes de stratégies de prédiction, l'une qui se contente de respecter les enchaînements de temps, à la (Grover et al., 1995), l'autre qui utilisait des mises à jour des points de perspectives temporelles plus élaborées.

Les résultats étaient généralement très bas pour toutes les méthodes, quelle que soit la mesure. Les meilleurs scores étaient obtenus avec la méthode comportant seulement des contraintes sur les enchaînements des temps, qui était aussi la plus vague.

Il était de toutes façons évident sur les mesures simples (précision et rappel des relations directement annotées) qu'aucune méthode n'était satisfaisante : les erreurs s'enchaînent et comme les événements sont tous reliés entre eux, elles se multiplient.

La conclusion à tirer était que les modèles d'interprétation de la littérature linguistique que nous avons repris supposaient un genre narratif idéalisé que l'on ne rencontrait jamais tel quel dans les textes étudiés. Ceux qui cherchaient à enrichir ce modèle faisaient appel à des notions discursives difficiles à mettre en oeuvre, et pour lesquelles on disposait d'encore moins de ressources.

Apprentissage des relations Au moment de l'étude de (Muller et Tannier, 2004), seuls (Li et al., 2001) et (Mani et Wilson, 2000; Mani et al., 2003) mentionnaient avoir essayé la tâche d'annotation de relations entre entités temporelles d'un texte dans un cadre similaire. Leurs scores étaient bons, autour de 75%, mais étaient soit restreints à des phrases seules contenant plusieurs événements (Li et al., 2001), soit supposaient connues les paires à relier, et incluait

les paires plus simples entre événements et dates, qui peuvent se résoudre pratiquement toujours à l'intérieur d'une seule phrase.

Un certain nombre de travaux sont apparus ensuite sur cette version restreinte de la tâche, d'autant plus que le développement de ressources comme la TimeBank (Pustejovsky et al., 2003) fournissaient des données normalisées, et permettaient d'appliquer des versions par apprentissage automatique, (Mani et al., 2006; Tatu et Srikanth, 2008; Chambers et Jurafsky, 2008a).

Ces méthodes utilisent soit le corpus tel quel pour l'entraînement, soit la version enrichie par inférence (Mani et al., 2006; Chambers et al., 2007), éventuellement la version corrigée de (Berthard et al., 2007). D'autres approches se contentent encore de contextes sélectionnés à l'intérieur d'une phrase (Li et al., 2004; Lapata et Lascarides, 2006).

Tous supposaient que les paires sont sélectionnées et que la tâche n'est plus que de trouver la relation qui tient entre deux événements que l'annotateur humain a choisi de relier temporellement. Certains vont jusqu'à préordonner la paire d'événements pour diviser le nombre de choix possibles par deux. Ces approches ne cherchent d'ailleurs pas à s'assurer de la cohérence des annotations produites, mais simplement à évaluer l'exactitude de la classification isolée.

Plus récemment, (Bramsen et al., 2006) puis (Chambers et Jurafsky, 2008a) ont proposé d'encoder le problème dans un solveur de contraintes en nombres entiers (ILP pour Integer Linear Programming) pour apprendre des décisions locales tout en maintenant la cohérence du réseau de relations temporelles. Ces travaux étaient néanmoins restreints aux seuls cas d'ordonnement temporel avec les relations avant/après. Cette piste intéressante pose des problèmes combinatoires non triviaux si l'on veut passer à l'algèbre de Allen complète, l'ensemble des disjonctions à considérer (et donc l'ordre du nombre de variables à considérer) étant de $2^{|R|}$ où R est l'ensemble des relations de base. Il n'existait donc pas de système prédictif qui ait un score acceptable sur la tâche générale.

Avec Pascal Denis, nous avons exploré plusieurs pistes pour contourner le problème de la taille de l'espace des hypothèses possibles pour construire une structure temporelle face au petit nombre de données disponibles pour l'induction, présentées dans (Denis et Muller, 2011). La première est de réduire la complexité de la description de cette structure, en convertissant l'ensemble des données en relations entre points, les extrémités des intervalles, et en utilisant l'algèbre des points de (Vilain et al., 1990). Cela permet non seulement de réduire la dispersion des classificateurs, en faisant l'hypothèse que les descripteurs linguistiques utilisés puissent aboutir à des généralisations intéressantes sur les liens entre limites des intervalles de temps dénotés par les événements, mais permet aussi d'étendre la stratégie d'optimisation du décodage sous contraintes en ILP. Notre hypothèse de travail, en l'absence de données différentes de TimeBank, est de travailler sur l'étape du décodage de la structure temporelle, c'est-à-dire la prédiction du graphe une fois apprises les régularités de liens entre événements, en testant des stratégies de résolution sous contraintes qui limitent les possibilités de façon plausible ou qui approximent la résolution exacte par des méthodes locales d'optimisation. Dans tous les cas cela suppose de trouver une façon de prédire un graphe au moins approximativement bon et cohérent à la fois. Ce problème est loin d'être trivial et explique l'absence actuelle de solutions au problème général. Sur la restriction consistant à prédire de façon cohérente des relations quelconques sur les paires de la

référence, l'exactitude atteint 50%. Une stratégie d'approximation du problème général a été de diviser le texte d'un point de vue temporel de différentes façons, soit en présupposant des groupes connexes d'événements reliés (donnés par une référence), soit en organisant un découpage autour des adverbiaux temporels qui servent souvent d'ancrage aux événements dans les phrases proches. La traduction en points et le découpage permet alors d'appliquer la résolution sous contraintes à base d'ILP, avec des résultats encourageants. Dans le cas où les sous-groupes idéaux d'événements sont connus à l'avance, la méthode obtient de meilleurs résultats (41% de F-score contre 25%) que la généralisation de la méthode NRO de [Bramsen et al. \(2006\)](#) consistant à ajouter les meilleures prédictions dans l'ordre du texte sous contrainte de cohérence. Dans le cas d'un découpage grossier centré sur les adverbiaux, les méthodes sont comparables. La méthode NRO ne peut guère profiter d'un prédécoupage du texte, il apparaît donc que notre approche ne manque plus que d'une approche plus élaborée de division du travail sur le texte, même si cette dernière étape est sans doute loin d'être triviale. Un certain nombre de caractéristiques des données que nous utilisons sont en effet source d'instabilité dont l'impact sur les expériences n'est pas clair. Les annotations semblent notamment très hétérogènes même sur des textes d'apparence similaire (même genre, même taille).

Nous avons constaté par ailleurs dans nos premières expériences que les liens temporels que l'on peut faire entre événements et adverbiaux temporels situés dans les mêmes phrases généraient un grand nombre de contraintes de précédence et d'inclusion, qui contraignent beaucoup les prédictions que l'on peut faire entre événements de phrases différentes, ce qui est la tâche finale à évaluer. De fait, un modèle général est plongé dans une incertitude qui le pousse à rester proche de ce squelette, et la part d'information temporelle prédite est alors réduite. Ceci pose le problème plus large de la faisabilité de cette tâche dans le cadre admis par les acteurs en TAL, et sans doute aussi interroge la pertinence des données collectées dans le corpus TimeBank. Nous revenons sur le premier point dans le chapitre 4, et sur le deuxième point plus méthodologique dans le chapitre 3. Pour finir sur l'extraction temporelle, nous allons voir comment elle est aussi liée aux problèmes de représentations d'informations spatiales.

2.3 Relations spatio-temporelles

Un prolongement naturel des travaux en extraction temporelle est en effet l'extraction ou l'annotation d'informations spatiales. Les travaux présentés au chapitre 1 forment un préalable au fondement sémantique d'une telle entreprise. Ils indiquent aussi qu'il y a quelque intérêt à considérer ensemble les aspects temporels et spatiaux pour prendre en compte notamment les mouvements des acteurs des textes que l'on veut annoter. Cette partie est encore à l'état d'embryon, mais a été spécifiée de façon préliminaire dans le projet Itipy¹¹, projet ARC INRIA, auquel je suis associé. L'orientation générale du projet, défini dans ([Asher et al., 2008](#)), est de spécifier l'annotation de données spatio-temporelles pour faciliter l'interfaçage de bases de données géographiques et de descriptions textuelles. Beaucoup de textes renferment en effet des localisations spatiales,

11. <http://imagine.enpc.fr/~marletr/LaBRI/projects/ITIPY/>

des descriptions d'itinéraires ou de trajectoires. L'objectif est donc extraire cette information automatiquement et de la mettre en rapport avec des bases de données géographiques (par l'intermédiaire de cartes notamment). Les membres du projet fournissent l'accès à de nombreux documents riches en descriptions d'itinéraires, notamment un ensemble de récits de voyages à travers les Pyrénées du XIXe siècle que la médiathèque de Pau veut valoriser, et dont on a vu des exemples en introduction, page 8.

La modélisation de ces informations nécessite d'impliquer à des degrés divers tous les niveaux de traitement de l'écrit : analyse syntaxique, ressources lexicales, sémantique, analyse discursive, au prix de quelques approximations pour pouvoir se focaliser sur l'information géo-spatiale et temporelle.

Les descriptions que l'on aborde sont du type de l'extrait suivant, tiré d'un guide d'escalade qui fait également partie du corpus considéré :

- (2.2) Laisser la voiture au parking de Sinsat et prendre le sentier du rocher école. Continuer après le secteur "de la dalle", vers le secteur "du lac" qui surplombe l'Ariege. (*Escalade en Haute Ariege*, Thierry et Colette Pouxviel, Publications Sicre, 1993)

Ce texte mobilise une séquence d'instructions permettant de se rendre sur différents secteurs d'un site d'escalade. Les instructions sont organisées pour pouvoir procéder dans le temps et l'espace, grâce à l'emploi de verbes de mouvements, et il n'y a pas de séparation entre le trajet spatial ou temporel. On voit l'usage que l'on peut faire ici des catégories de déplacement encodées par des verbes, et étudiées par (Asher et Sablayrolles, 1995; Muller et Sarda, 1999), et plus généralement utiliser les travaux sémantiques et ontologiques décrits au chapitre 1. Il est bien sûr nécessaire d'aller au-delà de modèles seulement topologiques pour représenter les aspects liés à l'orientation (points cardinaux, orientation frontale, latérale, etc), et aux distances. L'aspect inférentiel est sans doute à prendre en compte également, mais dans un cadre qualitatif qu'il reste à spécifier. Les descriptions d'itinéraires forment d'ailleurs un sujet d'étude en soi (Denis, 1997), et nous avons utilisé ce support dans l'étude du dialogue, chapitre 5.

Cette perspective rejoint par ailleurs l'effort actuel de normalisation ISO des annotations linguistiques spatiales¹², dans une veine similaire à TimeML, ces deux groupes de normalisation ayant comme animateur James Pustejovsky. Des schémas d'annotations précurseurs ont déjà été proposés (Cristani et Cohn, 2002), avec comme préoccupation de référencer des données géographiques. Il faut noter que c'est l'interface de descriptions linguistiques avec des données géographiques qui est visée, puisque ces dernières disposent déjà d'un standard de description, à travers le standard ISO 19136 :2007 qu'est le Geography Markup Language¹³, essentiellement pour des informations numériques.

La tâche associée à ce projet rejoint aussi naturellement d'autres efforts dirigés plus directement vers la modélisation des structures discursives, à travers la restriction à la structure spatio-

12. <http://sites.google.com/site/isospaceorg/home>

13. <http://www.opengeospatial.org/standards/gml>

2.3. RELATIONS SPATIO-TEMPORELLES

temporelle. L'effort de collecte de données est généralisé pour aller vers la modélisation explicite de structures discursives, sur lesquelles je concentre le chapitre 4.

Je reviens en détail dans le chapitre suivant sur les aspects méthodologiques posés par la détection d'informations temporelles, de par la nature structurée des prédictions qui sont faites.

Méthodologies d'évaluation pour structures sémantiques

Le chapitre précédent présentait des travaux à vocation empirique sur des objets discursifs particuliers. Que ce soit dans un but de modélisation ou pour des tâches de traitement automatique, l'étude des prédictions faites par un modèle sur des données empiriques sert de validation de ce modèle. Cette approche est déclinée sur des problèmes essentiellement pragmatiques, à l'échelle d'un discours, et cette perspective sera reprise plus généralement sur un texte écrit ou un dialogue dans les chapitres suivants. Avant cela, je reviens sur les questions d'évaluation qui sont posées par le caractère pragmatique, contextuel, des problèmes abordés, questions qui ne sont pas typiques, du moins sur le versant Traitement Automatique des Langues. En TAL, les tâches sont majoritairement liées à des problèmes de catégorisation, qui vont du plus local au plus contextuel, et du simple au plus ou moins structuré. Le contexte peut être seulement markovien (seule la décision précédant la décision courante est importante, comme dans les premières approches en étiquetage morpho-syntaxique). Il peut introduire des dépendances plus globales à une phrase (en analyse syntaxique) ; au-delà de la phrase les décisions peuvent intégrer des contraintes qui couvrent tout le discours : la résolution d'anaphores, l'analyse de relations temporelles, la segmentation thématique, le repérage de relations rhétoriques. Pour un problème différent d'une décision "locale" (une phrase, ou une décision de catégorisation isolée du contexte textuel) la prise en compte du contexte peut être délicate, mais n'entraîne pas nécessairement de complications au moment de l'évaluation si on peut isoler les résultats des décisions. C'est le cas de l'étiquetage morpho-syntaxique : l'évaluation peut se contenter de considérer la catégorisation réussie ou non de chaque token, de façon isolée.

Pour quelques sous-problèmes plus complexes, tels que l'anaphore, le temps, le discours, il faut considérer un paradigme différent, et prendre en compte les interactions entre les décisions séparées. Dans le cas de la résolution d'anaphore ou plus généralement la recherche de co-références, il s'agit par exemple de trouver des classes d'équivalence, et non plus des entités séparées. Dans ce cas précis, il ne s'agit pas seulement de trouver une liste de relations entre des entités, et de mesurer quelle est la proportion de la référence qui est correctement retrouvée, mais de définir un partitionnement correct de l'ensemble des référents d'un texte.

Ce problème méthodologique était soulevé pendant les campagnes d'extraction d'information Message Understanding Conference (Vilain et al., 1995; van Deemter et Kibble, 2000), et il présente l'essentiel des questions soulevées par les autres tâches que je vais considérer, dans leur version la plus élémentaire, et je le prends comme exemple introductif. Considérons un

texte avec un ensemble de mentions d'entités $E = \{x_1, \dots, x_n\}$. La tâche de détermination de la coréférence est de regrouper ces mentions en classes d'équivalence dénotant chacune un unique référent. Une mesure naïve du succès de la tâche est de considérer les liens de la forme " $x_i = x_j$ " induits par un partitionnement de E par un système, et d'appliquer des mesures de précision et rappel par rapport à un partitionnement de référence. Un effet pervers de cette mesure est alors de favoriser de gros regroupements, indépendamment de leur exactitude. Si on fixe que tous les x_i dénotent la même entité, le rappel sera toujours de 1, et la précision sera nécessairement non nulle si au moins deux mentions coréferent. Un ensemble de mentions coréférentes de taille N induit $N * (N - 1)/2$ relations, ce qui entraîne qu'un système manquant une seule mention de cet ensemble va manquer $N - 1$ relations de la référence. Ceci rend les mesures assez instables. On voit sur ce cas des aspects plus généraux et qui caractérisent aussi les approches que nous avons vues sur le temps et qui se généraliseront au discours : principalement, la mise en oeuvre d'un raisonnement, au sens où des informations implicites sont déduites à partir d'autres informations. Cela revient à dire qu'un ensemble réduit peut être considéré comme équivalent à un ensemble d'informations plus large, et qu'il faut pouvoir comparer des informations syntaxiquement distinctes mais sémantiquement équivalentes. On retrouve donc le problème de l'instabilité des mesures sur un ensemble de propositions interdépendantes.

Dans le cas du temps, une évaluation locale sur des instances indépendantes est donc inadaptée pour la même raison : il existe des variantes d'annotations équivalentes, que l'on peut retrouver avec un modèle d'inférence adéquat. Une autre question que cela soulève est la cohérence d'une prédiction globale par rapport à un modèle où l'on peut faire des inférences : quel est le statut d'une prédiction ou une annotation incohérente¹ ? Le dernier aspect à considérer est l'évaluation d'une tâche sur un corpus global où l'unité pertinente du sujet d'étude est un texte entier, et non pas les informations séparées qu'il contient. Très concrètement, sur une tâche discursive que l'on doit évaluer sur un ensemble de textes (par exemple trouver la structure temporelle), vaut-il mieux faire la moyenne d'un résultat sur le nombre de textes, ou sur le nombre de prédictions faites (par exemple le nombre de relations temporelles prédites) ?

Ce chapitre aborde mes travaux qui ont poussé cette réflexion en étudiant de nouvelles méthodes pour évaluer l'automatisation de tâches pragmatiques en TAL, avec comme objectifs :

- la réduction de différences en explorant les possibilités inférentielles ;
- la diminution des biais qui apparaissent dans les représentations enrichies par inférence ;
- la réduction de l'instabilité des mesures utilisées.

Ces questions ont une portée qui dépasse la simple évaluation de procédures d'annotation humaine ou automatique. La comparaison de structures linguistiques complexes a en effet un rôle à jouer également dans le problème de la prédiction de ces objets structurés. Les problèmes "locaux", abordés comme des problèmes de classification, se limitent à un ensemble fixé de décisions possibles, alors que les structures à prédire pour les tâches plus pragmatiques (des séquences, des arbres, des graphes) ne peuvent être bornées. Les modèles inductifs populaires en TAL doivent donc de plus en plus intégrer des contraintes d'interdépendance entre les décisions,

1. On verra malheureusement que certains corpus de référence contiennent des annotations incohérentes au niveau d'un texte, sans qu'il soit possible d'isoler une annotation fautive particulière.

au moment de l'apprentissage comme au moment de la prédiction. Se repose alors la question des choix de représentations des informations sémantiques que l'on veut annoter et l'on revient au point de départ de nos travaux sur les fondements ontologiques et formels de la représentation du langage naturel. Cette fois cependant nous pouvons nous poser la question avec des critères empiriques un peu différents : un schéma de représentation est adéquat s'il est plus ou moins facile à apprendre, s'il permet de coder une certaine somme d'informations ou d'en retrouver à partir des prédictions.

Considérons à nouveau ici la tâche d'interprétation temporelle d'un texte, où l'on cherche à prédire un graphe de relations entre entités temporelles introduites dans un discours écrit. Les relations ne sont pas indépendantes entre elles, et l'annotation humaine est notoirement difficile, en partie, on l'a vu, parce qu'il y a plusieurs façons différentes d'exprimer la même situation et qu'il est difficile de donner des consignes très directives. Prenons l'exemple suivant :

(3.1) Fortis a **investi** _{e_1} dans des junk bonds avant la **crise** _{e_2} financière, mais en a **revendu** _{e_3} la plupart pendant la **crise** _{e_4} . L'institution a quand même **fait faillite** _{e_5} un an plus tard.

Il est clair que e_1 est avant e_2 , e_3 est inclus dans e_2 , et e_3 est avant e_5 . La coréférence entre e_2 et e_4 implique l'égalité de leur référents temporels. On peut expliciter aussi que e_1 est avant e_3 , mais ce n'est pas nécessaire si une théorie temporelle peut expliquer le lien inférentiel entre cette proposition et les autres, et il est manifeste que les annotateurs humains peuvent le faire, que ce soit en raisonnant directement ou en construisant un modèle de la situation, comme des expériences psychologiques sur le raisonnement temporel semblent le montrer (Schaeken et Johnson-Laird, 2000). On voit là pourquoi on ne peut se contenter de comparer littéralement des descriptions pour faire l'évaluation d'une prédiction ou pour comparer deux annotations.

La pratique courante est donc de considérer une représentation pivot qui explicite un certain niveau d'information temporelle pour faciliter la comparaison. Sur le temps par exemple, les auteurs définissent une sorte de clôture transitive des contraintes temporelles à l'aide d'un formalisme d'enrichissement, comme le calcul des compositions de relations dans l'algèbre de Allen. Outre que cela permet de détecter certaines incohérences, on peut alors comparer les prédictions faites après inférence sur toutes les paires d'entités, y compris si on le souhaite les informations disjonctives créées par composition. Ainsi, en relations de Allen, on peut déduire de l'exemple et de $during(e_3, e_2)$, et $before(e_3, e_5)$, que l'une des relations $\{before, meet, overlap, finish_i, during_i\}$ est vraie entre e_2 et e_5 .

Il y a deux questions posées par cette méthodologie : (1) comment estimer la réussite de la tâche sans biaiser l'évaluation à cause de ces inférences (section 3.1) et (2) comment définir un formalisme de représentation qui permet d'exploiter au mieux les caractéristiques inférentielles (section 3.2).

3.1 La recherche de la bonne mesure

Mesures strictes ou relâchées

Une première chose à considérer dans la contribution apportée par un processus inférentiel qui enrichit une description basique est la présence d'information disjonctive.

On a vu que dans le cas général on peut avoir des prédictions disjonctives pour une relation entre deux événements. Dans des cas où on ne connaît pas la relation simple prévue, on peut réduire l'incertitude en excluant des cas impossibles, ou bien parce qu'on obtient après un calcul une relation de la forme par exemple $\alpha : before(e_1, e_2) \vee meet(e_1, e_2)$.

Une mesure stricte donnerait zéro pour une telle prédiction si ce qui est visé est par exemple $\beta : before(e_1, e_2)$, même si on peut considérer que la disjonction est partiellement correcte. Si l'on suppose que la prédiction est disjonctive, elle est plus vague, mais elle reste "vraie", dans le sens où elle est une conséquence logique de l'annotation de référence. A l'inverse, une prédiction trop précise β quand la référence est α est possiblement fausse, mais est un des modèles possibles de la situation temporelle décrite.

Avec Xavier Tannier, nous avons proposé d'incorporer une mesure plus graduelle de la réussite de la prédiction de structures temporelles, qui serait des équivalents de précision et rappel (Muller et Tannier, 2004). Si $S_{i,j}$ est la relation (disjonctive ou non) entre i et j prédite par un système et que $K_{i,j}$ est la relation (disjonctive ou non) de la référence alors on définit :

$$Prec_{temporelle\ i,j} = \frac{Card(S_{i,j} \cap K_{i,j})}{Card(K_{i,j})} \qquad Rappel_{temporel\ i,j} = \frac{Card(S_{i,j} \cap K_{i,j})}{Card(S_{i,j})}$$

Avec $Card(S)$ = le nombre de relations atomiques présentes dans la disjonction S . La précision totale (ou le rappel) est alors la moyenne sur le nombre de relations données par le système (ou la référence).

Des mesures similaires ont été indépendamment proposées pendant la campagne d'évaluation TempEval 2007 (Verhagen et al., 2007), avec un jeu de relation réduit à "before", "overlaps" et "before ou overlaps" et leurs inverses. Plus généralement, c'est la question de l'évaluation de réponses multiples qui est en jeu.

Estimer l'importance relative des informations

La "fermeture" temporelle d'une annotation est donc nécessaire pour pouvoir comparer deux graphes temporels. Mais comme pour les relations de co-référence, l'ajout de relations par inférence cache le nombre réel de relations nécessaires à décrire une situation de façon suffisante. Le cas temporel est plus complexe car il y a des relations différentes avec des propriétés inférentielles distinctes. Considérons un exemple assez minimal, montré figure 3.1, où le premier graphe est la référence visée, et les S_i sont des prédictions. La seule relation en jeu ici est "avant", notée b , qui est transitive. S_1 contient seulement deux relations, alors que K en contient six. Mais il suffit d'ajouter une relation ($B b C$) pour retrouver la référence complète, par transitivité. Idem pour S_3 . Si l'on pouvait toujours distinguer des relations "majeures" (en trait plein), des relations

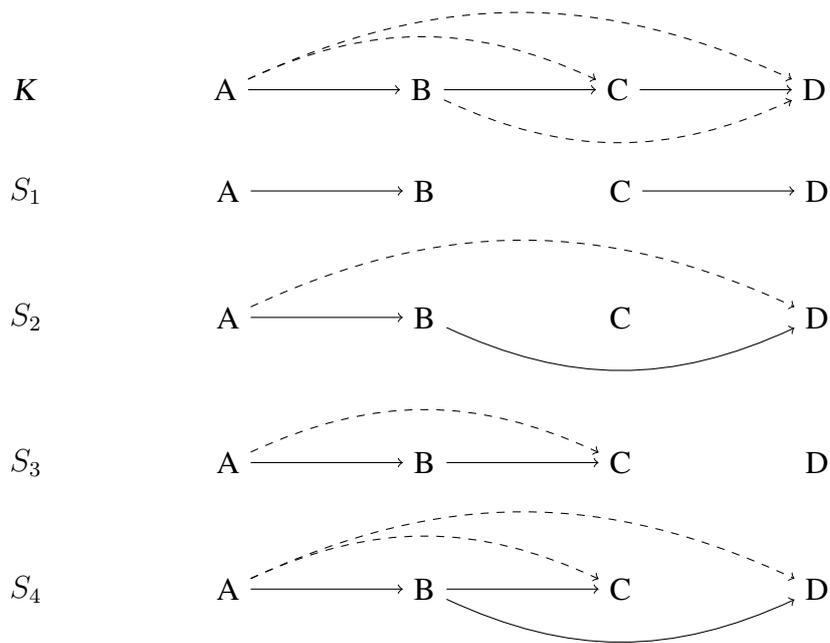


FIGURE 3.1 – Exemple d’annotations à distinguer dans une évaluation. K est la référence, les S_i sont des annotations candidates. S_4 est plus proche de K que S_1 et S_3 , plus proches que S_2 . Les relations annotées sont en trait plein, les relations inférables en pointillé. Toutes les entités sont reliées par la relation “avant” pour l’exemple.

“mineures”, déductibles (en pointillé), on aurait fait un pas. Par ailleurs, le graphe S_2 trouve la relation “ $B b D$ ”, qui est mineure dans la référence, mais qui ne peut être retrouvée autrement dans S_2 , et cette information devrait donc être valorisée. De la même façon, S_4 , à qui il manque une seule relation, comme S_1 et S_3 , trouve quand même plus d’information. Le cas général avec toutes les relations est rendu plus complexe par les compositions, mais l’idée reste la même.

On retrouve là le type de problème posé par la tâche de co-référence mentionnée au début du chapitre. Dans le cadre de MUC-6 pour cette tâche, les évaluations étaient faites sur l’ensemble minimal d’égalités (un arbre couvrant minimal des mentions) qui permettait de retrouver toutes les égalités d’une description (Vilain et al., 1995). Les mesures de précision et rappel étaient donc fonction du nombre minimal de liens à ajouter pour retrouver l’ensemble des coréférences. Si l’on pouvait définir un ensemble de relations minimal unique qui permet de générer l’ensemble des relations avec une procédure inférentielle fixe, on pourrait donc définir une mesure similaire, et ainsi faire le tri entre les relations importantes et celles simplement déductibles. Malheureusement, ce graphe minimal unique n’existe pas dans le cas général, et pas non plus dans le cas des relations de Allen, (Rodríguez et al., 2004), qui propose néanmoins de lister tous les

graphes minimaux et de définir une évaluation sur cet ensemble. Cette solution commence par trouver les relations qui sont communes à tous les graphes minimaux (relations “noyaux”), avant d’énumérer toutes les combinaisons de relations restantes pour trouver cet ensemble de graphes. Cette solution étant impraticable, nous avons proposé dans (Tannier et Muller, 2008) d’évaluer seulement les précision et rappels sur les relations noyaux, faciles à déterminer.

En poursuivant la démarche, une autre stratégie nous a paru plus prometteuse suivant (Dubois et Schwer, 2000), à savoir de considérer une représentation équivalente plus simple du graphe temporel, à travers l’ensemble des points de début et fin des événements, qui peuvent être alors reliés par les relations simples d’égalité, de précédence ou de postériorité. Les points reliés par une relation d’égalité sont ensuite fusionnés. Ce graphe peut lui aussi être saturé en utilisant l’algèbre des points (Vilain et al., 1990).

Nous avons montré que dans le cas d’annotations temporelles le graphe de points est semblable au graphe d’une relation transitive sans cycle. Ce dernier point est crucial, puisqu’alors on peut prendre la réduction transitive de ce graphe comme graphe minimal, qui est unique, cf (Aho et al., 1972). La réduction transitive du graphe est par définition l’ensemble minimal des arcs qu’il suffit de garder pour régénérer le graphe de départ par transitivité. Sa détermination est de plus algorithmiquement simple (La Poutré et van Leeuwen, 1988). On se retrouve alors dans le cas idéal de la figure 3.1.

