



HAL
open science

Sémantique des temps du français : une formalisation compositionnelle

Anais Lefeuvre

► **To cite this version:**

Anais Lefeuvre. Sémantique des temps du français : une formalisation compositionnelle. Informatique. Université de Bordeaux, 2014. Français. NNT : 2014BORD0065 . tel-01136420

HAL Id: tel-01136420

<https://theses.hal.science/tel-01136420>

Submitted on 27 Mar 2015

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.

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR DE
L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE MATHÉMATIQUES-INFORMATIQUE
SPÉCIALITÉ INFORMATIQUE

Par Anaïs LEFEUVRE

**SÉMANTIQUE DES TEMPS DU FRANÇAIS :
UNE FORMALISATION COMPOSITIONNELLE**

Sous la direction de : Christian RETORÉ
(co-directeur : Mauro, GAIO)

Soutenue le 23 Juin 2014

Membres du jury :

M. HANNUSSE, Nicolas	Directeur de Recherche	LaBRI	Président
M. ENJALBERT, Patrice	Professeur des Universités émérite	Université de Caen	Rapporteur
M. VERKUYL, Henk	Professeur des Universités émérite	Université d'Utrecht	Rapporteur
M. ANTOINE, Jean-Yves	Professeur des Universités	Université François Rabelais de Tours	Examineur
Mme BATTISTELLI, Delphine	Professeur des Universités	Université de Paris X	Examineur
M. MOOT, Richard	Chargé de Recherches	LaBRI	Examineur

Titre : Sémantique des Temps du français : une formalisation compositionnelle

Résumé : Cette thèse s'inscrit dans le cadre du projet Région Aquitaine - INRIA : ITIPY. Ce projet vise à terme l'extraction automatique d'itinéraires à partir de récits de voyage du XIX^{ème} et du début du XX^{ème} siècle. Notre premier travail fut de caractériser le corpus comme échantillon du français, par une étude contrastive d'une part de données quantitatives et d'autre part de la structure des récits de voyage. Nous nous sommes ensuite consacrée à l'étude du temps, et plus particulièrement à l'analyse automatique de la sémantique des temps verbaux du français. Disposant d'un analyseur syntaxique et sémantique à large échelle du français, basé sur les grammaires catégorielles et la sémantique compositionnelle (λ -DRT), notre tâche a été de prendre en compte les temps des verbes pour reconstituer la temporalité des événements et des états, notions regroupées sous le terme d'éventualité. Cette thèse se concentre sur la construction d'un lexique sémantique traitant des temps verbaux du français. Nous proposons une extension et une adaptation d'un système d'opérateurs compositionnels conçu pour les temps du verbe anglais, aux temps et à l'aspect du verbe français du XIX^{ème} siècle à nos jours. Cette formalisation est *de facto* opérationnelle, car elle est définie en terme d'opérateurs du λ -calcul dont la composition et la réduction, déjà programmées, calculent automatiquement les représentations sémantiques souhaitées, des formules multisortes de la logique d'ordre supérieur. Le passage de l'énoncé comportant une éventualité seule au discours, dont le maillage référentiel est complexe, est discuté et nous concluons par les perspectives qu'ouvre nos travaux pour l'analyse du discours.

Mots clés : Sémantique formelle - Temporalité des éventualités - Lexique compositionnel - Récits de voyage - λ -calcul simplement typé

Title : French Tenses Semantics: a Compositionnal Formalisation

Abstract : This work has been lead in the frame of the ITIPY project which goal was to automatically extract itineraries from travel novels from the XIXth century and from the beginning of the XXth. Our thesis work is close to the text understanding task in the information retrieval field and we aim at building a representation of meaning of linguistic utterances, leaning on the compositionality principle. More precisely, the itinerary extraction supposes to temporally represent displacement and localization events or states (that we actually call eventualities) of a traveler as far as we understand it through discourse. Working on an automatic parser for syntax (in categorial grammars) and semantics (in λ -DRT), we focused on the building of an semantic lexicon for tense in French. We actually characterized our corpora as a sample of French language, by the means of a quantitative and qualitative analysis as well as a study of the internal structure of this genre. The main contribution of this work deals with tense and aspect semantic processing of the event expressed by tensed verbs, and with its modelling. In this respect, we propose an adaptation and

an extension for French from XIXth century to nowadays of a lexicon originally produced to deal with English verbs. This formalisation is operational, for it is defined in λ -calculus which composition and réduction, already implemented, calculate automatically semantics représentations, high order logic formulas. Transition from a single event uterrance to a whole discourse which contains a complex referential network is discussed and allows us to define the limits of this hereby work.

Keywords : Formal Semantics - Eventualities temporality - Compositionnal Lexicon - Travel novels - Simply typed λ -calcul

Unité de recherche

[Laboratoire Bordelais de Recherche en Informatique, UMR 5800, 351 cours de la Libération 33405 Talence]

Remerciements

Je remercie très sincèrement Henk Verkuyl pour avoir accepté d'évaluer mon travail, et plus particulièrement pour les lumières qu'il a su porter sur un objet qui me fascine, le temps. Je ne trouverai pas les mots pour qualifier l'admiration que j'ai pour son travail. Je remercie également Patrice Enjalbert pour avoir accepté d'être rapporteur de mon manuscrit et pour l'intérêt qu'il a porté à mon travail dans la phase d'aboutissement. Je le remercie pour la précision et la qualité scientifique de ces remarques et questions qui ont permis d'élever mon point de vue sur mes recherches.

Je tiens ensuite à remercier mon directeur de thèse, Christian Retoré, qui m'a appris tellement pendant ces années de thèse. Je le remercie pour l'humanité avec laquelle il a guidé les pas de la jeune linguiste au travers de ce qui n'était pour moi alors qu'une forêt de symboles, m'a aidé à m'approprier ce territoire. Je remercie Richard Moot pour son temps et sa patience, pour avoir secondé parfaitement Christian dans mon apprentissage et dans la direction de ma thèse. Je tiens particulièrement à remercier Mauro Gaio sans qui tout ceci n'aurait pas été possible, et dans l'espoir de collaborations futures.

Je souhaite remercier Delphine Battistelli d'avoir accepté de relire ce travail et pour nos discussions à certains moments cruciaux de ma thèse. Je remercie Nicolas Hanusse pour les actions menées en faveur de tous les doctorants au sein de l'école doctorale et particulièrement en la mienne : j'en ai bénéficié directement à plusieurs moments importants de mon doctorat.

Je remercie vivement Jean-Yves Antoine, pour son soutien indéfectible et inquantifiable. Je le remercie pour son regard sur mes travaux et pour l'opportunité qu'il m'a donné de travailler avec lui à Blois.

Je souhaite ensuite adresser mes remerciements à tous ceux qui ont été présents à différents moments de ce cheminement¹, tout d'abord Miriam Bras et Philippe Muller de Toulouse qui ont su cerner très rapidement où j'en étais et m'orienter quand j'en avais besoin, je souhaite remercier Patrick Caudal pour nos brefs échanges parisiens, échanges qui ont aussi constitué un repère pour moi dans cette réflexion.

Je remercie ensuite les membres des équipes qui m'ont toujours très bien accueillie, pour la confiance qu'ils m'ont accordée ou m'accordent actuellement et pour tous les échanges enrichissants que nous avons pu avoir, je pense à Sylvain Salvati et Lionel

1. Je prie de m'excuser ceux que j'ai pu oublier, ou ceux que je n'ai pas la place de nommer...

Clément de SIGNES INRIA - LABRI, à Géraud Sénizergues et Pierre Casteran pour leur pédagogie et leurs enseignements, à Gilles Boyé, Ana Kupsc, Joan Busquets et toute l'équipe de l'ERSSàB, à Denis Maurel, Agata Savary et toute l'équipe du LI de François Rabelais, et merci à Emmanuel Schang, Iris Eshkol et Lotfi Abouda du LLL de l'université d'Orléans, sans oublier tous les personnels associés à ces équipes ou laboratoires.

Je pense également à Natalia Vinogradova avec qui j'ai eu l'honneur de réaliser mes premiers travaux de thèse et pour son amitié, à Pierre Bourreau et Livy-Maria Real-Coelho tant pour nos échanges amicaux que professionnels, à Noémie-Fleur Sandillon-Rezer, Farès Chukri, Yakub Waszczuk, Bruno Mery, Mathieu Morey, Jérôme Kirman, Ophélie Lacroix, Sandra Augendre et Christelle Delarue, docteurs ou sur le point de le devenir dans des domaines qui nous ont permis de nous rencontrer pour mon plus grand plaisir.

Je remercie ma famille : mes parents et mon frère pour m'avoir plus d'une fois soutenue quand il devenait difficile d'avancer, pour m'avoir inculqué tout ce dont j'avais besoin pour ne pas fléchir, et Luce Lefeuvre, ma cousine pour avoir été là et pour son aide le jour de ma soutenance. Bonne chance à elle pour la suite.

Je remercie Pierre Halftermeyer d'être là, simplement. Aucun mot ne pourrait qualifier l'impact de sa présence dans ma vie.

Je remercie Thomas Morsellino et sa compagne Laure, pour avoir accompagné mon arrivée dans le monde académique bordelais, et pour leur amitié qui m'est chère.

Je remercie Julien Aligon pour l'accueil blaisois et cette année partagée "sur la colline", bonne chance à lui aussi pour la suite.

Je remercie le "groupe du midi" d'alors, "Appellation Origine Campus", "Aquidoc", l'"Afodib", et l'équipe étendue du "13" : ils se reconnaîtront et sont vraiment trop nombreux pour être tous nommés. Ils savent tous qu'ils ont joué un rôle crucial pour moi.

J'ai une dernière pensée pour Julien Longhi et Michel Téguia qui m'ont donné la possibilité d'avancer dans des directions qui se rejoignent aujourd'hui, merci à eux.

Introduction générale

Les recherches en sémantique formelle au service du traitement automatique des langues naturelles sont animées par deux ambitions. La première est de déterminer la part automatisable et systématique de cette compétence si humaine qu'est la langue, tandis que la seconde est de connaître jusqu'à quel point on peut implanter à la machine un peu de cet outil si créatif. Quand pourrons-nous tous dialoguer avec notre ordinateur en langue courante comme si nous parlions à un de nos congénères ?

D'un point de vue linguistique, il semble impossible de saisir l'intégralité des subtilités d'une langue ne serait-ce qu'à un moment donnée de son évolution, tout simplement parce que nous ne dominons pas la totalité des phénomènes. Néanmoins, une partie des règles sous-jacentes à la production linguistique ont été dévoilées et sont aujourd'hui utilisées en traitement automatique du langage. A partir de ce constat on peut espérer continuer à étendre la partie automatisable de l'analyse du sens d'une production linguistique d'une part, et affiner les informations extraites des analyses automatiques d'autre part. C'est à ce dernier aspect que participent nos travaux de thèse, et plus généralement le projet dans lequel ils s'inscrivent.

Suite à la numérisation d'un grand nombre d'oeuvres littéraires sur la région paloise par la Médiathèque Intercommunale à Dimension Régionale (MIDR) de Pau, la volonté de valoriser cette ressource a permis de mettre en place la collaboration entre plusieurs équipes de recherche sous le nom de projet ITIPY, pour Itinéraires Pyrénéens. L'axe central de ce projet est l'extraction automatique d'itinéraires décrits dans un corpus de onze récits de voyage datant du XIXème et début XXème siècle formé à partir de ce fond patrimonial littéraire. Ce projet a été financé par INRIA et la région Aquitaine et rassemble les équipes suivantes :

- SIGNES (signes linguistiques, grammaire et sens : algorithmique logique de la langue), de l'INRIA Bordeaux – Sud-Ouest, localisée à Bordeaux,
- DESI (document électronique, sémantique et interaction), du LIUPPA (laboratoire d'informatique de l'université de Pau et des pays de l'Adour), localisée à Pau,
- LILaC (logic, interaction, language, and computation), de l'IRIT (institut de re-

cherche en informatique de Toulouse), localisée à Toulouse.

L'extraction d'itinéraires à partir de textes bruts est une vaste tâche qui s'inscrit dans le traitement automatique de la langue naturelle et nécessite de mobiliser tant des connaissances linguistiques sur l'objet à traiter, que des connaissances informatiques sur les méthodes à utiliser.

Un itinéraire, selon nous, est le fruit de l'interaction entre un voyageur, le temps et l'espace. Plus précisément, l'extraction d'itinéraire revient à ordonner temporellement les déplacements et localisations d'un voyageur tel qu'on peut se le représenter à travers le discours.

Nos travaux de thèse se rapprochent de la compréhension de texte en extraction d'information. Nous cherchons à construire une représentation du sens des énoncés, en nous appuyant sur les principes de la sémantique formelle compositionnelle. Nous entendons par là que la sémantique d'un énoncé est une représentation logique construite à partir de la forme syntaxique de celui-ci et du sens des unités qui le composent décrites au sein d'un lexique sémantique.

Cette thèse se situe donc en sémantique formelle et se concentre sur la construction d'un lexique sémantique de la temporalité des éventualités en français, éventualités étant le terme générique que nous utilisons pour désigner l'ensemble des événements et états.

La première étape de ce travail a été de caractériser le corpus comme échantillon du français, ce qui a nécessité une étude quantitative et qualitative dans laquelle les onze textes ont été confrontés à d'autres genres discursifs, ainsi qu'une approche de la structure interne de ce genre de discours afin d'observer le comportement du récit de l'itinéraire au sein de celui-ci. La seconde partie de ce travail est une contribution au traitement sémantique du temps et de l'aspect des éventualités et à leur modélisation. En effet, le traitement de la temporalité reste une question cruciale en extraction d'information, et trouve des applications tant en analyse qu'en génération. Par ailleurs, le système verbal en français étant relativement stable du point de vue aspectuo-temporel depuis la fin du XVII^{ème} siècle, le lexique est donc tout à fait approprié pour être utilisé sur du français contemporain.

Nous présentons dans les quatre premiers chapitres les principes en sémantiques formelles sur lesquels nous nous appuyons, la plateforme d'analyse syntactico-sémantique *Grail*, développée par Richard Moot. Plus précisément, cette plateforme offre une analyse syntaxique complètement automatique de la syntaxe des énoncés. Dans la tradition compositionnelle, pour chaque mot étiqueté syntaxiquement, on puise dans le lexique sémantique la représentation correspondant au mot sous la forme d'un λ -terme. Le lexique sémantique associe à chaque mot un ou plusieurs λ -terme dont le type correspond à la ou aux catégories syntaxiques de ce même mot dans le cadre des grammaires catégorielles de Lambek. Un tel λ -terme combiné aux autres suivant la syntaxe, donne la représentation sémantique de la phrase. Nos travaux s'inscrivent dans la lignée de nombreuses études sur le traitement temporel des éventualités dont nous tentons de présenter quelques unes des plus pertinentes. Nous proposons ensuite une introduction

à la théorie d'analyse de la temporalité verbale développée par Verkuyl dans son ouvrage *Binary Tense* et un modèle d'interprétation, sur lesquels nous basons nos travaux.

La seconde partie expose les résultats de ce travail, à savoir une analyse discursive généraliste du corpus et une analyse du comportement des segments dédiés à la description de l'itinéraire au sein de celui-ci. Est développée une adaptation et une extension du système *Binary Tense* afin qu'il soit utilisable pour le français du XIX^{ème} siècle à nos jours, tant au niveau temporel qu'au niveau aspectuel. Nous concluons par une évaluation de ce travail qui met en lumière les limites de celui-ci et ouvre les perspectives du passage d'un énoncé comportant une éventualité seule au discours qui présente un maillage référentiel complexe.

Table des matières

Remerciements	i
Introduction générale	iii
1 Sémantique et traitement automatique des langues naturelles	5
1.1 Introduction	5
1.2 L'informatique et la linguistique pour le traitement automatique de la langue	6
1.3 Définitions pour une sémantique formelle	8
1.3.1 Le discours	8
1.3.2 La dénotation et le sens	9
1.3.3 La compositionnalité	10
1.4 Une représentation logique du discours, la <i>Discourse Representation Theory</i> (DRT)	10
1.4.1 La notion de modèle	10
1.4.2 L'analyse d'un énoncé, la construction d'un univers du discours en <i>Discourse Representation Theory</i>	11
1.5 Conclusion	14
2 Grammaire Catégorielle et λ-calcul : Focus sur le point de vue compositionnel	15
2.1 Introduction	15
2.2 L'analyse syntaxique et les grammaires catégorielles	16
2.2.1 Les grammaires catégorielles classiques	16
2.2.2 Le calcul de Lambek	17
2.2.3 Une grammaire de Lambek	19
2.3 Interface syntaxe-sémantique	20
2.3.1 Sémantique de Montague et calcul de Lambek	20
2.3.2 Lexique et type	22
2.3.3 Construction des termes	22
2.3.4 Construction de l'analyse sémantique	22
2.3.5 Logique du premier ordre multisorte	23
2.3.6 Un exemple simple	26
2.4 Une implémentation : <i>Grail</i>	26

2.4.1	Architecture générale	26
2.4.2	Le traitement de la syntaxe	27
2.4.3	La représentation sémantique en DRT	29
2.5	Conclusion	33
3	De la temporalité des éventualités pour la sémantique formelle	35
3.1	Introduction	36
3.2	La temporalité dans le discours	36
3.2.1	Enonciation, discours et énoncé	36
3.2.2	Eventualité	37
3.2.3	Temps et aspect	38
3.3	Système de calcul de la temporalité verbale et modèle d'interprétation de la temporalité	39
3.3.1	Reichenbach et la description des temps verbaux	39
3.3.2	Les intervalles de Allen	40
3.4	La structure des éventualités	42
3.4.1	Une première typologie des éventualités	42
3.4.2	De l'interaction entre structure interne des éventualités et agenticité	44
3.4.3	Les transitions entre types d'éventualités	47
3.5	Deux modèles d'interprétation temporels et aspectuels pour le français	49
3.5.1	Une théorie de la sémantique temporelle du français : la Sémantique des Temps	49
3.5.2	Une typologie aspectuelle des éventualités du français	51
3.6	Du traitement de l'éventualité au traitement du discours : la DRT et la SDRT	55
3.6.1	Le dispositif pour la temporalité en DRT	55
3.6.2	Une nouvelle granularité en DRT, la SDRT (<i>Segmented Discourse Representation Theory</i>)	57
3.7	Conclusion	58
3.7.1	Synthèse	58
3.7.2	Spécifications de nos travaux	60
4	Présentation et interprétation d'un système bien adapté à la sémantique compositionnelle : <i>Binary Tense</i>	63
4.1	Introduction	63
4.2	Nos définitions	65
4.3	Le langage logique typé	70
4.3.1	Les éventualités et la λ -DRT	70
4.3.2	Le typage du langage	72
4.3.3	Les variables, constantes, fonctions et relations	74
4.3.4	Les termes	76
4.4	La correspondance entre combinaison d'opérateurs et temps verbal	82
4.4.1	Tableau des combinaisons	82
4.4.2	Le présent de l'indicatif	83
4.4.3	Le futur simple	84