Comment maintenant juger qu’une procédure d’évaluation répond aux critères que l’on s’est fixés ? Nous avons argumenté dans (Tannier et Muller, 2008), qu’une mesure de l’information produite sur un texte doit évoluer de façon linéaire en fonction de ce qu’elle récupère de l’information d’origine (mesure de rappel), et que l’on peut rapporter cela en fait au nombre d’événements présents dans un texte, de même que des mesures de coréférence doivent être en proportion du nombre de mentions, et non pas du nombre d’égalités possibles entre mentions. C’est pour cela que l’on considère qu’évaluer une structure temporelle sur le nombre total de relations inférables induit un biais, donnant de l’importance à un texte dans une évaluation globale en fonction du carré du nombre d’entités temporelles impliquées.

Une façon d’étudier le comportement d’une mesure de rappel est donc de comparer une annotation de référence à elle-même après qu’on lui a enlevé une partie de ses informations. On peut voir cela comme une illustration des conséquences de l’oubli de relations par rapport à la référence. Idéalement, on voudrait donc une mesure de rappel qui décroît au fur et à mesure que l’on ôte des relations de l’annotation. En pratique, on peut noter que les annotateurs humains produisent à peu près autant de relations qu’il y a d’événements : on peut faire l’hypothèse qu’ils produisent eux-mêmes une sorte de graphe minimal, mais nous n’avons même pas besoin de cette hypothèse dans notre démarche.

Expérimentations

Nous avons donc étudié le comportement des différentes mesures proposées : réduction transitive du graphe de points, noyau de relations, rappel strict ou relâché sur le graphe saturé. La mesure de

rappel qui porte sur le graphe de points réduits a été enrichie d'une fraction de l'information due aux relations mineures dans la référence mais non déductibles dans une annotation particulière. On peut également décider que l'évaluation d'un arc est strict ou relâché au sens vu plus haut, en l'adaptant aux relations sur des points.

L'expérimentation de ces mesures a été effectuée sur deux types de données. D'une part des graphes produits artificiellement pour pouvoir contrôler la taille des graphes et étudier l'influence de celle-ci sur les comportements de mesure. D'autre part, sur des données annotées réelles, issues du corpus TimeBank. Par rapport aux graphes synthétiques, le corpus humain est assez hétérogène, comporte des incohérences et introduit un certain bruit dans les données. Nous voulions aussi contrôler des biais éventuels de TimeBank, comme la préférence pour certaines relations, ou le fait que les textes du corpus soient plus ou moins précis dans leur structure temporelle. Les graphes artificiels ont été générés pour correspondre à des situations temporelles avec un degré d'incertitude sur certaines relations, qui permettent de simuler une structure issue d'un texte. Par ailleurs, on manque d'information sur les performances des annotateurs selon la taille des textes proposées, mais on a pu constater qu'elles se dégradent assez vite au-delà de 30 ou 40 événements à relier, une bonne partie des textes de TimeBank étant plus importants.

Les expériences ont montré que la seule mesure à suivre un comportement linéaire en fonction des informations qu'on en retire est comme prévu la réduction transitive du graphe de points. Les autres montrent des comportements paraboliques : enlever peu d'information ne dégrade pas la mesure (il reste beaucoup de redondances), mais passé un certain point, les scores s'écroulent. On constate des effets indésirables supplémentaires sur les mesures relâchées : celles-ci sont trop laxistes en donnant un score non nul à des annotations quasi vides, c'est-à-dire ne prédisant que des relations vagues. La mesure sur le graphe de relations noyaux suit le même comportement, contrairement à ce que semblaient montrer nos expériences de (Tannier et Muller, 2008), manifestement erronées. Ces biais apparaissent encore plus forts quand le nombre d'événements du texte augmente.

La même expérience sur le corpus TimeBank donne à peu près les mêmes résultats, avec beaucoup plus d'irrégularités, cf la figure 3.2. On observe le même comportement linéaire de la réduction de points, avec une grande instabilité quand il reste peu d'informations.

Cette étude des méthodologies d'évaluation avait pour but d'estimer les biais liés aux mesures courantes de la tâche d'extraction de la structure temporelle, et est présentée dans (Tannier et Muller, 2011). L'autre aspect de la tâche qui n'est considéré dans le domaine qu'avec un minimum de réflexion méthodologique est la représentation des relations temporelles. Je vais maintenant revenir sur cet aspect par une autre étude.

3.2 Niveaux de représentation pour la prédiction

On a vu l'importance du raisonnement temporel pour la détermination de l'ordre temporel des textes dans un objectif de comparaison et d'évaluation, à la fois au chapitre 2 et dans la section précédente. La question de son utilisation dans une approche par apprentissage laisse

3.2. NIVEAUX DE REPRÉSENTATION POUR LA PRÉDICTION

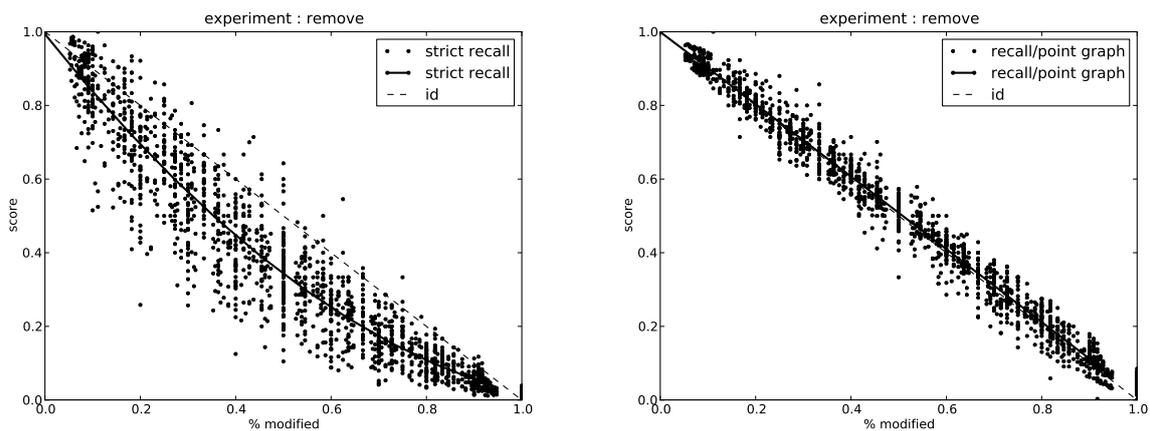


FIGURE 3.2 – Scores de deux mesures de rappel pour une annotation dégradée, en fonction du pourcentage de relations retirées de l’annotation de référence (corpus TimeBank) : précision et rappel strict (évaluation classique) à gauche et réduction transitive du graphe de points à droite. La ligne pointillée représente le comportement idéal ($y=x$), les courbes en trait plein indiquent la régression parabolique des nuages de points.

quelques aspects mal explorés. A la suite de (Mani et al., 2006), beaucoup de travaux utilisent les inférences permises par la théorie temporelle utilisée pour enrichir l’ensemble des instances servant à entraîner les systèmes à base d’apprentissage inductif. Au contraire (Bramsen et al., 2006) utilisent les compositions de relations pour fournir un ensemble de contraintes dans un problème d’inférence global, même si la tâche est légèrement différente puisqu’elle consiste à ordonner des paragraphes dans des historiques de patients en contexte médical. Tatu et Srikanth (2008) et Chambers et Jurafsky (2008a) combinent ces deux pratiques en utilisant le raisonnement pendant l’entraînement et le décodage de nouvelles instances de textes. Mais les deux projets n’utilisent pas le même ensemble de relations : pendant que Mani et al. (2006) reprennent les relations de la norme TimeML, Chambers et Jurafsky (2008a); Bramsen et al. (2006) en prennent un sous-ensemble très réduit : la précédence (avant/après) ou l’absence de relation (lien inconnu). Afin de clarifier le rôle des représentations dans ces questions, j’ai entrepris avec Pascal Denis une étude des différents jeux possibles de relations et leur impact sur la classification de relations temporelles entre événements (Denis et Muller, 2010). Nous avons comparé trois jeux de relations temporelles utilisés à des degrés divers dans la littérature, ainsi que les processus de calcul que l’on peut leur associer. Il s’agit de l’algèbre des 13 relations de Allen, de l’algèbre qui sous-tend les relations utilisées dans la campagne TempEval 2007, et de l’algèbre de Bruce, dont nous avons déjà parlé au chapitre 2, section 2.2.

Les relations de Allen sont pratiques pour mettre en place des inférences sur des contraintes

temporelles, mais elles sont peut-être trop précises pour la sémantique de la plupart des expressions temporelles. C’est pourquoi les campagnes d’évaluation TempEval 2007 et 2010 ont défini des cibles plus vagues. Elles utilisent trois relations de base, “before” (avant), “après”, et “overlap” (recouvrement, qui correspond ici à une intersection non vide entre les intervalles de temps), et les disjonctions $\{\text{before}, \text{overlap}\}$, $\{\text{after}, \text{overlap}\}$, “vague” (disjonction de toutes les relations). Elles correspondent toutes sémantiquement à des disjonctions de relations de Allen. Un ensemble de contraintes avec ces relations peut donc être exprimé en relations de Allen, sur lesquelles on peut appliquer les procédures bien établies de saturation de contraintes avant de les retraduire en relations TempEval. Celles-ci forment une sous-algèbre de l’algèbre de Allen.

Les relations de Bruce (1972) fournissent un niveau intermédiaire de représentation à 7 relations simples, également équivalentes à des disjonctions de relations de Allen.

Les équivalences sont résumées table 3.1².

TimeML	Allen	Bruce	Tempeval
BEFORE IBEFORE	before meet	before	before
(absente)	overlaps	overlaps	overlaps
STARTS IS_INCLUDED FINISHES	starts during finishes	included	
(absente)	overlaps _i	is-overlapped	
IS_STARTED INCLUDES IS_FINISHED	starts _i during _i finishes _i	includes	
IAFTER AFTER	meet _i before _i	after	after
SIMULTANEOUS	equals	equals	equals

TABLE 3.1 – Correspondances entre les algèbres temporelles. Une relation qui s’étend sur plusieurs cases est équivalente à la disjonction des relations contenues dans cette case. Le nom porté dans la norme TimeML est aussi indiqué.

Rappelons que l’objectif était de déterminer le meilleur compromis entre les critères que nous avons pu rencontrer pour la tâche en question :

- la facilité d’apprendre un jeu de relations ;
- la quantité d’information qu’il peut apporter ;
- quelle quantité d’information peut être retrouvée à partir des prédictions faites dans l’un ou l’autre de ces ensembles de relations.

Du point de vue de l’apprentissage vu comme une classification de liens entre événements, il est *a priori* plus facile d’apprendre sur un modèle qui a moins de classes différentes, à condition

2. Quand ce n’est pas évident par le contexte nous utiliserons des indices pour indiquer l’algèbre utilisée, par exemple b_t désigne la relation “before” dans le jeu TempEval.

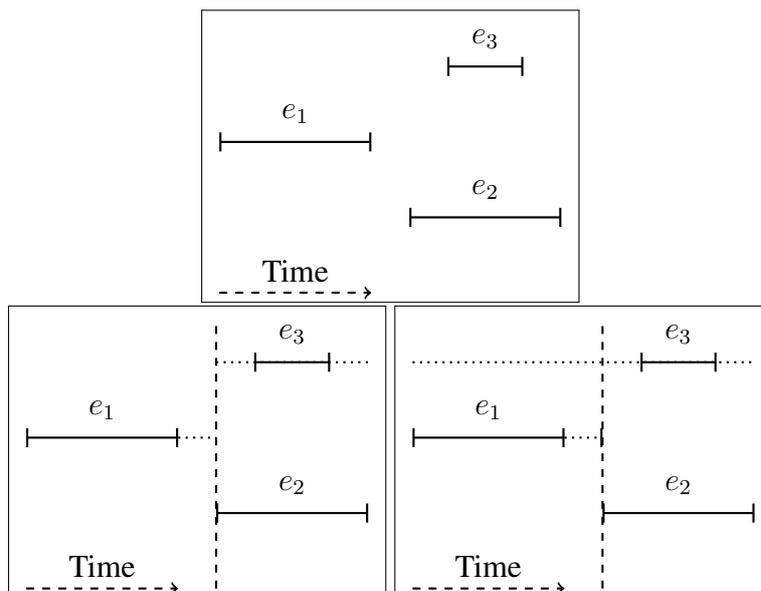


FIGURE 3.3 – Comparaison du pouvoir inférentiel dans les trois algèbres. En continu, le modèle le plus précis ; en pointillé, l’incertitude liée aux autres schémas, et les inférences qui peuvent encore être faites. Allen : $(e_1 \mathbf{b}_a e_2 \wedge e_3 \mathbf{d}_a e_2) \rightarrow e_1 \mathbf{b}_a e_3$ Bruce : $(e_1 \mathbf{b}_b e_2 \wedge e_3 \mathbf{d}_b e_2) \rightarrow e_1 \mathbf{b}_b e_3$ Tempeval : $(e_1 \mathbf{b}_t e_2 \wedge e_3 \mathbf{o}_t e_2) \rightarrow e_1 \{\mathbf{b}_t, \mathbf{o}_t\} e_3$

que ces classes soient représentatives de configurations généralisables. D’un point de vue de représentation, il est par contre plus informatif d’apprendre des relations les plus spécifiques possibles, car leur combinaison permet de retrouver plus d’informations implicites.

Un danger potentiel est cependant de prédire des relations trop spécifiques qui auront comme conséquences d’être incohérentes entre elles, alors qu’il est plus simple de rester cohérent en étant plus vague. A l’inverse, utiliser des relations trop vagues ne permet pas toujours d’encoder des informations implicitement. Reprenons l’exemple (3.1) :

(3.1) Fortis a **investi** _{e_1} dans des junk bonds avant la **crise** _{e_2} financière, mais en a **revendu** _{e_3} la plupart pendant la **crise** _{e_4} . L’institution a quand même **fait faillite** _{e_5} un an plus tard. .

On rappelle que e_1 est avant e_2 , e_3 est inclus dans e_2 , et e_3 est avant e_5 . La figure 3.3 montre les représentations équivalentes dans les trois jeux de relations introduits. Les lignes pointillées indiquent le flou introduit dans la référence aux intervalles par rapport au formalisme le plus précis (celui de Allen). On voit aussi ce que l’on peut déduire de l’annotation seule de “ e_1 est avant e_2 ”, et “ e_3 est inclus dans e_2 ”. Les formalismes de Bruce et d’Allen préservent au moins les propriétés inférentielles d’ordre et d’inclusion, puisque l’on peut inférer que “ e_1 est avant e_3 ”, ce qui n’est pas le cas en TempEval, qui ne permet des inférences que sur l’ordre temporel.

Pour évaluer l’impact de ces caractéristiques des algèbres sur la tâche effective d’ordonnance-

ment temporel, nous avons alors conçu un ensemble d'expériences sur le corpus Timebank, sur lequel nous apprenons à étiqueter des relations plus ou moins précises, et où nous évaluons les différences de résultats quand on applique les processus d'inférence adéquats.

Nous avons donc des classifieurs (comparables à la littérature sur cette tâche, que ce soit au niveau des méthodes ou des traits utilisés par les classifieurs) pour chaque jeu de relations, avec une base d'entraînement enrichie par inférence ou non. Nous avons testé deux types de décodage : un décodage simple, classique, qui étiquette chaque paire d'événements de la référence séparément, et le décodage NRO de (Bramsen et al., 2006) déjà mentionné au chapitre 2, qui vérifie à chaque étape la cohérence globale du réseau de contraintes temporelles induit avant d'ajouter une relation.

Les résultats ont à peu près confirmé l'hypothèse de base : le classifieur seul qui a la meilleure exactitude est celui avec le moins de classes (TempEval), mais il est dépassé par les deux autres algèbres quand on évalue sur le graphe temporel complet, que ce soit en mesure stricte ou en mesure relâchée.

Quand on rapporte les prédictions dans le formalisme TempEval, le meilleur système est celui basé sur les relations de Bruce, ce qui montre que l'inférence plus expressive permet de dépasser un schéma qui a pourtant moins de classes à différencier (TempEval), et que le niveau plus vague permet de dépasser quand même, légèrement, les classifieurs utilisant les relations plus exactes de Allen³. Les évaluations rapportées dans le formalisme de Allen sont favorables au schéma de Allen, ce qui est nécessaire dans une évaluation stricte, mais pas dans le cas relâché. Le schéma très précis génère par ailleurs de 7 à 9% de graphes incohérents, considérés comme une absence de prédiction dans l'évaluation. L'heuristique NRO rétablit la cohérence et fait gagner environ 10% de score de façon uniforme ou à peu près. Le meilleur compromis entre facilité d'apprentissage et pouvoir expressif semble donc être le formalisme de Bruce.

En conclusion, on peut noter que l'on propose une méthodologie pour aller au-delà du corpus TimeBank, dont les particularités ont été soulignées. Si les données annotées dans le cadre de la campagne TempEval sont produites dans un cadre que nous trouvons discutables, les données produites dans le cadre de TimeML ne sont pas affectées, puisqu'il suffit de les passer dans le formalisme moins précis des relations de Bruce. Il reste à voir si cela rend plus robustes les annotations, par exemple en générant moins d'incohérences, pas seulement parce que le formalisme perd de l'information, mais aussi hypothétiquement parce qu'il ne force pas l'annotateur à prendre des décisions trop précises sans nécessité. L'analyse détaillée de quelques exemples d'incohérences dans TimeBank nous a montré que les annotateurs ne visualisent pas toutes les conséquences de certaines relations précises (comme "immédiatement avant"), alors qu'une relation voisine ("avant") aurait créé un graphe cohérent.

Il est incontestable que l'on manque ici d'arrière-plan théorique sur la tâche réalisée par l'humain lui-même. Non seulement, la littérature psycho-linguistique est dépourvue de référence sur le processus d'interprétation temporel au niveau d'un texte, mais il semble bien que la tâche de-

3. Il faut noter que les annotateurs n'utilisent pas la relation d'overlap dans Timebank, ce qui diminue sans doute nettement les différences possibles entre Allen et Bruce ; il y a de plus peu de relation meet ("immédiatement avant"), mais un nombre non négligeable de starts/finishes, relations qui différencient Allen et Bruce.

mandée aux annotateurs soit très différente de la façon dont ils traiteraient la représentation temporelle d'un texte dans un cadre de lecture libre. On peut supposer que les informations traitées sur ce plan se situent à un niveau bien plus sous-spécifié. On pourrait objecter que toute tâche d'explicitation de représentations linguistiques est plus difficile que le traitement "inconscient" et est bien distincte de la compétence en tant que locuteur. Il est pourtant difficile de trouver un problème où il est aussi fréquent qu'un locuteur laissé à lui-même produise autant de données *sémantiquement* incohérentes : nous avons estimé que la première version du corpus TimeBank enrichi du corpus Aquaint contenait 20% d'annotations temporellement impossibles. Et le taux d'accord inter-annotateur cité généralement est assez bas pour être inquiétant (Mani et al., 2007) : sur les paires d'entités annotées par deux annotateurs, l'accord est de 77%. Les travaux de Setzer (2001) montrent par ailleurs la difficulté à faire produire des annotations exhaustives à des sujets, et le temps important consacré à chaque texte (une heure pour un texte comportant de 15 à 40 événements).

On ne peut s'empêcher de penser que l'avertissement donnée par Zaenen (2006) à propos des efforts d'annotations discursives peut s'appliquer aux projets d'annotation temporelle également :

One thing that is much less discussed is whether the annotation indeed helps isolate the property that motivated it in the first place. This is not the same as interannotator agreement. For interannotator agreement, it suffices that all annotators do the same thing. [...]

[...] language understanding systems have to be more than just language understanding systems : One expects them also to take care of some minimal understanding of the world the language is supposed to describe. But relations between linguistic entities and the world have not been studied in linguistics nor anywhere else in the systematic way that is required to develop reliable annotation schemas.

La conclusion est sans doute transposable également :

We should recognize that annotations are no substitute for the understanding of a phenomenon. They are an encoding of that understanding.

La manifestation de ce problème sur les projets de collecte de données reviendra au sujet du discours, mais il est frappant de voir à quel point ces avertissements sont ignorés sur la question du temps, alors que les données collectées montrent l'effort de réflexion méthodologique qu'il faudrait faire en retour des nombreuses expérimentations qu'elles supportent.

Ce n'est que récemment (Kolomiyets et al., 2012; Chambers, 2013) que les acteurs du domaine commencent à s'inquiéter du fossé qui sépare l'annotation humaine (collectée dans les corpus "classiques" TimeML) de ce que sont capables de produire les systèmes, quand les meilleures performances mesurées dans la méthodologie Tempeval sont atteintes par des heuristiques brutales, qui se cantonnent aux cas faciles de la prédiction. Une partie du problème est sans doute dans une meilleure définition des annotations visées, mais aussi une méthodologie plus claire en rapport avec les caractéristiques du domaine, ce que visent les travaux présentés dans ce chapitre. Il est à noter que la méthodologie d'évaluation des tâches Tempeval est mouvante : le protocole suivi par la récente campagne 2013 (UzZaman et al., 2013) reprend certaines idées présentées ici,

étant fondée sur l'approche de ([UzZaman et Allen, 2011](#)), qui s'inspire largement de notre papier ([Tannier et Muller, 2008](#)). Cette dernière approche présente néanmoins d'autres biais gênants, dont certains sont soulignés par ([Cherry et al., 2013](#)).

Du temps au discours

L'importance de l'organisation textuelle dans l'interprétation de la situation temporelle, des relations entre événements est reconnue, de même que l'interdépendance entre la compréhension de cette structure et les indices temporels explicites. Nous avons alors été amenés à élargir la portée des études précédentes pour prendre comme objet d'étude propre les phénomènes de structuration discursive, qui vont au-delà de l'organisation temporelle.

Ce constat est bien résumé dans (Zwaan et Razdvansky, 2001), avec un point de vue de psycholinguiste :

“ Comprehenders assume that the order in which events are reported in language matches their chronological order.

This is known as the iconicity assumption (Dowty, 1986, Fleischman, 1990). Narrative deviations from chronological order are possible only because a default order exists ; the default serving as a baseline from which all else can be compared and understood. When the iconicity assumption is not a valid guide, readers must use language cues to determine the order of events.”

Parmi les indices, on peut bien sûr lister les ancrages temporels explicites ainsi que l'utilisation des temps verbaux pour gérer des points de perspective différents. On a vu que d'un point de vue empirique, il est difficile de restreindre la question à un jeu entre ces critères seuls, et que l'organisation du discours joue un rôle :

(4.1) Elaine a refusé d'aller manger de la soupe pour deux raisons. Elle n'a pas aimé l'attitude du cuisiner et elle a attendu trop longtemps dans la queue.

Ici, les événements de la deuxième phrase sont présentés comme une explication de celui de la première, ce qui implique qu'ils aient eu lieu avant.

La compréhension des liens temporels du texte peut ne pas venir non plus de marques explicites, même sur des séquences prototypiques de narration au passé :

(4.2) Hier soir, Jerry est allé au restaurant avec Kenny et a mangé de la soupe.

e_2 pendant e_1

(4.3) Hier soir, Jerry est allé au restaurant et a vu le film 'Rochelle Rochelle'.

e_1 avant e_2

Rien ne distingue les deux exemples en surface, mais une connaissance des situations génériques

décrites (généralement on ne voit pas de films au restaurant) restreint les interprétations temporelles possibles.

C'est donc parfois la compréhension des liens temporels qui renseigne sur les intentions discursives de l'auteur, et qui peut alors conditionner la suite de l'interprétation. Voyons cela sur un exemple un peu plus long, toujours construit :

(4.4) Jerry a **passé**_{e6} une soirée formidable hier. Il a **fait**_{e7} un repas fantastique. Il a **mangé**_{e8} une soupe. Il a **englouti**_{e9} un homard. Il a **vu**_{e10} un film formidable, "Pronostic négatif".

On comprend ici que les événements des 3e et 4e phrases ont lieu pendant la 2e, mais pas le dernier événement, qui doit avoir lieu après. Par ailleurs, une contrainte discursive forte semble être qu'on ne peut revenir sur l'événement 2 après avoir énoncé l'événement 5.

Ce constat était déjà largement fait dans les travaux sur le temps de (Grover et al., 1995) et ceux sur le discours de (Asher et Lascarides, 1993; Webber, 1988). Cela rejoint une tradition pragmatique qui analyse le discours en termes d'actes de parole reliés entre eux rhétoriquement pour former une hiérarchie, dont les précurseurs sont Polanyi et Scha (1984), la structure attentionnelle de Grosz et Sidner (1986), et la Rhetorical Structure Theory (RST) de Mann et Thompson (1987). Dans une approche de ce genre, on peut structurer le discours entier, comme on peut le voir de façon schématique figure 4.1 pour l'exemple 4.4. Ici nous avons signalé graphiquement la hiérarchie par des groupes de segments emboîtés.

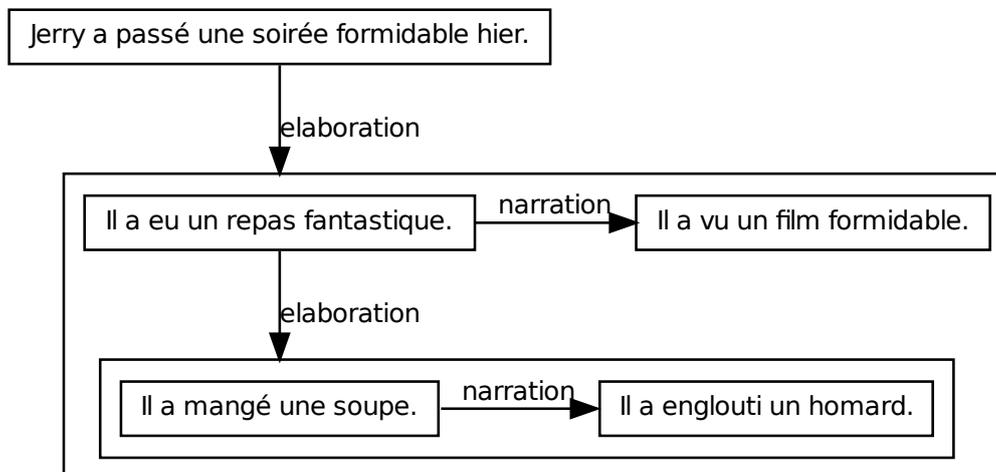


FIGURE 4.1 – Structuration schématique de l'exemple 4.4

des segments élémentaires à considérer et sur les typologies de relations mise en oeuvre dans

un discours, ces théories s'accordent sur le fait que les éléments du discours contribuent à une interprétation globale qui prend en compte les aspects sémantiques et pragmatiques au-delà de la simple concaténation des représentations des phrases. Ces approches ont été explicitement mises en rapport avec des théories sémantiques par [Asher \(1993\)](#), qui étend la DRT de [Kamp \(1981\)](#), ou par [Hobbs et al. \(1987\)](#). Dans tous les cas, on distingue au minimum des relations qui induisent une hiérarchie (comme "élaboration" dans l'exemple 4.4), ou une séquence à un niveau de la hiérarchie ("narration" dans l'exemple).

Dans ces théories, la détermination des relations rhétoriques est faite en intégrant plusieurs sources d'informations, depuis la sémantique lexicale jusqu'à des connaissances sur le monde, qui permettent d'expliquer des liens temporels ou de causalité entre événements introduits dans le discours. En SDRT ou chez [Hobbs et al. \(1987\)](#), la théorie discursive elle-même consiste en un ensemble de contraintes structurelles et des définitions logiques des relations rhétoriques dans un formalisme commun aux informations sémantiques et sur le monde. L'interprétation est faite par abduction révisable, pour produire comme représentation un ensemble de formules étiquetées, noeuds d'un graphe dont les arcs sont étiquetés par les relations rhétoriques, comme sur la figure 4.1.

Ce genre d'approche nécessite une grande quantité d'informations extra-linguistiques et des processus de raisonnement formellement complexes pour pouvoir être suivi à la lettre, au point qu'on a pu parler de problème IA-complet. Il est aussi difficile à évaluer de façon intrinsèque, faute de pouvoir disposer au moins d'une restriction testable impliquant tous les niveaux.

Avec le virage empirique qui a suivi la période logiciste dans les recherches en TAL, on a pu entrevoir une reconversion de cette approche représentationnelle sous forme d'approximations diverses qui peuvent se nourrir des approches inductives couramment employées sur des problèmes moins complexes. Par exemple, certaines associations lexicales typiques étaient considérées comme nécessaires en entrée de l'interprétation discursive (des liens de causalité par exemple), et étaient supposées provenir d'une base de données externe, de façon un peu irréaliste. On voit maintenant que ces données peuvent être collectées automatiquement, cf par exemple ([Chklovski et Pantel, 2004](#)), quitte à contenir des approximations, erreurs, oublis, qui ne sont pas rédhibitoires pour l'utilisation comme descripteurs d'approche par apprentissage automatique. Le point crucial était le manque de données pour mettre en oeuvre des évaluations.

En effet, de même que les données annotées temporellement étaient peu disponibles à l'époque des premiers modèles théoriques d'interprétation du temps et du discours, les corpus comportant des annotations discursives étaient alors rares et de petite taille. Il a fallu attendre ([Carlson et al., 2003, 2001](#); [Wolf et Gibson, 2005](#); [Miltsakaki et al., 2004](#); [Baldrige et al., 2007](#)) pour l'anglais et les autres langues étaient plutôt démunies. Ces corpus donnaient l'espoir d'étudier des régularités discursives en les mettant en relation avec les facteurs considérés jusque-là comme pertinents. Il paraissait alors justifié de vérifier comment ces présupposés se transposent en français. L'équipe dont je fais partie a donc monté un projet de collecte d'informations conjointes temps et discours en français, qui a évolué pour se centrer plus nettement sur les structures dis-

cursives pour donner le projet Annodis¹, (Péry Woodley et al., 2009), en partenariat avec le GREYC de l'Université de Caen et CLLE-ERSS de l'Université Toulouse le Mirail.

Je détaille ci-dessous les enjeux de ce projet empirique et l'utilisation que j'ai faite des données collectées pour l'analyse automatique du discours (section 4.1), puis les perspectives méthodologiques que cela ouvre, dans le cadre de la prédiction de structures de type sémantique (section 4.2).

4.1 Analyse de discours : projet empirique

Le projet Annodis a dans un premier temps donné une large place à la spécification de structures discursives annotables, pour rendre la collecte de données praticable avec des sujets extérieurs au projet. Le projet se décline sur deux échelles : l'analyse exhaustive de portions de textes assez courtes, et l'analyse ciblée de certains phénomènes sur des documents plus longs. Il a donné lieu à un volet orienté vers la collecte de données fiables, et un volet pratique qui utilise ces données pour développer et évaluer des outils d'analyse automatique de discours, détaillés ci-dessous.

Constitution de corpus de discours Suivant les approches d'analyse qui définissent le discours comme un ensemble d'éléments reliés entre eux par des liens fonctionnels qui expriment l'intention rhétorique de l'auteur, nous avons adapté les définitions des éléments proposés (appelés segments de discours) dans la littérature. Les relations rhétoriques que l'on peut définir sont en nombre variable selon la granularité et la précision que l'on veut choisir pour la sémantique des relations. Les projets inspirés de la RST peuvent choisir dans une taxonomie de relations plus ou moins précises, le standard contenant une trentaine de relations. Le corpus le plus volumineux, le Penn Discourse Treebank, propose une hiérarchie de relations à trois niveaux, et une trentaine de relations au niveau le plus détaillé (Miltsakaki et al., 2004). Certains travaux psycholinguistiques font état de fonctions repérables, notamment par l'usage de catégories de marqueurs implicites, au nombre de quatre (Spooren et Sanders, 2008), exprimant une causalité, un regroupement, un contraste, ou une séquence. D'autres théories font le choix d'avoir des sémantiques plus ciblées ou des étiquettes plus vagues, et pour des raisons pratiques, nous avons défini un jeu de 17 relations (Muller et al., 2012) adapté aux phénomènes visés (essentiellement visant des structures argumentatives ou narratives, avec par exemple les relations de narration, d'élaboration, de contraste, de résultat, d'explication pour les plus courantes).

La constitution du corpus lui-même visait à fournir des données exploitables et les moyens de développer des ressources comparables. Comme tel, il ne correspond qu'à peu près à la définition qu'en donne (Sinclair, 2005) :

A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research.

1. Projet ANR 2007-2010, <http://w3.erss.univ-tlse2.fr/annodis>

Il n'y pas eu en effet d'objectif de représentativité des textes choisis, le critère externe étant qu'ils soient libres de droit pour être diffusables avec leurs analyses, et la petite taille du corpus constitué, due à la complexité de la tâche d'annotation, est une source de particularismes sans doute dommageable (notamment pour le lexique couvert). Autant que possible, le choix des textes a été fait au hasard dans des catégories prédéfinies, pour éviter l'écueil principal de toute constitution, celui de mettre dans le corpus ce que l'on désire y trouver².

Chaque texte choisi a été annoté en double par des annotateurs extérieurs au projet mais formés à la tâche, pour pouvoir disposer d'une mesure d'accord entre annotateurs. Ensuite, les membres du projet ont repris ces annotations pour arbitrer entre les variantes proposées par les annotateurs, et constituer l'annotation de référence. L'accord inter-annotateur naïf était moyen au niveau de l'attachement de segments entre eux (0.64 de mesure kappa en moyenne sur les trois paires d'annotateurs naïfs), c'est-à-dire sur la décision ou non de relier deux segments dans le discours. L'accord est juste acceptable au niveau des relations sur les paires de segments que les annotateurs s'accordaient à relier (le meilleur kappa étant de 0.4). Une première analyse des causes de divergence d'attachement a montré en fait que certaines structures annotées respectaient les sémantiques des relations de façon parfois implicite ; en d'autres termes, les annotateurs décrivaient les mêmes structures de façons qui semblaient équivalentes ou partiellement en correspondance, mais de façon "syntaxiquement" différente, de la même façon qu'on peut décrire des situations temporelles de façons différentes comme on l'a vu plus haut. Au niveau des relations, les confusions courantes ont souligné le besoin de définition plus précise ou du moins d'un guide de décision plus strict (la confusion la plus courante est celle entre élaboration d'événement ou d'objet, qui a pourtant des critères syntaxiques très clairs). Au niveau structurel, on peut noter que l'accord inter-annotateur monte à 0.57 si on ne considère que la distinction la plus générale à toutes les théories entre relations hiérarchiques ou de séquence (distinction entre relations "subordonnantes" contre "coordonnantes" en terminologie SDRT). Les spécifications de l'annotation et les mesures d'accord sont présentées dans (Muller et al., 2012).

L'annotation a donc souligné l'importance d'approfondir les liens sémantiques entre relations, soulevant des problèmes méthodologiques sur lesquels nous revenons à la section 4.2.

A la collecte et la validation de ces données, le projet associe des procédures d'automatisation dont le but est de généraliser le type d'information présente dans le corpus, et d'aider à la collecte de futures données en prétraitant certains aspects de la tâche, en supposant que cela soulage au moins en partie les annotateurs. D'un point de vue de traitement des langues, une analyse discursive sur ces bases peut se décomposer en deux grandes tâches : repérer les unités de discours élémentaires (UDE), et établir l'ensemble des relations qui forment la structure.

Segmentation discursive La première tâche consiste à découper le texte en segments porteurs d'une intention rhétorique. Une phrase élémentaire est considérée comme une unité de base, et pour les phrases plus complexes un ensemble de règles a été établi (Muller et al., 2012). Un

2. Sinclair (ibid) : "Experts in corpus analysis are not necessarily good at building the corpora they analyse — in fact there is a danger of a vicious circle arising if they construct a corpus to reflect what they already know or can guess about its linguistic detail."

segment prototypique est une clause verbale indépendante. Typiquement un segment correspond à la description d'un événement ou d'un état unique. Les critères pertinents sont donc partiellement de nature syntaxique, partiellement sémantique. En plus d'être une étape préalable à l'analyse discursive, la segmentation peut aussi être utile en soi pour d'autres tâches où les unités élémentaires peuvent fournir des entrées plus simples que des phrases complètes. Cela peut être le cas du résumé automatique (Wan et Paris, 2008), ou de la compression de phrases (Clarke et Lapata, 2007; Sporleder et Lapata, 2005). Potentiellement cela pourrait s'envisager au niveau de l'alignement de bi-textes (textes parallèles en deux langues) pour la traduction, ou bien comme prétraitement en analyse syntaxique.

La segmentation discursive est assez largement dépendante de la théorie rhétorique que l'on vise, qui peut faire des hypothèses diverses sur les éléments discursifs qui peuvent y être reliés. Par exemple en RST, la théorie ne prévoit pas de segments discontinus, et donc pas d'emboîtements. Plus exactement, un segment coupé par une clause emboîtée constitue deux EDUs, qui seront reliées par une relation rhétorique artificielle "same unit"³. Cela permet de faire du problème de la détection de segments un problème de classification binaire où, à chaque séparateur potentiel, il faut décider si une transition a lieu d'un segment à un autre. Au contraire, la Segmented Discourse Representation Theory (SDRT) prévoit des structures emboîtées dans le cas de clauses en modifiant une autre, comme des appositions ou des relatives non restrictives. Le corpus collecté et présenté à la section 4.1 contient environ 10% de segments imbriqués. Par exemple, le segment 3 est emboîté dans le 2 :

[Bien que celle-ci soit géographiquement située en Afrique,]₁ [l'Art de l'Égypte antique, [né il y a environ cinq mille ans,]₃ est l'une des principales sources de l'art en Europe.]₂

Le problème est en fait structurellement assez proche de la détection de clauses grammaticales, et nous avons donc repris les procédures appliquées sur cet autre problème (Carreras et Màrquez, 2001), en adaptant les descripteurs à notre tâche, en collaboration avec Stergos Afantenos, Pascal Denis et Laurence Danlos (Afantenos et al., 2010).

Pour l'instant, la segmentation testée sur le corpus de développement de 45 documents atteint un F-score de 73% de repérage des segments, à comparer aux 80% de réussite du meilleur système sur le problème structurellement équivalent d'identification de clause syntaxique. Cette étape est plus un prérequis dans une perspective d'analyse discursive complète, préalable à la construction de la structure discursive elle-même.

Ceci ne concerne que la segmentation des unités élémentaires. Une autre échelle à laquelle on peut considérer la segmentation est celle en unités plus larges, notamment dans une approche top-down du discours où l'on cherche à découper le texte en unités thématiques. Je reviendrai sur cette échelle d'analyse au chapitre 6, qui fait jouer une dimension importante au lexique.

Analyse automatique de relations discursives La suite naturelle de l'analyse discursive est la prédiction des relations structurelles entre segments élémentaires, une fois ceux-ci détectés, et

3. Cf. (Carlson et Marcu, 2001), page 23.

c'est un des aboutissements prévus d'ANNODIS, sur le corpus constitué pendant le projet. J'ai travaillé sur cet aspect du projet au sein de l'IRIT avec Nicholas Asher, Stergos Afantenos, ainsi que Pascal Denis de l'équipe Alpage, et les résultats ont été publiés dans (Muller et al., 2012). Produire une structure discursive implique de prédire une structure relationnelle, et on peut considérer deux versants pour cette tâche : le premier concerne la détermination des segments à relier, ce qui correspond à la détermination de "l'attachement" d'un segment au contexte dans une perspective de construction incrémentale, comme c'est le cas de (Asher, 1993). Le deuxième versant consiste à identifier la nature de la relation entre les deux segments. Cette séparation est bien sûr artificielle si on considère qu'attacher deux segments ne peut se faire qu'en prenant en compte les relations qui peuvent les relier, et les contraintes afférentes.

Peu de travaux ont été réalisés sur des données empiriques, et ils sont pour l'instant restreints uniquement à l'anglais à ma connaissance. Sur la RST, après avoir conçu un analyseur symbolique partiel essentiellement à base de contraintes, Marcu et Echihabi (2002) ont appliqué des techniques de parsing statistiques (à base d'algorithmes de "shift-reduce") pour améliorer la résolution.

Dans une approche rhétorique mais sur des données dialogiques, et avec un jeu de relations très guidé par la tâche sous-jacente aux conversations (des prises de rendez vous téléphoniques) (Baldrige et Lascarides, 2005) ont induit une grammaire de discours à partir des données du projet Verbmobil (Wahlster, 2000).

Plusieurs études ont réutilisé les corpus RST ou PDTB pour tester de nouvelles approches prometteuses, là encore souvent inspirées de méthodes d'analyse syntaxique automatique⁴.

L'inconvénient majeur des approches existantes est l'absence de production de structures complètes (analyses limitées à l'intérieur des phrases), ou bien ne tenant pas compte de toutes les contraintes globales. Par exemple, duVerle et Prendinger (2009) s'attachent à produire des structures RST bien formées, mais suivent une heuristique d'attachement locale qui ne prend pas nécessairement en compte la structure globalement la plus probable.

Pour remédier à cela, nous avons visé à optimiser globalement le problème de l'analyse discursive : même en utilisant un modèle d'apprentissage purement local (c'est-à-dire qui considère les décisions d'attachements d'unités de façon indépendante), on construit une structure discursive qui maximise l'ensemble des décisions prises.

Un deuxième objectif était d'augmenter la généralité des modèles d'interprétation : les approches citées sont tournées vers le formalisme RST, qui produit des structures spécifiques, même si une traduction approximative en arbre est possible. Pour cela, il faut à la fois de la souplesse dans la construction d'une structure, pour pouvoir intégrer des contraintes spécifiques de bonne formation (arbres, graphes), et la possibilité d'intégrer des contraintes plus théoriques, comme le respect de l'attachement d'une unité sur la frontière droite (Polanyi, 1988). L'espace des solutions étant en général exponentiel, et l'ajout de contraintes susceptibles de rendre la recherche combinatoire, nous avons exprimé le décodage de la structure discursive comme un problème d'exploration avec un algorithme de type A*.

4. Le Thanh et al. (2004), Wolf et Gibson (2006) Subba et Di Eugenio (2009), duVerle et Prendinger (2009), Feng et Hirst (2012)

Un autre pas vers un modèle d'interprétation plus général a été effectué en visant comme structure des graphes de dépendances entre unités simples (graphe étiqueté orienté), similaires à des graphes de dépendance syntaxique. Dans un premier temps cela entraîne des simplifications par rapport à ce qui est encodé dans les corpus traités, mais fournit une approximation que l'on peut rapporter aux différents formalismes (SDRT, RST, PDTB). La question de l'expressivité du formalisme est en cours d'étude de façon distincte de la partie expérimentale, avec pour but de continuer à poser des ponts entre les théories et les corpus existants. Des processus de conversions sont notamment présentés dans les travaux effectués en collaboration avec Antoine Venant et les auteurs du papier Coling cité ci-dessus, cf. (Venant, Asher, Muller, Denis, et Afantenos, Venant et al.).

Le dispositif expérimental a permis de comparer l'adéquation de différents modèles d'apprentissage locaux, les performances des stratégies de décodage, et l'influence de certaines caractéristiques du processus d'interprétation : par exemple le fait de restreindre l'attachement à des petites distances dans le texte, forçant ainsi un modèle "local-étendu" ; ou encore l'impact d'une modularisation des tâches d'attachement et d'étiquetage de relations ; finalement l'importance de la granularité des relations rhétoriques choisies, la plupart des théories proposant en effet une hiérarchie de relations et de groupes de relations.

Comment pallier la rareté des données ? Les approches présentées sont limitées par la rareté des données, qui contraint le pouvoir de généralisation de l'approche inductive. Ce problème est bien plus critique que dans le cas de l'analyse syntaxique. L'équivalent d'une phrase en analyse syntaxique automatique est un discours entier, et la structure étant bien moins contrainte qu'en syntaxe, où on produit à peu près toujours un arbre de constituants ou de dépendances, l'espace des hypothèses est plus grand. Le plus gros corpus (RST) comprend 350 textes, ce qui fait très peu d'instances pour entraîner un modèle structurel. Par ailleurs, les descripteurs utiles sont mal connus ou en tout cas peu étudiés dans un contexte aussi empirique. On connaît bien sûr l'importance des marqueurs explicites de relations, qui sont au centre des études plus restreintes de (Sporleder et Lascarides, 2006; Pitler et al., 2008)⁵, et les travaux de Lascarides et Asher notamment ont maintes fois souligné l'importance du lexique et des liens typiques entre événements centraux des segments discursifs, mais il y a peu de résultats exploitables au-delà de cela.

Il me semble donc que l'effort doit porter sur deux directions. La première est l'étude réduite de facteurs pertinents de façon isolée sur les données disponibles pour l'instant, pour aider la précision des descripteurs utiles à une tâche d'apprentissage automatique. C'est cette direction qu'ont suivi par exemple (Afantenos et Asher, 2010), en étudiant la réalité d'une contrainte forte de hiérarchisation du discours, dont la perspective est de réduire les segments candidats à l'attachement à un autre. Etudier les attachements du corpus Annodis leur a permis de constater également que les segments étaient très rarement reliés à grande distance dans les textes.

La deuxième direction est motivée par le coût de production des annotations de référence, qui sont la source des analyses qualitatives et la source de données d'entraînement pour les algo-

5. On peut noter aussi que selon (Taboada, 2006), entre 60 et 70% des relations ne sont pas signalées par un marqueur explicite.

rithmes d'apprentissage statistique de la tâche. Nous avons dû réévaluer ce coût au cours du projet, car les données récoltées sont moins nombreuses et demandent plus de traitement que prévu, et leur pouvoir de généralisation sera donc diminué. Cela conduit à définir un nouveau contexte de recueil et d'automatisation, entre apprentissage semi-supervisé et apprentissage actif. Le cadre semi-supervisé utilise des données non annotées pour régulariser les modèles produits sur les données annotées, alors qu'en apprentissage actif, les premières données annotées servent à cibler les jugements demandés à des opérateurs humains, pour enrichir les descriptions là où les modèles sont les plus faibles. La mise en place de ce cycle différent est une nouveauté sur ce type de données, et il est déjà assez rare en TAL en général (voir l'atelier dédié récent de [Ringger et al., 2009](#)). On peut signaler les premiers travaux en ce sens sur le discours de [Braud et Denis \(2013\)](#).

Un aspect original de la vision rhétorique SDRT est la structuration en topiques, des segments complexes de taille arbitraire qui englobent un ensemble de segments et peuvent être reliés par les mêmes relations. On retrouve la structuration en segments complexes dans toutes les approches bien sûr, mais avec des contraintes généralement plus fortes. En RST par exemple, les segments complexes sont construits hiérarchiquement sur des paires de segments dont l'un est considéré comme la tête ("nucleus" dans la terminologie de la théorie), ce qui fait que la représentation est formellement équivalente à un arbre où un argument complexe peut être remplacé par cette tête. Les rares cas où il n'y a pas de nucleus peuvent d'ailleurs être ordonnés arbitrairement pour conserver cette structure d'arbre, plus simple à apprendre. Cette contrainte n'existe pas en SDRT, et complique donc la construction de la structure en ajoutant un problème de partitionnement. On pourrait se contenter d'introduire un nouvel élément pour chaque segment complexe et d'ajouter un lien de constitution entre celui-ci et les membres du segment, et se ramener à une structure de graphe pour l'ensemble, mais ce serait au prix d'un accroissement non négligeable de la complexité de la tâche de prédiction. Une autre solution serait de considérer une approximation de la structure en interrogeant les propriétés de ces segments complexes. On retrouve là la question des inférences à définir qui permettraient de réécrire la structure discursive de façon plus classique. Cela pose des questions de méthodologie pour définir la validité de ces inférences, que j'expose à la section suivante.

4.2 Evaluation de l'interprétation discursive

Inférence et relations de discours Les sujets abordés dans la section 2.3.1 sur les graphes temporels sont pertinents pour d'autres représentations à base de graphes ou de contraintes en traitement des langues, à partir du moment où il y a des questions de raisonnement, et un jeu de prédicats dont on peut spécifier la sémantique et qui définissent les possibilités de relier certains éléments ciblés. C'est le cas en particulier des théories rhétoriques du discours. Celles-ci décrivent des relations (qui dépendent de la théorie considérée), entre des éléments de base du discours que sont les segments élémentaires. Les propriétés inférentielles qui lient ces relations ne sont pas l'objet d'un consensus, mais la SDRT par exemple est à la source de beaucoup de

travaux visant à caractériser leurs liens sémantiques, et certaines propriétés importantes que l'on a vu sur le temps, comme d'être une relation d'ordre partiel, se retrouvent sur certaines structures discursives. Les chaînes narratives ou causales sont transitives, les élaborations imbriquées de thèmes le sont également, une forme que l'on retrouve également dans les séquences de dialogues emboîtées⁶.

Une perspective d'évaluation des tâches de traitement associées doit donc considérer les mêmes questions que pour l'extraction temporelle.

Par ailleurs, la question d'une représentation minimale intermédiaire qui évite certains biais d'évaluation comme pour le temps (cf section 3.1) peut se poser aussi pour le discours.

Les premières analyses des données récoltées dans le projet Annodis ont confirmé l'importance pour l'analyse discursive des aspects méthodologiques rencontrés sur l'analyse temporelle. Premièrement, la variété des annotations produites par les annotateurs d'un même texte (chaque texte est annoté en double), au-delà des désaccords attendus habituellement (erreurs, lacunes de certaines consignes, etc), a conduit à repenser le modèle de comparaison discursive, en intégrant la dimension inférentielle à la tâche elle-même. Deux annotateurs peuvent produire des annotations syntaxiquement différentes, mais sémantiquement équivalentes, car les descriptions produites ont pour conséquences logiques d'autres descriptions totalement ou partiellement retrouvées dans les annotations associées, comme pour le temps.

Considérons l'extrait suivant d'un article de l'Est Républicain, choisi pour le corpus Annodis :

(4.5) [Les services d'EDF-GDF ont pu rétablir l'électricité dans l'immeuble peu avant 21 h.]₂
[Pour le plus grand soulagement des locataires]₃ [qui ont pu regagner leurs logements.]₄

Deux annotateurs ont décrit cette partie du texte avec la relation `result(2,[3-4])` pour exprimer le lien de causalité entre la première et la deuxième phrase, vu comme un segment complexe de deux unités de discours élémentaires. Le premier annotateur a ajouté `explanation(3,4)`, précisant la chaîne de causalité en exprimant que 3 est expliqué par 4, alors que l'autre annotateur a seulement indiqué `continuation(3,4)`, une simple indication structurelle qui confirme le statut de segment complexe de l'ensemble 3-4. Comparées brutalement, la deuxième relation est différente pour les annotateurs, mais l'une des annotations peut-être considérée comme une version sous-spécifiée de l'autre, structurellement compatible.

On aurait pu imaginer ici une description valide `{result(2,4), explanation(3,4)}`. En supposant donc que les causalités exprimées soient transitives on a un lien de cause à effet entre 2 et 4, et entre 4 et 3, et donc un lien entre 2 et 3. On pourrait donc considérer qu'on a des formes équivalentes sémantiquement, au moins partiellement, et sans doute beaucoup plus que ne le laisserait voir un comptage brutal des annotations identiques entre annotateurs. On voit là qu'il faut spécifier des équivalences entre les relations rhétoriques, ou au moins d'une partie de leurs conséquences sémantiques, d'une façon similaire à l'usage qui est fait de la sémantique des relations temporelles pour comparer des annotations.

Ces équivalences demandent un travail important de spécification et d'étude des cas observés dans les données avant d'avoir un modèle général. Ce point est également crucial pour être ca-

6. Le cas du dialogue sera examiné au chapitre suivant.

pable d'évaluer de façon juste ce qui est la partie opératoire du projet : l'analyse automatique.

Les équivalences de description posent en fait deux problèmes distincts : (1) à partir du moment où sont admis des segments complexes dans le modèle, comment peut-on les relier à des descriptions plus ou moins équivalentes qui ne font appel qu'à des unités élémentaires ? et (2) quels sont les liens sémantiques, les lois de composition entre relations ?

Sur le point (1), une étude préliminaire des correspondances structurelles entre segments simples et complexes a été menée avec Nicholas Asher, Antoine Venant et Stergos Afantenos (Asher et al., 2011) et doit être poursuivie. Le point (2) est plus complexe. Si certaines règles semblent simples (les propriétés de la narration entre segments contigus sont transitives par exemple), la composition de relations dans le cas général reste largement inexplorée. Afin d'explorer les propriétés sémantiques des relations qui peuvent aider à déterminer ces compositions, nous avons construit une méthodologie, dans le cadre de la thèse de Charlotte Roze, encadrée par Laurence Danlos, et également en collaboration avec Pascal Denis, cf (Roze, 2013). Face à la rareté de données annotées, le travail consiste à trouver de bons indicateurs de relations (comme des marqueurs explicites) pour isoler un nombre significatif de contextes où une analyse sémantique peut dégager des principes généraux. A terme on peut viser la constitution d'un ensemble de règles de composition de relations qui ressemblerait à ce que l'on a vu sur les algèbres de relations temporelles.

Dans cette optique, le travail avec Charlotte Roze a consisté à étudier en détails des interactions typiques de relations, en les sélectionnant dans des groupes différents d'un point de vue de l'usage et de leur sémantique. Une opposition classique est faite par exemple entre des relations plutôt "présentationnelles" (argumentation parallèle, contraste, ...) et des relations plutôt "idéationnelles", de contenu. Dans les relations de contenu, la plupart des théories considèrent à part des relations exprimant des causalités, des relations qui supportent une narration temporelle. Les interactions entre ces différents groupes permettent d'avoir un échantillon des relations à la fois importantes, d'usage fréquent, et fournissant des situations variées.

Les exemples 1 et 2 illustrent un contexte typique de l'étude où les inférences sont relativement intuitives : dans les deux cas la structure de base est constituée d'une narration (introduite par *puis*) et d'une explication (introduite par *parce que*) mais dans le premier cas l'état décrit en (γ) est au moins concomitant avec l'événement décrit en (α) , alors que dans le deuxième cas l'événement de (γ) s'enchaîne directement à celui de (α) , ce qui change le lien rhétorique potentiel entre ces unités s'il y en a un, et contraint différemment les interprétations dans tous les cas.

- | | | | |
|-----|---|-----|---|
| (1) | a. Marie a fait une balade, (α) | (2) | a. Jean dormait en cours. (α) |
| | b. <i>puis</i> elle est allée boire un café (β) | | b. <i>Puis</i> il s'est réveillé (β) |
| | c. <i>parce qu'</i> il faisait très froid. (γ) | | c. <i>parce que</i> le prof lui a hurlé dessus.
(γ) |

Ce travail nécessite d'une part un bon investissement théorique linguistique pour considérer

l'éventail des constructions et des contextes envisageables, et une validation empirique sur des données où s'instancient les cas considérés. Pour contraindre les cas possibles, des contextes envisagés sont testés vis à vis des contraintes sémantiques imposées, avec une méthodologie systématique. Dans un second temps, il a fallu extraire de données linguistiques des cas représentatifs des contextes choisis. Les corpus annotés en relations de discours étant trop petits, il a fallu choisir une méthode alternative consistant à chercher des contextes dans un corpus étendu, quitte à ne considérer que des cas relativement explicites. Dans une forme de semi-supervision, des contextes ont été extraits automatiquement de corpus français en utilisant un ensemble de marqueurs possibles des relations visées, grâce au travail de recensement et de classification effectué dans (Roze et al., 2012). Une annotation manuelle a été faite en suivant la méthodologie de test mentionnée plus haut, afin d'obtenir des statistiques de réalisation des inférences.

La table 4.1 donne par exemple les résultats pour les contextes qui impliquent une relation de narration temporelle entre deux unités de discours, suivie d'une relation causale (explication) entre la 2e et une 3e unité discursive.

$R_z(\alpha, C)$	NB	%
<i>Pas de relation</i>	26	38.24
<i>Violation</i>	22	32.35
<i>Narration</i>	14	20.59
<i>Resultat</i>	6	8.82
<i>Arrière-plan</i>	3	4.41
<i>Explication</i>	2	2.94
<i>Flashback</i>	1	1.47
<i>Total</i>	68/74	108.83

TABLE 4.1 – Inférence de relation de discours sur un sous-échantillon annoté du corpus extrait automatiquement par Charlotte Roze, restreint à la composition d'une relation de temporalité avec une relation de causalité. Pour 68 instances, 74 relations (ou la relation "pas de relation") peuvent être inférées. Les proportions sont données par rapport au nombre d'instances de discours extraits.

A terme l'objectif est de fonder un modèle de prédiction automatique sur ces contextes, et d'étendre la couverture des compositions considérées. Au fur et à mesure de l'extension de cette couverture, les modèles de prédiction globale de la structure du discours présentés à la section 4.1 pourront être au moins contraints et au mieux informés par ces constructions.