4.4.4	L'imparfait	85
4.4.5	Le conditionnel présent	86
4.4.6	Le passé composé	87
4.4.7	Le futur antérieur	88
4.4.8	Le plus-que-parfait	89
4.4.9	Le conditionnel passé	90
4.4.10	Le passé simple	91
4.4.11	Le passé antérieur	92
4.4.12	Les adverbiaux	92
4.5	Conclusion	94
5	Mise en perspective de la méthode : l'itinéraire, le récit de voyage au XIX^{ème} siècle et analyse sémantique du segment	95
5.1	Introduction	95
5.2	Le projet ITIPY et l'itinéraire	96
5.2.1	Le projet ITIPY	96
5.2.2	Les travaux antérieurs sur le corpus Itipy : l'itinéraire dans le texte	97
5.3	Définitions	98
5.4	Notre corpus et le récit de voyage au XIX ^{ème} siècle	100
5.4.1	La méthode	100
5.4.2	Les types fonctionnels	107
5.5	La structure du discours au sein du récit de voyage	109
5.5.1	La segmentation du discours dans le cadre de la SDRT	109
5.5.2	Segmentation et relations discursives dans le RV	110
5.5.3	Proposition d'une nouvelle relation discursive	112
5.6	Le segment type : quelle approche ?	113
5.7	Narration, description et segment dédié à l'itinéraire dans récit de voyage du XIX ^{ème} siècle	116
5.8	Conclusion	118
6	Adaptation du système <i>Binary Tense</i>	119
6.1	Introduction	119
6.2	Définition de l'aoriste	120
6.3	Rappel du système <i>Binary Tense</i> et évaluation pour le français	122
6.3.1	<i>Binary Tense</i> , l'aspect et l'énonciation	122
6.3.2	Evaluation du système pour le français contemporain	128
6.3.3	Le récit de voyage du XIX ^{ème} siècle et le français contemporain	133
6.4	Adaptation du système : l'aoriste	136
6.4.1	Aoriste et énonciation	136
6.4.2	Aoriste et aspect	138
6.4.3	L'opérateur AOR et le système	138
6.4.4	La relation PERF, la relation AOR et l'enjeu adverbial	141
6.5	Interprétation	142
6.6	Conclusion	148

7	Extension du système	151
7.1	Introduction	151
7.2	Définitions	152
7.3	Les adverbiaux temporels	153
7.3.1	Spécification dans <i>Binary Tense</i>	153
7.3.2	Classification des expressions temporelles adverbiales	160
7.3.3	Le lexique des adverbiaux	162
7.4	Aspect lexical et information spatiale	169
7.4.1	Pertinence du calcul sémantique de la temporalité et informa- tion spatiale des verbes de déplacement et localisation	169
7.4.2	Eventualités, aspect lexical et structure sémantique spatiale	173
7.5	Conclusion	176
8	Evaluation et Limites	177
8.1	Introduction	177
8.2	Synthèse du système	178
8.2.1	Syntaxe, sémantique et compositionnalité	178
8.2.2	Couverture du système	183
8.3	Limites du système	186
8.3.1	Structure discursive et sous-spécification	186
8.3.2	Perspectives : l'inchoatif et le terminatif et l'itératif	187
8.4	Conclusion	190
	Conclusion	191
	Bibliographie	vi
	Résumé et mots-clefs	vii
	Titre	vii
	Résumé	vii
	Mots-Clefs	vii
	Title	viii
	Abstract	viii
	Keywords	viii

Chapitre 1

Sémantique et traitement automatique des langues naturelles

Ce premier chapitre nous permet d'introduire le traitement automatique des langues naturelles et de situer nos travaux au sein de celui-ci. Nous présentons la dimensions informatique et linguistique du traitement automatique des langues et donnons quelques définitions utiles à la caractérisation de notre objet d'étude. Nous clôturons en présentant brièvement le cadre de la *Discourse Representation Theory* dans laquelle se situent nos travaux en sémantique.

1.1 Introduction

Le premier chapitre de ce mémoire introduit le traitement automatique de la langue naturelle et le champ d'investigation qu'est la sémantique formelle. Il donne les définitions et considérations principales nécessaires à la compréhension de cette discipline sans pour autant entrer dans les détails techniques propres à chaque aspect présenté. Nous introduisons notre cadre théorique, la *Discourse Representation Theory* brièvement et illustrons par un exemple standard les possibilités d'un tel cadre.

1.2 L'informatique et la linguistique pour le traitement automatique de la langue

L'objectif de l'automatisation est d'opérer des actions, en un temps restreint et sur une quantité d'objets pouvant être massive. L'informatique est un cas particulier d'automatisation, c'est le traitement automatique d'informations. Par ailleurs, l'informatique est fortement attachée à la notion de machine, ainsi en traitement automatique de la langue, on demande à la machine informatique d'exécuter un programme structuré d'actions sur un objet linguistique, afin d'en obtenir une forme modifiée, et ce avec le moins de supervision humaine possible. On cherche alors à définir quel est l'objet à traiter, quel traitement opérer sur cet objet et par conséquent quelle forme obtenir après traitement. On parlera alors d'entrée, de traitement ou programme et de sortie.

L'origine du traitement automatique du langage, désormais TAL, remonte à l'origine de l'informatique, c'est à dire les années 40 [Cori et Léon, 2002]. Au départ c'est le rêve d'un traducteur complètement automatique qui pousse les chercheurs à approfondir l'automatisation de l'analyse linguistique. Le TAL a diversifié de manière générale ses domaines de recherche, la traduction automatique restant un de ces enjeux, on peut y ajouter la recherche d'information, l'extraction d'information, les systèmes question-réponse, la correction orthographique, le traitement de la parole et le dialogue homme-machine par exemple.

Les sciences du langage visent la description des mécanismes sous-jacents à la production langagière et sont hiérarchiquement organisés, nous donnons quelques pendants applicatifs TAL à cinq grand niveaux d'analyse en linguistique¹.

- pragmatique : l'extraction d'information, la recherche d'information, le dialogue homme-machine
- sémantique : l'extraction d'information, la recherche d'information, la traduction automatique
- syntaxe : la correction orthographique et grammaticale, l'extraction d'information, la recherche d'information, la traduction automatique
- morphologie : la correction orthographique, l'auto-completion
- phonétique /phonologie : traitement de la parole

Le langage est le produit de l'interaction entre tous ces niveaux. Chacun de ces niveaux ainsi que chacune des interfaces entre deux d'entre eux est le champs d'investigation de recherches en sciences du langage. Cette hiérarchie des niveaux offre des cadres pour décrire le plus pertinemment possible les règles régissant l'organisation de la langue et du discours.

Par ailleurs, il est primordial de distinguer langue et discours. Par langue on entend le système à proprement dit, le vocabulaire d'un code linguistique ainsi que les règles qui régissent ce code, et par discours on entend la réalisation linguistique d'énoncés en contexte, à partir de ce système. Ainsi pour traiter au mieux la langue naturelle, il convient de prendre en compte dans la mise au point d'un système artificiel (par opposition à naturel) les phénomènes discursifs. Par exemple si l'on tente d'accorder une représentation sémantique à l'énoncé :

(1.1) Pars d'ici tout de suite !

1. Cette liste n'a pas la prétention d'être exhaustive.

1.2. L'INFORMATIQUE ET LA LINGUISTIQUE POUR LE TRAITEMENT AUTOMATIQUE DE LA LANGUE

il convient d'accorder une valeur à *un locuteur à l'origine de l'énoncé, un interlocuteur recevant l'énoncé et un lieu désigné par ici*, ainsi qu'à la volonté pragmatique de l'énonciateur dans l'utilisation de l'impératif par exemple. Ici il faudra donc prendre en compte un niveau d'analyse pragmatique pour traiter correctement cet énoncé.

Nos travaux se situent dans le domaine sémantique, et ont pour visée de traduire le plus finement possible à partir d'un énoncé au sein d'un discours, sa représentation sémantique. La sémantique peut se définir comme étant l'étude du sens linguistique, c'est à dire l'observation des mécanismes à l'œuvre dans la production et l'interprétation d'un énoncé, d'un discours. Le « sens » étant porté par des formes linguistiques il convient d'observer plusieurs niveaux d'analyse de ces formes, allant du morphème aux expressions complexes, et de leur localisation en système à leur utilisation en discours.

L'étude de ces unités linguistiques nécessite de prendre en compte leur situation en système, c'est à dire au sein de la langue, ce qui les oppose les uns aux autres, on parlera alors d'axe paradigmatique. Tandis que leur mise en discours, c'est à dire sur l'axe syntagmatique, doit être prévue dans la construction du lexique compositionnel afin d'obtenir une interprétation correcte une fois l'analyse d'un énoncé terminée. La sémantique compositionnelle s'appuie sur l'interface syntaxe-sémantique et par conséquent sur la syntaxe pour que les deux axes, paradigmatique et syntagmatique, soient pris en compte dans la représentation du sens.

Une autre distinction utile à la présentation du TAL concerne la méthode utilisée pour spécifier les actions à exécuter pendant le traitement. Il est possible d'utiliser une approche "centrée données", qui demande à la machine d'apprendre par elle-même, en lui donnant des informations formalisées en entrée, et en lui fournissant ce qu'on souhaite obtenir en sortie pour qu'elle déduise par elle-même les règles à observer ou plutôt comment les observer. Ici nous parlerons d'approche statistique, probabiliste ou encore d'apprentissage. La seconde approche consiste à fournir à la machine exactement les actions à réaliser en lui donnant en entrée des informations pour obtenir en sortie ces informations traitées. On désigne cette approche par "système symbolique" ou encore "système à base de règles". Un tel système permet de décrire finement les mécanismes linguistiques sous-jacents sous forme de règles et semble approprié pour traiter de la sémantique.

Les systèmes à base de règles, sont construits sur :

- un ensemble de faits qui sont constitués dans notre cas par le lexique sémantique et la représentation syntaxique de la phrase à représenter sémantiquement,
- un ensemble de règles, qui, dans notre cas sont les règles communes à l'analyse syntaxique et à la représentation sémantique

Notre projet est un projet d'analyse, plus précisément localisé en sémantique formelle et s'appuyant sur la syntaxe de l'énoncé. Notre approche en sémantique propose d'enrichir le lexique de *Grail*, analyseur syntaxique et sémantique à large échelle du français, permettant d'obtenir entièrement automatiquement la représentation sémantique d'un énoncé à partir du texte brut. *Grail*, développé et maintenu par Richard Moot [Moot, 1999], prend en entrée du texte brut pour atteindre un niveau de représentation logico-sémantique, en passant par les niveaux morphologique et syntaxique. On peut donc dire que notre contribution au sein de ce projet se situe plus précisément en

sémantique, s'appuyant sur les traitements syntaxiques déjà finalisés de l'outil. Nous reviendrons dans le chapitre suivant sur la présentation précise des méthodes de calcul employées.

1.3 Définitions pour une sémantique formelle

Afin de montrer au mieux les enjeux de notre proposition, nous donnons tout d'abord quelques définitions des objets de notre étude et des concepts utiles à la construction de celle-ci.

1.3.1 Le discours

Le discours est un ensemble de mots qui peut être étudié sous divers points de vue, nous ne nous intéressons pas ici à la dimension sociale, psychologique ou encore phonologique, mais bien à la création linguistique de sens et à l'organisation interne du discours. Nous reprendrons à titre introducteur la définition de Ramsay :

"A discourse is the extended sequence of sentences produced by one or more people with the aim of conveying or exchanging information." [Ramsay, 2005]

Notre conception du discours regroupe différentes propriétés dont celle présentée par Ramsay qu'est la visée du discours. En effet pour chaque discours, on reconnaît une visée communicative, plus ou moins spécifique à un auditoire, ou à un objectif poursuivi, mais pas seulement, on pense par exemple à l'élaboration d'une forme poétique du discours, ou à l'intérêt phatique de s'adresser à un interlocuteur (nous ne rappellerons pas ici les fonctions du langage de Jakobson).

Il y a, en réalité, deux manières de définir le discours, la première conçoit le discours comme on définirait le langage d'une grammaire formelle ou l'intégralité des énoncés pouvant être produits à partir de la langue et perçus comme appartenant à celle-ci. Nous en appelons pour cette définition à la notion de dialogie ou polyphonie de Bachtine [Bachtine, 1970] ou de Ducrot [Ducrot, 1994] qui considèrent que derrière chaque négation, il y a une affirmation enrichie d'une marque, que sous chaque énoncé, il y a le réservoir des énoncés déjà reçus et émis.

Le discours dans cette perspective ne fait aucune différence entre celui qui a été énoncé et celui qui ne l'a pas été, celui qui ne le sera jamais et celui qui ne restera que prononcé à soi-même. Au sein de cette matière, l'énonciateur décide de faire émerger des formes disponibles, et de les organiser. Dans notre utilisation du concept "discours", nous assumons que la composante majeure du discours est "le dit" ou "l'énoncé" qui se distingue des formes non marquées (non énoncées).

Nous restreignons donc la définition du discours pour le présent travail, ne considérant que ce qui est dit, et en supposant une cohérence et une cohésion entre les énoncés qui le composent. Ainsi le discours est l'unité regroupant tout ce qui est dit par un énonciateur dans un lieu et à un moment donnés. Notre objet est donc le discours que l'on peut décomposer en énoncés, ayant la propriété d'être bien formés en regard du code linguistique à partir duquel ils sont construits et forment, ensemble, une unité cohérente.

Néanmoins, les deux notions de discours sont intimement liées, en s'appuyant sur un ensemble d'énoncés attestés, on entend décrire le plus finement possible les unités du lexique, les rendant effectives à terme dans l'analyse de tout énoncé pouvant être prononcé ou écrit.

1.3.2 La dénotation et le sens

A cette définition du discours, il faut ajouter la référénciation, ou dénotation. Ici, nous observons la différence entre *sens* et *dénotation* au sens de Frege [Black, 1960].

La dénotation d'une expression linguistique est l'objet du monde (c'est-à-dire la portion de réalité intersubjective) que cette expression désigne.

[Roussarie, 2006]

Par réel intersubjectif nous entendons un univers composé de tous les référents possibles, accessibles par le langage, et non un réel extralinguistique, car les univers décrits par le discours ne se limitent pas au réel appréhendé par les sciences du vivant. Ce réel intersubjectif est donc le plus grand ensemble des référents dénotables par l'utilisation d'une langue, et sur lequel est construite toute communication langagière.

L'appareil le plus à même de traiter du sens et de la dénotation est la logique vériditionnelle, ainsi, la dénotation d'un individu est portée par son marqueur référentiel, une variable au sein d'une formule logique. Frege avance que la dénotation d'un énoncé est le caractère vrai ou faux de la prédication portée par cet énoncé, ainsi, la dénotation d'une phrase ou d'un énoncé sera la valeur de vérité attribuée cette prédication représentée comme une formule logique au sein de laquelle sont décrites des relations entre les individus.

La dénotation étant le référent dans un réel intersubjectif d'une unité linguistique, il faut admettre que le *vrai* et le *faux* ont eux aussi des référents au sein d'une réalité intersubjective.

Le sens quant à lui est le "mode de donation" de la dénotation d'une unité linguistique. On peut aussi dire que c'est l'ensemble des conditions nécessaires à déterminer si un énoncé est vrai ou s'il est faux.

Autrement dit, la représentation sémantique d'un énoncé sera donc une formule logique dans laquelle sont regroupées toutes les conditions nécessaires à déterminer qu'il est vrai.

Le célèbre exemple repris par Kripke dans [Kripke, 1999], "Hesperus est Phosphorus", soit "l'étoile du soir est l'étoile du matin", soit encore au niveau dénotationnel : "Vénus = Vénus", proposent des sens différents pour une seule dénotation car "Hesperus", "Phosphorus", "l'étoile du soir", "l'étoile du matin" et "Venus" ont le même référent. Alors on comprend que cet énoncé est doté de sens par la diversité des dénominations, même si au niveau dénotationnel, c'est une tautologie.

Notre approche propose des passerelles entre l'univers du discours et un modèle spatio-temporel, tout en formalisant cette passerelle par l'appareil logique, que nous considérons comme portant la garantie d'une construction stable du sens. Notre cadre théorique se situe dans une linguistique du discours, utilisant une représentation du sens par la logique telle qu'elle a été développée ces dernières années en sémantique formelle, et ce à des fins de représentations spatio-temporelles.

1.3.3 La compositionnalité

La définition standard de la compositionnalité que l'on doit à Gottlob Frege [Black, 1960] et Richard Montague [Partee, 1996] est la suivante [Nicolas, 2006] :

Le sens d'une expression complexe est une fonction du sens de ses parties et de leur mode de combinaison.

C'est en s'appuyant sur le principe de compositionnalité qu'il devient évident que la construction d'une représentation sémantique d'un énoncé doit reposer sur :

- un **lexique sémantique** dans lequel chaque unité linguistique² est associé à une représentation sémantique, soit, dans notre cas, une formule logique partielle.
- l' **analyse syntaxique** de l'énoncé, c'est à dire la structure organisationnelle interne à la phrase³.

Autrement dit, la représentation sémantique est une formule logique construite à partir de formules partielles puisées dans le lexique, et construite dans l'ordre indiqué par la structure syntaxique.

Nous montrons dans le chapitre suivant le détail théorique de notre approche permettant d'obtenir automatiquement ces représentations logico-sémantiques à partir de ces deux éléments, le lexique étant construit et saisi tandis que l'analyse syntaxique est obtenue complètement automatiquement.

1.4 Une représentation logique du discours, la *Discourse Representation Theory* (DRT)

1.4.1 La notion de modèle

La représentation sémantique d'un énoncé dans le cadre d'une sémantique formelle est une formule logique qui nécessite, si on veut montrer qu'elle est valide, de définir un modèle d'interprétation dans lequel elle est vraie.

Un modèle de l'univers discursif sera noté \mathcal{M} , composé d'un ensemble de marqueurs référentiels, noté \mathcal{X} et d'un ensemble de fonctions d'interprétation \mathcal{F} , on aura donc :

$$\mathcal{M} = \langle \mathcal{X}, \mathcal{F} \rangle$$

Le domaine \mathcal{X} est l'ensemble des marqueurs référentiels extraits du discours tandis que l'ensemble des fonctions d'interprétation \mathcal{F} est constitué des relations entre les marqueurs référentiels précédemment introduits.

Dans un objectif de traitement automatique, la représentation logico-sémantique d'un discours doit s'inscrire dans une structure incrémentale permettant d'être enrichie au fur et à mesure de la production et de l'analyse des énoncés le constituant. Cette représentation sémantique de forme logique peut être interprétée dans un modèle, si elle est valide dans un modèle alors, cette formule est vraie. La nature incrémentale du discours est définie par sa linéarité et par la volonté du locuteur que son interlocuteur comprenne et croit en l'univers décrit par son discours. La somme des informations

2. Considérée comme une *partie* dans la précédente définition.

3. Soit ce qui est désigné comme le *mode de combinaison* dans la précédente définition.

délivrées augmente avec le nombre de phrases au sein du discours. Le modèle d'interprétation d'un énoncé est donc un modèle d'interprétation du discours partiel, qui est enrichi, énoncé après énoncé.⁴

Nous retiendrons que la représentation sémantique d'un discours dans un modèle doit témoigner de la construction d'un univers discursif dynamique, c'est à dire qui accepte une forme de mise à jour.