Une autre façon de procéder est d'utiliser les données annotées d'Annodis pour trouver des caractéristiques de contextes pertinents (par exemple dans l'expression de la causalité), qui permettent de trouver de nouveaux exemples. Cette méthodologie est proche de celle qui est suivie

en extraction de relations, cf par exemple les travaux déjà mentionnés de [Chklovski et Pantel \(2004\)](#).

4.3 Perspectives

L'analyse automatique discursive correspond à un cas difficile en TAL. En tant que dernier étage de l'analyse du langage naturel elle mobilise potentiellement toutes les sources d'informations à l'oeuvre dans les étages inférieurs. Les entrées du problème sont donc multiples, et dans une perspective d'automatisation les données intermédiaires, comme les analyses syntaxique et sémantique, sont partielles ou bruitées. Surtout, le problème principal est la forme de ce que l'on cherche à prédire : dans le cas le plus favorable, ce sont des arbres étiquetés, au pire ce sont des graphes. Ces graphes obéissent à des contraintes, mais celles-ci sont mal connues. Le problème le plus proche de l'analyse discursive d'un point de vue structurel est bien sûr l'analyse syntaxique, et certaines de ces techniques ont commencé à être adaptées au problème discursif. Principalement ce sont les méthodes qui construisent des dépendances, les plus faciles à modifier pour produire des graphes obéissant à d'autres contraintes que les contraintes syntaxiques (cf section 4.1).

On a vu dans le cas de la prédiction de structures temporelles qu'il fallait pouvoir exprimer des contraintes pour guider l'analyse, et que ces contraintes avaient peu de chance de correspondre aux contraintes assez fortes de l'analyse grammaticale. Dans le domaine temporel, il n'est pour l'instant pas clair qu'il y ait des restrictions plus fortes que celles de maintenir une cohérence entre les relations prédites. Pour le discours, un certain nombre de théories prédisent des contraintes structurelles qui sont plus ou moins bien définies selon les cadres.

Dans tous les cas, on se trouve à la frontière du TAL, avec (a) des sorties complexes, (b) des données annotées en petit nombre (difficiles à collecter, et à la fiabilité variable), ce qui limite les approches inductives, et (c) des critères d'évaluation mal spécifiés.

On peut alors considérer qu'une stratégie similaire à l'analyse syntaxique ne peut réussir telle quelle.

Pour affronter le premier point, qui implique un grand nombre de prédictions possibles qu'on ne peut énumérer de façon pratique, une approche raisonnable est de générer une sélection de structures hypothèses plausibles (les meilleures selon un premier score), et d'apprendre à les comparer ("ranking"). Cela permet de séparer le problème en deux : l'un consiste à utiliser des informations locales et un guidage de la construction globale plus ou moins fruste (une optimisation gloutonne de l'agglomération des décisions locales par exemple), et l'autre consiste à apprendre ou à définir des préférences sur les structures a posteriori. Un exemple typique de problème discursif traité avec succès de cette manière est le problème de la résolution de coréférence, cf ([Denis et Baldridge, 2007](#)).

L'apprentissage de préférences ou si l'on veut de comparaisons de prédictions de structures implique notamment de pouvoir comparer des structures pour définir des similarités entre elles. Une avenue prometteuse en apprentissage sur des données structurées, qu'elles soient utilisées

en entrée ou comme résultat des prédictions est en effet d'utiliser des distances entre structures, et de considérer l'apprentissage dans l'espace qui correspond à ces structures. Un certain nombre de méthodes, dites méthodes à noyau, comme les SVM, n'ont pas besoin d'une représentation explicite de cet espace, qui est ici de dimension inconnue, mais seulement d'un produit vectoriel entre les structures concernées⁷, qui correspond ici à une mesure de similarité.

L'un des intérêts des travaux menés à la section 4.2 est de définir de telles mesures de similarité. L'autre intérêt de la définition de comparaison entre structures est aussi de résoudre certaines questions posées par le point (c) ci-dessus, à savoir l'évaluation du succès d'une prédiction. Si on compare avec les évaluations en analyse syntaxique, par exemple (Black et al., 1991), on voit la nécessité d'évaluer de façon graduelle, en comparant des sous-structures communes et un degré de recouvrement de la structure.

Enfin, il reste les difficultés posées par le point (b). J'ai déjà mentionné la possibilité d'aller vers des approches avec moins de supervision. On a aussi vu sur le temps l'intérêt de définir des restrictions du problème qui échappent en partie aux contraintes posées par la pauvreté des données disponibles, et notamment des stratégies fondées sur le principe "diviser pour régner". Il est alors crucial de préserver la possibilité de recombinaison ou de généraliser les restrictions. Cette généralisation n'est pas toujours possible avec un certain nombre d'approches de la littérature sur le temps et le discours n'est sans doute pas à l'abri de stratégies réductrices à l'excès.

Les pistes proposées pour l'interprétation temporelle, à savoir l'hybridation de contraintes locales/globales et l'exploration de divisions du problème semblent raisonnables aussi pour le discours. Notamment l'approche essentiellement ascendante présentée jusqu'ici peut rencontrer des approches où le découpage du texte est plus macroscopique, et je reviens sur cet aspect au chapitre 6, où le lexique joue un rôle primordial.

La démarche présentée ici tente de rapprocher les approches du temps et du discours à la fois formalistes et empiriques. L'adaptation de concepts théoriques, nés par ailleurs de l'observation empirique, à des cadres d'évaluation précis permet de réévaluer certaines hypothèses de constructions discursives. Par exemple, la possibilité d'avoir un modèle incrémental pour l'interprétation temporelle semble trop optimiste au-delà de quelques phrases ; ou encore, on a pu voir que les contraintes théoriques qui régissent les possibilités d'attachement de relation de discours ne sont pas toujours respectées en pratique.

Cela se fait bien sûr sous des contraintes techniques qui ne permettent pas nécessairement de mettre à l'épreuve toutes les options. Parmi les facteurs qui restent à étudier, le rôle du lexique est resté jusqu'ici un peu en retrait, ainsi que les aspects liés à la taille des textes considérés : la petite taille des textes assure une certaine homogénéité thématique à l'intérieur d'un texte, et fait l'impasse sur des phénomènes de structuration plus évidents à une échelle plus large. C'est l'autre aspect du projet Annodis, dit "macroscopique", d'étudier les formes de structuration du discours sur des textes plus importants, la mise en forme par exemple, afin de voir à quel niveau l'analyse rhétorique qu'on a présentée ici devient pertinente. Cette perspective du projet a commencé sur

7. (Collins et Duffy, 2001), cf aussi le chapitre "Measuring Similarity with Kernels" de (Bakir et al., 2007)

l'aspect de la structuration thématique à gros grain, et fait intervenir pour l'instant surtout des informations de nature lexicale, ce qui est l'objet du chapitre 6.

Par ailleurs, les évaluations présentées sur ces efforts sont intrinsèques, dans le sens où elles reposent sur des annotations de référence des concepts qu'on étudie, et non pas sur une tâche externe qui mettrait en jeu les informations étudiées. Cela pose un problème de méthode quand les jugements demandés à des sujets humains qui produisent la référence ne sont plus seulement linguistiques comme ici, mais font appel à des concepts externes, comme des représentations temporelles, causales, etc. En l'absence de modèle cognitif consensuel, la constitution d'annotations de référence navigue donc entre l'écueil d'une spécification très précise, opératoire, mais peu naturelle et déconnectée du sujet réel de l'étude, et celui de directives floues conduisant à trop de variations entre les juges.

Les chapitres suivants vont exposer ce que cette démarche hybride, le développement de modèles formels sur la base d'indices empiriques, peut apporter comme éclairage sur l'objet discursif.

Du discours au dialogue

Les chapitres précédents ont présenté des travaux mêlant une modélisation de la structure discursive et des questions de sémantique spatiale et temporelle, l'aspect temporel étant le plus développé. Parallèlement, j'ai étudié l'échange d'informations de même nature dans ce qui peut être considéré comme un cadre discursif étendu : une conversation entre deux locuteurs. Il existe une tradition de recherche sur l'information spatiale dans le dialogue, avec une coloration psycho-linguistique (Taylor et Tversky, 1992), notamment dans l'étude des objets spatiaux-temporels particuliers que sont les itinéraires. Un itinéraire est la description ou la réalisation d'un déplacement dans l'espace, organisé par rapport à un ensemble de repères et déterminés par des actions de la part d'un sujet (Denis, 1997). La communication sur cet ensemble de déplacements pose de nombreuses questions à la fois sur la sémantique des expressions utilisées, la façon dont elles sont reliées, et sur la façon dont elles peuvent être partagées entre deux locuteurs. Cela fournit en retour un cas d'étude idéal dans une perspective plus générale de modélisation du partage d'informations pendant une conversation. On retrouve donc des aspects discursifs déjà rencontrés, dans l'organisation des informations par rapport à un locuteur, avec des aspects à la fois sémantiques et pragmatiques. Cependant, les intentions qui sont à la base des productions langagières proviennent de deux agents, et vont donc interagir, se compléter ou se contredire, et vont faire intervenir une dimension conventionnelle différente du discours écrit. Dans cette perspective également, j'ai ménagé les dimensions à la fois formelle et empirique. Aux études de modèles qui permettent de représenter les phénomènes dialogiques liés à l'échange d'information (sections 5.1 et 5.2), j'ai ajouté des études empiriques détaillées après collecte d'un corpus restreint à des dialogues focalisés sur une tâche spatiale (sections 5.3 et 5.4). Les aspects formels abordent les contraintes qui portent sur la structuration thématique d'un dialogue orienté tâche, et sur la dynamique des engagements des locuteurs au cours d'une conversation. Le volet empirique se concentre sur les marques d'acquiescements, une fonction relativement ignorée par la littérature sur le dialogue, et de leurs différentes fonctions dialogiques dans le processus d'établissement de l'information partagée et des engagements. Cet aspect est complété par une analyse du rapport entre certaines marques linguistiques et la structure topicale de la conversation. La dimension empirique est ici plus analytique et ne comporte pas de volet prédictif.

5.1 Représentation sémantique structurée du dialogue

Dans une approche sémantique du dialogue, la conversation est vue comme un ensemble d'actions réalisées linguistiquement, dotée d'un contenu propositionnel, qui ont des conséquences sur l'état de l'interaction entre les deux locuteurs. L'état du dialogue ou son "contenu" est généralement associé à une forme d'attitude mentale : croyance, intention, objectif, obligation. On peut distinguer deux problèmes principaux abordés par les théories qui modélisent l'interaction dans cette perspective :

- quelles sont les séquences correctes d'actes dialogiques qui constituent un échange cohérent, admissible ?
- comment peut-on interpréter ces séquences en termes d'échange d'information, d'évolutions des croyances des locuteurs, d'interaction sociale ?

Selon qu'elles insistent plus sur les effets, sur les états mentaux des participants (croyances, intentions) ou sur les contraintes sociales qui gouvernent les interactions (obligations, engagements), on parle d'approches mentalistes (Cohen et Perrault, 1979) ou conventionnelles (Traum et Allen, 1994). L'analyse des énoncés eux-mêmes peut alors être effectuée à la lumière du paradigme de reconnaissance d'intentions initié par (Grice, 1975), et implémenté plus tard dans des contributions influentes comme (Grosz et Sidner, 1986), ou bien dans une perspective plus sociale dans laquelle les énoncés répondent à une forme de pression interactive et à des obligations sociales. Une autre ligne est suivie par les approches plus centrées sur la formalisation sémantique du discours, qui prennent les énoncés comme point de départ de la construction du dialogue comme discours structuré, de la même façon qu'un discours écrit, et se focalisent alors sur l'interface sémantique-pragmatique, notamment les travaux de (Asher et Lascarides, 2003), qui relie une théorie sémantique à une modélisation des actes de dialogue. J'ai commencé à faire le pont entre ce type de cadre formel, qui permettent d'expliquer l'organisation des actes de dialogue (les structures de questions-réponses notamment, voir ci-dessous) et le souci d'intégrer les contraintes conventionnelles, sociales, qui portent sur les participants, par le biais de la notion d'engagement propositionnel (section 5.2), définie par Hamblin (1970). Le modèle dialogique se rapproche alors d'un enregistrement à la Lewis (1979), avec des engagements ("commitments") dans la lignée de Walton et Krabbe (1995). On trouve là une notion similaire à la cohérence d'un discours écrit : des liens pragmatiques entre unités discursives, ici des actes de parole dans une conversation.

Dans ce contexte, j'ai travaillé à la spécification des représentations de séquences questions-réponses et des séquences évaluatives, à travers la modélisation de topique de conversation, en encadrant la thèse de Laurent Prévot (Prévot, 2004) avec Joan Busquets, et dont les directeurs étaient Andrée Borillo et Andreas Herzig.

Le rôle rhétorique des questions est un aspect important du traitement de la structure d'une conversation. A l'origine une théorie rhétorique comme la SDRT ne considérait que le versant cognitif de ce problème (le niveau intentionnel, notamment la réalisation d'un plan). Les questions étaient alors reliées au contexte par une fonction d'élaboration ("question-elaboration"), une version interrogative de la fonction d'élaboration du plan lié à la conversation. Elles sont

donc concernées par la structure intentionnelle telle que la conçoivent [Grosz et Sidner \(1986\)](#). Pourtant les approches du dialogue dont l'unité d'analyse est l'acte de dialogue font la distinction entre les requêtes à visée informative, liées aux buts sous-jacents des locuteurs, et les actes qui concernent la gestion de l'interaction elle-même, comme les demandes de confirmation ([Bilange, 1991](#); [Poesio et Traum, 1997](#)). Nous avons défendu l'hypothèse qu'une question peut remplir des fonctions différentes ([Muller et Prévot, 2002, 2008](#)). Nous ajoutons que toutes les relations importées du monologue peuvent être réalisées par une unité question-réponse. Nous avons alors étudié tous les cas possibles de séquences et d'imbrications de questions-réponses et la façon dont elle sont liées aux diverses fonctions rhétoriques.

Dans ([Asher et Lascarides, 2003](#)), un autre niveau d'attachement des questions est venu modifier les versions précédentes de la modélisation dialogique dans un cadre SDRT : le niveau du contenu. Les questions peuvent alors être reliées sur ce plan avec des liens du même type qu'en monologue : une question dont la réponse formera un tout narratif avec un segment précédent est liée au contexte par une relation appelée *narration_q* par exemple, et ceci est valable pour les autres relations monologiques. C'est le segment complexe question-réponse qui joue en fait le rôle rhétorique, avec un contenu propositionnel qui est la combinaison des contenus sémantiques de la question et de la réponse.

Cela pose cependant un problème structurel, puisque certaines de ces fonctions rhétoriques sont (a) subordonnantes et d'autres (b) coordonnantes, c'est-à-dire (a) induisent une construction hiérarchique des segments de discours ou des tours de parole, ou bien (b) correspondent à des séquences à un même niveau de hiérarchie¹.

Nous avons étudié un certain nombre de cas dans le cadre des dialogues d'itinéraires ([Muller et Prévot, 2008](#)), et montré que cela pose la question du rôle sémantique de la paire question-réponse, au moins dans le cas d'un dialogue orienté par une tâche, et sans doute dans des cas plus généraux. Une façon de résoudre cette tension est de considérer qu'il y a toujours un topique global dans un dialogue, et que celui-ci consiste en un ensemble de sous-topiques emboîtés, dont certains peuvent avoir un certain ordre, ce qui correspond plus ou moins à la structure de "chaînes de topiques" de [Polanyi \(1988\)](#). Dans un dialogue orienté tâche, ce topique est la tâche elle-même, qui peut être divisée en sous-tâches, qui ont parfois un ordre à respecter. Finalement, on doit conclure que ces sous-tâches, introduites par des questions, doivent pouvoir être reliées au contexte de façon coordonnante (satisfaction-précédence).

5.2 Gestion du fonds commun conversationnel

L'analyse des structures rhétoriques discursives m'a mené à la relativisation de l'importance de la modélisation des croyances et intentions, pour se focaliser sur l'établissement du partage

1. Rappelons que ce typage des relations, similaire à la distinction de ([Grosz et Sidner, 1986](#)) entre relation de satisfaction-précédence et relation de dominance, impose des contraintes sur la continuation du discours, notamment du point de vue de l'accessibilité des référents aux reprises anaphoriques. Or la paire résolue question-réponse est considérée comme attachée au contexte de façon subordonnante, donc plus bas dans la hiérarchie.

d'information. Dans cette perspective, il suffit de modéliser les actes langagiers dans ce qu'ils ont de social, suivant en cela les approches dans la tradition de (Hamblin, 1970) qui préfèrent insister sur les engagements entre locuteurs, c'est-à-dire les propositions auxquelles leurs actes de parole les tiennent s'ils respectent une certaine cohérence. La notion de fonds commun de la conversation est alors réduite à l'ensemble des propositions sur lesquelles les deux locuteurs sont engagés au cours de la conversation. La modélisation de l'échange revient donc en priorité à la question de la dynamique de ces engagements et aux transactions langagières qui mènent à leur établissement.

Nous avons tenté dans cette perspective de rapprocher ce genre d'approche socio-linguistique avec un cadre sémantique tel que celui de la SDRT, en collaboration avec Nicolas Maudet et Laurent Prévot. Si une approche rhétorique du dialogue peut commencer avec une représentation unique censée coder le fonds commun de façon sémantique, cela ne peut suffire dans une approche centrée sur la notion d'engagements. Il faut entretenir deux jeux de propositions et de relations entre elles pour chaque locuteur et expliquer les processus d'évolution de ces engagements au cours du dialogue. L'engagement tient à ce qui est public, et ne suit pas nécessairement les croyances des locuteurs, qui deviennent ici non pertinentes².

Dans une approche rhétorique, ce qui caractérise un acte de parole, sa fonction, se retrouve dans le lien avec un autre énoncé. Si on reprend le jeu de relations proposé dans (Asher et Lascarides, 2003), il faut alors définir l'apport d'un acte relié au contexte par une relation en détaillant ce qui constitue une introduction d'un nouvel engagement, une révision, ou une acceptation d'une proposition en suspens.

Le critère le plus pertinent pour distinguer les phénomènes nous a semblé être la véridicalité d'une relation rhétorique : ce qu'elle implique sur la véracité des propositions reliées, assumée par le locuteur qui les a formulées. Pour une relation véridique, où les deux propositions reliées sont affirmées comme vraies, les engagements du locuteur doivent inclure les propositions. Les relations dialogiques sont typiquement non-véridiques : une question qui élabore un point de la discussion présuppose seulement la véracité de la première proposition, la question seule ne correspondant pas à une proposition. Une correction d'un locuteur par un autre met en cause un engagement précédent, et engage alors sur le contraire. Nous avons explicité tous les cas de relation en fonction de cet éclairage dans (Maudet et al., 2006).

On peut l'illustrer avec un exemple classique (Asher et Lascarides, 2003; Kreutel et Matheson, 2002) :

- (5.1) (A_{1a}) John went to jail. (A_{1b}) He was caught embezzling funds.
 (B_{2a}) No. (B_{2b}) He went to jail (B_{2c}) because he was convicted of tax evasion.
 (A_3) oh... OK.

Et l'évolution prévue des tableaux d'engagement \mathcal{CS}_A et \mathcal{CS}_B des locuteurs A et B est alors présentée table 5.1. On note \mathcal{DS} la structure discursive rhétorique interprétée, que nous avons limitée pour l'exemple aux liens entre actes de paroles (π_{1a} est par exemple l'acte de parole

2. Notons que de nombreux travaux logiques explorent cette distinction sous l'étiquette d'annonce publique (public announcements), ou de logique d'établissement (grounding), cf par exemple (Gaudou et al., 2006).

Tour	\mathcal{DS}	\mathcal{CS}_A	\mathcal{CS}_B
(A _{1a})	π_{1a}	$+\mathcal{F}_{1a}$	
(A _{1b})	$+\pi_1 : \text{expl}(\pi_{1a}, \pi_{1b})$	$+\mathcal{F}_{1b}, +\text{expl}(\pi_{1a}, \pi_{1b})$	
(B _{2a})	$+\pi_A : \text{rejet}(\pi_1, \pi_{2a})$		
(B _{2b})	$+\pi_B : \text{ack}(\pi_{1a}, \pi_{2b})$		$+\text{ack}(\pi_{1a}, \pi_{2b}), +\mathcal{F}_{1a}$
(B _{2c})	$+\pi_C : \text{expl}(\pi_{2b}, \pi_{2c})$ $+\pi_D : \text{corr}(\pi_A, \pi_{2b})$		$+\mathcal{F}_{2c}, +\text{expl}(\pi_{2b}, \pi_{2c})$ $+\neg\text{expl}(\pi_{1a}, \pi_{1b})$ $+\text{corr}(\pi_1, \pi_2)$
(A ₃)	$+\text{ack}(\pi_2, \pi_3)$	$+\text{ack}(\pi_2, \pi_3)$ $+\mathcal{F}_{2b}, +\mathcal{F}_{2c}$ $+\text{corr}(\pi_1, \pi_2)$ $\oplus \neg\text{expl}(\pi_{1a}, \pi_{1b})$	

TABLE 5.1 – Évolution de tableau conversationnel à base d’engagements, sur l’exemple (5.1). Les abbréviations *expl*, *ack*, *corr* dénotent les relations d’explication, d’accusé de réception et de correction. Les opérateurs $+$ et \oplus désignent l’ajout simple d’information ou l’ajout avec révision.

correspondant à l’énoncé (A_{1a})), les contenus propositionnels étant simplement listés (notés \mathcal{F}_i).

L’intérêt ici est de faire l’intégration de contraintes conversationnelles dans un cadre sémantique. Une étape à considérer ensuite est de rapprocher un peu plus les approches conventionnelles au niveau structurel supérieur, où les contraintes d’enchaînement sont souvent modélisés par des protocoles ou “jeux de dialogue” cf. (Mann, 1988). Plus précisément on peut penser que les éléments de la cohérence thématique dans une approche rhétorique sont les topiques, qui fournissent une sorte de synthèse de certaines séquences d’actes dialogiques. Les jeux de dialogue gèrent l’interaction d’une façon un peu similaire en posant des contraintes sur les règles d’enchaînement cohérents. Les définitions de tels jeux font intervenir des séquences d’actions spécifiques et un thème, par exemple voir (Lewin, 2000).

Ces travaux doivent être vus comme une étape de modélisation nécessaire pour spécifier la représentation sémantique d’un échange, avec comme objectif une conversation guidée par une tâche particulière. Pour l’instant situés à un niveau assez théorique, ils forment une partie d’un projet plus large de mise à l’épreuve pratique des théories dialogiques soit entre agents conversationnels, soit dans des cas de dialogues humains avec un intermédiaire automatique, comme dans le cadre de négociation multi-agents³.

3. Ceci fait l’objet d’une European Research Council advanced grant, le projet Stac, dirigé par Nicholas Asher, auquel je suis associé.

5.3 Rôle de l'acquiescement dans l'évolution du fonds conversationnel

A un niveau moins général, dans le cadre des travaux de thèse de Laurent Prévot, nous nous sommes restreints à un type de dialogue particulier, les dialogues d'orientations ou d'explication d'itinéraire, pour une étude empirique.

Afin de disposer de données contrôlées, nous avons procédé à la collecte de conversations dont la préoccupation est l'explication d'un itinéraire d'un point à un autre dans un espace urbain⁴. Pour collecter un corpus de cette nature, deux groupes de sujets étaient situés à deux endroits différents de la ville, mais accessibles à pied l'un de l'autre en moins de vingt minutes. Chaque membre du premier groupe (le "fournisseur", noté F par la suite) devait alors expliquer par téléphone à un membre du deuxième groupe (le "receveur", noté R par la suite) comment venir à l'endroit où se trouvait le premier groupe. Les conversations étaient enregistrées puis transcrites. Cette tâche est proche des expériences Maptask (Anderson et al., 1991), où deux locuteurs doivent s'accorder sur un itinéraire décrit graphiquement sur un support papier, mais où les versions données à chaque locuteur sont légèrement différentes, et donnent donc lieu à des négociations sur le chemin à suivre.

Les annotations en actes de discours de projets de collecte typiques (Core et Allen, 1997; Carletta et al., 1996) ignorent généralement la dimension relationnelle des énoncés, pour parvenir à des étiquetages de la nature suivante, par exemple sur un extrait de notre corpus :

(5.2) — Extrait du Dialogue 2.9 —

(F_{11.1}) t'as pas l'air branchée trop bars – *commenter*

(R_{12.1}) euh non – *rejet*

(R_{12.2}) mais je connais pas très bien Toulouse – *informer*

(F_{13.1}) ah ouais d'accord – *accusé de réception*

(F_{13.2}) donc les Carmes tu vois où c'est ? – *demander de l'information*

(R_{14.1}) oui – *inform*

(F_{15.1}) bon ben voilà – *accusé de réception*

(F_{15.2}) donc là tu continues sur sur cette rue – *instruction*

(F_{15.3}) et tu arrives aux Carmes – *instruction*

On peut voir ces schémas comme un peu trop restrictifs d'un point de vue sémantique, car chaque énoncé est vu isolé de son contexte dialogique⁵. Nous avons proposé un niveau d'annotation intermédiaire pour relier ces approches par actes de discours et les relations rhétoriques d'une théorie plus sémantique. Classiquement, les annotations d'actes ont une dimension orientée vers la conversation (question, réponse) et une dimension liée à la tâche que réalisent les locuteurs, ici la construction d'un itinéraire, et qui sera spécifiée précisément à partir du contenu propositionnel lié aux itinéraires et leur description spatiale et temporelle. Ainsi les données collectées

4. Disponible sur <http://crdo.fr/crdo000742/fr>.

5. Même si Core et Allen (1997) admet deux catégories d'actes, "vers l'avant" ou "vers l'arrière", pour indiquer qu'un acte est toujours relié à un autre dans l'historique, les arguments ne sont pas reliés.

permettent d'étudier comment interagissent ces différents niveaux, les informations spatiales, et de suivre l'établissement du fonds commun par la conversation.

Denis et Briffault (1997) donnent cette définition d'une description d'itinéraire, quand elle est produite de façon purement monologique par un "émetteur" qui explique à un "récepteur" :

1. localisation du récepteur au point de départ, et orientation ;
2. démarrage de la progression (instruction) ;
3. introduction d'un repère, qui indique la fin d'une étape ;
4. réorientation du récepteur ;
5. répéter 2-4 jusqu'à l'arrivée.

L'étude des dialogues collectés a permis de raffiner les types d'actions propres à la tâche d'orientation définis par Denis (1997), qui comprenait des prescriptions avec repère(s) explicite(s), prescriptions sans repères, introductions de repères, descriptions de repères. Nous avons ajouté les actes qui apparaissent en plus dans les dialogues, positionnement et raffinement d'une description déjà introduite. Par ailleurs, certains actes sont uniquement des commentaires sur la tâche, qui ne fournissent pas d'indication pour la faire avancer (comme "c'est facile, tu verras").

Pour la dimension dialogique, un certain nombre de typologies sont envisageables Carletta et al. (1996); Traum (1994); Jurafsky et al. (1997), mais on peut reprendre l'exemple 5.2 pour illustrer ce que l'on recherche. Nous nous sommes attachés à étudier le lien que l'on peut faire entre le niveau linguistique d'un énoncé et l'acte qu'il accomplit dans la globalité de la conversation. Notre corpus est de taille limitée (il comprend une vingtaine de dialogues, 7000 mots et environ 1200 actes de paroles), et sert de point de départ à une étude qui se veut plus large. Dans un premier temps, notre objectif était d'étudier les processus par lesquels deux locuteurs s'accordent sur les référents introduits dans la conversation. Dans le cadre d'une explication d'itinéraire, ceci revient à s'accorder sur la localisation d'un certain nombre de référents géographiques. Peu de travaux ont étudié les processus de partage d'information à un niveau plus fin que la proposition logique. Dans cette perspective propositionnelle, les référents sont absents et leur ancrage n'est pas considéré, ce qui occulte un certain nombre de processus dialogiques. Nous considérons que c'est une information essentielle à traiter qui peut servir de guide pour l'interprétation structurelle rhétorique.

Le dialogue suivant illustre notre perspective :

(5.3) — Dialogue 1.1 —

(F_{7.1}) heum tu vois le rond point de St-Michel ? – *introduction d'un référent, question*

(F_{7.2}) là où il y a le commissariat – *localisation, information*

(R_{8.1}) ok ouais – *réponse*

(F_{9.1}) d'accord donc – *accusé de réception*

(F_{9.2}) tu vas par là – *instruction utilisant un repère*

(F_{9.3}) tu continues tout droit tout droit tout droit – *instruction sans repère*

Le cas prototypique présenté ci-dessus est constitué de l'introduction d'un repère, une négociation (ici réussie) sur ce référent, et la poursuite du dialogue en considérant le référent comme établi (partagé).

5.3. RÔLE DE L'ACQUIESCEMENT DANS L'ÉVOLUTION DU FONDS CONVERSATIONNEL

Parmi les actes étudiés, le cas des acquiescements est particulièrement significatif, puisqu'il constitue une grande partie des actes de notre corpus (Muller et Prévot, 2003).

Les travaux centrés sur les rôles que peuvent avoir les assertions, les questions et leurs réponses dans une conversation sont légion et les schémas d'annotation du dialogue (Core et Allen, 1997; Carletta et al., 1997) prennent grand soin à distinguer leurs fonctions. Par ailleurs, l'étude des interactions entre sémantique et pragmatique dans les enchaînements de questions et de réponses a reçu beaucoup d'attention au niveau théorique, (Asher et Lascarides, 1998; Ginzburg et Sag, 2001), on l'a vu aussi à la section 5.1. Il y a eu moins d'intérêt pour le rôle des tours de parole chargés de s'assurer que l'information échangée est reçue et interprétée correctement : confirmations, accusés de réception divers forment une part importante des énoncés d'une conversation (Allwood et al., 1992), et (Clark, 1996) a montré l'importance de considérer qu'il existe plusieurs niveaux d'acceptation d'un acte de parole. Le travail central de Poesio et Traum (1998) sur l'établissement a établi avec minutie le rôle de ces actes pour décider le status de l'information échangée, acceptée ou en discussion.

Les fonctions de feedback sont associées à l'établissement de deux sortes d'objets : d'une part des propositions (et leur valeur de vérité) et d'autre part des référents (et leur identité). On distingue généralement quatre phénomènes :

- l'établissement (grounding) est le signal de la compréhension d'une proposition (indépendamment de l'acceptation de sa vérité)
- l'acceptation (agreement) par contre caractérise l'acceptation de la vérité de la proposition énoncée, au moins pour le but courant de la conversation
- l'ancrage (d'un référent) est l'attribution d'une ancre interne pour un référent mis en jeu dans la conversation (un référent commun)
- la clôture est la fermeture d'une partie du discours

Déterminer la fonction exacte d'un énoncé de feedback donné n'est pas toujours une tâche facile. Dans notre étude, nous avons étudié plusieurs marqueurs courants qui apparaissent dans des énoncés d'acquiescements, et des facteurs mis en jeu qui permettent d'identifier leur fonctions. Un des facteurs examinés était notamment la fonction de l'énoncé précédent dans la structure globale. Nous avons regroupé les fonctions déjà mentionnées relatives à la tâche d'orientation en : (1) introduction de repères et descriptions de repères, (2) instructions, (3) autres et (4) considéré également les tours de gestion de la conversation, souvent déjà des marques de feedback. Ces distinctions sont importantes pour différencier l'acceptation de l'ancrage, ce dernier étant principalement lié à la gestion des repères. Les instructions semblent plus naturellement destinées à être acceptées ou rejetées.

Les marqueurs de discours sont un sujet d'étude bien balisé (Schiffrin, 1987) mais les marqueurs spécifiques de feedback sont plutôt ignorés, si l'on excepte (Gardner, 2001; Bangerter et Clark, 2003; Fetzer et Fischer, 2006). Sur le français, Roulet et al. (1985) et Rossari et al. (2004) ne sont pas focalisés sur les problèmes de feedback. Nous avons donc isolé un jeu de marqueurs à partir de notre corpus, et nous avons ajouté dans les énoncés pertinents ceux qui reprennent tout ou partie d'un énoncé précédent, dont la pertinence est soulignée par Walker (1992).

La corrélation entre les marqueurs et les rôles de feedback a été testée statistiquement, en mettant

en rapport chaque marqueur et l'énoncé qu'il cible.

L'hypothèse est que chaque marqueur remplit une fonction distincte dans le processus d'établissement, et a des chances de cibler des énoncés de type différent par rapport aux notions de feedback introduites.

Les résultats ont montré que l'hypothèse était partiellement valide : la plupart des marqueurs sont associés significativement à certaines classes d'actes sémantiques. Ceux qui impliquent des référents ("ouais", "d'accord", "je vois"), ceux qui correspondent plutôt à des instructions ("ouais", "mmhm") et ceux qui ciblent plutôt des feedbacks ("voilà", "bon"). Il reste à voir si ces groupes peuvent recouper les niveaux d'établissement présentés plus haut. Les marqueurs ciblant plutôt des feedbacks sont par exemple des bons candidats pour détecter la clôture d'un sous-échange.

L'asymétrie entre les deux participants aux dialogues permet d'étudier le lien entre chaque marqueur et l'initiative des locuteurs dans le dialogue (même si des inversions étaient localement possibles). Evidemment, le receveur produit bien plus de marques de feedback que le fournisseur, mais on a pu observer que la répartition n'était pas homogène, certains marqueurs étant au contraire autant ou plus employés par le fournisseur, comme 'voilà' et 'bon', ce qui tend à confirmer leur rôle dans la clôture d'une sous-partie du dialogue, ce qui est un indice important dans l'analyse.

Cette étude empirique venait en complément des travaux structurels présentés auparavant, pour indiquer comment on peut spécifier des propositions théoriques et utiliser des indices surfaciques pour construire les structures visées. Les indices lexicaux ne sont bien sûr pas les seuls pertinents pour cet objectif, et nous avons étendu ceci en prenant en considération d'autres facteurs qui permettent d'aider à trouver la structure du dialogue.

Plus généralement, des indices lexicaux et syntaxiques relativement simples permettent de prédire les types d'actes de dialogue dans le cadre bien contraint des conversations considérées, ce qui a constitué la matière du DEA d'Ana Hernandez ([Hernandez, 2006](#)) que j'ai supervisé, où des techniques de classification permettent d'apprendre à étiqueter les actes, à la manière de [Jurafsky et al. \(1998\)](#), avec des résultats satisfaisants, quoique difficilement généralisables à un autre type de dialogue.

5.4 Acquiescements, intonation et structure topicale

Les indices lexicaux présentés à la section précédente permettent de contraindre les enchaînements possibles dans le dialogue et donc d'aider à l'interprétation de l'échange et de l'établissement d'information. Nous avons vu aussi à la section 5.1 que l'on pouvait organiser un dialogue en termes de hiérarchie de topiques et de sous-topiques, au moins dans le cas d'un dialogue orienté par une tâche. Un autre indicateur potentiel de la nature des actes de parole et partant, de la construction de cette structure, est la présence de marques prosodiques. Par exemple, la montée finale sur un tour de parole est considérée comme un indicateur classique de questions en général. On peut alors se poser le problème de chercher les marques prosodiques

pour les autres fonctions étudiées ici, et leur rapport avec les sous-parties de la conversation. On peut vouloir attester la structure de topique étudiée plus haut en cherchant des indices réguliers dans la prosodie, notamment à travers l'étude encore une fois des acquiescements.

En collaboration avec Laurent Prévot et Marie Šafářová, j'ai entrepris une étude empirique qui cherche à établir le sens des montées intonatives dans le dialogue pour répondre à ces interrogations, et compléter l'étude des facteurs pertinents de l'organisation topicale (Šafářová et al., 2005). Si des études empiriques (Grundstrom, 1973; Fónagy et Bérard, 1973) confirment qu'il y a une corrélation entre des contours montants ou descendants et les questions et les assertions, elles montrent aussi qu'une intonation montante n'est pas toujours une intonation de question. Notre objectif était de tester s'il y avait une corrélation avec des fonctions discursives, et d'étudier la nature de celles-ci.

En dehors des questions, le corpus a aussi été annoté avec les types d'actes de dialogue caractérisant un dialogue d'explication d'itinéraires, suivant Prévot (2004), cf section 5.3. Le corpus présenté section 5.3 a donc été repris, avec les types d'actes de dialogue caractérisant un dialogue d'explication d'itinéraires, en ajoutant une annotation des questions polaires, qui attendent une réponse oui/non, et sont souvent proches d'assertions simples syntaxiquement dans l'oral courant.

Nous avons alors évalué statistiquement la présence ou l'absence d'une montée en fonction des catégories principales d'actes de dialogue, ici acquiescement, instruction, information, question, réponse à une question, combinée à l'introduction ou non d'un nouveau repère.

Peu de corrélations convaincantes sont apparues entre les types d'actes et la présence d'une montée intonative, sauf pour les instructions introduisant un nouveau repère, associées à des montées, les réponses à une question (absence de montée), et les questions polaires avec un nouveau référent (présence d'une montée).

Pour évaluer le lien entre ces montées et la structure de discours à travers l'organisation topicale, nous avons aussi annoté le corpus avec une forme simplifiée d'information structurelle, qui reprend seulement des ouvertures ou fermetures de topiques.

La structure résultante fournit une hiérarchie de sous-dialogues, incluant des fermetures de sous-topiques qui font revenir à un sujet en suspens bien plus haut dans la hiérarchie que l'énoncé précédent.

Les résultats montrent que les montées sont associés significativement aux ouvertures de topiques, et leur absence à une fermeture. Les données étaient par contre probablement insuffisantes pour trouver un lien entre ces caractéristiques et le rôle du locuteur entre fournisseur ou receveur d'instructions, et aucun biais n'a été détecté. La compétence par rapport au topique courant d'une discussion semblait en effet a priori reliée aux définitions de jeux de dialogues selon Kowtko (1996).

Il est difficile d'abstraire une telle étude du genre choisi pour l'expérience, notamment parce que le niveau des règles d'interaction est défini essentiellement par type de dialogue, et les contraintes entre actes de dialogues également. Tout au plus peut-on prétendre à une certaine généralité sur les dialogues impliquant la réalisation conjointe d'une tâche.

5.5 Perspectives

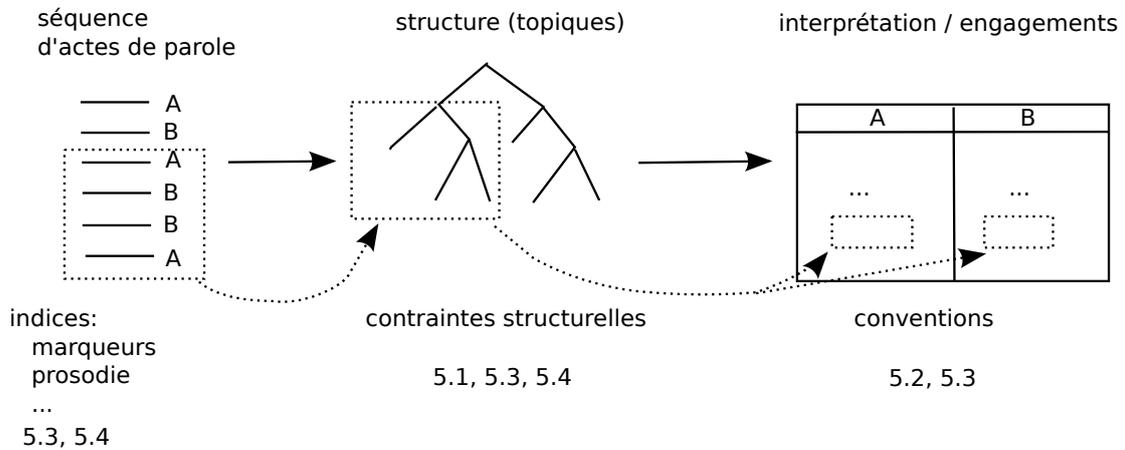


FIGURE 5.1 – Synthèse des contributions sur les structures dialogiques, en prenant comme point d'entrée une séquence de tours de parole et de leurs producteurs. L'interprétation part de la séquence d'actes de parole pour construire une structure en topiques et produire une interprétation en termes d'engagements. En bas sont indiquées les facteurs qui nourrissent chaque partie et les sections concernées.

La figure 5.1 synthétise l'approche que je considère pour la modélisation de la conversation, dans le cadre d'une tâche : depuis une séquence de tours de parole, l'interprétation consiste à construire une structure des enchaînements regroupés en topiques, et à utiliser cette structure pour interpréter l'évolution de certains états mentaux ou attitudes propositionnelles des participants (ici les engagements propositionnels, mais on pourrait généraliser à d'autres types d'états en adaptant les contraintes). D'un point de vue formel, j'ai essayé d'ajuster les modèles de structures comme ceux proposés par la SDRT pour les rendre plus adéquats à certains phénomènes, comme les séquences de question(s)-réponse(s)-acquiescement(s). En parallèle, ces structures peuvent être utilisées pour la représentation de la dynamique de la conversation, pour faire le pont entre une gestion du dialogue par état et une approche sémantique, guidées par des conventions socio-linguistiques. D'un point de vue empirique, ces études sont soutenues par l'étude précise sur corpus de certains indices, des marqueurs lexicaux et des phénomènes prosodiques, et de leurs corrélations avec les structures modélisées.

Dans un contexte de traitement des langues, le dialogue a d'abord été associé à des questions d'interface homme-machine, où la machine devait comprendre les interventions humaines, dans le cadre d'une tâche à réaliser (par exemple une recherche d'information interactive), et gérer l'évolution du dialogue pour parvenir au but recherché. Dans cette perspective les travaux présentés concernent essentiellement le deuxième aspect, qui nécessite des modèles de l'interaction humaine réalistes. Un autre aspect qui intéresse le TAL est l'analyse de conversations et la recherche dans celles-ci d'informations spécifiques. Par exemple pour faire des synthèses de

réunions de travail (Shriberg et al., 2004), analyser la validité d'argumentations (Georgila et al., 2011), et en étendant un peu le concept de dialogue au cas d'échanges écrits multi-locuteurs, à l'analyse de fil de courriers électroniques, comme le célèbre corpus Enron, cf (Prabhakaran et al., 2012), ou encore la recherche de réponses dans des forums techniques (Wang et al., 2011). Ici c'est l'analyse des liens entre parties du dialogue, par exemple l'association de réponses aux questions, et l'établissement de l'accord sur les réponses qui peut informer des processus de recherche, et peut être vu comme une spécialisation des problématiques présentées au chapitre 4.

Les dialogues envisagés sont tous ici fortement restreints par la tâche considérée, même si le modèle sous-jacent ne l'est pas. Une avenue de recherche prometteuse consiste donc à tester ce genre de méthodologie sur des dialogues contraints différemment, par des argumentations ou des négociations sous-jacentes. Il manque par ailleurs aux travaux de cette section à être véritablement intégrés dans une approche opérationnelle qui permettrait d'évaluer leur cohérence, les travaux présentés ici étant nettement moins computationnels que les chapitres précédents. Une partie de ce travail consisterait à adapter le travail d'analyse discursive du chapitre précédent, une autre à insérer l'analyse dans un gestionnaire de dialogue qui prendrait en charge un agent dans les types de dialogue considérés. Ce travail est en cours dans le cadre d'un projet plus large dirigé par Nicholas Asher, financé par une ERC advanced grant (STAC) déjà mentionnée. Une partie des enjeux de ce projet vis-à-vis de la modélisation des interactions dialogiques a été présentée dans (Afantenos et al., 2012). L'adaptation de ces approches sur l'analyse de fils de discussion dans des forums est aussi une piste prometteuse qui combine analyse automatique et modélisation de conversations au sens large. Cela avait fait l'objet d'un dépôt de projet ANR autour de la recherche d'informations sur des forums techniques (en 2011), qui n'a pas été retenu.

Lexique et Discours

Les analyses sémantiques présentées jusqu'ici font la part belle aux aspects pragmatiques, en soulignant tout de même des ensembles d'éléments lexicaux qui initient le processus d'interprétation : lexique propre aux domaines temporel et spatial ou marqueurs de structure, comme les acquiescements dans le dialogue, ou les indicateurs de relations rhétoriques dans des textes. Je reviens maintenant sur la dimension lexicale et son rôle dans l'organisation discursive, qui dépasse l'étude de marqueurs dont la fonction principale est de signaler cette organisation. Ce chapitre revient sur l'importance des liens lexicaux comme indices, éventuellement secondaires, de la cohérence textuelle.

Les chapitres 4, 5 présentaient une vision du discours où celui-ci est un ensemble de parties et de relations fonctionnelles entre ces parties, qui se superposent, voire s'opposent à la linéarité de la présentation du texte. Les relations rhétoriques, reflétant les intentions des producteurs du discours, et les conventions pragmatiques, décrivent la cohérence entre les parties. A l'intérieur de cette organisation, on trouve des ruptures locales avec l'ordre du texte, comme des liens anaphoriques ou temporels à longue distance, ou bien des changements d'échelle comme l'élaboration de certains éléments.

On peut rapprocher deux types d'établissement de la cohérence en question, qui se reflètent dans l'usage lexical : une première forme qui emploie des marqueurs spécifiques des fonctions discursives et de l'organisation du texte (par exemple "parce que" est une marque explicite d'explication entre deux assertions), et une deuxième forme qui traduit la cohérence par les liens sémantiques entre les segments, l'organisation thématique, qui se retrouve aussi dans les liens entre éléments lexicaux (une explication peut être ainsi implicitement indiquée par des liens causaux typiques entre deux items lexicaux dénotant des événements). La première forme est plus explicite quand la seconde, la cohésion lexicale, est plus diffuse.

Les deux formes de cohésion peuvent se voir comme l'expression de l'organisation en parties distinctes, c'est-à-dire du point de vue de la segmentation du discours (cf section 6.1), ou comme l'expression des liens entre ces parties (cf section 6.2).

En poursuivant ces questionnements du lexique dans une perspective discursive, j'en suis donc venu à considérer de plus en plus l'importance des *liens* entre éléments lexicaux comme une question en soi. Mes premiers travaux se rapportant à la sémantique lexicale¹ se situaient dans une approche classique de décomposition du sens en éléments primitifs, définis par leurs pro-

1. Mentionnés au chapitre 1, (comme Muller et Sarda, 1999).

priétés formelles où les liens entre éléments sont induits indirectement par leur sémantique. Elle s'oppose à une sémantique ici plus relationnelle. La recherche de primitives sémantiques qu'il faut spécifier pour ensuite coder le lexique comme une structure formelle utilisant ces primitives laisse alors la place à une recherche de dimensions sémantiques sur lesquelles on peut projeter des relations lexicales pour rendre compte du sens.

Je reviens à la section 6.3 sur mes travaux qui ont mis à l'épreuve quelques façons de mettre au jour ces informations, et les directions sémantiques qu'elles indiquent.

6.1 Lexique et segmentation discursive

J'ai présenté jusqu'ici une vision de la cohérence d'un texte essentiellement relationnelle : un discours est un ensemble de parties plus ou moins structurées reliées par des intentions discursives plus ou moins précises. Les parties sont éventuellement imbriquées, ce qui donne une structure proche d'un arbre, avec éventuellement des liens supplémentaires, et des structures complexes qui peuvent recouvrir des unités élémentaires. On a vu qu'au minimum une approche structurée distingue entre deux types de fonctions, une relation de dominance et une relation de séquence (figure 4.1). La première fonction relie un segment et ses descendants dans une approche arborescente, ce qui est une forme de regroupement de parties. La deuxième fonction relie des segments ayant un ancêtre commun dans la hiérarchie. Dans cette optique, le lexique intervient par des marqueurs spécifiques qui signalent l'organisation de façon explicite, en distinguant des parties, et en indiquant la fonction qui les relie. Ces signaux peuvent être d'ailleurs ambigus entre plusieurs fonctions. Mes travaux sur les marques d'acquiescement du chapitre 5 se placent dans cette perspective.

Il existe une tradition pragmatique complémentaire qui insiste sur les régularités d'usage lexical dans un discours cohérent, et on parle alors de cohésion lexicale, un terme popularisé par Halliday et Hasan (1976). Dans cette perspective, ce ne sont pas seulement les liens entre parties qui sont la source de l'attention mais aussi les ruptures entre des parties différenciées par leurs thèmes, dont une indication est le vocabulaire utilisé et les changements au cours du texte. La nature des liens sémantiques qui permettent de parler de cohésion lexicale reste théoriquement assez floue. Chez Halliday la cohésion consiste en l'usage de répétitions lexicales (incluant synonymes et hyperonymes) et de termes en collocations, ce qui recouvre un éventail large et assez vague d'associations régulières entre des termes dans des contextes d'usage.

Le lien thématique peut s'exprimer en effet de façons diverses, illustrées dans l'exemple figure 6.1, où l'on a projeté des liens entre termes de distributions similaires.

Dans ce texte on trouve des répétitions (*avion, tour, détourné, New York, etc*) des coréférences (*twin towers, tours jumelles*) et parmi les termes associés il y a des paires d'hyperonymes/hyponymes plus ou moins proches (*tour/immeuble, impact/événement*). On y trouve aussi des éléments lexicaux sémantiquement reliés : *économique/commercial, frapper/endommager*, même s'ils sont reliés sans doute par des sens différents de ceux qu'ils réalisent dans cet exemple particulier (*ligne/série*), ou de façon contingente (*commercial/militaire*). Cet exemple ne montre par ailleurs que les associations les plus fortes, et on peut potentiellement

Les attentats du 11 septembre 2001 frappèrent New York et Washington à l'aide d'avions de ligne détournés, dans la matinée du jour éponyme. Le terme regroupe une série d'évènements synchronisés qui se déroulèrent dans le nord-est des États-Unis d'Amérique : trois avions commerciaux (sur quatre détournés) furent précipités sur des immeubles représentatifs de la puissance américaine, économique pour les tours jumelles du World Trade Center à Manhattan, New York, et militaire pour le Pentagone, siège du ministère de la Défense des États-Unis, à Washington. Les "Twin Towers" s'effondrèrent spectaculairement moins de deux heures après les impacts, ainsi qu'une troisième tour proche dite WTC7, le Pentagone fut endommagé.

FIGURE 6.1 – Exemple de liens lexicaux possibles projetés dans un paragraphe.

trouver un nombre immense de liens plus ou moins lâches (ce qui aurait rendu la figure illisible). Pour être une notion prédictive, et ne pas devoir accepter n'importe quelle suite de phrases comme lexicalement cohésive, il faut donc pouvoir constater des contrastes de cohésion lexicale, indicateurs de ruptures thématiques dans un texte.

L'étude des relations lexicales comme indice de l'organisation thématique a fait l'objet de la thèse de Clémentine Adam, que j'ai co-encadrée avec Cécile Fabre et dont les directeurs sont Marie-Paule Péry-Woodley et Nicholas Asher, dans le cadre d'un projet interdisciplinaire du PRES toulousain, Voiladis.

Le point de départ de ce travail est une base de similarité lexicale pour le français, les "voisins de wikipedia"², constituée par Didier Bourigaut et Frank Sajous. Cette base est dérivée d'une analyse distributionnelle dans la lignée de (Lin, 1998), et rapproche des termes qui se retrouvent significativement dans les mêmes contextes syntaxiques³. Cette similarité entre termes est utilisée comme indice potentiellement bruité de la cohésion lexicale, et la problématique est de confirmer son utilité pour prédire des ruptures thématiques ou signaler des liens rhétoriques implicites. Dans le premier cas, on retrouve des préoccupations liées à un problème spécifique d'analyse du discours qui est la segmentation thématique. La segmentation thématique (ST) a pour but le découpage linéaire d'un texte en unités présentant une cohérence autour d'un sujet. C'est une forme simplifiée des analyses que l'on vise en général sur le discours, où sont ignorées les structures internes des unités thématiques⁴ et où il n'y a pas de relation entre les unités autre que la succession. Elle repose sur l'hypothèse que les textes s'organisent selon un plan thématique, chaque thème se singularisant par le recours à un vocabulaire suffisamment spécifique pour le distinguer des autres.

2. <http://redac.univ-tlse2.fr/applications/vdw.html>

3. Je reviens sur les différents moyens de calculer une similarité lexicale à la section 6.3, cette question étant transversale à la problématique de la cohésion discursive.

4. Peu d'auteurs vont au-delà d'un niveau de découpage, mais on peut signaler l'approche récursive de Eisenstein (2009).

La grande majorité des approches de ST se fondent sur la méthode initiée par [Hearst \(1997\)](#) : le texte est divisé en blocs contigus correspondant à une unité fixée à l’avance (un nombre n de mots, de phrases ou de paragraphes), puis on définit une fenêtre glissante qui parcourt le texte linéairement et permet de calculer un score de similarité à chaque intersection entre blocs dans le texte. Une méthode de segmentation s’attache alors à trouver les points où la similarité présente des évolutions fortes, interprétées comme des indications de rupture de la continuité thématique. Une alternative à cette approche par pavage (*tiling*) est de supposer des thèmes sous-jacents qu’il s’agit de détecter : chaque unité du texte considéré est rapportée à un ou plusieurs thèmes, et la segmentation consiste à trouver ces thèmes⁵. Une fois les thèmes identifiés pour chaque unité de texte, les segments correspondent aux blocs d’unités contiguës partageant le même thème.

Au vu des approches sémantiques que nous avons décrites plus haut, fondée sur l’analyse de l’organisation rhétorique avec des fonctions liées à des intentions de communication plus précises, l’hypothèse d’une organisation fondée purement sur des distributions lexicales est évidemment très simplificatrice, et fait l’abstraction de particularités d’organisation propres à des types de textes différents. Il est possible que cette approximation soit effective dans des contextes spécifiés, et elle fournit un cadre algorithmiquement assez simple à mettre en oeuvre, ce qui a motivé de nombreux travaux. L’impact des types de textes sur la procédure de ST n’a pas été pris en considération par les travaux qui mettent en oeuvre cette tâche, au point que les mêmes algorithmes sont appliqués indifféremment à la tâche de segmentation de textes ou de délimitation de textes concaténés, pour lesquels les contrastes de vocabulaire sont évidemment bien plus marqués. Les expériences de ST qui sont menées sur des “vrais” textes le sont sur des types qui semblent se prêter intuitivement à cette approche – par exemple des exposés scientifiques ou des articles encyclopédiques comme chez [Chen et al. \(2009\)](#) –, sans qu’on cherche à déterminer la nature des textes qui sont adaptés à la tâche.

Nous avons testé et confirmé dans ([Adam et al., 2010](#)) l’hypothèse que les performances de la tâche de ST sont effectivement très sensibles au type de text considéré. Nous avons pour cela constitué deux sous-ensembles de textes issus de Wikipédia, en prenant le niveau des catégories définies dans l’encyclopédie comme critère de répartition. Un sous-corpus rassemble des textes consacrés à la description de pays, de villes et d’animaux, dont on sait qu’ils se prêtent généralement bien à une organisation thématique. La catégorie opposée réunit des biographies, dont l’organisation est typiquement temporelle, et des textes présentant des notions abstraites, des concepts, pour lesquels l’approche thématique semblerait moins adaptée.

On constate que l’hypothèse globale d’une différence entre les deux types de textes se vérifie assez nettement, quelle que soit la métrique considérée, et que les algorithmes de segmentation choisis sont meilleurs sur les textes du sous-corpus plus thématique, même si les variances sont importantes. Par ailleurs, alors qu’on observe des différences significatives entre les *baselines* et les algorithmes de segmentation sur l’expérience thématique, la différence n’est plus significative pour les textes du sous-corpus non thématique. Étant donnée la forte variance que nous avons

5. Les thèmes peuvent être prédits par des distributions lexicales différentes ([Chen et al., 2009](#); [Misra et al., 2008](#)) ou bien par des associations lexicales calculées à partir des textes, par exemple par un clustering en amont ([Ferret, 2007](#)).

observée sur les résultats, y compris sur le sous-corpus thématique, nous avons aussi évalué les résultats par catégories Wikipédia à l'intérieur des corpus. Les résultats des deux sous-corpus se reportent sur les catégories qui les composent, avec encore une fois des variances fortes. Il s'avère que notre découpage volontairement grossier *a priori* pourrait s'affiner – à condition de poser clairement les paramètres de ce que nous avons appelé pour l'instant le caractère thématique fort ou faible des textes –, mais qu'il semble valide.

Une hypothèse secondaire de ce travail était que les liens induits par similarité distributionnelle étaient plus informatifs et devaient avoir un impact sur l'évaluation globale. Concernant cet aspect, la proximité des scores sur les deux sous-corpus et les tests statistiques que nous avons utilisés n'indiquent pas de différences significatives.

L'évaluation des liens lexicaux par l'utilisation dans une tâche complexe doit donc être affinée en étudiant plus directement les rapports entre ces liens et l'organisation discursive, focalisée sur des aspects plus facilement testables.