1.4.2 L'analyse d'un énoncé, la construction d'un univers du discours en *Discourse Representation Theory*

La *Discourse Representation Theory*[Kamp et Reyle, 1993a], à présent DRT, est une théorie dans laquelle on crée des univers discursifs, grâce auxquels on peut représenter dynamiquement la forme logico-sémantique du discours. Ces structures représentent l'univers du discours et s'appellent des DRS (*Discourse Representation Structure*). Elles sont construites à partir de quatre ensembles disjoints :

- l'ensemble *Ref* des référents du discours
- l'ensemble *Name* des relations unaires définies pour l'attribution d'une variable à l'individu en le désignant par son nom
- l'ensemble *Relⁿ* des relations n-aires entre référents
- l'ensemble *Sym* des connecteurs logiques classiques

Les éléments significatifs ou saillants sont positionnés les uns par rapport aux autres dans une structure, une DRS. Cette théorie permet de construire des DRS, interprétables dans un modèle, reflétant un état du monde. Ainsi on peut dire qu'une DRS est une représentation logico-sémantique intermédiaire entre l'énoncé et le modèle.

On parlera ici d'un ensemble de référents introduits par le discours dans un domaine, D et d'un ensemble de conditions d'interprétation, F , une DRS est un couple $\langle D, F \rangle$. Cette DRS est interprétable dans un modèle \mathcal{M} tel que décrit plus tôt s'il existe une fonction f définie telle que $f : U \rightarrow \mathcal{X}$ vérifiant toutes les conditions F .

Les DRS permettent de déduire le modèle construit par le discours tout en respectant sa nature incrémentale.

Pour illustrer cette présentation de la DRT, nous nous inspirons des DRS de l'article [Amsili et Bras, 1998] Ainsi une structure du discours en DRT typique aura cette forme :

Ici, x et y appartiennent à l'ensemble *Ref*, le predicat unaire *Pedro* appartient à l'ensemble *Name*, *âne* et *possède* à l'ensemble *Relⁿ*, et nous le verrons, $=$ à l'ensemble *Sym*. Les ensembles *Name* et *Relⁿ* permettent de relier les marqueur et les conditions d'interprétation au modèle.

Les DRS sont construites pour recevoir tout type d'information sémantique. L'ajout d'information sémantique change les contraintes de validité de l'énoncé et enrichissent sa représentation sémantique. Par exemple, lorsque les référents sont en relation les uns avec les autres au travers de prédicats, on peut distinguer aussi les rôles thématiques.

4. Si l'on prend le cas particulier du dialogue, chacun construit sa propre représentation du monde, son propre modèle par le discours, il y a négociation pour que les conditions de vérité d'un modèle commun s'accordent. Toutes les marques phatiques du type « tu vois ? », « tu comprends ce que je veux dire ? », témoignent de cette négociation.

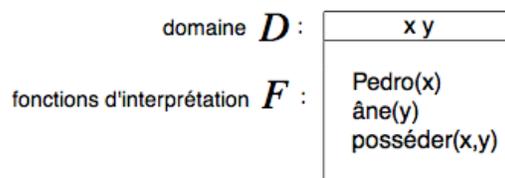


FIGURE 1.1 – Constitution d'une DRS

Les rôles thématiques essentiels sont l'agent et le patient, le premier initiant l'action, et le second la subissant. Néanmoins, il existe d'autres rôles décrivant d'autres relations selon les théories, comme par exemple le thème (qui caractérise une entité déplacée, conséquence de l'action dénotée par le prédicat) ou encore l'expérienteur (qui fait l'expérience d'un état psychologique)⁵. Ces rôles thématiques expriment un lien très fort entre syntaxe et sémantique, et donnent une propriété sémantique au référents en fonction de leur place dans la structure syntaxique. Par exemple :

(1.2) Pedro bat l'âne.

(1.3) L'âne bat Pedro.

Les rôles d'agent et de patient sont déterminés par la position propre au sujet ou à l'objet des référents. Le rôle syntaxique de chacun détermine son rôle thématique.

La logique permet d'exprimer pleinement les spécificités a priori de l'objet et les déductions attenantes afin de rendre compte des coréférences au sein de ce discours. La cohérence et la cohésion d'un discours sont portés par ces marques de reprises anaphoriques, qui une fois décontextualisées sont ambiguës. La résolution d'anaphore est l'exemple le plus parlant. Considérons l'exemple suivant souvent utilisé pour illustrer ces problématiques :

(1.4) Pedro possède un âne. Il le bat.

La seconde DRS créée pour la phrase « Il le bat » est la suivante :

Il nous faut nous arrêter un instant sur l'opération qui a lieu entre la première DRS en figure 1.1 et la seconde 1.2 afin d'obtenir une DRS seule pour le discours constitué de ces deux phrases. Les DRS sont fusionnées (opération de *merge*, et c'est à ce moment là que la résolution de l'anaphore peut avoir lieu.

Ainsi lorsque la seconde DRS 1.2 est fusionnée avec la première 1.1, un lien d'identification entre u et x et entre v et y doit être fait, ce qui est possible par la disponibilité des variables x et y dans le domaine.

5. Nous empruntons ces deux définitions à [Saint-Dizier, 2006].

1.4. UNE REPRÉSENTATION LOGIQUE DU DISCOURS, LA DISCOURSE
REPRESENTATION THEORY (DRT)

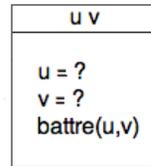


FIGURE 1.2 – DRS de « Il le bat »

Une fois les deux DRS créés, il ne reste plus qu'à les fusionner en une seule et à résoudre les coréférences comme vu en K_2 dans la figure 1.4. Le système consiste en l'assimilation successive d'une DRS après l'autre en partant de la gauche vers la droite. L'opération de fusion \uplus permet d'unir les domaines des deux DRS et d'unir les conditions d'interprétation de celles-ci, plusieurs types de fusion ont été mis au point pour gérer les situations où l'intersection entre l'ensemble des variables de la première DRS et la seconde n'est pas vide.

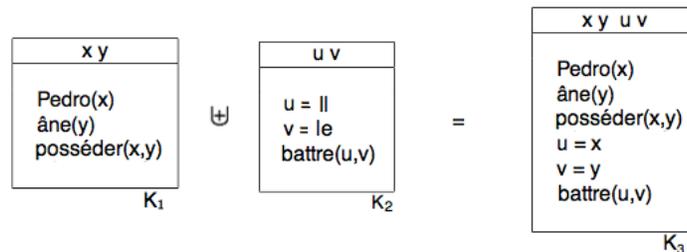


FIGURE 1.3 – DRS mise à jour du discours « Pedro possède un âne. Il le bat. »

La DRS construite en figure 1.3 à partir de ces deux phrases permet de montrer le stade premier ou la DRS est vide K_0 , la seconde étape ou la DRS est enrichie de l'analyse de la première phrase, K_1 , et enfin la DRS une fois les deux phrases analysées, K_2 . Cette méthode permet de mettre à jour le contexte discursif en fusionnant les univers du discours K_1 et K_2 , ce qui garantit une construction dynamique reflétant la linéarité du discours. Sous cette contrainte de dynamique, la DRT présente une représentation sémantique compositionnelle, comportant individus et conditions d'interprétation d'un discours.

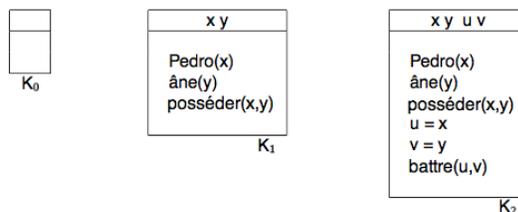


FIGURE 1.4 – 3 étapes de construction d’une DRS pour l’exemple 1.4

1.5 Conclusion

Nous avons montré qu’au sein du vaste champs d’investigation qu’est le traitement automatique de la langue naturelle, la sémantique formelle entend traiter le discours par le biais de la logique, restriction formelle de la langue, afin de lui donner une représentation sémantique logique, interprétable dans un modèle. La construction d’un tel modèle repose sur un lexique sémantique et une analyse syntaxique de chaque énoncé composant le discours, respectant par conséquent le principe de compositionnalité du sens. Par ailleurs, la nature incrémentale et par conséquent dynamique des méthodes employées au sein du cadre théorique choisit, la DRT, permet de respecter la linéarité du langage naturel. Ce cadre théorique présente donc nombres de qualités pour permettre de représenter les relations entre référents du discours ayant trait à l’expression des déplacements au sein d’un discours, essentiels à l’extraction d’itinéraires dans le domaine de la compréhension de texte.

Chapitre 2

Grammaire Catégorielle et λ -calcul : Focus sur le point de vue compositionnel

Nous présentons dans ce chapitre les méthodes de calcul des représentations sémantiques de *Grail*, parser pour grammaire logique de types avec un ensemble restreint de règles fixes et utilisant un lexique riche, qui est développé et maintenu par Richard Moot[Moot, 2010b]. En premier lieu sont présentés les grammaires catégorielles puis le calcul de Lambek, afin d'expliquer le cadre formel dans lequel nous nous situons, à savoir une grammaire de Lambek pour obtenir automatiquement une analyse syntaxique. Nous abordons la correspondance Curry-Howard afin de montrer que l'analyse syntaxique en grammaire de Lambek, qui est une preuve de la logique intuitionniste permet d'obtenir la représentation sémantique de manière compositionnelle et automatique en puisant dans un lexique sémantique en λ -termes.

Nous présentons ensuite *Grail*, dont le fonctionnement s'appuie sur cette correspondance. Nous montrons les étapes allant du texte brut à son analyse syntaxique, et une fois cette première analyse obtenue, nous présentons en dernier lieu la version en λ -DRT du lexique implémentée dans *Grail*.

2.1 Introduction

Nous abordons dans ce chapitre l'analyse syntaxique avec une grammaire catégorielle utilisant les règles du calcul de Lambek, puis la correspondance Curry-Howard nous permettant de calculer la représentation sémantique à partir de l'analyse syntaxique, nous proposons ici de présenter les développements menés dans [Moot et Retoré, 2012]. Nous donnons ensuite une présentation du parseur de gram-

maires logiques de types *Grail*, dont nous aurons alors présenté le cadre théorique.

2.2 L'analyse syntaxique et les grammaires catégorielles

L'intérêt des grammaires formelles pour le traitement de la syntaxe des langues naturelles provient de la volonté de reproduire avec un système à base de règles, appelé aussi système symbolique, la compétence linguistique humaine. Les grammaires formelles entendent décrire la part automatisable des mécanismes de construction des énoncés et nous nous attachons ici à présenter un exemple de grammaire formelle, les grammaires catégorielles. Nous montrons ensuite en quoi l'utilisation de ces grammaires permet de dépasser la syntaxe en devenant une base fondamentale à l'analyse sémantique.

2.2.1 Les grammaires catégorielles classiques

Les grammaires catégorielles classiques ou grammaires AB sont définies à partir de catégories ou types de la forme suivante [Adjukiewicz, 1935][Bar-Hillel, 1953] :

$$T ::= P \mid T \backslash T \mid T / T$$

P est l'ensemble des catégories de base, tandis que les catégories complexes ont une notation fractionnaire. Le lexique L , permet d'attribuer aux mots une ou plusieurs catégories : $L(m_i) = \{t_i^1, \dots, t_i^{k_i}\}$. On donne l'exemple suivant pour P :

- n pour les noms (ex : *montagne*)
- sn pour les syntagmes nominaux (ex : *la montagne*)
- S pour les phrases (ex : *Azais aime la montagne*)

Ces catégories atomiques forment un ensemble fermé.

Pour une partie des éléments du lexique, elles peuvent se combiner pour former des catégories composées :

- sn/n pour les déterminants (ex : *la* attend un n à sa droite pour former un sn)
- $(sn \backslash S) / sn$ pour les verbes transitifs (ex : *aime* attend un sn objet à sa droite puis un sn à sa gauche pour former une phrase de type S)

Ainsi, les règles de calcul sur ces catégories sont les suivantes :

$$\begin{aligned} B \quad (B \backslash A) &\longrightarrow A \\ (A/B) \quad B &\longrightarrow A \end{aligned}$$

Le langage engendré par cette grammaire est défini comme l'ensemble des suites de mots du lexique formant des phrases $m_1 \dots m_n$ telles que pour chaque m_i , il existe une catégorie $t_i \in L(m_i)$, tel que $t_1 \dots t_n \xrightarrow{*} S^1$, S étant l'équivalent de l'axiome dans

1. On note $\xrightarrow{*}$ la clôture transitive de la relation \longrightarrow .

2.2. L'ANALYSE SYNTAXIQUE ET LES GRAMMAIRES CATÉGORIELLES

une grammaire syntagmatique, la catégorie pour la phrase. Pour nos exemples, on aura donc le lexique suivant :

Mots	Catégories
Azaïs	sn
Cauterets	sn
montagne	n
la	sn/n
aime	$(sn \setminus S)/sn$
atteint	$(sn \setminus S)/sn$

La phrase *Azaïs aime la montagne* appartient donc au langage de cette grammaire :

$$\begin{aligned}
 & \text{Azaïs aime la montagne} \\
 & sn \quad (sn \setminus S)/sn \quad sn/n \quad n \\
 & \longrightarrow sn \quad (sn \setminus S)/sn \quad sn \\
 & \longrightarrow sn \quad sn \setminus S \\
 & \longrightarrow S
 \end{aligned}$$

Cette réduction de la catégorie pour atteindre S est une analyse syntaxique dans le cadre des grammaires catégorielles. Ainsi la suite de mots *Azaïs aime la* dont le type sera : $sn \quad (sn \setminus S)/sn \quad sn/n$ ou encore *la Azaïs aime* dont la catégorie est $sn/n \quad (sn \setminus S)/sn$ n'appartiennent pas au langage car ne peut être réduit à S .

Un des atouts des grammaires catégorielles réside dans le fait que le mot est considéré comme une unité contextuelle, autrement dit dépendante du contexte et par l'attribution même des catégories aux mots, on obtient la structure syntaxique de la phrase. Ce calcul nous permet de construire une analyse bien formée. Un verbe transitif tel que *aimer* attend deux arguments, un sujet à sa gauche et un objet à sa droite tous deux de catégorie sn , ce qui donne la catégorie $sn \setminus s/sn$, tandis que *dormir*, qui est intransitif, n'attend qu'un sujet à sa gauche et obtient par conséquent la catégorie $sn \setminus s$.

2.2.2 Le calcul de Lambek

Le calcul de Lambek est une restriction des règles de la logique intuitionniste telles que décrites dans [Retoré, 2000], les deux règles déjà présentées pour le calcul sur les catégories peuvent être réécrites comme deux règles d'élimination et on ajoute des règles d'introduction pour lesquelles il faut décharger l'hypothèse notée $[B]$ pour introduire B dans la preuve :

(Elimination)

$$\frac{A/B \quad B}{A} \text{ (Elim. droite)} \quad \frac{B \quad B \backslash A}{A} \text{ (Elim. gauche)}$$

(Introduction)

$$\frac{\begin{array}{c} [B] \\ \vdots \\ A \end{array}}{A/B} \text{ (Intro. droite)} \quad \frac{\begin{array}{c} [B] \\ \vdots \\ A \end{array}}{B \backslash A} \text{ (Intro. gauche)}$$

La concaténation dans le calcul de Lambek est associative, pour trois catégories A , B et C , on a :

$$(AB)C \equiv A(BC)$$

On donne un premier exemple de preuve en calcul de Lambek ne mettant en jeu que des éliminations.

$sn \quad (sn \backslash S) / sn \quad sn$ peut donc se réduire de deux manières par associativité :

$$\frac{sn \quad \frac{(sn \backslash S) / sn \quad sn}{sn \backslash S} \text{ (Elim. droite)}}{S} \text{ (Elim. gauche)} \quad \frac{sn \quad \frac{sn \backslash (S / sn)}{S / sn} \text{ (Elim. gauche)}}{S} \quad sn \text{ (Elim. droite)}$$

Nous montrons maintenant un exemple mettant en jeu une introduction. Ici, nous déchargeons l'hypothèse² lors de l'utilisation de la règle d'introduction droite pour que la preuve soit complète.

$sn / (S / sn) \quad sn \quad (sn \backslash S) / sn \quad sn \backslash S$ se réduit de la manière suivante :

$$\frac{sn / (S / sn) \quad \frac{\frac{(sn \backslash S) / sn \quad \cancel{sn}}{sn \backslash S} \text{ (Elim. gauche)}}{S} \text{ (Elim. gauche)}}{S} \text{ (Intro. droite)} \quad \frac{sn \backslash S}{S} \text{ (Elim. droite)} \quad sn \backslash S \text{ (Elim. gauche)}$$

2. Notée \cancel{sn} .

2.2.3 Une grammaire de Lambek

Une grammaire de Lambek est une grammaire catégorielle obéissant aux règles du calcul de Lambek. On donne maintenant un lexique pour illustrer la grammaire de Lambek ensuite par trois exemples d'énoncés mettant en jeu ses différentes règles.

Mots	Catégories
Cauterets	sn
Azaïs	sn
le	sn/n
sommet	n
que	$(sn \backslash sn) / S / sn$
l'on	sn
atteint	$(sn \backslash S) / sn$
a accueilli	$(sn \backslash S) / sn$
repart	$sn \backslash S$
gravir	$(sn \backslash S) / sn$
vient_de	$(sn \backslash S) / sn \backslash S$

On peut représenter l'analyse syntaxique de l'énoncé *Azaïs atteint Cauterets* par la preuve suivante :

$$\begin{array}{c}
 \text{Azaïs} \quad \text{atteint} \quad \text{Cauterets} \\
 sn \quad (sn \backslash S) / sn \quad sn \\
 \\
 \frac{\frac{\text{Azaïs}}{sn} \quad \frac{\frac{\text{atteint}}{(sn \backslash S) / sn} \quad \frac{\text{Cauterets}}{sn}}{sn / S}}{S} \quad (\text{Elim. gauche})
 \end{array}$$

L'énoncé *Azaïs que Cauterets a accueilli repart* nécessite la règle d'introduction dans son analyse syntaxique :

$$\begin{array}{c}
 \text{Azaïs} \quad \text{que} \quad \text{Cauterets} \quad \text{a accueilli} \quad \text{repart} \\
 sn \quad (sn \backslash sn) / S / sn \quad sn \quad (sn \backslash S) / sn \quad sn \backslash S \\
 \\
 \frac{\frac{\text{Azaïs}}{sn} \quad \frac{\frac{\frac{\text{que}}{(sn \backslash sn) / S / sn} \quad \frac{\frac{\frac{\text{Cauterets}}{sn} \quad \frac{\frac{\text{a accueilli}}{sn \backslash S / sn} \quad [sn]}{sn \backslash S}}{S}}{S / sn}}{S}}{sn \backslash S}}{S} \quad (\text{Elim. gauche}) \quad \frac{\text{repart}}{sn \backslash S}}{S} \quad (\text{Elim. gauche})
 \end{array}$$

Le calcul de Lambek permet de représenter les catégories syntaxiques tout en permettant de montrer dans chaque entrée du lexique syntaxique, les catégories attendues par

Et pour les catégories complexes, on définit que :

$$(x \setminus y)^* = x^* \rightarrow y^*$$

$$(y/x)^* = x^* \rightarrow y^*$$

Ce qui permet de déduire pour deux exemples de catégories complexes :

Types syntaxiques	Types sémantiques
$(sn/n)^*$	$(e \rightarrow t) \rightarrow e$
$((sn \setminus S)/sn)^*$	$e \rightarrow (e \rightarrow t)$

Pour tout mot m le lexique sémantique associe un λ -terme τ_k pour chaque type syntaxique $t_k \in L(m)$ de sorte que le type de τ_k soit précisément t_k , le pendant syntaxique de τ_k .

En d'autres termes, le lexique sémantique associe à chaque mot un ou plusieurs λ -termes dont le type correspond à la ou aux catégories syntaxiques de ce même mot.

On a donc un homomorphisme entre la démonstration des types syntaxiques et sémantiques pour :

$$\begin{array}{ccc} sn & (sn \setminus S)/sn & sn \\ e & e \rightarrow (e \rightarrow t) & e \end{array}$$

$$\frac{sn \quad \frac{(sn \setminus S)/sn \quad sn}{sn \setminus S} \text{ (Elim. droite)}}{S} \text{ (Elim. gauche)} \quad \frac{e \quad \frac{e \rightarrow (e \rightarrow t) \quad e}{e \rightarrow t} \text{ (Elim.)}}{t} \text{ (Elim.)}$$

Nous donnons un second exemple avec la règle d'introduction :

$$\frac{\frac{\frac{sn \quad \frac{(sn \setminus S)/S/sn \quad S/sn}{S} \text{ (Intro. droite)}}{S/sn} \text{ (Elim. droite)}}{sn \setminus sn} \text{ (Elim. gauche)}}{sn} \quad \frac{\frac{sn \setminus S/sn \quad \overline{[sn]} \text{ (Elim. droite)}}{sn \setminus S} \text{ (Elim. gauche)}}{S} \text{ (Elim. gauche)}$$

$$\frac{e \quad \frac{e \rightarrow t \rightarrow e \rightarrow e \quad \frac{t}{e \rightarrow t} \text{ (Intro.)}}{e \rightarrow e} \text{ (Elim.)}}{e} \quad \frac{\frac{e \rightarrow e \rightarrow t \quad \overline{[e]} \text{ (Elim.)}}{e \rightarrow t} \text{ (Elim.)}}{t} \text{ (Elim.)} \quad \frac{e \rightarrow t}{t} \text{ (Elim.)}$$

2.3.2 Lexique et type

Le lexique sémantique associé au mot dépend de la catégorie syntaxique de celui-ci, en effet par la correspondance Curry-Howard, on considère que l'analyse syntaxique est isomorphe à l'analyse sémantique et nous avons montré que pour un mot donné, le type du ou des λ -termes dans le lexique sémantique doit être construit en fonction du type syntaxique donné par la catégorie.

On présente maintenant les types tels que nous les utilisons de manière standard en λ -calcul simplement typé.

- Sont des types de base :
 - t les *valeurs de vérité*, e les *entités* ou *individus*,
 - plusieurs autres types constants correspondant aux *différentes sortes d'individus*, par exemple *humains* que nous noterons h , *region* : r , et *intervalles temporels* : i , par exemple.
- Lorsque T_1 et T_2 sont des types, $T_1 \rightarrow T_2$ est aussi un type.

Ainsi la version Montagovienne de la compositionnalité est respectée par la présence des types propres aux valeurs de vérité et aux individus avec une variation multi-sorte³ permettant une sémantique plus fine des phénomènes linguistiques, on désire par exemple empêcher qu'un énoncé tel que *La montagne parle* puisse être interprété correctement, ainsi, le verbe *parler* aura un λ -terme de type $h \rightarrow t$, tandis que la montagne est de type r (même si certains glissements de sens sont possibles, pour plus de détails à ce sujet voir [Mery, 2011] et [Moot *et al.*, 2011]).

2.3.3 Construction des termes

Maintenant pour définir les termes, on se donne une infinité dénombrable de variables de chaque type, ainsi que des constantes en nombre fini pour chaque type⁴ :

- Une variable de type T c'est-à-dire $x : T$ (ce qu'on écrit aussi x^T) est un *terme* de type T .
- Une constante de type T c'est-à-dire $c : T$ (ce qu'on écrit aussi c^T) est un *terme* de type T .
- $f(\tau)$ est un terme de type U quand τ est de type T et f est de type $T \rightarrow U$.
- $\lambda x^T. \tau$ est un terme de type $T \rightarrow U$ quand x est une variable de type T , et τ un terme de type U .

2.3.4 Construction de l'analyse sémantique

La β -réduction notée \rightarrow_β pour le λ -calcul est définie comme suit :

3. Nous donnons une définition de multisorte du point de vue de la logique du premier ordre, en section section 2.3.5

4. Cette finitude n'est pas nécessaire, mais raisonnable aussi bien d'un point de vue applicatif que cognitif : les constantes sont introduites dans le lexique, qui comporte un nombre fini d'entrées, chacune ne contenant qu'un nombre fini de termes finis : par opposition aux modélisations en termes de mondes possibles, nous restons ici dans le champ des règles et du calculable.[Lefevre *et al.*, 2012]

– $\lambda x. \tau(u) \rightarrow_{\beta} \tau[x := u]$

On substitue dans τ toutes les occurrences liées de x par u ⁵.

Le typage permet d'appliquer un terme à un autre terme. Ce dernier peut être vu comme une fonction si le terme auquel on l'applique est du type attendu par le terme-fonction. Par exemple pour deux types α et β quelconques on a :

Le terme suivant

$$(\lambda x^{\alpha}. \tau)^{\alpha \rightarrow \beta} (u)^{\alpha}$$

se β -réduit en

$$\tau[x := u]^{\beta}$$

La forme normale d'un λ -terme est la forme d'un terme qui ne peut plus être β -réduit. On appelle alors forme normale d'un terme M tout terme N qui est une forme normale et tel que $M \xrightarrow{*}_{\beta} N$ ⁶.

La correspondance Curry-Howard peut donc être illustrée par les règles du calcul de Lambek qui donnent les types des termes dans le λ -calcul simplement typé. f, u, v et x sont des termes en λ -calcul, tandis que les catégories sont désormais les types de ces termes. Le principe de compositionnalité est observé car la construction de la représentation sémantique est fonction du lexique sémantique et de l'analyse syntaxique :

(Elimination)

$$\frac{f : A/B \quad u : B}{f(u) : A} \text{ (Elim. droite)} \quad \frac{u : B \quad f : B \setminus A}{f(u) : A} \text{ (Elim. gauche)}$$

(Introduction)

$$\frac{\begin{array}{c} [x : B] \\ \vdots \\ f : A \end{array}}{\lambda x. f(x) : A/B} \text{ (Intro. droite)} \quad \frac{\begin{array}{c} [x : B] \\ \vdots \\ f : A \end{array}}{\lambda x. f(x) : B \setminus A} \text{ (Intro. gauche)}$$

2.3.5 Logique du premier ordre multisorte

Nous montrons ici le lien entre représentations sémantiques en λ -termes simplement typés et formules logiques du premier ordre multisorte. Tout comme la sémantique de Montague, la logique du premier ordre standard a deux types, e pour les individus et t pour les valeurs de vérité. Nous voulons utiliser une variante, la logique du premier ordre multisorte, considérant plusieurs sortes de e_i avec $1 \leq i \leq N$.

Définissons le langage d'une logique multisorte :

5. Ceci suppose qu'aucune occurrence de x ne soit libre dans u .

6. On note $\xrightarrow{*}_{\beta}$ la clôture transitive de la β -réduction.

- Les variables : pour chaque sorte $\alpha = e_{q_0}$, nous avons un ensemble de variables x_n^α de sorte α .
- Les constantes : chaque sorte $\alpha = e_{q_0}$, nous avons un ensemble de constantes a_n^α de sorte α .
- Les symboles de fonctions : pour chaque type fonctionnel $\phi = e_{q_1} \rightarrow (e_{q_2} \rightarrow (\dots \rightarrow (e_{q_{sq}} \rightarrow e_{q_0})))$, nous avons un ensemble dénombrable de fonctions f_q^ϕ de ce type ϕ .
- Les symboles relationnels : pour chaque type relationnel $\psi = e_{q_1} \rightarrow (e_{q_2} \rightarrow (\dots \rightarrow (e_{q_{sq}} \rightarrow t)))$, nous avons un ensemble dénombrable de symboles de relation R_k^ψ de type ψ .

Les termes sont définis comme suit :

- Une variable x_n^α est un terme de sorte α .
- Une constante a_n^α est un terme de sorte α .
- Si (f_q) est un symbole de fonction de type $\phi = e_{q_1} \rightarrow (e_{q_2} \rightarrow (\dots \rightarrow (e_{q_{sq}} \rightarrow e_{q_0})))$ et si t_i pour $i = 1, \dots, sq$ sont des termes de sortes e_{q_i} alors $f_q(t_1, \dots, t_{sq})$ est un terme de sorte e_{q_0} .
- Rien d'autre n'est un terme. Les termes sans variables sont appelés terme clos ; chaque occurrence d'une variable dans un terme est une occurrence libre.

Si R est un symbole relationnel de type $\psi = e_{q_1} \rightarrow (e_{q_2} \rightarrow (\dots \rightarrow (e_{q_{sq}} \rightarrow t)))$ et si t_i pour $1 \leq i \leq sq$ termes de sortes respectives e_{q_i} , alors la formule suivante est une formule atomique :

$$R(t_1, \dots, t_{sq})$$

Les variables libres d'une formule atomique sont les variables libres du terme.

Une formule logique peut être définie ainsi :

- Une formule atomique est une formule.
- si F est une formule alors $\neg F$ est une formule.
- si F et G sont des formules alors $F \wedge G, F \vee G, F \implies G$ sont des formules.
- si F est une formule, et si x est une variable de sorte e_i alors $\forall x : e_i.F$ et $\exists x : e_i.F$ sont des formules.
- Rien d'autre n'est une formule.

Donnons les types des symboles logiques en commençant par les constantes logiques :

- \neg est de type $t \rightarrow t$
- \implies, \wedge, \vee sont de type $t \rightarrow (t \rightarrow t)$

Les constantes du langage :

- R_q de type $e_{q_1} \rightarrow (e_{q_2} \rightarrow (\dots \rightarrow (e_{q_{sq}} \rightarrow t)))$
- f_q de type $e_{q_1} \rightarrow (e_{q_2} \rightarrow (\dots \rightarrow (e_{q_{sq}} \rightarrow e_{q_0})))$

Les quantificateurs du langage :

- \forall_{e_q} et \exists_{e_q} de type $((e_q \rightarrow t) \rightarrow t)$

Dans un premier temps, on donne la traduction d'une formule logique en un λ -

terme de forme normale de type t .

Tout d'abord, nous rappelons que l'on appelle forme normale d'un terme M tout terme N qui est une forme normale et tel que $M \xrightarrow{*}_\beta N$, c'est à dire qui ne peut plus être β -réduit. Nous ajoutons une spécificité, la forme η -longue.

L' η -expansion pour un terme t sans occurrence libre de x est :

$$t^{A \rightarrow B} \rightarrow_\eta \lambda x^A t(x)$$

Un terme de forme normale η -longue peut être défini en utilisant la notion de forme atomique :

- les variables sont des formes atomiques
- $(M N)$ est une forme atomique si M de type $A \rightarrow B$ est une forme atomique et N de type A est un terme de forme normale η -longue.
- les termes atomiques dont les types sont des types de base sont de forme normale η -longue.
- $\lambda x^A t$ est un terme de forme normale η -longue si t l'est.

Tout terme n'a qu'une seule forme normale η -longue, qui est obtenue par différentes étapes de β -réduction et d' η -expansion. Autrement dit tout terme de type $A \rightarrow B$ sans argument de type A est de forme $\lambda x^A t$ avec t de type B . Pour une forme η -long, f de type $A \rightarrow B$ et sans argument, doit être remplacée par $(\lambda x^A f(x))^{A \rightarrow B}$.

Soit F une formule logique, cette formule peut être :

- une formule atomique de forme $R(t_1, \dots, t_{sq})$, les arguments t_1, \dots, t_{sq} de R sont des termes logiques de sorte e_{q_i} avec $1 < i < sq$, tandis que R est de type $\psi = e_{q_1} \rightarrow (e_{q_2} \rightarrow (\dots \rightarrow (e_{q_{sq}} \rightarrow t)))$, ils se traduisent par des λ -termes de type e_{q_i} . La traduction de cette formule atomique est un λ -terme normal de type t .
- si F est un λ -terme de type t alors $\neg F$ est un λ -terme de type t .
- si F et G sont des λ -termes de type t alors $F \wedge G, F \vee G, F \implies G$ sont des λ -termes de type t .
- si F est un λ -terme normal de type t , et si x est une variable de sorte e_i alors $\forall x : e_i. F$ et $\exists x : e_i. F$ sont des λ -termes de type t .

Nous cherchons maintenant la traduction d'un λ -terme de forme normale de type t en une formule de la logique du premier ordre.

Proposition : Un λ -terme de forme normale et de type t correspond à une formule de la logique du premier ordre.

Cette traduction repose sur la structure d'un λ -terme de forme normale. Remarquons qu'un λ -terme contient des λ -abstractions $\lambda x_1 \dots \lambda x_n$ avec $1 < i < n$ mais peut ne pas en avoir pour $i = 0$ et un sous-terme f qui est successivement appliqué à différents arguments, des λ -termes $t_1 \dots t_p$, mais peut là aussi ne pas en avoir avec $p = 0$. Si le terme est de forme normale, f ne peut être une λ -abstraction, si c'était le cas, il serait de forme $\lambda x_0 u$ et parce que le terme est normal, il n'y aurait donc pas d'application de f (avec $p = 0$), dans ce cas, λy serait $\lambda x_{n+1}. f$ est donc soit un constante soit une variable.

Les λ -termes de forme normale peuvent être vus comme des formules de logique du premier ordre multisorte d'après la preuve de Rétoré et Moot dans [Moot et Retoré, 2012].

2.3.6 Un exemple simple

On donne ensuite la correspondance entre catégories, typage et λ -termes pour l'exemple *Azaïs atteint Cauterets*, ainsi que la construction de la représentation sémantique en λ -calcul.

Mots	Lexique syntaxique	Types	Lexique sémantique
Cauterets	sn	r	$(Cauterets)^r$
Azaïs	sn	h	$(Azaïs)^h$
atteint	$(sn \setminus S) / sn$	$r \rightarrow h \rightarrow t$	$(\lambda x^r \lambda y^h . atteindre(y, x))^{r \rightarrow (h \rightarrow t)}$

$$\begin{array}{c}
 \frac{\frac{atteint}{(sn \setminus S) / sn} \quad \frac{Cauterets}{sn}}{sn/S} \quad (Elim. \text{ gauche}) \\
 \frac{Azaïs}{sn} \quad (Elim. \text{ droite})
 \end{array}
 \quad
 \begin{array}{c}
 \frac{r \rightarrow (h \rightarrow t) \quad r}{h \quad h \rightarrow t} \quad (Elim.) \\
 t
 \end{array}$$

$$\begin{aligned}
 & \left((\lambda x^r \lambda y^h . atteindre(y, x))^{r \rightarrow (h \rightarrow t)} (Cauterets)^r \right) (Azaïs)^h \\
 & (\lambda y^h . atteindre(y, Cauterets))^{(h \rightarrow t)} (Azaïs)^h \\
 & atteindre(Azaïs, Cauterets)^{(t)}
 \end{aligned}$$

La catégorie S est obtenue en correspondance avec la forme normale du λ -terme de type t . On obtient la forme normale du terme en λ -calcul une fois les applications entre les termes correspondant au verbe et à ses deux arguments réalisés dans l'ordre imposé par le parenthésage. La représentation obtenue est un terme de type t et donc une formule de la logique du premier ordre multisorte.

2.4 Une implémentation : *Grail*

2.4.1 Architecture générale

Grail est une plateforme pour le développement et le parsing des grammaires catégorielles multimodales [Moot, 1998] développée et maintenue par Richard Moot. Il a surtout été utilisé pour développer des grammaires spécifiques à des phénomènes linguistiques, en exploitant la transparence de l'interface syntaxe/sémantique des grammaires catégorielles. Nous détaillons tout d'abord les étapes pour obtenir automatiquement l'analyse syntaxique d'un énoncé.

Il est nécessaire de préciser que la grammaire pour le français a été semi-automatiquement extraite à partir du corpus de Paris 7 Treebank [Abeillé *et al.*, 2003]. L'acquisition de cette grammaire a permis d'obtenir un lexique pour le français donnant une catégorie syntaxique dépendante de son contexte à chacun des mots du corpus. Nous présentons dans le schéma 2.1 l'architecture générale de *Grail*.

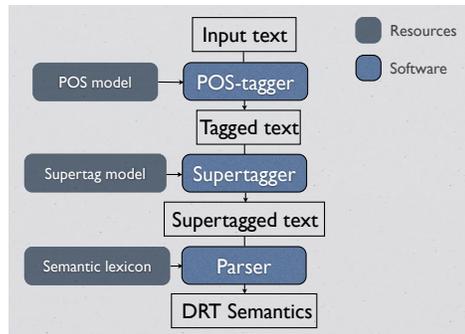


FIGURE 2.1 – Architecture générale de Grail

- L'entrée dans notre cas est le texte brut.
- En ce qui concerne le traitement, plusieurs étapes d'analyses syntaxiques statistiques, une étape de tagging et de supertagging, ont lieu. Puis, en puisant dans un lexique sémantique basé sur le λ -calcul et la DRT, la correspondance entre l'analyse syntaxique et l'analyse sémantique permet de construire la représentation logico-sémantique.
- En sortie nous avons une somme d'informations, extraite à partir de l'objet fourni en entrée, et modifiée en conséquence par le programme : une représentation sémantique.

2.4.2 Le traitement de la syntaxe

Le traitement démarre donc par une étape d'étiquetage morpho-syntaxique ou tagging, chaque mot se voit associer une catégorie syntaxique (part-of-speech), puis le supertagging permet de proposer pour chaque phrase plusieurs analyses syntaxiques en arbre. Le tagger et le supertagger atteignent une précision de 98,4% pour le tagger et de 90,5% pour le supertagger. Le corpus d'entraînement est l'ensemble des 12.440 phrases (371.033 mots) du Treebank de Paris VII [Moot, 2010b]. A partir du lexique syntaxique appris et de l'étiquetage, les catégories sont attribuées dans le cadre d'une grammaire de Lambek. Une catégorie est attribuée pour chaque mot, elles sont proposées en fonction des catégories environnantes afin d'obtenir la catégorie S pour la phrase. Nous précisons qu'ici la catégorie sn pour *syntagme nominal* est nommée np pour *noun phrase*.