6.2 Lexique et relations rhétoriques

On a vu que l'organisation discursive peut être traduite en termes de relations entre unités de discours, exprimant certaines intentions de l'auteur, au-delà de la sémantique des unités considérées isolément. Ces intentions peuvent être explicitées directement par des marqueurs dits discursifs, mais peuvent également être plus implicites, ce qui pose alors la question du rôle de la cohésion lexicale dans l'établissement de ces intentions par le lecteur.

Considérons quelques exemples typiques, comme celui d'une proposition expliquée explicitement par une autre :

(6.1) *Jacques vient d'être condamné à 10 ans de prison **parce qu'**il a fraudé aux élections municipales.*

Ou encore l'élaboration d'une proposition par une autre :

(6.2) *Tout concourt à ce que le procès dure longtemps. **Notamment**, le procureur ne lui fait pas de cadeaux.*

L'interprétation rhétorique peut aussi être induite par la seule juxtaposition en discours, et inférée par le lecteur à partir de la situation décrite :

(6.3) *Jacques vient d'être condamné à 10 ans de prison. Il a fraudé aux élections municipales.*
(explication)

(6.4) *Le procès dure depuis des semaines. Le procureur ne lui fait pas de cadeaux.*
(élaboration)

Dans ce cas là l'interprétation peut être plus ambiguë (notamment entre les notions d'explication et les autres relations plausibles), mais il est clair que l'information lexicale joue un rôle important. Dans les situations où les relations ne sont pas explicitées par un connecteur, les théories discursives supposent que le locuteur peut inférer la relation par des liens typiques entre événements : une *explication* intervient quand un événement est une cause possible d'un

autre (frauder/être condamné), une *narration* quand deux événements sont les éléments probables d'une séquence temporelle, sans causalité nécessairement évidente, une *élaboration* quand un événement peut être un sous-événement d'un autre. Une théorie comme la SDRT par exemple présuppose un ensemble de règles d'interprétation nécessitant une certaine quantité d'information de sens commun, déclenchées par les éléments lexicaux.

Dans un cadre sémantique tel que celui-ci, l'interprétation peut aussi se faire en intégrant des contraintes sémantiques en l'absence de marque explicite, en cherchant la fonction la mieux à même d'expliquer la cohérence d'un passage. Des contraintes évidentes sont données par exemple par des indices temporels, qui expriment certaines relations entre événements susceptibles de circonscrire les relations discursives possibles. Un lien de causalité entre deux événements, une séquence narrative impliquent des relations de précédence entre les événements mis en jeu. Une élaboration implique par contre des inclusions entre certains événements et sous-événements. Asher et Lascarides introduisent dans leur théorie un ensemble de règles sémantiques, semblables à des postulats de signification mais exprimés dans une logique révisable, qui relie comme ci-dessus des causes et des enchaînements possibles. Ceux-ci peuvent guider l'interprétation de diverses fonctions rhétoriques.

L'hypothèse sous-jacente est qu'un discours cohérent est constitué d'événements reliés, et qu'on ne peut expliquer la compétence d'un locuteur sans mettre en jeu sa connaissance du monde, à travers un ensemble de situations typiques, une approche dans la tradition des scripts de Schank. Si l'on voit le lien avec la branche de l'IA préoccupée de représentation des connaissances l'objectif paraît difficilement atteignable sans des avancées significatives dans cette branche même.

Ce goulet d'étranglement est cependant en train d'être contourné par les travaux sur l'extraction de relations lexicales, qui cherchent des associations typiques entre termes à partir de corpus de grande taille (Chklovski et Pantel, 2004). On retrouve une forme de cohésion lexicale, éventuellement susceptible de raffinements, comme support des liens discursifs. Les liens de voisinage distributionnel sont une forme encore mal dégrossie de ce genre d'information, où les liens ne sont pas typés, et où certaines associations accidentelles sont mal filtrées.

Il est quand même possible d'en évaluer l'intérêt, et Clémentine Adam a commencé l'étude du rôle de ces relations lexicales pour la caractérisation de relations implicites. La relation la plus courante dans le corpus Annodis, l'élaboration, est aussi une relation peu marquée explicitement, et forme le sujet de l'étude de (Adam et Vergez-Couret, 2010). Cette relation est très courante et importante parce qu'elle donne l'indication de la structure hiérarchique de base du discours.

Deux événements en relation d'élaboration doivent partager certains actants, et si les théories supposent que les événements doivent être des sous-types l'un de l'autre, on constate en pratique une diversité de liens lexicaux à l'oeuvre : synonymes, hyperonymes, ou simplement appartenance à un schéma ou un cadre narratif.

En l'absence de données annotées en nombre suffisant, l'étude s'est restreint à un échantillon de liens intra-sentenciaux, présélectionnés par la présence d'un marqueur d'élaboration potentiel mais peu précis, à savoir la présence d'un gérondif, comme par exemple :

(6.5) *[Les] villages contribuaient également à ce grand projet religieux en envoyant des vivres.*

emplacement du lien	relation	moyenne	±	échantillon	t-test / (ligne X)
même corpus	docs. diff.	40.0	2.0	2000	
même document	non reliés	61.5	4.26	2000	$p < 10^{-17}$ (1)
même document	reliés	73.7	4.86	1927	$p < 10^{-3}$ (2)

TABLE 6.1 – Comparaison de la cohésion lexicale ($\times 1000$) entre des unités reliées et des unités non-reliées rhétoriquement dans le corpus Annodis, soit dans le même document soit dans des documents distincts, avec la moyenne, l'intervalle de confiance à 95%, la taille de l'échantillon et un test statistique de significativité. La mesure de similarité est distributionnelle, utilise les noms, verbes et adjectifs et agrège par la méthode de Mihalcea.

L'étude montre une grande corrélation entre la présence de liens lexicaux entre les verbes principaux des deux unités de discours, encore renforcée si les objets sont eux-mêmes reliés, bien plus qu'en considérant les gérondifs seuls (60% environ des gérondifs seraient des élaborations, contre respectivement 80 et 95% pour les autres cas), mais le nombre d'exemples concernés est faible (moins de 2% du total des contextes extraits). Cela encourage néanmoins à étendre un peu ce genre de méthode où l'on peut combiner des associations lexicales et éventuellement d'autres traits informatifs (le gérondif en est un), afin de classifier des relations spécifiques. Il est sans doute nécessaire d'explorer l'utilisation non pas de la mesure globale d'association distributionnelle, mais d'explorer les relations syntaxiques porteuses d'informations pertinentes pour chaque relation. Cela rejoint notamment les travaux qui cherchent à améliorer l'utilité des thesaurus distributionnels, sur lesquels je reviens à la section 6.3.

Nous avons poursuivi cette étude en l'étendant aux autres relations, et en utilisant le corpus Annodis comme jeu d'évaluation pour le lien entre cohésion lexicale, du moins dans sa version introduite ici, et cohérence de discours. Pour cela nous avons testé plusieurs hypothèses, la première étant que des unités discursives reliées rhétoriquement sont plus susceptibles d'exhiber une bonne cohésion lexicale. Nous avons aussi voulu étudier la dépendance des fonctions discursives avec un niveau plus ou moins élevé de cohésion lexicale. La cohésion dont nous parlons n'est pour l'instant qu'une relation entre deux unités lexicales. Pour comparer à la cohérence discursive, l'échelle d'analyse doit être les unités élémentaires⁶, et il se pose alors la question de l'agrégation des liens entre les items lexicaux des deux propositions. Nous avons repris la proposition de (Mihalcea et al., 2006) qui pour comparer deux textes moyenne les similarités maximales entre deux mots des deux unités.

Les résultats de similarité entre diverses unités sont résumées table 6.1, validant bien la première hypothèse, et le détail par relation rhétorique présente dans le corpus Annodis est présenté figure 6.2. Il est difficile de conclure pour les relations faiblement représentées, mais on peut voir pour les autres un lien possible entre les relations souvent implicites (élaboration par exemple) et une

plus forte cohésion lexicale telle que nous l'avons définie.

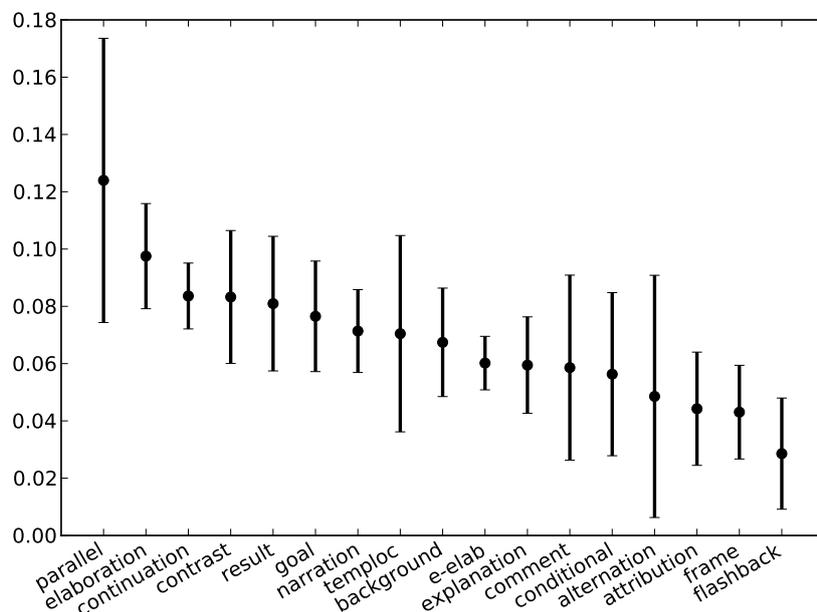


FIGURE 6.2 – Cohésion moyenne et intervalles de confiance entre unités discursives reliées, par type de relation rhétorique.

Ces travaux restent à un stade préliminaire, et évaluer la pertinence de la cohésion lexicale pour l'analyse discursive implique d'avoir amélioré le rapport signal/bruit dans ce que produisent les similarités de distribution. On a vu dans l'exemple de la figure 6.1 qu'un certain nombre de liens lexicaux plausibles ne sont pas pertinents en contexte, soit parce que des sens différents des termes sont mis en jeu, soit parce que le lien sémantique réel est contingent à la cohérence du passage.

Les travaux avec Clémentine Adam et Cécile Fabre ont donc poursuivi l'étude des conditions auxquelles ces liens peuvent être considérés comme prédictifs d'une certaine cohérence selon le contexte dans lequel ils apparaissent, en mettant au point d'une part une méthodologie de validation manuelle des liens lexicaux pertinents en contexte, et d'autre part un système automatique prédictif de cette pertinence. Ces aspects ont été publiés dans un article récemment accepté (Adam et al., 2013). La question principale est la suivante : étant donnés deux éléments lexicaux présents dans un même paragraphe, peut-on considérer qu'ils ont un lien de proximité sémantique ? Le protocole d'annotation manuelle a permis de montrer que l'accord inter-annotateur est bon sur cette question, et bien meilleur que si la question est posée sans le contexte

6. Cf. chapitre 4.

du paragraphe. Ensuite, en utilisant des caractéristiques des éléments lexicaux à la fois liés au corpus d'étude (fréquences relatives par exemple), aux propriétés distributionnelles (similarité de Lin, mais aussi propriétés du réseau lexical), et donc aussi à certains facteurs contextuels (fréquence dans le texte, éloignement des items, caractéristiques du réseau lexical du texte, notamment), nous avons entraîné plusieurs modèles inductifs, ce qui permet de retrouver les paires jugées pertinentes avec un F-score de 46%, quand la similarité distributionnelle seule n'atteint que 24%.

La taille réduite du corpus annoté en relations de discours permet de toutes façons difficilement de conclure quand il s'agit de voir l'impact des liens lexicaux sur la prédiction de telle ou telle relation, tant il y a de variations sur chacune des fonctions discursives. Réduite pour chaque relation à des sous-échantillons d'un ensemble déjà réduit, l'expérience appelle à un contexte plus large. Ce constat a donné lieu à un projet de thèse, commencée à l'automne 2012, par Juliette Conrath, dirigée par Nicholas Asher, et dont je partage le co-encadrement avec Stergos Afantenos. Le projet consiste à rechercher des contextes bien marqués correspondant à certaines relations (notamment en utilisant des connecteurs de discours non ambigus), pour extraire des associations lexicales à partir d'un corpus de grande taille⁷. Ces associations seront alors testées sur la prédiction de liens rhétoriques dans un corpus distinct. Le travail devrait également porter sur l'affinage des types de relations lexicales.

6.3 Similarités lexicales

L'utilisation de similarités lexicales pose la question de la nature de ce que l'on calcule comme rapports entre items lexicaux. Depuis (Grefenstette, 1994; Gazdar, 1996; Lin, 1998) jusqu'à plus récemment (Clark et al., 2008) de nombreux auteurs défendent l'idée d'une représentation du sens des mots en termes purement relationnels, comme un vecteur défini sur l'ensemble du lexique, à partir d'une mesure d'association, souvent la collocation.

Ainsi au degré 1, un mot est défini comme le vecteur de ses cooccurrences avec l'ensemble du lexique (cooccurrence graphique ou syntaxique etc). Bien sûr, on peut aussi considérer une similarité au degré 2, où l'on mesure une distance entre ces vecteurs pour redéfinir des coordonnées sur l'ensemble du lexique.

Ces représentations se prêtent bien à des réductions de dimensionnalité de l'espace de représentation, car il est naturel de considérer que chaque terme n'est pas une dimension lexicale indépendante. Ainsi des variantes d'analyse en composantes principales ou de décomposition en valeurs singulières ont été appliquées pour définir des dimensions significatives (un analogue du modèle de document en recherche d'information) qui peuvent à leur tour être utilisées pour des tâches sémantiques, comme la LSA pour la segmentation thématique par exemple (Choi et al., 2001).

7. Les travaux préliminaires utilisent le corpus Wacky, de près de 2 milliards de mots, cf <http://wacky.sslmit.unibo.it>.

On retrouve l'idée de primitives sémantiques sous la forme de directions sémantiques qui définissent des sous-espaces suffisant à décrire le lexique. Plus récemment, des chercheurs s'intéressent à définir des composantes plus facilement interprétables que des vecteurs quelconques dans un espace vectoriel de mots, notamment en restreignant les vecteurs au quadrant naturel des valeurs positives sur les axes lexicaux (pour un résumé, cf les travaux de [de Cruys, 2010](#)).

Chercher des dimensions sémantiques dans les espaces lexicaux créés revient à chercher des combinaisons significatives d'éléments particuliers, et non plus des éléments lexicaux privilégiés qui servent de support primitif. La façon dont tout le lexique est projeté sur ses dimensions essentielles donne l'organisation du lexique, ou du moins son reflet dans l'usage discursif. Une question cruciale se pose alors sur ces dimensions sémantiques : comment caractériser leur nature pour pouvoir les évaluer, comment comparer des propositions différentes qui peuvent être issues de données différentes, et construites selon des modes variés, fonctions de paramètres d'ajustement plus ou moins arbitraires. L'outil fondamental pour cela est la comparaison d'éléments lexicaux, notamment les relations que l'on peut déduire entre items à partir de leur représentation. Le concept le plus généralement appliqué est celui de "similarité" des représentations vectorielles, qui peut être défini de façons diverses, mais qui ne correspond pas nécessairement à un concept linguistique unique. Dans ce cadre, les fonctions lexicales relationnelles "classiques" (synonymie, antonymie, hyperonymie, hyponymie, méronymie principalement) peuvent être vues comme une partie observable dans ces données. Elles n'épuisent pas les liens que l'on peut trouver dans des rapprochements de distribution bien sûr, et on peut s'interroger sur la nature de ce qui reste.

Cette vision du lexique est par ailleurs limitée par au moins deux problèmes importants : d'une part le problème de l'ambiguïté lexicale, la polysémie des termes devant être retrouvée a posteriori de l'observation des régularités d'usage. D'autre part, le lien avec la phrase et la construction de sa sémantique interroge aussi la notion de sens attaché aux mots ; même si certains auteurs comme [Clark et al. \(2008\)](#) sont à la recherche d'une généralisation du sens d'une phrase sur cette base relationnelle du lexique, il manque encore au moins un équivalent de la notion d'inférence. Par ailleurs, les régularités de distribution ne sont pas le seul moyen de dériver des similarités entre termes. Un dictionnaire électronique peut aussi permettre des comparaisons par son organisation relationnelle, selon les mots utilisés dans les définitions, d'autant que le dictionnaire permet d'avoir accès à la structure des sous-sens de ses entrées⁸. Une autre source de rapprochement, plus semblable aux régularités de distribution, est donnée par le nombre croissant de corpus de textes parallèles, traductions en différentes langues. Ceux-ci permettent de rapprocher des items apparaissant souvent dans des traductions similaires, au prix d'un alignement des phrases traduites et d'un alignement sous-phrastique dont on peut calculer des régularités⁹.

Reste le problème de l'évaluation de ce qui est issu de ces processus. Une évaluation entièrement intrinsèque est délicate puisque par définition la notion de lien doit se faire dans un contexte

8. Cf. ([Ide et Véronis, 1990](#); [H.Kozima et Furugori, 1993](#)).

9. Cf. ([Weeds et al., 2004](#); [Van der Plas, 2008](#); [Gorman et Curran, 2005](#)), et une application du même principe à la construction d'un réseau lexical à la Wordnet dans ([Sagot et Fišer, 2008](#)).

textuel, discursif. D'autres formes de validation de cette notion peuvent être néanmoins d'étudier directement la place des fonctions lexicales classiques.

J'ai indiqué des pistes à la section 6.1 sur l'utilisation de liens indirects comme indices textuels, comme une forme d'étude indirecte de pertinence des liens de similarité. On l'a vu par exemple sur la segmentation thématique, avec un succès sans doute discutable, mais dans une voie qui reste encore à explorer. On a vu aussi les perspectives que cela ouvre en liaison avec l'analyse discursive.

Parmi les autres tâches envisageables, on peut trouver la désambiguïsation lexicale, la substitution lexicale, la détection de locutions, l'identification de préférences de sous-catégorisation, de rôles sémantiques¹⁰, ou bien leur utilisation en analyse syntaxique¹¹.

On peut aussi se retourner vers des tâches plus essentiellement lexicales avec des contours définis, comme la constitution de thesaurus, et l'organisation de bases lexicales en lien avec des ontologies. La construction automatique de thesaurus par collecte d'un ensemble de relations entre unités lexicales (synonymie, antonymie, méronymie, etc) est un objectif relativement ancien en traitement automatique des langues. Il est parfois étendu par l'ajout d'associations thématiques, qui correspondent aux concepts de similarité introduits plus haut, comme par exemple dans le thesaurus Moby¹².

Quand ils sont évalués dans cette perspective, les rapprochements lexicaux par similarités de distribution se révèlent plutôt décevants : les similarités ainsi définies rapportent un mélange hétérogène de fonctions lexicales et de termes sémantiquement apparentés sans que les contours de cette parenté soient évidents à délimiter, et les fonctions classiques sont minoritaires.

Dans la perspective d'approfondissement de ces similarités sémantiques, j'ai poursuivi des travaux d'évaluation de plusieurs types d'entrées pour les calculs qui fondent ces approches : d'une part l'utilisation de structures de dictionnaires¹³, et d'autre part l'utilisation de corpus parallèles multi-lingues.

J'ai exploré l'utilisation de dictionnaires avec Bruno Gaume et Nabil Hathout, en testant une similarité lexicale induite par la structure de graphe de type "petit monde" d'un réseau lexical (Gaume et al., 2004). Les années 2000 ont vu en effet une résurgence du dictionnaire comme ressource sémantique, due notamment à l'émergence de travaux sur ces structures de graphes particulières dont les dictionnaires sont un exemple¹⁴. L'apparition de dictionnaires électroniques avait initié un certain nombre de travaux sur leur utilisation pour faire des rapprochements sémantiques, même si l'évaluation était déjà un aspect difficile à délimiter.

Cette similarité fondée sur des marches aléatoires dans le graphe, similaire à celle induite par un algorithme comme PageRank, avait été évaluée pour son apport dans la désambiguïsation des définitions du dictionnaire. Pour chaque nom et chaque verbe d'une définition donnée, on

10. Cf les synthèses dans (de Cruys, 2010; Turney et Pantel, 2010).

11. Par exemple (Candito et Crabbé, 2009).

12. <http://icon.shef.ac.uk/Moby/>

13. Ces travaux ont en grande partie eu lieu dans le cadre du projet Dilan, dirigé par Bruno Gaume, cf <http://erss.irit.fr/Dilan/>.

14. (Watts et Strogatz, 1998)

essaie de retrouver l'entrée du dictionnaire qui correspond à son usage dans la définition, avec plus ou moins de précision en distinguant parmi les éventuels homographes, les sens principaux ou bien les distinctions les plus fines admises par le dictionnaire. Évalué sur petit échantillon d'éléments (une centaine) pris au hasard dans le dictionnaire les résultats étaient très bons sur les homographes (90%), moyens sur les sens principaux (54%) et moins bons sur toutes les distinctions (38%). En prenant un sens au hasard, les scores seraient respectivement 49, 36 et 18%. Cette méthode est difficile à généraliser à une tâche générique de désambiguation, où l'on perdrait le premier lien entre un mot et la définition dans laquelle il apparaît. Par ailleurs, la taille des définitions joue un rôle crucial. Dans un dictionnaire général les nombreuses définitions assurent une bonne connectivité du graphe, alors que des définitions succinctes rendent la méthode très instable (nous avons par exemple tenté la même opération sur les gloses d'entrée WordNet, généralement assez courtes, avec des résultats à peine mieux que le hasard). Dans tous les cas, la distribution des sens d'un mot polysémique est très concentrée sur les sens les plus courants et il est quasiment impossible de faire mieux qu'une méthode qui prédirait le sens le plus courant sans prendre un peu en compte cette distribution des sens.

Une façon plus intrinsèque d'évaluer la nature des liens de similarité calculés est de vérifier que des termes similaires entretiennent des relations lexicales établies, liées à une proximité sémantique, essentiellement les relations de synonymie, ou d'hyponymie et d'hyperonymie. Les liens lexicaux issus du réseau lexical d'un dictionnaire (le TLFi) ont ainsi été évalués manuellement (et les résultats présentés dans Muller et al., 2006), sur un échantillon aléatoire d'entrées du dictionnaire. Les sorties du système avaient aussi été comparées à des dictionnaires de synonymes de référence¹⁵. On a alors constaté que la méthode produisait plus de bons candidats que les baselines selon les juges (qui ne connaissaient par la source des données à évaluer), alors que la différence était nettement moins sensible avec les ressources de référence. La couverture d'un dictionnaire de synonymes est en effet assez problématique et on a là un début de confirmation de l'utilité de méthodes qui permettent de les compléter. Le problème est alors d'évaluer ce complément, qui par définition est absent de toute référence. Seule l'évaluation humaine, coûteuse, permet d'apporter cette information¹⁶, mais elle n'est pas sans obstacle non plus : un jugement hors contexte peut confondre un juge qui oublie certains sens des mots à évaluer, alors que des tests de substituabilité restreignent eux aussi les contextes, tout en brouillant la différence avec l'emploi d'hyperonymes, qui sont par ailleurs nombreux dans les dictionnaires de référence. A titre d'illustration, le kappa mesuré entre les deux juges n'atteignaient que 0.5, même si 85% des synonymes acceptés par l'un étaient acceptés par l'autre. Cette évaluation avait été reproduite avec deux lexicographes de l'Atilf, avec des résultats similaires. Il faut noter aussi que les recouvrements entre les différents thesaurus dont l'ensemble constitue la référence du Crisco ne dépassent pas 60%.

L'autre type de données que j'ai considéré pour définir une notion de similarité lexicale est celui qui résulte des corpus parallèles multilingues. Parmi les ressources déjà mentionnées, l'utilisation de données d'alignement bilingue a été la moins exploitée, sans doute parce que ce sont des

15. Le dictionnaire du Crisco, <http://www.crisco.unicaen.fr/>.

16. C'est également la méthode choisie par (Falk et al., 2009).

données plus récemment disponibles en quantité et destinées à une tâche différente. On peut citer quelques précurseurs expérimentaux, tels que [van der Plas et Tiedemann \(2006\)](#), qui comparent deux items avec une mesure de similarité entre leurs “vecteurs d’alignement” (la fréquence d’alignement d’un mot avec les autres mots du lexique) dans différentes langues, et [Wu et Zhou \(2003\)](#) qui ont tenté de combiner linéairement des classifieurs regroupant tous les types de ressources mentionnés : similarité d’alignement, de distribution syntaxique, et de proximité dans un graphe de dictionnaire.

Il a été aussi proposé, de façon plus marginale, d’utiliser des corpus parallèles multilingues pour retrouver une similarité entre termes ([Dyvik, 2002](#)) en se fondant sur l’hypothèse que des termes proches doivent partager largement leurs traductions. Dans cette optique, on considère les termes associés par des traductions en “miroir” : les traductions d’un terme d’une langue source 1 dans une langue cible 2 sont mis en rapport avec leur traduction dans la langue 1. L’hypothèse est que les termes obtenus après cet aller-retour 1-2-1 sont des bons candidats à la synonymie avec le terme de départ. Cette méthode est parfois utilisée par exemple pour trouver des locutions paraphrastiques ([Bannard et Callison-Burch, 2005](#)). Cette idée n’ayant pas été validée empiriquement pour les fonctions lexicales, j’ai commencé à l’explorer dans le cadre de similarités sémantiques en collaboration avec Philippe Langlais ([Muller et Langlais, 2011](#)). Dans ces premières expériences, nous avons considéré un échantillon d’éléments lexicaux, noms et verbes, et retenu pour chacun un certain nombre d’éléments similaires d’un point de vue distributionnel, ou similaires de par leur probabilité d’être la traduction d’un même terme dans une autre langue¹⁷. L’objectif à terme est d’utiliser diverses sources pour constituer des associations lexicales typées, et reconnaître notamment des fonctions classiques de synonymie et d’hyponymie, en espérant que les régularités distributionnelles et de traduction sont complémentaires. Mais dans un premier temps nous nous sommes contentés d’évaluer la présence des fonctions classiques dans des listes de candidats constituées à partir des termes ayant la plus forte similarité avec un terme donné. La méthode d’évaluation est la même que celle mentionnée plus haut : comparer avec une ressource existante pour avoir un premier point de comparaison sur l’ensemble du jeu de test, et analyser manuellement un échantillon pour raffiner l’évaluation. Le but de cette première étape est de produire des candidats intéressants concentrés le plus possible au début de la liste. Pour cela on évalue la précision des candidats en coupant la liste à des rangs n variables, et en utilisant des mesures globales, classiques en recherche d’information : Mean Average Precision (MAP), Mean Reciprocal Rank (MRR). La fréquence naturelle des termes candidats est un paramètre additionnel important dans ce genre d’approche, et nous avons aussi différencié les évaluations en fonction de cette caractéristique. Les résultats ont montré que les deux sources ont des précisions similaires en début de liste (précision à 1 autour de 20-25% sur WordNet, précision à 5 autour de 5-8%), légèrement meilleurs pour les termes issus de la méthode miroir pour les termes de fréquence plus élevée, et meilleurs pour les voisins distributionnels sinon,

17. Ces expériences ont été faites sur l’anglais, à partir des données distributionnelles de [Lin \(1998\)](#), et des données alignées du corpus anglo-français Hansard. La référence principale était le thesaurus WordNet. Des résultats semblables ont été constatés sur le français, en utilisant la base distributionnelle des voisins de Wikipedia déjà mentionnée, et de Dicosyn du Crisco pour l’évaluation.

alors que le rappel global de la méthode des miroirs semble bien meilleur. Les mesures globales (MAP, MRR) sont elles uniformément meilleures pour les miroirs de 2 à 3%. Les résultats relatifs sont comparables sur les noms et les verbes, mais globalement meilleurs sur les verbes pour les deux méthodes. La combinaison par intersection des listes candidates des deux méthodes améliore tous les scores sauf le rappel, et fait gagner de 5 à 10% pour P1, 2-4% pour P5, soit des améliorations relatives de 40 à 50% à chaque fois. Il faut noter que WordNet a une couverture probablement assez réduite vu le nombre moyen de synonymes par lexème¹⁸, et que cela rend l'évaluation sévère. En faisant une analyse d'erreur détaillée, sur les premiers candidats hors WordNet proposés pour un échantillon par une approche miroir par exemple, nous avons trouvé 18% de synonymes manifestes (en utilisant d'autres thesaurus) et 13% d'hyponymes.