Nous noterons que les items lexicaux les plus représentés ont souvent plusieurs formules associées, par exemple "et" ou "est". Dans le cas où le mot est inconnu, une catégorie est proposée par le supertagger ce qui permet l'acquisition de nouveaux items.

Pour illustrer le résultat du supertagging sur le discours, voici deux exemples, le premier est un exemple construit, tandis que le second est un exemple du corpus :

(2.1) Azais atteint Cauterets.

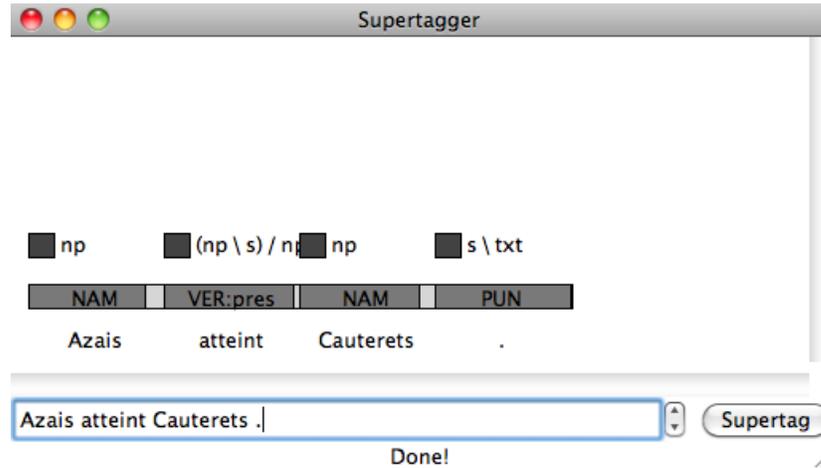


FIGURE 2.2 – Sortie du supertagger, analyse de la phrase : *Azais atteint Cauterets*

(2.2) La reine de Hollande traverse le Vignemale pour aller de Cauterets à Gavarnie par la Hourquète.

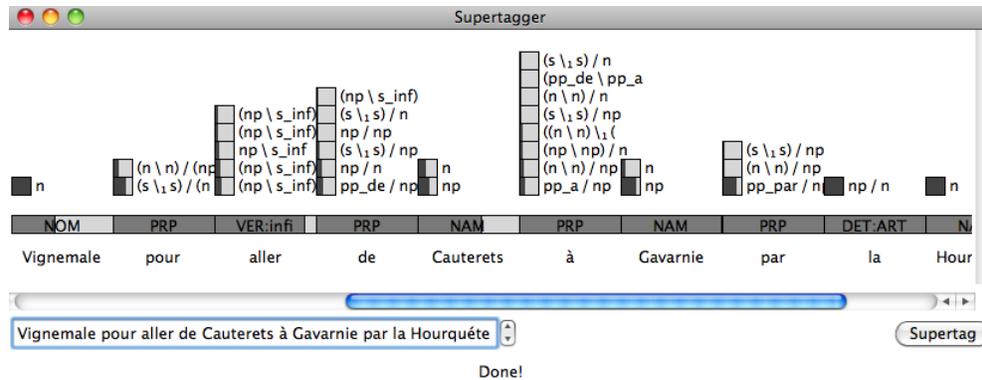


FIGURE 2.3 – Sortie du supertagger, analyse de la phrase : *La reine de Hollande traverse le Vignemale pour aller de Cauterets à Gavarnie par la Hourquète.*

La figure 2.2 montre la sortie du tagger et du supertagger pour la phrase *Azais atteint Cauterets*. Le tagger donne les catégories syntaxiques suivantes : **NAM** pour

Azais et *Cauterets* et *VER : pres* pour *atteint*. Une seule catégorie est proposée par mot en fonction de sa catégories syntaxique étiquetée, par exemple, pour *atteint*, la catégories proposée est $(np \setminus s) / np$ car il est étiqueté *VER : pres*.

En revanche, la figure 2.3 montre une partie des sorties du tagger et du supertagger. Les probabilités de chaque supertag en fonction de leur contexte proche sont indiquées de manière non exhaustive dans les boîtes associées aux catégories, les plus probables étant les plus foncées.

Le nombre de supertags par mot reste raisonnable : 2,3 tags par mot pour les 98,4% de supertags corrects[Moot, 2010b]. Ici on voit clairement que plusieurs supertags sont proposés pour chaque catégorie syntaxique étiquetée : six supertags pour *de*, étiqueté *PRP*, soit préposition. Par exemple, *la* n'a qu'une seule catégorie associée, celle d'un déterminant (*DET : ART*) attendant un *n* à sa droite pour former une *np*, la forme pronominale *la* n'est pas prise en compte. Lorsque les catégories sont plus difficiles à discriminer pour un même item, on aura donc plusieurs formules associées, on pense ici aux prépositions *PRP*, et aux différentes formes de verbes comme *VER : infi* par exemple.

L'utilisation des probabilités permet de choisir la meilleure analyse globale. Par ailleurs on voit pour les prépositions *de*, *à* et *par* que ce n'est pas le supertag le plus probable qui est choisit mais en remontant du plus au moins probables les propositions de supertags, un des suivants.

C'est à partir de la structure syntaxique trouvée que l'on va pouvoir associer une représentation sémantique du discours. Automatiquement, *Grail* sélectionne donc la meilleure (la plus probable) analyse syntaxique pour l'énoncé. Une fois les deux étapes d'analyse syntaxique achevées, le calcul de la représentation sémantique est immédiat, par la transparence de l'articulation entre la catégorie de Lambek et le type du λ -terme associé dans le lexique sémantique.

Nous avons désormais l'analyse syntaxique sous la forme d'une dérivation, il nous reste à présenter le lexique sémantique et à montrer en quoi il est compatible avec la structure syntaxique obtenue.

- L'entrée ici considérée est la représentation syntaxique en catégories de Lambek [Lambek, 1958] de la phrase.
- En ce qui concerne le traitement, on puise pour chaque mot dans le lexique sémantique en λ -DRT (qui est une version du λ -calcul) et on réduit la structure sémantique dans l'ordre de la structure syntaxique en suivant les règles du calcul de Lambek dont la correspondance en λ -calcul a été démontré par Curry-Howard[Howard, 1980].
- En sortie on obtient des DRS, représentations logico-sémantiques du discours dans le cadre de la *Discourse Representation Theory*[Kamp et Reyle, 1993a].

2.4.3 La représentation sémantique en DRT

Comme nous l'avons présenté dans le chapitre précédent, la DRT est une théorie proposant de représenter la sémantique d'un discours comme une boîte (Discourse Representation Structure) dans laquelle on trouve premièrement le domaine, composé

des individus, puis les conditions d'interprétation sémantique de ce modèle. Les DRS créées se fusionnent les unes avec les autres par l'opération de "merge" et permettent d'interpréter les phénomènes de cohérence du discours comme par exemple la résolution des anaphores pronominales. La description en λ -DRT des mots de la langue est simplement une variante de la description en λ -termes comme montré jusqu'ici.

Nous rappelons que la catégorie syntaxique S (qui correspond au type sémantique des phrases, les valeurs de vérités t) est un λ -terme, ou une λ -DRS dont les variables libres correspondent aux mots, et le lexique fournit des λ -termes du même type sémantique : en les substituant et en réduisant le terme obtenu on obtient un terme normal de type t . C'est une formule logique, la représentation sémantique, et dans notre cas la λ -DRS.

Nous avons représenté de manière simplifiée le calcul en correspondance entre la dérivation dans la grammaire catégorielle de Lambek utilisée et la réduction en λ -DRT.

On utilise dans cet exemple deux fois la règle d'élimination, une première fois à droite entre *atteint* et *Cauterets* puis une fois à gauche entre *Azaïs* et le reste de la phrase dans la figure 2.4.

Par ailleurs, il est important de montrer que la DRS finale peut exprimer une formule logique. Nous avons montré la traduction entre λ -terme et formule logique du premier ordre multisorte, ici il faut ajouter deux éléments pour que la traduction soit effective. Les éléments du domaines sont quantifiés par des existentiels, tandis que les conditions d'interprétation sont articulées par des conjonctions :

$$\exists x \exists y. \text{atteindre}(x, y) \wedge \text{Cauterets}(x) \wedge \text{Azaïs}(y)$$

Puis on donne la représentation sémantique en DRS en figure 2.5. Le traitement des entités nommées est quelque peu divergente de celle introduite par notre présentation de la DRT. Ici les entités sont notées comme peuplant le domaine d'une DRS chacun mais qui est accessible pour la DRS contenant le prédicat *atteindre*. Par ailleurs c'est un traitement spécifique aux entités nommées qui dans chacune de ces DRS lie le marqueur référentiel x_0 et y_0 à l'unité linguistique nom lui correspondant. Ici, *Cauterets* et *Azaïs* ne sont plus des prédicats unaires appartenant à l'ensemble *Name* présenté plus tôt mais des arguments du prédicat *nommé*(\cdot).⁷ Par ailleurs, la variable z_0 représente l'évènement, terme auquel nous préférons *éventualité* afin de désigner sans distinction pour le moment les évènements des états par exemple. Nous dirons simplement que cette variable est une réification de l'éventualité qui permet de la manipuler et ainsi de la situer temporellement⁸

La syntaxe garantit que la représentation sémantique est bien formée conformément au principe de la compositionnalité. Néanmoins, c'est dans l'entrée du lexique sémantique (du côté système, reflet du lexique du français) qu'il faut injecter les informations que l'on souhaite extraire de la représentation sémantique (en discours, des énoncés).

7. Néanmoins, c'est bien un seul univers du discours qui nous intéresse ici, les deux DRS entre parenthèse exprimant les présuppositions, que l'on peut se représenter comme appartenant au contexte. On peut les lire comme Il existe x_0 un individu dans le contexte nommé *Azaïs*. Par ailleurs avant de fusionner deux DRS, il faut d'abord résoudre les présuppositions si elles posent certaines ambiguïtés, que nous ne détaillons pas ici.

8. Nous ne précisons pas pour le moment la signification de $\text{temps}(z_0) \circ \text{maintenant}$ que nous introduisons dans les chapitres suivants.

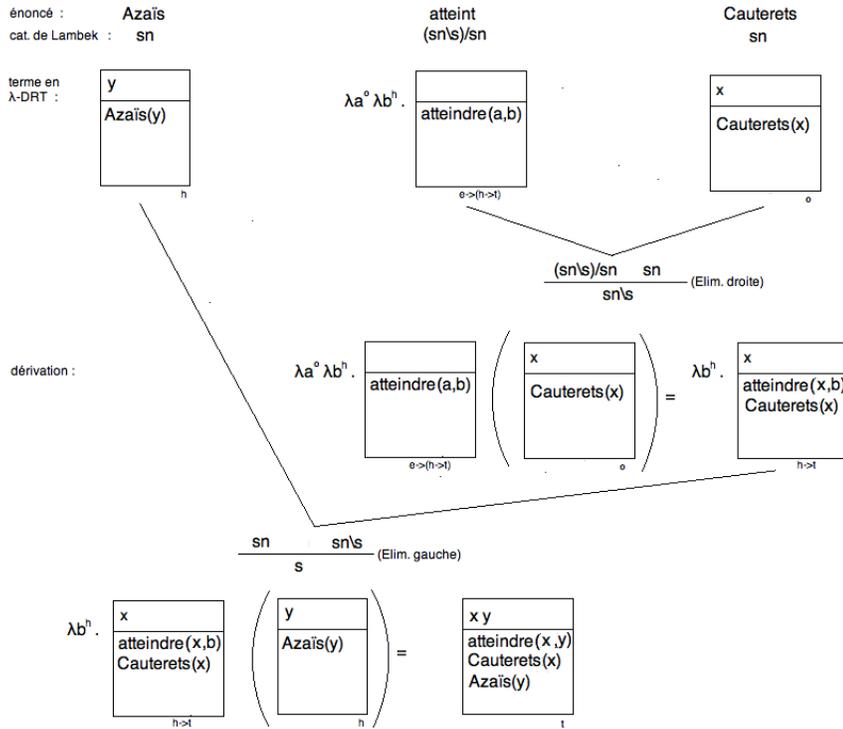


FIGURE 2.4 – Exemple de réduction sémantique simplifiée à partir de la structure syntaxique pour l'exemple *Azaïs atteint Cauterets*

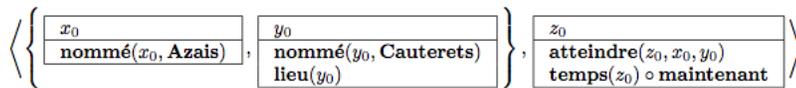


FIGURE 2.5 – Représentation sémantique par Grail de la phrase : *Azaïs atteint Cauterets*

Par exemple pour les positions sujet et objet dans la phrase, pour une bonne partie des verbes transitifs tels que *manger* on veut distinguer qui mange et qui est mangé, au-

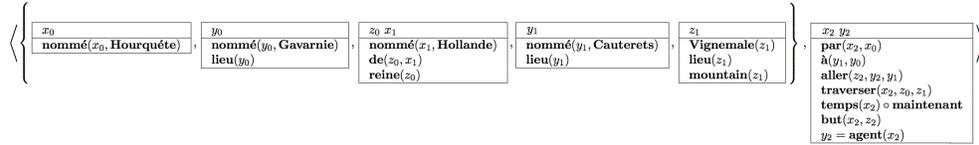


FIGURE 2.6 – Sortie sémantique de Grail, analyse de la phrase : *La reine de Hollande traverse le Vignemale pour aller de Caunterets à Gavarnie par la Hourquète.*

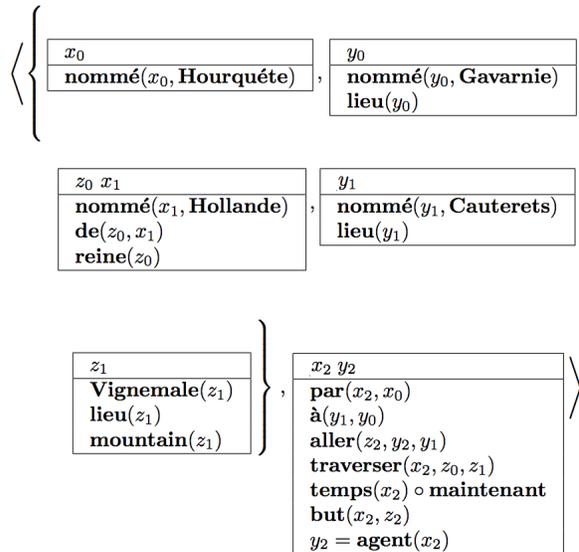


FIGURE 2.7 – Détail de la sortie sémantique de Grail, analyse de la phrase : *La reine de Hollande traverse le Vignemale pour aller de Caunterets à Gavarnie par la Hourquète.*

trement dit les rôles thématiques de l'*agent* et du *patient* peuvent être respectivement représentés au sein de la structure syntaxique de la phrase par le sujet et l'objet à la voie passive.

On donne en figure 2.7 une autre sortie de *Grail* pour l'exemple 5.3, introduit plus tôt. Dans cette figure, les prédicats *aller* et *traverser* établissent une relation entre trois arguments, le premier est une réification de l'éventualité (événement) qui lui est attachée, et les second et troisième arguments sont des marqueurs référentiels propres aux personnes et lieux. Plus précisément, *la reine* dont le marqueur référentiel est z_0 est le second argument du prédicat *traverser*, ici, son sujet. z_1 , marqueur référentiel pour le

Vignemale, est le troisième argument du prédicat *traverser*, et dans le cas présent, son objet.

Pour ce qui est du prédicat *aller*, il établit une relation entre un sujet y_2 dont on donne l'information qu'il est l'agent de l'évènement x_2 , évènement associé à *traverser*. Autrement dit, l'*agent* de l'évènement *traverser* est aussi impliqué dans la relation proposée par le prédicat *aller*. On comprend ici que la *reine de Hollande* est à la fois la personne qui "traverse" et qui "va". Le dernier argument du prédicat *aller* est y_1 c'est-à-dire le marqueur référentiel pour *Cauterets*, on comprend qu'ici, le dernier argument de la relation *aller* est donc *Cauterets*. Malheureusement, cet argument devrait être l'origine mais code en réalité la destination, le lexique au moment de ce travail n'ayant pour entrée pour le verbe *aller* que deux possibilités, aucun des deux ne codant l'origine comme une de ses arguments. On représente dans un petit tableau la correspondance entre premier, second et troisième argument des prédicats *traverser* et *aller* :

traverser	(x_2 ,	z_0 ,	z_1)
		"la reine"	"le Vignemale"	
aller	(z_2 ,	y_2 ,	y_1)
		=agent(x_2)	"Cauterets"	
		= z_0		
		"la reine"		

La préposition *à* prend deux argument un lieu origine et un lieu destination, à savoir y_1 et y_0 , respectivement *Cauterets* et *Gavarnie*. "La reine" "traverse le Vignemale" et "va de Cauterets" "à Gavarnie". Il y a aussi une relation de *but* entre l'évènement "traverser" et l'évènement "aller", qui est porté par la préposition *pour*. Le *Vignemale*, *Cauterets* et *Gavarnie* sont tous trois dans le lexique sémantique pour lequel on a spécifié qu'ils étaient des lieux, voire on a même spécifié que le Vignemale était une montagne (*mountain*), tandis que la *Hourquette*, n'est pas dans le lexique, néanmoins, il a été analysé comme étant une entité nommée à juste titre.

Le typage du lexique permet de vérifier de manière stricte la bonne formation de la représentation. Nous souhaitons intégrer à ce lexique sémantique les informations nécessaires à l'extraction des itinéraires dans le discours. Nous avons donc choisit de travailler sur les représentations sémantiques de la temporalité des événements de déplacement ponctuant le récit de voyage. Les exemples développés jusqu'ici montrent les possibilités de représenter des relations entre personnes et région par exemple, ce qui permet de donner les localisations d'un voyageur mais que faire d'une phrase telle que *Azaïs quitta Cauterets* ou encore *Azaïs quittera tous les ans Cauterets pour quelques jours* ?

2.5 Conclusion

Nous avons présenté les bases du calcul syntaxique et sémantique qui régissent les analyses de l'un et de l'autre. Nous avons montré en quoi le principe de compositionnalité était respecté dans le détail du traitement opéré par *Grail*. Et enfin nous avons

montré le rôle prépondérant d'un lexique sémantique sophistiqué dans le cadre d'une telle analyse. Le traitement étant entièrement basé sur un système logique typé, dont les règles sont implémentées, il permet une analyse complètement automatique, qui nécessite un enrichissement du lexique pour affiner les représentations sémantiques et extraire les informations désirées. Néanmoins les seules informations extraites pour le moment dans le lexique tel que présenté sont les relations et les individus, nous abordons dans le chapitre suivant les méthodes pour traiter la temporalité des événements en sémantique formelle.