Les évaluations présentées dans cette section sont complémentaires de l'utilisation dans des tâches discursives présentées dans les sections précédentes. Les tâches supplémentaires tentent de préciser l'intérêt de l'utilisation de similarités sémantiques, et les évaluations intrinsèques permettent d'améliorer la conception de ces similarités en fournissant des comparaisons informatives. Il est incontestable que ce domaine, malgré ses promesses, est encore en phase de défrichage.

6.4 Perspectives

J'ai insisté ici sur l'utilisation du lexique pour signifier la cohérence d'un discours, ou son rôle observable dans cette cohérence à travers la notion de cohésion lexicale. L'importance de la structuration sémantique du lexique, à travers des relations significatives entre items, va bien au-delà de la seule construction discursive. Les dernières années ont vu une montée des travaux cherchant à formaliser les liens entre la structuration de connaissances, sous le label d'ontologie(s), dont nous avons déjà parlé, et les liens avec le lexique sont forts. Nous en avons vu ici deux aspects : les liens "purement lexicaux" que sont les synonymes, et des liens sémantiques avec une contrepartie conceptuelle (hypéronymes, méronymes, éléments impliqués dans des frames sémantiques communes). Cette problématique chevauche plusieurs projets dans lesquels je suis ou j'ai été impliqué¹⁹.

On retrouve un écho de l'utilisation du lexique dans les axiomes d'interprétation des premières théories sémantiques du discours de Hobbes ou Asher : les liens typiques entre événements ou acteurs qui sont encodés par des règles par défaut pourraient être collectés par des associations typiques en corpus, dont les similarités sémantiques sont un exemple particulier. A mi-chemin de ces deux perspectives, on retrouve l'idée de Verbocean ([Chklovski et Pantel, 2004](#)), qui est de collecter explicitement des relations sémantiques typiques entre certains verbes (causalité, succession temporelle, etc) à partir de repérages de patron dans des grandes masses de textes et filtrage avec des mesures de cooccurrences significatives, une idée raffinée dans les travaux de [Chambers et Jurafsky \(2008b\)](#) sur les "schémas narratifs".

18. Sur notre échantillon, 3,5 en moyenne par nom, et 5.5 par verbe.

19. Les projets déjà mentionnés Annodis, Voiladis, Dilan.

A notre connaissance, cette approche n'a pas été intégrée ou même reliée à des approches rhétoriques, même si l'objectif sémantique poursuivi est similaire, et ouvre des perspectives communes aux approches sémantiques et discursives, ce qui a motivé la thèse de Juliette Conrath, mentionnée plus haut.

L'étude des similarités lexicales a consisté pour l'instant à construire une sémantique lexicale essentiellement relationnelle, sans que les relations soient complètement spécifiées. Je compte maintenant faire évoluer ces travaux au-delà de cette base pour faire le lien vers la sémantique à un niveau plus construit, idéalement celui d'une phrase complète ou du moins des composantes de sens plus complexes où se pose la question de la combinaison des éléments lexicaux. Cette question est au centre de nombreux travaux récents fondés sur des approches distributionnelles, et repose pour l'instant sur des représentations vectorielles dans l'espace du lexique entier, par exemple quand un mot est représenté par ses cooccurrences avec chaque élément du lexique, cf la discussion de la section 6.3. Ces sens, relationnels, doivent être alors combinés pour produire une représentation qui peut permettre au moins certains aspects liés à la sémantique traditionnelle, équivalences, subsomptions, implications... J'ai commencé à explorer cette voie à travers des petites tâches de la campagne Semeval, en collaboration avec Tim van de Cruys et Stergos Afantenos, en reprenant les travaux de Tim van de Cruys qui utilisent des techniques de réduction de dimensionalité pour abstraire ce qui pourrait être des dimensions de sens, et en combinant cette fondation non supervisée avec des éléments supervisés, cf ([Van de Cruys et al., 2013a](#)), ([Van de Cruys et al., 2013b](#)).

Conclusion : représentation des connaissances et traitement des langues

La thématique qui unifie les travaux présentés est au premier abord le traitement d'informations temporelles et spatiales exprimées en langage naturel. Si leur importance les justifie comme sujet d'étude en eux-mêmes, il me paraît crucial de les considérer aussi comme terrain idéal pour l'investigation de problèmes plus généraux, liés à la compréhension de textes ou de dialogues. Cela mène à des problèmes rangés à l'étage "pragmatique" en traitement des langues, où l'essentiel des sujets abordés concerne la prise en compte du contexte et son interaction avec une interprétation incrémentale, par exemple phrase par phrase, ou acte de langage par acte de langage. Le raisonnement et sa modélisation prennent alors une place centrale dans la gestion du contexte, comme on l'a vu par exemple pour l'interprétation d'informations temporelles, et nécessite d'appliquer des techniques de TAL en intégrant des thématiques d'intelligence artificielle que l'on rencontre moins souvent dans les autres tâches du domaine, comme la résolution de systèmes de contraintes. D'un point de vue méthodologique, on peut voir aussi une alternance entre études formalistes et études empiriques, et la volonté de concilier ces deux versants dans la modélisation des phénomènes linguistiques. Le traitement automatique des langues a (re)pris un virage empirique, depuis les années 1990, qui a un peu éclipsé les approches rationalistes précédentes, à la fois parce que l'évaluation des travaux l'a rendu nécessaire, mais aussi parce que les progrès des techniques statistiques ont poussé à leur application de façon très large. La mise au point de modèles à base de règles ou de théories axiomatiques, qui soutiennent la résolution de tâches essentiellement par déduction, a laissé la place à des modèles inductifs, fondés sur des exemples caractérisés par un ensemble de descripteurs pertinents. Sur des problèmes qu'on peut réduire à des décisions locales plus ou moins indépendantes, on ne peut questionner leur efficacité. Sur des problèmes structurés, on constate une variété dans la facilité d'appliquer ce paradigme. Les objets que l'on considère ici, structures temporelles ou discursives, se prêtent encore mal à ces méthodes, d'autant qu'il est très coûteux d'obtenir des données annotées, et j'ai consacré une grande part de mes travaux à essayer d'intégrer des techniques symboliques à un cadre inductif. Ces aspects relativement techniques ne doivent pas cacher que l'horizon visé est la prise en compte du contexte de plus en plus large dans les tâches de TAL, appelées sans doute un peu vite "discursives", de quelques phrases à un texte entier ou des textes reliés entre eux, en dehors des tâches naturellement considérées comme discursives (comme le résumé automatique). La fouille d'opinions et de sentiments par exemple a fait prendre conscience de l'importance

de considérer l'analyse sémantique à un niveau global²⁰. La traduction automatique, longtemps retranchée au niveau de la phrase, s'est récemment ouverte à des contextes plus larges, jusqu'à susciter un atelier à l'ACL 2013 (Webber et al., 2013)²¹. Les enjeux majeurs sont naturellement la gestion de la co-référence, mais aussi la cohésion lexicale et la prise en compte des procédés argumentatifs, et de l'expression de relations discursives. La notion de document écrit s'est dans le même temps élargi à des formes de communication plus variées que le texte mono-auteur, en prenant comme objet d'étude, comme on l'a vu, les fils de discussion, les échanges par courriel, mais aussi les conversations par chat ou le micro-blogging. Cette variété d'objets largement en friche va remettre en cause un certain nombre de présupposés de l'analyse textuelle pour la mêler à des thématiques plus proches du dialogue, par la prise en compte d'auteurs multiples, et de phénomènes pragmatiques mal fixés par des conventions, et dont l'étude en est à ses débuts (Herring et al., 2013).

L'autre enseignement de l'étude des données temporelles et spatiales est la nécessité pour les approches sémantiques en TAL de considérer les problèmes de représentation de connaissances. L'émergence du "Web sémantique" et de la structuration des connaissances que sont les ontologies, nécessaires pour l'indexation d'informations massives pose de plus en plus crucialement la question de l'appariement entre des expressions en langage naturel et une représentation normalisée, exploitable en aval des applications de recherche ou d'extraction d'information. Cet effort au niveau sémantique trouve un début de réalisation dans les projets de représentation lexicale tels que FrameNet (Baker et al., 1998), mais il pose encore des problèmes importants de couverture et de spécification, notamment dans son rapport avec une ontologie claire²². On voit ici quel rôle peut jouer l'extraction de connaissances à partir de textes pour nourrir les bases conceptuelles, qui facilitent à leur tour l'exploitation de nouvelles données textuelles.

20. Voir par exemple les travaux de et dans l'équipe Melodi ceux de .

21. <http://www.idiap.ch/workshop/DiscoMT>

22. Je participe d'ailleurs à l'ANR Asfalda, pilotée par Marie Candito sur la spécification d'un lexique sémantique français à la Framenet, dont un objectif est de mieux spécifier la structure ontologique, et de mieux prendre en compte les effets contextuels entre propositions.

Il est évident, d'après ce qui précède, que l'affaire du poète, ce n'est pas de parler de ce qui est arrivé, mais bien de ce qui aurait pu arriver et des choses possibles, selon la vraisemblance ou la nécessité.

Aristote.

Bibliographie

- Adam, C., C. Fabre, et P. Muller (2013). Évaluer et améliorer une ressource distributionnelle : protocole d’annotation de liens sémantiques en contexte. *TAL* 54(1), in press. 84
- Adam, C., P. Muller, et C. Fabre (2010). Une évaluation de l’impact des types de textes sur la tâche de segmentation thématique. In M. Gagnon et P. Langlais (Eds.), *Traitement Automatique des Langues Naturelles (TALN)*, Montréal, <http://www.atala.org/>, pp. (support électronique). Association pour le Traitement Automatique des Langues (ATALA). 80
- Adam, C. et M. Vergez-Couret (2010). Signalling elaboration : Combining gerund clauses with lexical cues. In *Proceedings of Multidisciplinary Approaches to Discourse (MAD 2010)*, Moissac. 82
- Afantenos, S. et N. Asher (2010, August). Testing SDRT’s right frontier. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, Beijing, China, pp. 1–9. Coling 2010 Organizing Committee. 56
- Afantenos, S., N. Asher, F. Benamara, A. Cadilhac, C. Dégremont, P. Denis, M. Gühe, S. Keizer, A. Lascarides, O. Lemon, P. Muller, S. Paul, V. Popescu, V. Rieser, et L. Vieu (2012). Developing a corpus of strategic conversation in the settlers of catan. In N. Tomuro et J. Zagal (Eds.), *Proceedings of the 1st Workshop on Games and NLP (GAMNLP-12)*, pp. 7 p. 76
- Afantenos, S. D., P. Denis, P. Muller, et L. Danlos (2010). Learning recursive segments for discourse parsing. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010*, Valletta, Malta. 54
- Ahn, D., J. van Rantwijk, et M. de Rijke (2007, April). A cascaded machine learning approach to interpreting temporal expressions. In *Human Language Technologies 2007 : The Conference of the North American Chapter of the Association for Computational Linguistics ; Proceedings of the Main Conference*, Rochester, New York, pp. 420–427. Association for Computational Linguistics. 27
- Aho, A., M. Garey, et J. Ullman (1972). The Transitive Reduction of a Directed Graph. *SIAM Journal on Computing* 1(2), 131 – 137. 40

- Allen, J. F. (1984). Towards a general theory of action and time. *Artificial intelligence* 23, 123–154. 14, 16
- Allwood, J., J. Nivre, et E. Ahlson (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics* 9. 72
- Anderson, A., M. Bader, E. Bard, E. B. and G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, et R. Weinert (1991). The hrc map task corpus. *Language and Speech* 34, 351–366. 70
- Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Kluwer. 51, 55
- Asher, N., M. Aurnague, M. Bras, P. Sablayrolles, et L. Vieu (1995). De l'espace-temps dans l'analyse du discours. *Sémiotique* (9), 11–63. 24
- Asher, N. et A. Lascarides (1993). Temporal interpretation, discourse relations, and common-sense entailment. *Linguistics and Philosophy* 16, 437–493. 23, 28, 50
- Asher, N. et A. Lascarides (1998). Questions in dialogue. *Linguistics and Philosophy*. 72
- Asher, N. et A. Lascarides (2003). *Logics of conversation*. Cambridge University Press. 66, 67, 68
- Asher, N., P. Muller, et M. Gaio (2008). Spatial entities are temporal entities too : the case of motion verbs. In G. Katz, I. Mani, et T. Tenbrink (Eds.), *Methodologies and Resources for Processing Spatial Language (Workshop LREC'2008), Marrakech, 31/05/08*. pp 16-21. 32
- Asher, N. et P. Sablayrolles (1995, June). A typology and discourse semantics for motion verbs and spatial PPs in French. *Journal of Semantics* 12(1), 163–209. 19, 33
- Asher, N., A. Venant, P. Muller, et S. Afantenos (2011). Complex discourse units and their semantics. In *Proceedings of Constraints in Discourse 2011*, pp. 9p. 59
- Asher, N. et L. Vieu (1995). Toward a geometry of common sense : A semantics and a complete axiomatization of mereotopology. In *Proc. 14th Int. Joint Conf. on Artificial Intelligence (IJCAI'95)*, pp. 846–852. Morgan Kaufmann. 14, 15
- Bach, E. (1986). The algebra of events. *Linguistics and Philosophy* 9, 5–16. 20
- Baker, C., C. Fillmore, et J. Lowe (1998). The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pp. 86–90. Association for Computational Linguistics. 94
- Bakir, G., T. Hofmann, B. Schölkopf, A. Smola, B. Taskar, et S. Vishwanathan (Eds.) (2007). *Predicting structured data*. The MIT Press. 62

- Baldrige, J., N. Asher, et J. Hunter (2007). Annotation for and Robust Parsing of Discourse Structure on Unrestricted Texts. *Zeitschrift für Sprachwissenschaft* 26, 213–239. 51
- Baldrige, J. et A. Lascarides (2005, June). Probabilistic head-driven parsing for discourse structure. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, Ann Arbor, Michigan, pp. 96–103. Association for Computational Linguistics. 55
- Baldwin, J. A. (2002). Learning temporal annotation of French news. Master's thesis, Georgetown University, Washington, DC. 26
- Bangerter, A. et H. H. Clark (2003). Navigating joint projects with dialogue. *Cognitive Science* (27), 195–225. 72
- Bannard, C. et C. Callison-Burch (2005, June). Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, Ann Arbor, Michigan, pp. 597–604. Association for Computational Linguistics. 89
- Bateman, J. A. (2010a). Language and space : a two-level semantic approach based on principles of ontological engineering. *International Journal of Speech Technology* 13(1), 29–48. 21
- Bateman, J. A. (2010b). Situating spatial language and the role of ontology : issues and outlook. *Linguistics and Language Compass* 4(8), 639–664. 21
- Bateman, J. A., R. Henschel, et F. Rinaldi (1995). Generalized upper model 2.0 : documentation. Rapport technique, GMD/Institut für Integrierte Publikations und Informationssysteme, Darmstadt, Germany. 13
- Bernard, P. et D. Maurel (1993). Interrogation en langage naturel d'une base de données : interprétation des adverbiaux de localisation temporelle. *Mathématiques et Sciences Humaines* 123, 45–52. 26
- Bethard, S. et J. H. Martin (2006, July). Identification of event mentions and their semantic class. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, Sydney, pp. 146–154. 27
- Bethard, S., J. H. Martin, et S. Klingenstein (2007). Timelines from text : Identification of syntactic temporal relations. In *International Conference on Semantic Computing*, Los Alamitos, CA, USA, pp. 11–18. IEEE Computer Society. 31
- Bilange, E. (1991). *Modélisation du dialogue oral finalisé personne-machine par une approche structurelle*. Ph. D. thesis, Université de Rennes I. 67
- Bittar, A. (2009). Annotation of events and temporal expressions in French texts. In *Proceedings of the Third Linguistic Annotation Workshop*, pp. 48–51. Association for Computational Linguistics. 28

- Bittar, A. (2010). *A Corpus Project for the Temporal Annotation of French Texts*. Ph. D. thesis, Université Paris Diderot. 25, 26, 27
- Black, E., S. Abney, D. Flickenger, C. Gdaniec, R. Grishman, P. Harrison, D. Hindle, R. Ingria, F. Jelinek, J. Klavans, et al. (1991). A Procedure for Quantitatively Comparing the Syntactic Coverage of English Grammars. In *Speech and natural language : proceedings of a workshop, held at Pacific Grove, California, February 19-22, 1991*, pp. 306. Morgan Kaufmann Pub. 62
- Boguraev, B. et R. Ando (2005). Timeml-compliant text analysis for temporal reasoning. In L. P. Kaelbling et F. Giunchiglia (Eds.), *Proceedings of IJCAI05*, pp. 997–1003. 26
- Bramsen, P., P. Deshpande, Y. K. Lee, et R. Barzilay (2006, July). Inducing temporal graphs. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Sydney, Australia, pp. 189–198. Association for Computational Linguistics. 25, 30, 31, 32, 42, 45
- Braud, C. et P. Denis (2013). Identification automatique des relations discursives "implicites" à partir de données annotées et de corpus bruts. In *TALN - 20ème conférence du Traitement Automatique du Langage Naturel 2013*, Volume 1, pp. 104–117. 57
- Bruce, B. (1972). A model for temporal references and its application in a question answering program. *Artificial Intelligence* 3(1-3), 1–25. 29, 43
- Candito, M. et B. Crabbé (2009). Improving generative statistical parsing with semi-supervised word clustering. In *Proceedings of the 11th International Conference on Parsing Technologies*, pp. 138–141. Association for Computational Linguistics. 87
- Carletta, J., A. Isard, S. Isard, J. Kowtko, et G. Doherty-Sneddon (1996). Hrc dialogue structure coding manual. Rapport Technique TR-92, HCRC. 70, 71
- Carletta, J., A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, et A. Anderson (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*. 72
- Carlson, G. N. (1980). *Reference to Kinds in English*. New York : Garland Publishing. 19
- Carlson, L. et D. Marcu (2001). Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*. 54
- Carlson, L., D. Marcu, et M. Okurowski (2001). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue-Volume 16*, pp. 10. Association for Computational Linguistics. 51
- Carlson, L., D. Marcu, et M. E. Okurowski (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In J. van Kuppevelt et R. Smith (Eds.), *Current Directions in Discourse and Dialogue*, pp. 85–112. Kluwer Academic Publishers. 51

- Carreras, X. et L. Màrquez (2001). Boosting trees for clause splitting. In W. Daelemans et R. Zajac (Eds.), *Proceedings of CoNLL-2001*, pp. 73–75. Toulouse, France. 54
- Caselli, T., V. Bartalesi Lenzi, R. Sprugnoli, E. Pianta, et I. Prodanof (2011, June). Annotating events, temporal expressions and relations in italian : the it-timebank experience for the it-timebank. In *Proceedings of the 5th Linguistic Annotation Workshop*, Portland, Oregon, USA, pp. 143–151. Association for Computational Linguistics. 25
- Chambers, N. (2013, June). Navytime : Event and time ordering from raw text. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2 : Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, Atlanta, Georgia, USA, pp. 73–77. Association for Computational Linguistics. 25, 46
- Chambers, N. et D. Jurafsky (2008a, October). Jointly combining implicit constraints improves temporal ordering. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Honolulu, Hawaii, pp. 698–706. Association for Computational Linguistics. 31, 42
- Chambers, N. et D. Jurafsky (2008b, June). Unsupervised Learning of Narrative Event Chains. In *Proceedings of ACL-08 : HLT*, Columbus, Ohio, pp. 789–797. 90
- Chambers, N., S. Wang, et D. Jurafsky (2007). Classifying temporal relations between events. In *ACL*. 25, 31
- Chen, H., S. Branavan, R. Barzilay, et D. Karger (2009). Content Modeling Using Latent Permutations. *Journal of Artificial Intelligence Research* 36, 129–163. 80
- Cherry, C., X. Zhu, J. Martin, et B. de Bruijn (2013). À la recherche du temps perdu : extracting temporal relations from medical text in the 2012 i2b2 nlp challenge. *Journal of the American Medical Information Association*. 47
- Chklovski, T. et P. Pantel (2004). Verbocean : Mining the web for fine-grained semantic verb relations. In *Proceedings of EMNLP*, Volume 4, pp. 33–40. 51, 61, 82, 90
- Choi, F. Y. Y., P. Wiemer-Hastings, et J. Moore (2001). Latent semantic analysis for text segmentation. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*, Pittsburgh, pp. 109–117. 85
- Clark, H. H. (1996). *Using Language*. Cambridge University Press. 72
- Clark, S., B. Coecke, et M. Sadrzadeh (2008). A compositional distributional model of meaning. In *Proceedings of the Second Quantum Interaction Symposium (QI-2008)*. College Publications. Citeseer. 85, 86
- Clarke, B. (1981). A calculus of individuals based on “connection”. *Notre Dame Journal of Formal Logic* 22(3), 204–218. 14

- Clarke, J. et M. Lapata (2007). Modelling compression with discourse constraints. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 1–11. 54
- Cohen, R. et C. Perrault (1979). Elements of a plan-based theory of speech-acts. *Cognitive Science* 3, 177–212. 66
- Cohn, A. (1996). Calculi for qualitative spatial reasoning. In J. Calmet, J. Campbell, et J. Pfalzgraf (Eds.), *Artificial Intelligence and Symbolic Mathematics*, Volume 1138 of *LNCS*, pp. 124–143. Heidelberg : Springer Verlag. 13
- Collins, M. et N. Duffy (2001). Convolution kernels for natural language. In *NIPS*, pp. 625–632. 62
- Core, M. et J. Allen (1997). Coding dialogs with the damsl annotation scheme. In *Working Notes of the AAAI Fall Symposium on Communicative actions in Humans and Machines*, Cambridge, MA, pp. 28–35. 70, 72
- Cristani, M. et A. G. Cohn (2002). Spaceml : A mark-up language for spatial knowledge. *Journal of Visual Languages and Computing* 13(1), 97 – 116. 33
- Cui, Z., A. Cohn, et D. Randell (1992). Qualitative simulation based on a logical formalism of space and time. In *Proceedings of AAAI'92*, Cambridge, MA. AAAI/MIT Press. 15
- Daraux, J.-P. (2000). *Le Voyage Aux Pyrénées De James David Forbes En 1835*. Cairn. 8
- de Cruys, T. V. (2010). A non-negative tensor factorization model for selectional preference induction. *Natural Language Engineering* 16(4), 417–437. 86, 87
- de Saint-Cyr, F. et J. Lang (1997). Reasoning about unpredicted change and explicit time. *Qualitative and Quantitative Practical Reasoning*, 223–236. 18
- Denis, M. (1997). The description of routes : A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive* 16(4), 409–458. 33, 65, 71
- Denis, M. et X. Briffault (1997). Les aides verbales à la navigation. In M. Denis (Ed.), *Langage et cognition spatiale*, Sciences Cognitives, pp. 127–154. Masson, Paris. 71
- Denis, P. et J. Baldrige (2007). A ranking approach to pronoun resolution. In *Proc. IJCAI*. 61
- Denis, P. et P. Muller (2004). A semantics for temporally dependent referring expressions. In O. Bonami et P. C. Hofherr (Eds.), *Empirical Issues in Syntax and Semantics 5, Papers from Cssp 2003*, pp. 45–62. <http://www.cssp.cnrs.fr/eiss5/>. 19, 20
- Denis, P. et P. Muller (2010). Comparison of different algebras for inducing the temporal structure of texts. In *Proceedings of Coling 2010*, Beijing, pp. 250–258. 42

- Denis, P. et P. Muller (2011). Predicting globally-coherent temporal structures from texts via endpoint inference and graph decomposition. In *Proceedings of IJCAI 2011*, pp. 1788–1793. 31
- Dubois, M. F. et S. R. Schwer (2000). Classification topologique des ensembles convexes de Allen. In *Proceedings of Reconnaissance des Formes et Intelligence Artificielle (RFIA)*, Volume III, pp. 59–68. 40
- duVerle, D. et H. Prendinger (2009, August). A novel discourse parser based on support vector machine classification. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, Suntec, Singapore, pp. 665–673. Association for Computational Linguistics. 55
- Dyvik, H. (2002). Translations as semantic mirrors : From parallel corpus to wordnet. In *The Theory and Use of English Language Corpora, ICAME 2002*. <http://www.hf.uib.no/i/LiLi/SLF/Dyvik/ICAMEpaper.pdf>. 89
- Eisenstein, J. (2009, June). Hierarchical text segmentation from multi-scale lexical cohesion. In *Proceedings of Human Language Technologies : The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Boulder, Colorado, pp. 353–361. Association for Computational Linguistics. 79
- Falk, I., C. Gardent, E. Jacquy, et F. Venant (2009). Sens, synonymes et définitions. In *Conférence sur le Traitement Automatique du Langage Naturel - TALN'2009*, Senlis France. 88
- Fellbaum, C. (Ed.) (1998). *WordNet : An Electronic Lexical Database*. MIT Press. 27
- Feng, V. W. et G. Hirst (2012, July). Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, Jeju Island, Korea, pp. 60–68. Association for Computational Linguistics. 55
- Ferret, O. (2007, June). Finding document topics for improving topic segmentation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague, Czech Republic, pp. 480–487. Association for Computational Linguistics. 80
- Fetzer, A. et K. Fischer (2006). *Lexical markers of common grounds*. Elsevier. 72
- Fónagy, I. et E. Bérard (1973). Questions totales simples et implicatives en français parisien. In A. Grundstrom et P. Léon (Eds.), *Interrogation et Intonation*, Number 8, pp. 53–98. Paris : Didier. 74
- Forbus, K. (1983). Qualitative reasoning about space and motion. In D. Gentner et A. Stevens (Eds.), *Mental models*, pp. 53–73. Hillsdale (NJ) : Erlbaum. 14

- Galton, A. (1993). Towards an integrated logics of space, time and motion. *IJCAI*. 15
- Galton, A. (1997). Continuous change in spatial regions. In A. Frank et W. Kuhn (Eds.), *Spatial Information Theory - Proceedings of COSIT'97*, Number 1329 in Lecture Notes in Computer Science, Berlin. Springer. 16
- Gambarotto, P. et P. Muller (2003). Ontological problems for the semantics of spatial expressions in natural language. In J. S. E. van der Zee (Ed.), *Representing direction in language and space*, Number 2 in Language and Space, pp. 144–165. Oxford University Press. 20
- Gardner, R. (2001). *When listeners talk*. Pragmatics and Beyond. John Benjamins. 72
- Gaudou, B., A. Herzig, et D. Longin (2006). A Logical Framework for Grounding-based Dialogue Analysis1. *Electronic Notes in Theoretical Computer Science* 157(4), 117–137. 68
- Gaume, B., N. Hathout, et P. Muller (2004). Word sense disambiguation using a dictionary for sense similarity measure. In *Proceedings of Coling 2004*, Volume II, Genève, pp. 1194–1200. 87
- Gazdar, G. (1996). Paradigm merger in natural language processing. pp. 88–109. 85
- Georgila, K., R. Artstein, A. Nazarian, M. Rushforth, D. Traum, et K. Sycara (2011, June). An annotation scheme for cross-cultural argumentation and persuasion dialogues. In *Proceedings of the SIGDIAL 2011 Conference*, Portland, Oregon, pp. 272–278. Association for Computational Linguistics. 76
- Gerevini, A. et B. Nebel (2002). Qualitative spatio-temporal reasoning with rcc-8 and allen's interval calculus : Computational complexity. In *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI'02)*. 17
- Ginzburg, J. et I. Sag (2001). *Interrogative Investigations*. The university of Chicago press. 72
- Goodman, N. (1977). *The structure of appearance* (3rd ed.). Pallas Paperbacks, 3. Dordrecht : Reidel. 20
- Gorman, J. et J. Curran (2005, June). Approximate searching for distributional similarity. In *Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition*, Ann Arbor, Michigan, pp. 97–104. Association for Computational Linguistics. 86
- Grefenstette, G. (1994). *Explorations in automatic thesaurus discovery*. Springer. 85
- Grenon, P. et B. Smith (2004). Snap and span : Towards dynamic spatial ontology. *Spatial Cognition and Computation* 4(1), 69–103. 13
- Grice, P. (1975). Logic and conversation. In P. Cole et J. Morgan (Eds.), *Syntax and Semantics*, Volume 3. London : Academic Press. 66

- Grosz, B. et C. L. Sidner (1986, July-September). Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3), 175–204. 50, 66, 67
- Grover, C., J. Hitzeman, et M. Moens (1995). Algorithms for analysing the temporal structure of discourse. In *Sixth International Conference of the European Chapter of the Association for Computational Linguistics*. ACL. 23, 28, 30, 50
- Grundstrom, A. (1973). L'intonation des questions en français standard. In A. Grundstrom et P. Léon (Eds.), *Interrogation et Intonation*, Number 8, pp. 19–51. Paris : Didier. 74
- Halliday, M. et R. Hasan (1976). *Cohesion in English*. London : Longman. 11, 78
- Hamblin, C. L. (1970). *Fallacies*. Methuen. 66, 68
- Harper, L., I. Mani, et B. Sundheim (Eds.) (2001). *ACL Workshop on Temporal and Spatial Information Processing*, 39th Annual Meeting and 10th Conference of the European Chapter. Association for Computational Linguistics. 29, 106
- Hayes, P. (1985). Naive physics i : Ontology for liquids. In J. Hobbs et R. Moore (Eds.), *Formal theories of the commonsense world*, pp. 71–107. Norwood (NJ) : Ablex. 14
- Hearst, M. A. (1997, March). TextTiling : segmenting text into multi-paragraph subtopic passages. *Computational Linguistics* 23(1), 33–64. 80
- Heller, M. (1990). *The Ontology of Physical Objects : Four-Dimension Hunks of Matter*. Cambridge University Press. 20
- Hernandez, A. (2006). La détection de types d'actes de langage dans un dialogue orienté tâche. Master's thesis, Université Paul Sabatier. 73
- Hernández, D. (1994). *Qualitative Representation of Spatial Knowledge*, Volume 804 of *Lecture Notes in Artificial Intelligence*. Springer. 13
- Herring, S., D. Stein, et T. Virtanen (Eds.) (2013). *Pragmatics of Computer-Mediated Communication*. De Gruyter Mouton. 94
- H.Kozima et T. Furugori (1993). Similarity between words computed by spreading activation on an english dictionary. In *Proceedings of the conference of the European chapter of the ACL*, pp. 232–239. 86
- Hobbs, J., W. Croft, T. Davies, D. Edwards, et K. Laws (1987). Commonsense metaphysics and lexical semantics. *Computational Linguistics* 13(3-4), 241–250. 51
- Hwang, C. et L. Schubert (1992). Tense Trees as the fine structure of discourse. In *Proceedings of the 30th annual meeting on Association for Computational Linguistics*, pp. 232–240. Association for Computational Linguistics. 28

- Ide, N. et J. Véronis (1990). Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING'90)*, Volume 2, pp. 389–394. 86
- Jurafsky, D., R. Bates, N. Coccaro, R. Martin, M. Meteer, K. Ries, E. Shriberg, A. Stolcke, P. Taylor, et C. V. Ess-Dykema. (1997). Switchboard discourse language modeling project final report. Rapport Technique Summer Research Workshop Technical Reports 30, Johns Hopkins University, Baltimore, MD. 71
- Jurafsky, D., E. Shriberg, B. Fox, et T. Curl (1998). Lexical, prosodic and syntactic cues for dialog acts. In *ACL/COLING-98 Workshop on Discourse Relation and Discourse Markers*. 73
- Kameyama, M., R. Passonneau, et M. Poesio (1993). Temporal centering. In *Proceedings of ACL 1993*, pp. 70–77. 23, 28
- Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk et M. Stockhof (Eds.), *Formal methods in the study of Language*. Mathematical Centre Tract, Amsterdam. 51
- Kamp, H. et U. Reyle (1993). *From Discourse to Logic*. Kluwer Academic Publishers. 23, 24, 28
- Katz, G. et F. Arosio (2001). The annotation of temporal information in natural language sentences. See [Harper et al. \(2001\)](#), pp. 104–111. 28
- Kolomiyets, O., S. Bethard, et M.-F. Moens (2012, July). Extracting narrative timelines as temporal dependency structures. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, Jeju Island, Korea, pp. 88–97. Association for Computational Linguistics. 46
- Kowtko, J. (1996). *The function of intonation in task-oriented dialogues*. Ph. D. thesis, University of Edimburgh. 74
- Kreutel, J. et C. Matheson (2002). From dialogue acts to dialogue act offers : building discourse structure as an argumentative process. In Bos, Fox, et Matheson (Eds.), *Proceedings of the sixth workshop on the semantics and pragmatics of dialogue (Edilog02)*, pp. 85–92. 68
- La Poutré, J. et J. van Leeuwen (1988). Maintenance of transitive closures and transitive reductions of graphs. In *Proceedings of the International Workshop WG '87 on Graph-theoretic concepts in computer science*, pp. 106–120. 40
- Lang, J. et P. Muller (2001). Plausible reasoning from spatial observations. In M. Goldszmidt (Ed.), *Uncertainty in AI : Proceedings of the 17th Conference*, pp. 285–292. Morgan Kaufmann. 18

- Langacker, R. (1987). *Foundations of Cognitive Grammar*, Volume 1. Stanford University Press. 18
- Lapata, M. et A. Lascarides (2006). Learning sentence-internal temporal relations. *J. Artif. Intell. Res. (JAIR)* 27, 85–117. 31
- Le Thanh, H., G. Abeysinghe, et C. Huyck (2004, Aug 23–Aug 27). Generating discourse structures for written text. In *Proceedings of Coling 2004*, Geneva, Switzerland, pp. 329–335. COLING. 55
- Lenat, D. (1995). Cyc : A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38(11). 14
- Lewin, I. (2000). A formal model of conversational game theory. In *Proceedings of the 4th workshop on the semantics and pragmatics of dialogue (Gotalog)*. 69
- Lewis, D. K. (1979). Scorekeeping in a language game. *Journal of philosophical logic* 8, 339–359. 66
- Li, W., K.-F. Wong, G. Cao, et C. Yuan (2004, July). Applying machine learning to chinese temporal relation resolution. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, Barcelona, Spain, pp. 582–588. 31
- Li, W., K.-F. Wong, et C. Yuan (2001). A model for processing temporal reference in chinese. See [Mani et al. \(2005\)](#). 30
- Lin, D. (1998). Automatic retrieval and clustering of similar words. Volume 2, Montreal, pp. 768–774. 79, 85, 89
- Link, G. (1983). The logical analysis of plurals and mass terms : A lattice-theoretical approach. In R. Bäuerle, C. Schwarze, et A. von Stechow (Eds.), *Meaning, Use and Interpretation of Language*. de Gruyter. 20
- Mani, I., J. Pustejovsky, et R. Gaizauskas (Eds.) (2005). *The Language of Time : A Reader*. Oxford University Press. 107, 111
- Mani, I., B. Schiffman, et J. Zhang (2003). Inferring temporal ordering of events in news. In *Proceedings of the Human Language Technology Conference (HLT-NAACL'03)*. (short paper). 30
- Mani, I., M. Verhagen, B. Wellner, C. M. Lee, et J. Pustejovsky (2006, July). Machine learning of temporal relations. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia, pp. 753–760. Association for Computational Linguistics. 31, 42

- Mani, I., B. Wellner, M. Verhagen, et J. Pustejovsky (2007). Three approaches to learning links in time. Rapport technique, Technical Report CS-07-268, Brandeis University. 46
- Mani, I. et G. Wilson (2000). Robust temporal processing of news. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000)*, New Brunswick, New Jersey, pp. 69–76. 23, 26, 30
- Mann, W. et S. Thompson (1987). Rhetorical structure theory : a theory of text organization. Rapport technique, Information Science Institute. 50
- Mann, W. C. (1988). Dialogue games : Conventions of human interaction. *Argumentation* 2(4), 511–532. 69
- Marcu, D. et A. Echiabi (2002). An unsupervised approach to recognizing discourse relations. In *40th Annual Meeting of the Association for Computational Linguistics*. 55
- Masolo, C., S. Borgo, A. Gangemi, N. Guarino, et A. Oltramari (2003). Wonderweb deliverable d18, ontology library (final). Rapport technique, Laboratory for Applied Ontology, Trento, Italy. 13
- Masolo, C., S. Borgo, A. Gangemi, N. Guarino, et A. Oltramari (2009). WonderWeb Deliverable D18, Ontology Library (final). 21
- Maudet, N., P. Muller, et L. Prévot (2006). Social constraints on rhetorical relations in dialogue. In C. Sidner, J. Harpur, A. Benz, et P. Kühnlein (Eds.), *Proceedings of the Workshop on Constraints in Discourse*, Maynooth, Ireland, pp. 133–139. 68
- Mazur, P. et R. Dale (2007). The DANTE temporal expression tagger. In *Proceedings of the 3rd Language & Technology Conference (LTC)*, Poznan, Poland. 26, 27
- Mihalcea, R., C. Corley, et C. Strapparava (2006). Corpus-based and knowledge-based measures of text semantic similarity. In *AAAI*, pp. 775–780. 83
- Miltsakaki, E., R. Prasad, A. Joshi, et B. Webber (2004). The penn discourse treebank. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*. Citeseer. 51, 52
- Misra, H., O. Cappé, et F. Yvon (2008). Using lda to detect semantically incoherent documents. In *CoNLL '08 : Proceedings of the Twelfth Conference on Computational Natural Language Learning*, Morristown, NJ, USA, pp. 41–48. Association for Computational Linguistics. 80
- Molines, F. (1989). Acceptabilité et interprétation des adverbiaux de localisation temporelle, grammaire ou dictionnaire ? Master's thesis, Université de Toulouse le Mirail. Mémoire de DEA. 26

- Muller, P. (1998). A qualitative theory of motion based on spatio-temporal primitives. In A. Cohn, L. Schubert, et S. Shapiro (Eds.), *Principles of Knowledge Representation and Reasoning : Proceedings of the Sixth International Conference (KR'98)*. Morgan Kaufmann. 15, 17
- Muller, P. (2002). Topological spatio-temporal reasoning and representation. *Computational Intelligence* 18(3), 420–450. 15, 16, 17
- Muller, P. (2007). The temporal essence of spatial objects. In M. Aurnague, M. Hickmann, et L. Vieu (Eds.), *Categorization of Spatial Entities in Language and Cognition*, pp. 285–306. John Benjamins. 20
- Muller, P., S. Afantenos, P. Denis, et N. Asher (2012, December). Constrained decoding for text-level discourse parsing. In *Proceedings of COLING 2012*, Mumbai, India, pp. 1883–1900. The COLING 2012 Organizing Committee. 55
- Muller, P., N. Hathout, et B. Gaume (2006). Synonym extraction using a semantic distance on a dictionary. In D. Boguraev et R. Mihalcea (Eds.), *TextGraphs : the Second Workshop on Graph Based Methods for Natural Language Processing, New York, E.-U., 09/06/06-09/06/06*, <http://www.aclweb.org>, pp. 65–72. Association for Computational Linguistics (ACL). 88
- Muller, P. et P. Langlais (2011). Comparing distributional and mirror translation similarities for extracting synonyms. In C. Butz et P. Lingras (Eds.), *Proceedings of the 24th Canadian Conference on Artificial Intelligence*, LNAI, St John's, Canada, pp. 323–334. Springer. 89
- Muller, P. et L. Prévot (2002). Conversation sous les topiques, du contenu propositionnel à la structure du dialogue. *Information - Interaction - Intelligence* (Hors série), 179–196. 67
- Muller, P. et L. Prévot (2003). An empirical study of acknowledgement structures. In *Proceedings of Diabruck, 7th workshop on semantics and pragmatics of dialogue, Saarbrücken*. <http://www.irit.fr/~Laurent.Prevot/>. 72
- Muller, P. et L. Prévot (2008, march). The rhetorical attachment of questions and answers. In K. Korta et J. Garmendia (Eds.), *Meaning, Intentions, and Argumentation*, Volume 186 of (CSLI-LN) *Center for the Study of Language and Information - Lecture Notes*, pp. 0–0. <http://www.journals.uchicago.edu/> : University of Chicago Press. 67
- Muller, P. et A. Reymonet (2005). Using inference for evaluating models of temporal discourse. In *12th International Symposium on Temporal Representation and Reasoning*, pp. 11–19. IEEE Computer Society Press. 24, 28
- Muller, P. et L. Sarda (1998, Juin). Représentation de la sémantique des verbes de déplacements transitifs du français. In *Actes de la Conférence Traitement Automatique du Langage Naturel (TALN'98)*, Paris. 19

- Muller, P. et L. Sarda (1999). Représentation de la sémantique des verbes de déplacements transitifs du français. *Traitement Automatique des Langues* 39(2), 127–147. 19, 33, 77
- Muller, P. et X. Tannier (2004). Annotating and measuring temporal relations in texts. In *Proceedings of Coling 2004*, Volume I, Genève, pp. 50–56. 24, 26, 28, 29, 30, 38
- Muller, P., M. Vergez-Couret, L. Prévot, N. Asher, B. Farah, M. Bras, A. L. Draoulec, et L. Vieu (2012). Manuel d’annotation en relations de discours du projet annodis. Rapport Technique 21, Carnets de Grammaire — CLLE. 52, 53
- Musan, R. (1999). Temporal interpretation and information-status of noun phrases. *Linguistics and Philosophy* 22, 621–661. 19
- Negri, M. et L. Marseglia (2004). Recognition and normalization of time expressions : ICTirst at TERN 2004. In *TERN 2004 Evaluation Workshop*, Alexandria, VA. Technical Report. 26
- Niles, I. et A. Pease (2001). Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems - Volume 2001*, FOIS '01, New York, NY, USA, pp. 2–9. ACM. 9, 13
- Noonan, H. W. (1976). The four-dimensional world. *Analysis* 37, 32–39. 20
- Parent, G., M. Gagnon, et P. Muller (2008). Annotation d’expressions temporelles et d’événements en français. In F. Béchet (Ed.), *Traitement Automatique des Langues Naturelles (TALN)*, Avignon, 09/06/08–13/06/08, <http://www.atala.org/>, pp. (support électronique). Association pour le Traitement Automatique des Langues (ATALA). 24
- Passonneau, R. (2006). Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation. In *Proc. 5th International Conference on Language Resources and Evaluation (LREC-06)*, pp. 831–836.
- Passonneau, R. J. (1988). A computational model of the semantics of tense and aspect. *Computational Linguistics* 14(2), 44–60. 23, 28
- Péry Woodley, M.-P., N. Asher, P. Enjalbert, F. Benamara, M. Bras, C. Fabre, S. Ferrari, L.-M. Ho Dac, A. Le Draoulec, Y. Mathet, P. Muller, L. Prévot, J. Rebeyrolle, L. Tanguy, M. Vergez Couret, L. Vieu, et A. Widlöcher (2009, 06). ANNODIS : une approche outillée de l’annotation de structures discursives. In *Actes de TALN 2009 TALN 2009 (Conférence sur le Traitement Automatique des Langues Naturelles)*, Senlis France. 52
- Pitler, E., M. Raghupathy, H. Mehta, A. Nenkova, A. Lee, et A. Joshi (2008). Easily identifiable discourse relations. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, Manchester, UK, August. Citeseer. 56
- Poesio, M. et D. R. Traum (1997). Conversational actions and discourse situations. *Computational Intelligence* 13(3), 309–347. 67

- Poesio, M. et D. R. Traum (1998). Towards an axiomatization of dialogue acts. In *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues (13th Twente Workshop on Language Technology)*, pp. 207–222. 72
- Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics* 12. 55, 67
- Polanyi, L. et R. Scha (1984, July). A syntactic approach to discourse semantics. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics*, Stanford, California, USA, pp. 413–419. Association for Computational Linguistics. 50
- Prabhakaran, V., O. Rambow, et M. Diab (2012, June). Predicting overt display of power in written dialogs. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, Montréal, Canada, pp. 518–522. Association for Computational Linguistics. 76
- Prévoit, L. (2004). *Structures sémantique et pragmatiques pour la modélisation de la cohérence dans des dialogues finalisés*. Ph. D. thesis, Université Paul Sabatier. 66, 74
- Pustejovsky, J., P. Hanks, R. Saurí, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, et M. Lazo (2003, March). The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics*, Lancaster University, UK, pp. 647–656. 31
- Pustejovsky, J., R. Ingria, R. Sauri, J. Castano, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, et I. Mani (2005). The specification language TimeML. See [Mani et al. \(2005\)](#). 25
- Pustejovsky, J., J. Littman, R. Saurí, et M. Verhagen (2006). *TimeBank 1.2 Documentation*. <http://www.timeml.org/site/timebank/documentation-1.2.html>. 27
- Pustejovsky, J., J. Moszkowicz, et M. Verhagen (2010). The recognition and interpretation of motion in language. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing*, Volume 6008 of *Lecture Notes in Computer Science*, pp. 236–256. Springer Berlin / Heidelberg. 19
- Pustejovsky, J., R. Saurí, A. Setzer, R. Gaizauskas, et B. Ingria (2002, july). *TimeML Annotation Guidelines*. TERQAS Annotation Working Group. 23
- Quine, W. O. (1960). *Word and Object*. MIT Press. 20
- Reichenbach, H. (1947). *Elements of Symbolic Logic*. New York : McMillan. 23
- Renz, J. et B. Nebel (1997). On the complexity of qualitative spatial reasoning : A maximal tractable fragment of the region connection calculus. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI-97)*, San Francisco, pp. 522–527. Morgan Kaufmann Publishers. 16

- Ringger, E., R. Haertel, et K. Tomanek (Eds.) (2009, June). *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*. Boulder, Colorado : Association for Computational Linguistics. 57
- Rodríguez, A., N. V. de Weghe, et P. D. Maeyer (2004, October). Simplifying Sets of Events by Selecting Temporal Relations. In *Geographic Information Science, Third International Conference, GIScience 2004*, Volume 3234/2004 of *Lecture Notes in Computer Science*, Adelphi, MD, USA, pp. 269–284. Springer Berlin / Heidelberg. 39
- Rossari, C., A. Beaulieu-Masson, C. Cojocariu, et A. Razgouliaeva (2004). *Autour des connecteurs : réflexions sur l'énonciation et la portée*. P. Lang. 72
- Roulet, E., A. Auchlin, J. Moeschler, M. Schelling, et C. Rubattel (1985). *L'articulation du discours en français contemporain*. Berne, Lang (Collection Sciences pour la communication). 3ème éd. 1991. 72
- Roze, C. (2013). *Vers une algèbre des relations de discours*. Ph. D. thesis, Université Paris Diderot, Paris-7. 59
- Roze, C., L. Danlos, et P. Muller (2012). Lexconn : A french lexicon of discourse connectives. *Discours* (10). 60
- Russell, B. (1914). *Our Knowledge of the External World*. London and New York : Routledge. 20
- Šafářová, M., P. Muller, et L. Prévot (2005, 7-9 september). The discourse function of final rises in French dialogues : an extended study. In *International Symposium on Discourse and Prosody as a complex interface*. 74
- Sagot, B. et D. Fišer (2008). Building a free French wordnet from multilingual resources. In *OntoLex*, Marrakech, Maroc. 86
- Saurí, R., R. Knippen, M. Verhagen, et J. Pustejovsky (2005). Evita : A robust event recognizer for QA systems. In *Proceedings of HLT/EMNLP*, Vancouver, pp. 700–707. 27
- Schaeken, W. et P. N. Johnson-Laird (2000). Strategies in temporal reasoning. *Thinking and Reasoning* 6, 193–219. 37
- Schiffrin, D. (1987). *Discourse Markers*. Cambridge University Press. 72
- Setzer, A. (2001, September). *Temporal Information in Newswire Articles : an Annotation Scheme and Corpus Study*. Ph. D. thesis, University of Sheffield, UK. 28, 46
- Shafer, G. (1976). A mathematical theory of evidence. *Princeton University Press*. 18

- Shriberg, E., R. Dhillon, S. Bhagat, J. Ang, et H. Carvey (2004, April 30 - May 1). The icsi meeting recorder dialog act (mrda) corpus. In M. Strube et C. Sidner (Eds.), *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, Cambridge, Massachusetts, USA, pp. 97–100. Association for Computational Linguistics. 76
- Sider, T. (2001). *Four-Dimensionalism : An Ontology of Persistence and Time*. Oxford : Clarendon Press. 20
- Sinclair, J. (2005). Corpus and text-basic principles. In M. Wynne (Ed.), *Developing Linguistic Corpora : a Guide to Good Practice*, Chapitre 1, pp. 1–16. Oxford, UK : Oxbow Books. 52
- Singh, M. et M. P. Singh (1995). The temporal structure of narratives : A semantic approach. In *2nd Conference of the Pacific Association for Computational Linguistics (PACLING)*, Brisbane, Australia. 28
- Song, F. et R. Cohen (1991). Tense interpretation in the context of narrative. In *Proceedings of AAAI'91*, pp. 131–136. 23, 28
- Spooren, W. et T. Sanders (2008). The acquisition order of coherence relations : On cognitive complexity in discourse. *Journal of pragmatics* 40(12), 2003–2026. 52
- Sporleder, C. et M. Lapata (2005). Discourse chunking and its application to sentence compression. In *HLT/EMNLP*. 54
- Sporleder, C. et A. Lascarides (2006). Using automatically labelled examples to classify rhetorical relations : An assessment. *Natural Language Engineering* 14(03), 369–416. 56
- Steedman, M. (1995). Dynamic semantics for tense and aspect. In *IJCAI*, pp. 1292–1298. 28
- Subba, R. et B. Di Eugenio (2009, June). An effective discourse parser that uses rich linguistic information. In *Proceedings of Human Language Technologies : The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Boulder, Colorado, pp. 566–574. Association for Computational Linguistics. 55
- Taboada, M. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics* 38(4), 567 – 592. Focus-on Issue : The Pragmatics of Discourse Management. 56
- Talmy, L. (1983). How language structures space. In H. P. et L. Acredolo (Ed.), *Spatial Orientation : theory, research and application*, pp. 225–282. Plenum, New York. 18
- Tannier, X. et P. Muller (2008, May). Evaluation Metrics for Automatic Temporal Annotation of Texts. In E. L. R. A. (ELRA) (Ed.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. 40, 41, 47
- Tannier, X. et P. Muller (2011). Evaluating temporal graphs built from texts via transitive reduction. *Journal of Artificial Intelligence Research* (40), 375–413. 41

- Tatu, M. et M. Srikanth (2008, August). Experiments with reasoning for temporal relations between events. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, Manchester, UK, pp. 857–864. Coling 2008 Organizing Committee. 31, 42
- Taylor, H. et B. Tversky (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language* (31), 261–292. 65
- Teissèdre, C., D. Battistelli, et J.-L. Minel (2010, may). Resources for calendar expressions semantic tagging and temporal navigation through texts. In N. C. C. Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, et D. Tapias (Eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA). 26
- Tonhauser, J. (2002). A dynamic semantic account of the temporal interpretation of noun phrases. In B. Jackson (Ed.), *Proceedings of Semantics and Linguistics Theory XII (SALT XII)*, Cornell. CLC Publications. 19
- Traum, D. (1994). *A computational theory of grounding in natural language conversation*. Ph. D. thesis, University of Rochester. 71
- Traum, D. et J. F. Allen (1994). Discourse obligations in dialogue processing. In *Proceedings of the 32nd annual meeting of the Association for Computational Linguistics*, pp. 1–8. 66
- Turney, P. D. et P. Pantel (2010). From frequency to meaning : Vector space models of semantics. *J. Artif. Intell. Res. (JAIR)* 37, 141–188. 87
- UzZaman, N. et J. Allen (2011, June). Temporal evaluation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies*, Portland, Oregon, USA, pp. 351–356. Association for Computational Linguistics. 47
- UzZaman, N., H. Llorens, L. Derczynski, J. Allen, M. Verhagen, et J. Pustejovsky (2013, June). Semeval-2013 task 1 : Tempeval-3 : Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2 : Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pp. 1–9. Association for Computational Linguistics. 23, 46
- Van de Cruys, T., S. Afantenos, et P. Muller (2013a). MELODI : A Supervised Distributional Approach for Free Paraphrasing of Noun Compounds. In *International Workshop on Semantic Evaluation (SEMEVAL)*, Atlanta, GA. Association for Computational Linguistics (ACL). 91
- Van de Cruys, T., S. Afantenos, et P. Muller (2013b). MELODI : Semantic Similarity of Words and Compositional Phrases using Latent Vector Weighting . In *International Workshop on Semantic Evaluation (SEMEVAL)*, Atlanta, GA. Association for Computational Linguistics (ACL). 91

- van Deemter, K. et R. Kibble (2000). On coreferring : Coreference in MUC and related annotation schemes. *Computational Linguistics* 26(4), 629–637. 35
- Van der Plas, L. (2008). *Automatic lexico-semantic acquisition for question answering*. Ph. D. thesis, University of Groningen. 86
- van der Plas, L. et J. Tiedemann (2006, July). Finding synonyms using automatic word alignment and measures of distributional similarity. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, Sydney, Australia, pp. 866–873. Association for Computational Linguistics. 89
- Vandeloise, C. (1986). *L'espace en français*. Paris : Seuil. 18
- Venant, A., N. Asher, P. Muller, P. Denis, et S. Afantenos. Expressivity and comparison of models of discourse structure. In *Proceedings of SIGDIAL 2013*. 56
- Verhagen, M., R. Gaizauskas, F. Schilder, M. Hepple, G. Katz, et J. Pustejovsky (2007). SemEval-2007 - 15 : TempEval Temporal Relation Identification. In *Proceedings of SemEval workshop at ACL 2007*. 23, 38
- Verhagen, M., R. Sauri, T. Caselli, et J. Pustejovsky (2010). Semeval-2010 task 13 : Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 57–62. Association for Computational Linguistics. 23
- Vieu, L. (1991). *Sémantique des relations spatiales et inférences spatio-temporelles : une contribution à l'étude des structures formelles de l'espace en langage naturel*. Ph. D. thesis, Université Paul Sabatier, Toulouse. 14
- Vilain, M., J. Burger, J. Aberdeen, D. Connolly, et L. Hirschman (1995). A model-theoretic coreference scoring scheme. In *MUC6 '95 : Proceedings of the 6th conference on Message understanding*, Columbia, Maryland, USA, pp. 45–52. 35, 39
- Vilain, M., H. Kautz, et P. van Beek (1990). Constraint propagation algorithms for temporal reasoning : a revised report. In *Readings in qualitative reasoning about physical systems*, pp. 373–381. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc. 31, 40
- Wahlster, W. (Ed.) (2000). *Verbmobil : Foundations of Speech-to-Speech Translation*. Artificial Intelligence. Springer. 55
- Walker, M. A. (1992). Redundancy in collaborative dialogue. In *COLING'92*. 72
- Walton, D. N. et E. C. W. Krabbe (1995). *Commitment in dialogue : basic concepts of interpersonal reasoning*. State University of New York Press. 66
- Wan, S. et C. Paris (2008). Experimenting with clause segmentation for text summarization. In *Proceedings of the 1st TAC*, Gaithersburg, MD. 54

- Wang, L., M. Lui, S. N. Kim, J. Nivre, et T. Baldwin (2011, July). Predicting thread discourse structure over technical web forums. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, UK., pp. 13–25. Association for Computational Linguistics. 76
- Watts, D. et S. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature* (393), 440–442. 87
- Webber, B., A. Popescu-Belis, K. Markert, et J. Tiedemann (Eds.) (2013, August). *Proceedings of the Workshop on Discourse in Machine Translation*. Sofia, Bulgaria : Association for Computational Linguistics. 94
- Webber, B. L. (1988). Tense as discourse anaphor. *Computational Linguistics* 14(2), 61–73. 23, 24, 28, 50
- Weeds, J., D. Weir, et D. McCarthy (2004, Aug 23–Aug 27). Characterising measures of lexical distributional similarity. In *Proceedings of Coling 2004*, Geneva, Switzerland, pp. 1015–1021. COLING. 86
- Whitehead, A. (1929). *Process and reality*. New-York : Mac Millan. 20
- Wolf, F. et E. Gibson (2005). Representing discourse coherence : A corpus-based study. *Comput. Linguist.* 31(2), 249–288. 51
- Wolf, F. et E. Gibson (2006). *Coherence in Natural Language : Data Structures and Applications*. The MIT Press. 55
- Wu, H. et M. Zhou (2003). Optimizing synonyms extraction with mono and bilingual resources. In *Proceedings of the Second International Workshop on Paraphrasing*, Sapporo, Japan. Association for Computational Linguistics. 89
- Zaenen, A. (2006). Mark-up barking up the wrong tree. *Computational Linguistics* 32(4), 577–580. 46
- Zwaan, R. A. et G. A. Razdvansky (2001). Time in narrative comprehension. In D. Schram et G. Steen (Eds.), *Psychology and Sociology of Literature*. Amsterdam : John Benjamins. 49