Chapitre 3

De la temporalité des éventualités pour la sémantique formelle

Dans ce chapitre nous introduisons différents aspects du traitement de la temporalité des éventualités en sémantique formelle. Nous posons tous les concepts utiles à la compréhension des enjeux d'une telle tâche et contrastons les réponses apportées par les théories présentées. Plus précisément, nous abordons tout d'abord deux propositions phares, la première émanant d'*Elements of Symbolic Logic* de Reichenbach et la seconde de *Maintaining knowledge about temporal intervals* d'Allen. Le premier permet d'introduire la notion de système compositionnel du temps d'un verbe conjugué, et le second, les relations bases du modèle d'interprétation du temps le plus utilisé dans la communauté scientifique, les relations minimales entre intervalles. Nous introduisons ensuite les théories de Vendler, Pustejovsky et Moens et Steedman, trois propositions traitant de la structure des éventualités, c'est à dire trois typologies des éventualités basées sur leur sémantique en langue ainsi que leur sémantique une fois contextualisées au sein d'un énoncé. Ensuite, nous présentons les bases de deux systèmes de traitement de la temporalité des éventualités pour le français que proposent Gosselin et Battistelli. A la lumière de ces différents auteurs, nous présentons le traitement classique de la temporalité en DRT (Kamp) et son articulation avec la SDRT pour *Segmented Discourse Representation Theory* (Asher). Nous finissons ce chapitre par une synthèse de ces différentes théories et mettons en relief les spécificités de notre proposition.

3.1 Introduction

Dans le chapitre précédent, nous avons montré la méthode nous permettant d'obtenir automatiquement une représentation de la sémantique d'un énoncé. Ces représentations sont construites de manière compositionnelle, à partir de l'analyse syntaxique de l'énoncé et du lexique sémantique.

Nous avons présenté jusqu'à maintenant des exemples en sémantique formelle dans lesquels on observe des relations entre individus, il nous faut désormais aborder la temporalité de ces relations afin de leur accorder une validité en fonction du temps. Les conditions d'interprétations d'un énoncé tel que *Azaïs quitte Cauterets* ne sauraient être par défaut toujours vraies ou toujours fausses, particulièrement lorsqu'il s'agit d'un déplacement. Pour ceci, nous devons introduire les éléments essentiels à la présentation de la sémantique de la temporalité de ces relations en premier lieu, puis présenter ensuite différentes méthodes existantes pour un traitement de celle-ci en sémantique formelle.

Nous définissons dans ce chapitre plusieurs concepts importants pour le traitement de la temporalité en discours. Nous présentons tout d'abord deux propositions phares, de Reichenbach et d'Allen, permettant d'introduire d'une part la notion de système compositionnel du temps d'un verbe conjugué et de l'autre un modèle standard du temps, les intervalles temporels. Nous abordons ensuite trois théories traitant de la structure des éventualités, que nous aurons alors définies, c'est à dire trois typologies des éventualités basées sur leur sémantique en langue ainsi que leur sémantique une fois énoncés. Par la suite, nous présentons les bases de deux systèmes de traitement de la temporalité pour le français. A la lumière des différents travaux cités, nous présentons le traitement classique de la temporalité en DRT et son articulation avec la SDRT pour *Segmented Discourse Representation Theory*. Nous finissons ce chapitre par une synthèse de ces différentes théories et mettons en relief les spécifications de notre proposition.

3.2 La temporalité dans le discours

3.2.1 Énonciation, discours et énoncé

L'énonciation est le principe selon lequel la personne à l'origine du discours, autrement dit l'énonciateur ou locuteur, s'inscrit lui-même dans l'univers qu'il construit avec ses propres dires. Cette notion a tout d'abord été proposée par Emile Benveniste dans [Benveniste, 1966] comme actualisation de la langue en discours au sein de laquelle on peut voir émerger *Je* comme "instance énonciative unique et mobile". L'énonciateur se manifeste par le biais de cette instance dans une "réalité du discours" que nous appelons univers discursif. Nous reprenons la définition de la problématique de l'énonciation selon Kerbrat-Orecchioni dans *L'énonciation. De la subjectivité dans le langage* [Kerbrat-Orecchioni, 1999] : 36 :

C'est la recherche des procédés linguistiques (shifters, modalisateurs, termes évaluatifs, etc.) par lesquels dans le message (implicite ou explicite) le locuteur imprime sa marque à l'énoncé, s'inscrit dans le message (implicite ou explicitement) et se situe par rapport à lui.

La langue offre des possibilités illimitées pour l'énonciateur de construire son discours et c'est sur ce terrain fertile que chaque énonciateur construit son univers discursif, appelé aussi scénographie selon Maingueneau [Maingueneau, 1991]. Dans cette scénographie, il se représente lui-même et s'auto-définit, on parlera alors de construction de l'ethos. Cette construction d'ordre identitaire met en jeu une représentation de l'énonciateur telle qu'il l'envisage dans l'interaction. Ce phénomène de construction est intrinsèque à la notion d'énonciation, puisqu'elle implique systématiquement la manifestation de l'énonciateur dans son propre discours.

La subjectivité prégnante de l'énonciateur dans son discours a donc deux propriétés qu'il faut grader à l'esprit, tout d'abord c'est la seule condition d'accès au sens des objets du discours et ensuite, par l'étude du rapport de l'énonciateur à son discours, on peut observer aussi l'énonciateur et la construction de son ethos.

On entend par discours ici la production linguistique d'un énonciateur dans un lieu donné et à un moment donné. Nous appuyons la distinction introduite dans le premier chapitre entre l'ensemble exhaustif des discours pouvant être produits dans une langue donnée et ceux qui sont émergents, marqués, attestés, autrement dit énoncés. C'est en ce sens que nous restreignons donc l'emploi du terme "discours".

D'un point de vue énonciatif, le discours est le produit de l'acte d'énonciation qui lui-même est temporalisé dans le temps réel, extralinguistique. Il permet de créer un univers dans lequel on interprète ce qui est énoncé. Il existe une passerelle entre le temps décrit par le discours par le biais de la dénotation (nous rappelons là l'appel à un réel intersubjectif mais auquel on ne peut accéder que par langage) et un temps réel, ou extra-linguistique dans lequel s'inscrit le moment d'énonciation. Ainsi la langue permet de créer un univers situé par rapport à cet acte d'énonciation ou bien en rupture avec lui [Culioli, 1999].

L'énoncé quand à lui revêt deux propriétés importantes, c'est en premier lieu le produit de l'acte d'énonciation, ensuite on peut dire qu'il constitue une unité cohérente syntaxiquement, il se rapproche de la notion de proposition. Le discours est donc l'ensemble cohérent des énoncés produits par un énonciateur dans un lieu et un moment donné. La cohésion du discours quant à elle est assurée par tous les phénomènes cités dans [Charolles, 1995], tels que les connecteurs entre énoncés et anaphores ou encore les cadratifs.

3.2.2 Eventualité

Nous définissons un évènement, auquel nous préférons le terme éventualité [Caudal, 2006a], comme référant à tout type de relation entre individus se déroulant dans le temps, et codée au sein d'un énoncé. On distingue donc dans la représentation sémantique d'un énoncé ou d'un discours, le domaine peuplé d'individus, et de relations entre ceux-ci, dont certaines sont temporalisées, et dénotent donc pour ces cas particuliers, des éventualités. En d'autres termes la classe la plus évidente des éventualités est exprimée dans la langue française par différents "types de procès ou de situation que dénotent les verbes" [Recanati et Recanati, 1999].

Notons par ailleurs que toute proposition, tout prédicat, renvoie à des unités de temps pendant lesquels ils sont vrais, ce qui inclut d'autres formes linguistiques que le verbe pour exprimer une éventualité. Par ailleurs, remarquons qu'un prédicat ou une

proposition issus d'un énoncé dans lequel il n'y aurait pas de marque linguistique de temporalité pourrait être interprété dans un modèle incluant une dimension temporelle.

3.2.3 Temps et aspect

Le temps linguistique permet de saisir une localisation des éventualités les unes par rapport aux autres. Nous dirons que le temps linguistique est la localisation d'éventualités par rapport à une éventualité exceptionnelle : le moment d'énonciation. Nous pouvons dire que le temps linguistique est l'organisation des extensions temporelles des éventualités et des expressions temporelles énoncées dans le discours. L'aspect quand à lui est la manière, la perspective ou bien encore l'angle selon lequel l'énonciateur montre ou expose une éventualité dans son énoncé. En ceci, il participe à l'expression de sa subjectivité au sein de ses propres énoncés, l'énonciateur construit son ethos en même temps qu'il construit la temporalité des éventualités dans son discours. Il est néanmoins nécessaire de faire une distinction entre deux dimensions de l'aspect, nous citons Caudal [Caudal, 2006a] à ce propos :

Le terme d'aspect désigne dans la tradition linguistique romane et germanique la manière dont la langue encode le déroulement une éventualité (c'est-à-dire, d'un état ou d'un événement).

Historiquement identifiée pour les langues slaves, la catégorie de l'aspect y est incarnée d'une façon tout à fait différente, car exclusivement liée à la morphologie dérivationnelle (hormis, par exemple, dans le cas du bulgare, qui a conservé une morphologie aspectuelle flexionnelle). Pour cette raison, on s'en tiendra ici aux données et analyses concernant les langues germaniques et romanes, au moins dans un premier temps.

Dans la linguistique des langues romanes et germaniques, donc, il est d'usage d'opposer l'information aspectuelle exprimée par le noyau verbal, ses auxiliaires ou semi-auxiliaires, et l'ensemble de sa complémentation et de ses modificateurs, d'une part, et l'information aspectuelle exprimée par la morphologie flexionnelle verbale, d'autre part. La première sorte d'information aspectuelle est généralement traitée en termes de "types de procès/d'événements", ou Aktionsarten il s'agit de classer les descriptions d'événements en fonction de différents propriétés aspectuelles. La seconde sorte d'information aspectuelle est en revanche traitée tantôt en termes de simple outil de recatégorisation de l'Aktionsart, soit en des termes lui attribuant un statut aspectuel à la fois différent et plus complexe, selon les approches suivies.

Dans cette section il sera question d'Aktionsart, appelé communément aspect lexical et d'aspect grammatical, qui permet la recatégorisation en contexte de l'aspect initialisé lexicalement.

Nous défendons le point de vue selon lequel énonciation et aspect sont des composants du temps linguistique. On dira que la temporalité d'une éventualité est le fruit de l'interaction entre l'éventualité objectivée, telle qu'elle s'est déroulée dans le réel et le regard qu'on pose sur elle, ce qui inclut un calcul d'ordre aspectuel et énonciatif.

3.3 Système de calcul de la temporalité verbale et modèle d'interprétation de la temporalité

Nous présentons dans cette section deux auteurs phares dans le traitement de la temporalité en sémantique formelle. Le premier, Reichenbach, a mis en évidence des relations élémentaires entre indices permettant de distinguer les temps verbaux les uns des autres. Allen quant à lui a proposé un modèle en intervalles muni de 13 relations minimales permettant de représenter le temps comme un axe orienté, ce qui a fourni les premiers outils à nombres de travaux sur la temporalité. Nous décrivons brièvement le système du premier et le modèle du second.

3.3.1 Reichenbach et la description des temps verbaux

Reichenbach dans [Reichenbach, 1947] propose un modèle d'interprétation du temps verbal sur une structure ordonnée, un axe orienté. Il présente la temporalité comme étant construite à partir du moment d'énonciation qui aura pour représentation sur l'axe un point :

- *S* : le point d'énonciation (Point of Speech)

qu'il met en relation avec l'éventualité sur cet axe grâce à

- *E* : le point de l'éventualité (Point of Event)

et il détermine une dernière coordonnée, le repère, ou point d'anaphore, qui permet de savoir où se place le dernier repère dans le cours du récit :

- *R* : le point de repère (Point of Reference)

Les relations élémentaires sont l'égalité et la précédente et l'on peut par exemple utiliser des intervalles à la place des points pour les formes progressives.

Ces trois points ou indices, sont chacun mis en relation deux à deux afin de dresser les représentations de chaque temps conjugué (la virgule représente la simultanéité, le tiret l'antériorité) :

Structure	Reichenbach Name	Traditional Name
E-R-S	Anterior Past	Past Perfect
E,R-S	Simple Past	Simple Past
R-E-S R-S,E R-S-E	Posterior Past	
E-S,R	Anterior Present	Present Perfect
S,R,E	Simple Present	Present
S-R-E S,E-R E-S-R	Anterior Future	Future Perfect
S-R,E	Simple Future	Simple future
S-R-E	Posterior Future	

TABLE 3.1 – Sémantique des temps conjugués de l'anglais de Reichenbach 1948

Nous donnons un exemple de discrimination standard entre les deux temps que sont le *preterit* dans et le *present perfect* dans la figure 3.1 :



FIGURE 3.1 – Le prétérit, appelé *Simple Past* par Reichenbach et le *Present Perfect*

Les deux temps présentent une action qui s'est déroulée avant le moment d'énonciation, on a donc la relation $E < S$ dans les deux cas. La discrimination entre les deux temps se situe dans le positionnement du repère utilisé pour décrire la temporalité de l'éventualité. Dans le premier cas, le prétérit, appelé *simple past* par l'auteur, propose un évènement qui est situé par rapport à un repère passé $R = E$ tandis que le *present perfect* localise son repère dans le présent $R = S$. Ainsi on a réussi à distinguer les deux énoncés suivant :

- (3.1) I ate an apple.
- (3.2) I have eaten an apple.

Pour le premier et le second exemples, le moment où j'ai mangé la pomme est situé avant le moment où je l'énonce : $E < S$, en revanche, le point de repère lui se situe soit dans le passé et simultanée avec l'évènement pour le *prétérit (simple past)*, $R = E$, ce qui signifie que le cours du récit continue dans le passé à partir de cette éventualité. Pour le second exemple au *present perfect*, le point de vue est simultanée avec l'énonciation, $R = S$, la suite du récit est ancré par rapport à ce S .

La mise en évidence d'un point de repère, permettant d'exprimer les différentes perspectives et l'anaphoricité dans le récit (son cours) est le point de départ de nombreux travaux, il permet de discriminer la sémantique des temps de l'anglais et du français et de beaucoup d'autres encore (Reichenbach s'est intéressé par ailleurs aussi au turc et à l'allemand). En revanche la définition de l'aspect est inexistante et même si les trois relations entre point ou intervalles permettent de différencier des temps dont l'aspect est le caractère discriminant majeur. Comme nous le verrons plus loin, le modèle ne suffit pas à distinguer des concepts plus fins tels que l'aspect aoriste propre au passé simple en français ou encore l'itératif. Ce modèle manque de spécification quant à la structure orientée et ordonnée utilisée, et nous introduisons dans la section suivante une proposition de Henk Verkuyl résolvant une partie des problèmes rencontrés et servant comme base à nos travaux.

3.3.2 Les intervalles de Allen

Dans [Reichenbach, 1947], l'auteur propose des intervalles particulièrement pour les évènements (E) dans certains cas comme les progressifs par exemple. En revanche, il ne propose pas d'alternative aux points pour R ou encore S. Or il semble légitime d'utiliser un modèle en intervalles pour le calcul de la temporalité mettant en jeu ces

3.3. SYSTÈME DE CALCUL DE LA TEMPORALITÉ VERBALE ET MODÈLE
D'INTERPRÉTATION DE LA TEMPORALITÉ

objets tels que le moment d'énonciation, un repère ou encore les éventualités entre elles. Nous dirons simplement que les points posent en premier lieu des problèmes de granularité : comment représenter des éventualité les unes par rapports aux autres si, temporellement l'une est incluse dans l'autre par exemple. Par ailleurs les points s'ils deviennent nécessaires pourraient être portés par des intervalles (minimaux en l'occurrence) tandis que l'inverse n'est pas vrai, et il va sans dire que la richesse des relations entre intervalles est sans comparaison avec les possibles relations entre points, c'est pourquoi nous choisissons de présenter ici le principal modèle d'interprétation utilisé pour la modélisation du temps, les intervalles d'Allen[Allen, 1983].

Chaque intervalle possède une borne initiale d et une borne finale f , tels que $d < f$ et Allen a dégagé treize relations minimales permettant de représenter toutes les relations possibles entre deux intervalles, voir figure 3.2.

Relations		Interprétations
$x < y$	$y > x$	
$x m y$	$y mi x$	
$x o y$	$y oi x$	
$x d y$	$y di x$	
$x s y$	$y si x$	
$x f y$	$y fi x$	
$x = y$		

FIGURE 3.2 – Les 13 relations entre intervalles de Allen 1983

Chaque relation possible entre deux intervalles est définie par un opérateur et son inverse (i ajouté à l'opérateur pour m (meet), o (overlap), d (during), s (start), et f (finish)). L'antériorité s'oppose à la postériorité et seule l'égalité n'a pas de pendant.

Nous ne nous attardons pas sur la logique temporelle proposée par Allen, qui fait entrer dans la sémantique beaucoup d'éléments pragmatiques tels que l'intention du locuteur dans ses dires (dans le sens des objectifs poursuivis par ce locuteur lors de l'énonciation)¹. Nous dirons simplement que l'apport principal de cette présentation est de fournir un modèle d'interprétation du temps linguistique en intervalle sur lequel beaucoup de travaux se sont basés. Nous présentons dans la section 3.5 deux d'entre eux sur le français plus particulièrement.

1. Nous reprochons principalement à celle-ci son caractère peu automatisable.

3.4 La structure des éventualités

Nous avons montré que Reichenbach propose d'interpréter n'importe quelle éventualité comme un point ou un intervalle sur lequel l'énonciateur pose un regard, lui-même situé par rapport au moment de l'énonciation. Néanmoins, nombre de théories distinguent plusieurs types sophistiqués d'éventualités à partir de l'observation de leur comportement temporel en discours et proposent une modélisation de leur structure interne.

3.4.1 Une première typologie des éventualités

Vendler [Vendler, 1967] s'intéresse à une classification des verbes de l'anglais exprimant des éventualités. Cette classification a pour objectif de distinguer différents comportements temporels au sein des discours. Les quatre catégories proposées sont les *states*, les *activities*, les *accomplishments* et les *achievements*².

En premier lieu il convient de rappeler que Vendler considère l'éventualité comme étant définie par le verbe et une partie de son contexte, son objet syntaxique, ou un adverbial par exemple. Ainsi, *courir cinq kilomètres* dénote une éventualité d'une catégorie différente de celle de l'éventualité *courir*. La classification ne concerne pas véritablement une classification des verbes pour et par eux-mêmes mais bien une classification des éventualités en fonction de leurs propriétés syntaxiques et sémantiques. Ces catégories ont été éprouvées et analysées dans [Recanati et Recanati, 1999] et nous livrons ici une synthèse partielle des travaux de Vendler à la lumière de [Recanati et Recanati, 1999] (nous utilisons par ailleurs certains exemples et le tableau présenté dans ce dernier).

On peut observer différents critères pour déterminer à quelle classe un verbe appartient selon deux tests majeurs.

Le premier concerne les éventualités pouvant être exprimées avec un progressif, forme conjuguée en anglais, que nous traduisons par une périphrase en français :

(3.3) *He is singing* / il chante ou il est en train de chanter

(3.4) *He is writing a letter* / il écrit une lettre ou il est en train d'écrire une lettre

tandis que

(3.5) *He knows he is late* / il sait qu'il est en retard

(3.6) **He is knowing he is late* / *il est en train de savoir qu'il est en retard

(3.7) *I recognize your voice* / je reconnais ta voix

(3.8) **I am recognizing your voice* / ? je suis en train de reconnaître ta voix

Ce critère met en évidence une distinction entre un premier groupe formé des *activities* et *accomplishments*, et un second, regroupant les *states* et *achievements* et qui intuitivement s'expliquerait par le caractère défini ou indéfini de l'extension temporelle de l'éventualité. *Chanter* et *écrire* dans ces exemples, dénotent des éventualités dont

2. Nous conservons la terminologie anglosaxonne originale.

3.4. LA STRUCTURE DES ÉVENTUALITÉS

l'extension temporelle est définie et unique, tandis que pour le second groupe, l'extension temporelle est indéfinie, voire insaisissable dans le cas de *reconnaître* qui ajoute une dimension ponctuelle.

Le second concerne la possibilité d'utiliser ou non un syntagme prépositionnel introduit par *pendant* en français (*during* en anglais) :

(3.9) J'ai cru cela pendant quatre jours

(3.10) J'ai marché pendant trois heures

tandis que

(3.11) ? Il a fait le portrait de Marie pendant dix minutes

(3.12) Il a fait le portrait de Marie en dix minutes

(3.13) *Il a atteint le sommet pendant trois heures

(3.14) Il a atteint le sommet en trois heures

Cette seconde distinction propose une répartition en deux autres groupes, le premier formé des *states* et des *activities* tandis que le second cette fois regroupe les *accomplishments* et les *achievements*. Cette distinction peut s'interpréter par le caractère homogène ou non d'une éventualité. Par caractère homogène on entend que cette éventualité poursuit un but, une finalité qui doit être atteinte pour que l'éventualité soit complète. Par exemple, *dessiner un cercle* aurait pour finalité le cercle sur la feuille de papier, et *gagner une course*, le nouveau statut de vainqueur pour quiconque effectue l'action. Tandis que *croire* ou *courir* dénotent des éventualités sans apogée.

On représente donc la classification de cette manière :

	- progressif	+ progressif
+ 'pendant' (homogène)	<i>states</i> croire, savoir, être	<i>activities</i> écrire, courir, marcher
- 'pendant' (hétérogène)	<i>achievements</i> reconnaître, gagner la course	<i>accomplishments</i> faire le portrait, dessiner un cercle

TABLE 3.2 – Classification des éventualités de Vendler 1967

Nous ajoutons une remarque de Vendler intéressante quant à sa classification, il affirme qu'entre *achievements* et *accomplishments*, si on utilise une durée, alors c'est la durée de l'éventualité complète pour l'un et la durée d'une phase préparatoire pour le second :

(3.15) J'ai fait le portrait de Sam en dix minutes.

(3.16) J'ai gagné la course en dix minutes.

L'éventualité *faire le portrait* a bien duré dix minutes, tandis que *gagner la course* implique de la courir d'abord pendant les dix minutes, ce qui a abouti à l'éventualité *gagner la course*. Nous ajoutons que cette distinction semble fragile, et sera discutée dans le chapitre suivant.

3.4.2 De l'interaction entre structure interne des éventualités et agentivité

Pustejovsky, dans son projet de lexique sémantique génératif entend donner une représentation logico-sémantique des éventualités en adjoignant deux structures. La première est une structure interne complexe en fonction de la catégorie du verbe hors contexte. La seconde est une structure conceptuelle et lexicale nommée LCS pour *Lexical Conceptual Structure* dans laquelle, entre autres, la structure agentive est décrite [Pustejovsky, 2005].

Il distingue trois types hérités des réflexions de Vendler auxquels correspondent des structures primitives nommées ES pour *Event Structure*, par ailleurs, la structure interne d'une éventualité est compositionnelle (comme nous le verrons pour *closed*). Le contexte joue un rôle dans la LCS et dans l'ES. Ceux qu'il appelle *states*, de type *S* et correspondant à la définition donnée précédemment, ont une structure interne illustrée en figure 3.3. Les *states* sont composés d'une seule éventualité atomique.



FIGURE 3.3 – Structure interne des *states*

La classe des *processes* de type *P* correspondent aux *activities* pour Vendler et ont la structure interne illustrée en figure 3.4.

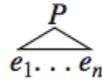


FIGURE 3.4 – Structure interne des *processes*

Les *processes* sont composés d'une séquence d'éventualité $e_1 \dots e_n$ pouvant n'être constitué que d'une seule. Tandis que les *transitions*, nommé *T* regroupent les *accomplishments* et les *achievements* sous la forme suivante :

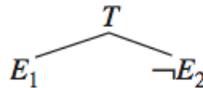


FIGURE 3.5 – Structure interne des *transitions*

Les *transitions* sont structurées sur une opposition entre deux variables avec E_1 et

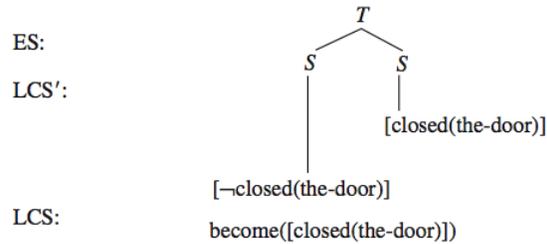


FIGURE 3.7 – Structure interne (ES) et agentivité (LCS) de l'exemple *The door closed*

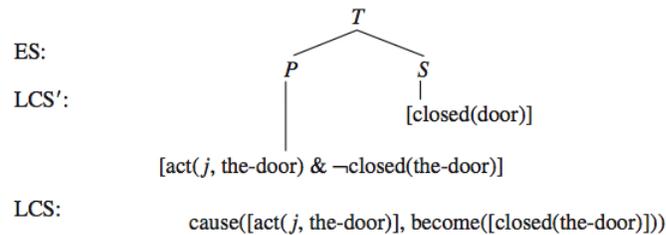


FIGURE 3.8 – Structure interne (ES) et agentivité (LCS) de l'exemple *John closed the door*

brancher sur la structure déjà en place avec par exemple le type fonction $\langle P, T \rangle$ pour *to* :

(3.20) *Mary ran to the store.*

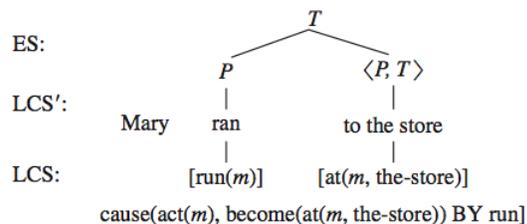


FIGURE 3.9 – Structure interne (ES) et agentivité (LCS) de l'exemple *Mary ran to the store*

Le syntagme prépositionnel *to the store* aura pour entrée un *process ran* premièrement considéré comme tel et en fait une *transition* (permettant à *Mary* de passer de

l'état $\neg at(m, the - store)$ à l'état $at(m, the - store)$ ici porté par $become(at(m, the - store))$).

Cette conception compositionnelle des éventualités construite sur des primitives permet de mettre en évidence les traits minimaux communs aux éventualités issus du même verbe et se trouvant dans des contextes différents. Néanmoins, il est quelque peu difficile de connaître la structure interne ES sans le contexte. Les opérations de transformations sous-jacentes au passage d'un *state* à une *transition* entre les deux premiers exemple *The door is closed* et *The door closed* permettrait d'exploiter complètement la compositionnalité de telles structures, mais ne sont pas détaillées à notre connaissance dans ces travaux. Pour une théorie exploitant davantage les mécanismes de transformation d'une éventualité d'une catégorie à une autre, nous présentons l'automate de Moens et Steedman dans la section suivante.

3.4.3 Les transitions entre types d'éventualités

Moens et Steedman [Moens et Steedman, 2005] présentent un modèle dans lequel la représentation d'une éventualité prend en entrée une structure *nucleus* standard à toute éventualité sur laquelle influent toutes les catégories concernées par la temporalité en sémantique telles que l'aspect, les adverbiaux aspectuels, les propositions elles-mêmes.

La structure d'un *nucleus* permet de représenter l'archétype d'une éventualité telle que suit :

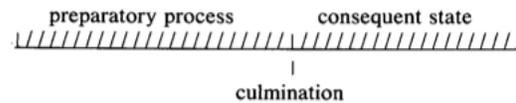


FIGURE 3.10 – Nucleus de Moens et Steedman 1988

Les critères pour déterminer la typologies sont différents de ceux de Vendler et par conséquent ils obtiennent une classification en cinq catégories présentée en figure 3.11.

La différenciation des états, ou *states*, du reste des éventualités n'est pas remise en cause, par contre les deux critères utilisés pour discriminer les éventualités restantes sont l'opposition entre la ponctualité (que l'on peut rapprocher d'une propriété d'atomicité) et l'extension temporelle (on parlera alors d'éventualité étendue) et, l'association ou non à un état conséquent.

L'ajout majeur de cette théorie réside dans la modélisation de l'interaction entre le verbe et chacun des éléments du contexte, dont les aspects grammaticaux portés par les conjugaisons. La figure suivante représente l'ensemble des transitions possibles recensées par Moens et Steedman sur l'anglais.

Par exemple, l'utilisation d'un adverbial commençant par *for* en anglais pour exprimer une durée ne peut être appliquée qu'à un *process*. *arrived* est considéré comme

	EVENTS		STATES
	atomic	extended	
+conseq	CULMINATION recognize, spot, win the race	CULMINATED PROCESS build a house, eat a sandwich	understand, love, know, resemble
-conseq	POINT hiccup, tap, wink	PROCESS run, swim, walk, play the piano	

FIGURE 3.11 – Classification des éventualités de Moens et Steedman 1988

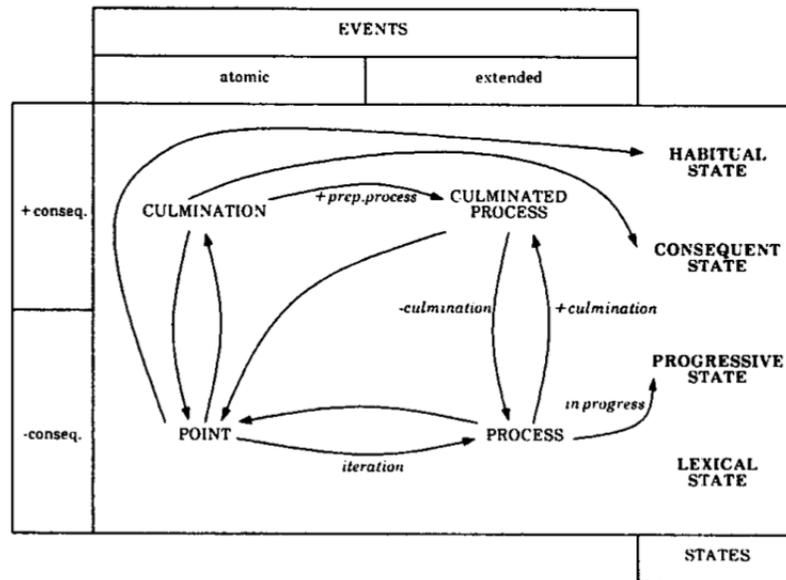


FIGURE 3.12 – Transitions entre catégories d'éventualités, Moens et Steedman 1988

culminatif, et une fois *late* pris en compte, la transition vers *point* est assurée. Si l'éventualité à laquelle l'adverbial est appliqué est un *point*, alors il force une transition vers *process* en lui donnant une sémantique itérative :

(3.21) John arrived late at work for several days

Nous donnons une schématisation simplifiée des transitions d'une catégorie à l'autre

pour l'exemple *John arrived late at work for several days* :

1. $(John\ arrive) : culmination$
2. $late(John\ arrive) : culmination \rightarrow point$
3. $for_several_days(late(John\ arrive)) : culmination \rightarrow point \xrightarrow{iteration} process$

Ces trois théories présentent chacune une typologie des éventualités intéressante et une conception différente de l'éventualité. Vendler [Vendler, 1967] a été le premier à proposer une typologie complète et a ouvert beaucoup de questions autour du verbe *too see* qu'entendent résoudre Pustejovsky et Moens et Steedman, en proposant une lecture compositionnelle pour le premier et transitionnelle pour le second des éventualité en contexte. Ces deux propositions prennent en compte les éléments environnants de l'éventualité dans l'énoncé mais n'abordent pas le temps conjugué de manière systématique, c'est pourquoi nous introduisons les auteurs Gosselin et Battistelli dans la section suivante.

3.5 Deux modèles d'interprétation temporels et aspectuels pour le français

Gosselin et Battistelli, présentent chacun un système de calcul inspiré de Reichenbach ainsi qu'un modèle d'interprétation inspiré d'Allen. Ce modèle est une structure ordonnée linéaire, munie d'intervalles.

3.5.1 Une théorie de la sémantique temporelle du français : la Sémantique des Temps

Gosselin [Gosselin, 2005] utilise les mêmes catégories d'éventualités que Vendler, à savoir les états, les activités, les accomplissements et les achèvements. Dans le modèle de Gosselin, les intervalles sont définis par leurs bornes initiales et finales et on retrouve en partie les valeurs proposées par Reichenbach originellement :

- $[01,02]$ est l'intervalle temporel du moment d'énonciation : S chez Reichenbach
- $[B1,B2]$ est l'intervalle temporel de l'éventualité : E chez Reichenbach
- $[I,II]$ est l'intervalle temporel de référence, une "fenêtre" qui délimite ce qui est perçu, montré de l'éventualité : serait R chez Reichenbach mais les deux théories ne lui font pas porter exactement les mêmes valeurs.
- $[ct1,ct2]$, un quatrième intervalle est donné pour les intervalles circonstanciels qui permettent la représentation des expressions temporelles adverbiales.

Ce modèle est muni de relations entre les bornes des intervalles qui permet de décrire les temps et aspects du français. Les relations entre intervalles suivantes sont définies par leur relation entre bornes : Soient l'intervalle $x = [x_1, x_2]$ et l'intervalle $y = [y_1, y_2]$

Un première distinction s'opère sur une classification en temps "absolu" :

relation	relation	correspondance chez Allen
x ANT y	$x_2 < y_1$	$x < y$
x POST y	$x_2 > y_1$	$x > y$
x SIMUL y	$x_1 \leq y_2 \wedge y_1 \leq x_2$	$x \text{ d } y \vee x = y \vee x \text{ s } y \vee y \text{ f } x \vee$
x CO y	$x_1 = y_1 \wedge x_2 = y_2$	$x = y$
x RE y	$x_1 < y_1 \wedge x_2 > y_2$	$x \text{ di } y$

TABLE 3.3 – Relations entre intervalles Gosselin 2005

- Passé : [I,II] ANT [01,02]
- Présent : [I,II] CO [01,02]
- Futur : [I,II] POST [01,02]

Tout comme introduit dans la section 3.2.2, l'aspect grammatical se distingue de l'aspect lexical, cette seconde distinction est introduite dans [Gosselin, 2005] (l'aspect grammatical dont il s'agit, sera appelé "visée aspectuelle" dans [Gosselin, 2010]³) :

L'aspect lexical correspond au "type de procès" marqué par le verbe et son environnement actanciel. Il s'agit du procès tel qu'il est "conçu" alors que l'aspect grammatical définit la façon dont il est "montré/perçu".

Nous donnons ici des exemples de Gosselin lui-même :

- Aoristique : [B1,B2] CO [I,II]
(3.22) Il traversa le carrefour.
- Inaccompli : [B1,B2] RE [I,II]
(3.23) Il traversait le carrefour
- Accompli : [B1,B2] ANT [I,II]
(3.24) Il a terminé son travail depuis dix minutes
- Prospectif : [B1,B2] POST [I,II]
(3.25) Il allait traverser le carrefour

Gosselin envisage les phénomènes de recatégorisation de l'éventualité, tels que présentées comme des transitions dans l'automate de Moens et Steedman, comme des conflits à résoudre. L'itération par exemple peut-être le produit de la résolution d'un conflit entre l'aspect imperfectif et la présence d'un circonstanciel de durée, d'autres éléments peuvent déclencher l'interprétation itérative⁴.

La modalité a un impact sur la temporalité et l'aspectualité des éventualités, elle désigne le changement entre ce qui est irrévocable et ce qui devient possible mais il

3. Par ailleurs, dans [Gosselin, 2010], l'auteur nuance cette opposition entre aspect lexical et grammatical, notamment dans ses travaux sur les périphrases, qu'il situe « entre lexicale et grammaticale », les unités minimales aspectuelles du type « commencer à » ou « aller VInf » pouvant être exprimées soit lexicalement soit grammaticalement d'une langue à une autre

4. Nous remercions Patrice Enjalbert pour l'avoir souligné.

3.5. DEUX MODÈLES D'INTERPRÉTATION TEMPORELS ET ASPECTUELS POUR LE FRANÇAIS

distingue la modalité temporelle (posée en borne finale de l'intervalle d'énonciation) de la modalité aspectuelle (posée en borne finale de l'intervalle de référence ou fenêtre de monstration).

Nous donnons ici deux exemples de l'auteur pour illustrer ce propos :

(3.26) La cavalerie attaquera à l'aube (exemple emprunté lui-même à Fauconnier)

Dans ce premier cas, la modalité du possible indiquée par trait discontinu, commence

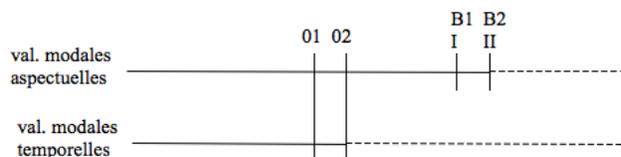


FIGURE 3.13 – Interprétation de la modalité temporelle et aspectuelle pour *La cavalerie attaquera à l'aube* Gosselin 2009

à partir de la fin de l'éventualité ([I, II], elle même en correspondance exacte avec la fenêtre [B1,B2] pour le référentiel aspectuel, tandis que la modalité temporelle quant à elle démarre à l'issue de l'énonciation. Si l'on veut, pour le référentiel temporel, la modalité du possible se situe toujours tout de suite après maintenant, (je ne peux être sûr de l'instant à venir, la seule connaissance irrévocable du monde que j'ai concerne ce que j'ai déjà vécu). Tandis que la modalité aspectuelle permet de montrer l'incertitude quant à la finalité d'un imparfait :

(3.27) Il traversait la route

Ici, si la phrase suivante dans le discours est *mais il s'arrêta au milieu*, alors B2 ne

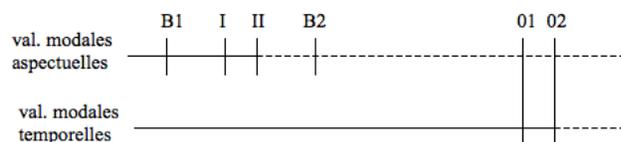


FIGURE 3.14 – Interprétation de la modalité temporelle et aspectuelle pour *Il traversait la route* Gosselin 2009

sera jamais atteinte, c'est pourquoi l'intervalle [I, II] qui permet un regard particulier sur l'éventualité permet de marquer la fin de l'irrévocable.

Ces deux référentiels sont constitutifs de la temporalité des éventualités et l'on construit les deux dimensions, d'un côté l'éventualité et ce qui en est montré, et de l'autre ce qui est vécu.

3.5.2 Une typologie aspectuelle des éventualités du français

La version de la théorie de Battistelli [Battistelli, 2009] que nous présentons se pré-occupe de modéliser l'aspect d'une éventualité afin de la représenter temporellement.

Néanmoins, ces travaux ont pour visée applicative l'extraction d'informations à partir de récits, qui imposent un mode narratif qui ne nécessite pas dans les exemples de placer les éventualités par rapport au moment d'énonciation. Tout comme Vendler, Pustejovsky, Moens et Steedman, et Gosselin, Battistelli [Battistelli, 2009] présente une théorie distinguant la catégorie de l'éventualité ou sa structure interne, appelé aussi aspect lexical, de son pendant grammatical, c'est à dire la recatégorisation de l'éventualité en fonction du contexte dans lequel il se trouve. Néanmoins, la représentation sémantique d'une éventualité est une synthèse des deux plutôt que le résultat d'une transition qui effacerait l'état précédent. En ceci elle rejoint Pustejovsky qui distingue une sémantique minimale grâce à laquelle on compose une structure plus complexe en gardant en son sein cette structure atomique⁵.

Battistelli propose une opération \otimes , entre deux référentiels, un référentiel prédicatif (qui témoigne de ce qu'on appelle l'aspect lexical, autrement dit "en système") et un référentiel énonciatif (qui témoigne de l'aspect grammatical, c'est-à-dire "en discours").

L'auteur argumente en faveur d'un calcul temporel dont la base est l'aspectualité énonciative et prédicative, et propose, pour ce faire, le raffinement du modèle en référentiels mis au point par Declés et Guentcheva [Declés et Guentchéva, 2000]. Nous citons ici Delphine Battistelli qui présente les différents outils utiles à la création du modèle :

- trois notions aspectuelles de base (état, évènement, processus) représentées et visualisées par des intervalles topologiques ;
- l'introduction de différents référentiels temporels ;
- les notions plus temporelles de simultanéité, de différenciation (antériorité/postériorité) ou de rupture entre référentiels, qui permettent de repérer les procès énoncés à l'intérieur de chaque référentiel.

Les trois catégories aspectuelles choisies par Declés et Guentcheva [Declés et Guentchéva, 2000] et reprises par Battistelli [Battistelli, 2009] sont :

- les états communément reconnus, représentés par un intervalle ouvert
- les évènements, qui se voient attribuer un intervalle fermé dont les bornes initiale et finale sont respectivement la borne finale et initiale de deux états (préparatoire et conséquent)
- et les processus, représentés par un intervalle semi-ouvert à droite.

Les processus sont la classe la plus complexe qui permettent d'introduire la coupure modale [Battistelli, 2009]. En effet, le processus est vu comme un intervalle fermé en sa borne initiale et ouvert en sa borne finale, auquel est attribué un état initial le précédent, et un état postérieur possible.

L'axe orienté du temps dans le modèle de Declés-Battistelli ne représentant que la modalité irrévocable, la modalité possible est un axe branchant voir figure 3.16

La prédication et l'énonciation sont définies comme des niveaux de représentation permettant respectivement de modéliser l'aspect au sein de la prédication, au sens lexical, et de la modéliser au sein de l'énonciation, c'est à dire en représentant les degrés de prise en charge de son discours l'énonciateur à l'origine de celui-ci.

5. C'est par ailleurs le cas dans les trois exemples de Pustejovsky contenant l'ES de l'état *closed*

3.5. DEUX MODÈLES D'INTERPRÉTATION TEMPORELS ET ASPECTUELS POUR LE FRANÇAIS

La visée aspectuelle de chacun des niveaux pouvant être respectivement représentée par ASP_E et ASP_{RP} , le second est argument du premier : $ASP_E(ASP_{RP})$. Ainsi si l'on analyse

(3.28) Il gagnait la course.

On peut dire du côté prédicatif que *gagner une course* un évènement, qui a une finalité et plutôt ponctuel, et donc a un état conséquent (*il est le vainqueur de cette course*), tandis que sur le plan énonciatif, on a à faire à un processus plutôt homogène dans ce qui est présenté comme irrévocable.

La catégorie attribuée en lexique s'appelle schème aspectuel prédicatif et appartient à l'ensemble :

$$\{SC_EVEN_{RP}, SC_PROC_{RP}, SC_ETAT_{RP}\}$$

La catégorie attribuée en discours s'appelle schème aspectuel énonciatif et appartient à l'ensemble :

$$\{SC_EVEN_E, SC_PROC_E, SC_ETAT_E\}$$

Chacun des schèmes énonciatifs peuvent être appliqués à chacun des schèmes prédicatifs, ce qui donne 9 possibilités, nous introduisons le tableau de l'auteur :

	PROC_E	EVEN_E	ETAT_E	
PROC _P	processus sans terme visé	évènement simple	état d'activité sans terme visé	état résultant
EVEN _P	processus avec terme visé	évènement achevé	état d'activité sans terme visé	état final
ETAT _P	∅ ?	état conséquent	état descriptif	

FIGURE 3.15 – Combinatoire aspectuelle des éventualités selon Battistelli 2011

Les référentiels sont des axes du temps comme nous le voyons dans l'exemple $SC_EVEN_E(SC_PROC_{RP})$ suivant repris de [Battistelli, 2009] :

La représentation de l'éventualité résulte donc dans la synthèse des deux référentiels.

Cette partie des travaux de Battistelli présente ainsi un système spécifique pour une représentation sémantique de l'aspect d'une éventualité. La subjectivité de l'énonciateur dans la construction de l'univers discursif est portée par le référentiel énonciatif qui est le représentant de cette prise en charge des dires de l'énonciateur.

Tout comme Gosselin, Battistelli présente un système de calcul aspectuo-temporel de l'éventualité et propose un modèle dans lequel interpréter ce calcul. Ces deux sys-

CHAPITRE 3. DE LA TEMPORALITÉ DES ÉVENTUALITÉS POUR LA
SÉMANTIQUE FORMELLE

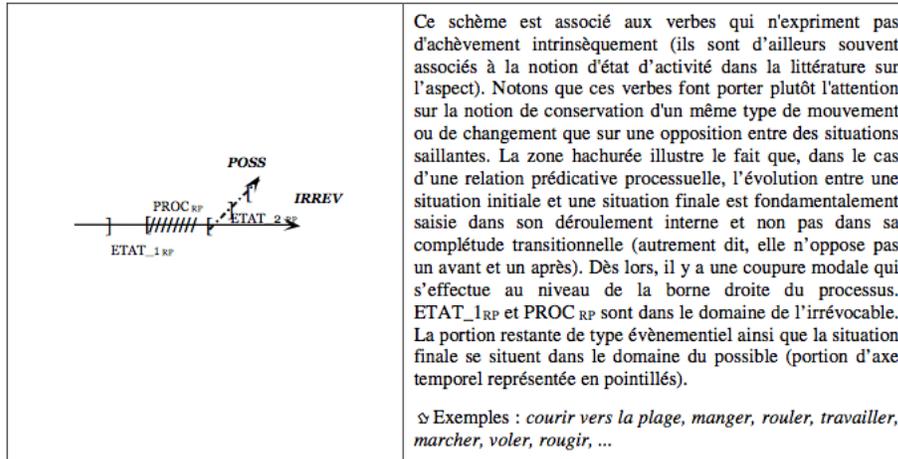


FIGURE 3.16 – Schème prédicatif processus : SC_PROC_{RP} , Battistelli 2009

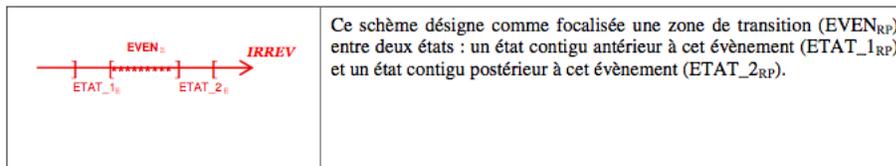


FIGURE 3.17 – Schème énonciatif événement : SC_EVEN_E , Battistelli 2009

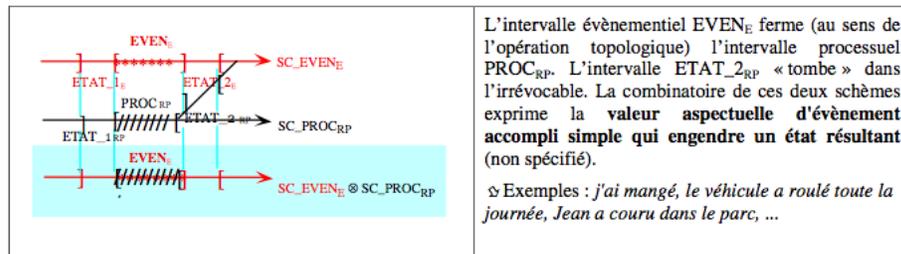


FIGURE 3.18 – Représentation sémantique d'une éventualité selon Battistelli 2009

tèmes ont pour avantage de représenter une dimension énonciative au calcul temporel ainsi qu'une composante modale entre *irrévocable* et *possible*.

3.6 Du traitement de l'éventualité au traitement du discours : la DRT et la SDRT

3.6.1 Le dispositif pour la temporalité en DRT

Ayant montré quelques auteurs théories pertinentes pour le calcul et le modélisation du temps des éventualités, nous pouvons désormais évaluer les outils de la DRT tels qu'ils ont été construits pour traiter le discours. Nous donnons ici rapidement les principes de base de calcul de la temporalité des éventualités en DRT classique davidsonienne [Kamp *et al.*, 2011], qui permet de réifier une éventualité par une variable individuelle au même titre que les marqueurs référentiels pour les personnes. Les deux seules catégories événementielles sont les *états*, et les *événements*. Dans notre langage logique multisorte présenté dans le chapitre précédent, il convient donc de rajouter deux types, *s* pour *state* et *e* pour *event*.

Nous empruntons ici l'exemple de [Kamp *et al.*, 2011] :

(3.29) Yesterday, Fred bought a lawn mower.

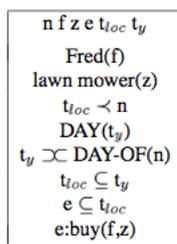


FIGURE 3.19 – La DRS pour l'exemple *Yesterday, Fred bought a lawn mower* Kamp 2011

Les calculs temporels en DRT standard sont faits à partir d'une première relation à la manière de Reichenbach, situant un indice t_{loc} en relation avec la constante n qui permet de pointer le moment d'énonciation, un équivalent de la relation entre *E* et *S* chez Reichenbach.

La différence entre le traitement à la Reichenbach et en DRT tel que le fait Kamp, est une dissociation des tâches originellement portées par *R*, maintenant déléguées à différentes variables. Ainsi en DRT, le *R* de Reichenbach est divisé en deux variables, *r* pour *reference time* et *p* pour *perspective time*. Le premier permet d'assurer le caractère cohérent du discours, de porter l'anaphoricité du fil du récit tandis que le second permet une expression de l'aspect, donnant le point de vue sur l'éventualité *e*. Par exemple :

(3.30) Now, he was worried

r_T est la variable *reference time* qui permet donc de situer le cours du récit, tandis que p_T est la variable *perspective time*, elle montre ici que l'éventualité se déroule dans

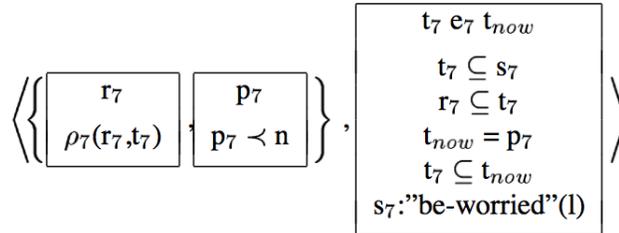


FIGURE 3.20 – La DRS pour l’exemple *Now, he was worried* Kamp 2011

le passé du moment d’énonciation ($<$ est un opérateur pour l’antécédence posée comme étant la relation entre p_7 et n). Le type de l’éventualité est l’état (s_7)⁶, l’anaphoricité du récit est donc explicitée par la localisation de t_{now} imposée par l’adverbe *now* en p_7 (avec la relation $=$), t_7 (*location time* de l’éventualité elle-même) est en relation \subseteq avec t_{now} .

Ces deux variables sont introduites dans la présupposition propre à l’énoncé traité et permettent d’être traitées dans une première phase lors de la fusion entre deux DRS. r change au fil du discours tandis que p est constant.

La DRT étant basée sur une théorie observant la dynamicité du discours, les mises à jour du contexte prennent en compte les présuppositions portées par l’énoncé, ici entre accolades.

Néanmoins certaines limites se posent, nous pensons plus précisément à deux exemples, le premier introduit par Patrick Caudal [Caudal, 2006b] et le second par Nicholas Asher [Lascarides et Asher, 1993] :

(3.31) We reached Dorset at last. A thick blanket of snow covered the fields.

(3.32) John fell. Max pushed him.

Dans le premier exemple, il est impossible de savoir si *covered* exprime un état déjà en place lors de l’éventualité portée par *reached*, ou bien si c’est un évènement qui débute au moment précis où s’achève l’éventualité portée par *reached*. Pour remédier à ce problème, la sous-spécification est une possibilité intéressante de la DRT permettant de garder les deux représentations dans la DRS jusqu’à ce qu’un élément du contexte puisse permettre de trancher.

Pour le second exemple, les deux seuls éléments permettant de résoudre l’ordre temporel entre les deux éventualités sont d’une part des connaissances du monde spécifiant le lien de cause à effet entre *pousser/push* et *tomber/fall*, et de l’autre la relation discursive qui articule ces deux phrases. Nous présentons donc rapidement dans la section suivante la SDRT qui entend répondre à cette problématique.

6. Par ailleurs nous assumons que e_7 devrait être s_7 , et que ce décalage est dû à une coquille dans l’article original.

3.6.2 Une nouvelle granularité en DRT, la SDRT (*Segmented Discourse Representation Theory*)

L'intégration des relations discursives à la DRT a été motivée par l'impossibilité d'exprimer la temporalité d'une suite d'événements en s'appuyant simplement sur la syntaxe et sur le lexique, et par la nécessité de faire appel à des connaissances du langage et du monde pour une représentation sémantique du discours juste. La SDRT [Lascarides et Asher, 1993] a permis de répondre à ce besoin et a intégré les relations rhétoriques entre segments au sein du discours dans un formalisme hérité de la DRT.

L'unité utilisée est l'EDU *elementary discourse unit*, définie comme proposition élémentaire. Autrement dit, l'EDU correspond à un segment textuel minimal du discours représentant soit un événement, soit une période temporelle. Pour se rapprocher de l'analyse textuelle, on dira qu'une phrase peut être composée de plusieurs segments. On peut aussi vouloir relier un bloc de plusieurs segments à un seul segment par une relation, on parlera alors de CDU, *complex discourse unit* pour désigner le bloc. Les CDU contiennent plusieurs unités élémentaires, en préservant la cohérence thématique et rhétorique. La définition formelle de ce concept est proposée dans ([Asher, 2011]). Nous proposons un exemple inspiré de notre corpus et analysé en SDRT :

- (3.33) Notre arrivée à Bagnères ne fût pas facile. [a]
- (3.34) Nous avançons à un rythme irrégulier depuis un moment [b]
- (3.35) et nous entrâmes tard dans la rue Saint-Blaise. [c]
- (3.36) Depuis là-bas, nous fîmes une promenade à Superbagnères. [d]
- (3.37) Nous suivions d'abord un chemin d'une pente peu rapide [e]
- (3.38) et ensuite nous nous élevâmes presque sans nous en apercevoir jusqu'à Superbagnères. [f]

Les EDU [a], [b], [c], [d], [e] et [f] sont des boîtes, des DRS, que l'on traduit par des labels, respectivement π_a , π_b , π_c , π_d , π_e et π_f . π_A et π_D sont les labels de CDU, contenant π_b et π_c pour le premier, et π_e et π_f pour le second. Ici π_a est élaboré par π_A et π_d est élaboré par π_D .

La relation *elaboration* entre π_a et π_A n'est pas donnée par la syntaxe, et seule la connaissance de la géographie de Bagnères et de ses rues permet de s'assurer que π_A contenant l'entrée dans la rue Saint-Blaise est une description des étapes utiles à l'arrivée à Bagnères. L'*elaboration* étant définie comme une relation subordonnante, permettant de relier les deux événements respectivement dénotés dans les segments π_a et π_b par une inclusion thématique et temporelle. La relation temporelle entre les trois événements portés par π_a , π_b et π_c n'est pas une succession stricte portée éventuellement par une relation de narration, mais bien un détail des étapes, π_b et π_c , de l'événement porté par π_a , c'est pourquoi la relation discursive est *elaboration*, impliquant une inclusion temporelle au moins partielle de π_A dans π_a . Les relations discursives *elaboration* et *narration* sont deux relations incompatibles, deux événements ne pouvant par ailleurs pas être successifs et inclus l'un dans l'autre.

En accord avec le traitement en DRT, la SDRT propose une granularité supérieure dans le traitement temporel des éventualités, les regroupant en unité complexe (CDU) si une relation rhétorique du discours le justifie. Ainsi le traitement temporel n'est plus

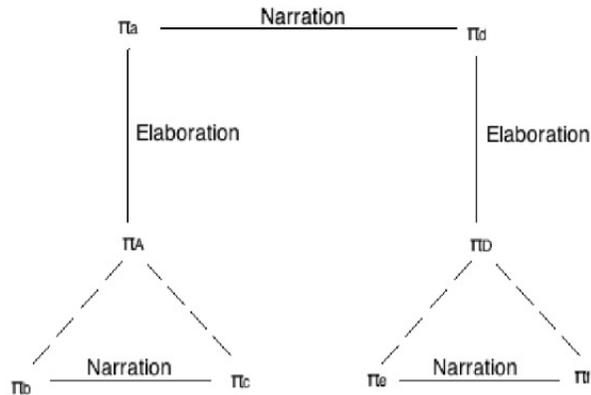


FIGURE 3.21 – une représentation en SDRS des exemples de 3.33 à 3.38

guidé directement par l'éventualité mais par la relation discursive entre le segment précédent et le segment en cours, ainsi que par une base de connaissances du monde posée par une *glue logic*. Ainsi au moment de la fusion en DRT, la SDRT, de son côté, cherche à intégrer les effets temporels des relations rhétoriques. Notons à ce sujet que la construction de la base de connaissances nécessaire à la prédiction de la bonne relation discursive revient à la capture de l'intégralité des relations possibles entre objets de l'intersubjectif (abordé au début de ce chapitre). Cette base de connaissances ne saurait être partielle en vue de l'automatisation totale de cette méthode, et la rendre exhaustive est une entreprise peu réaliste d'un point de vue empirique.

3.7 Conclusion

3.7.1 Synthèse

En guise de synthèse, nous avons aligné dans des tableaux les différents objets utiles aux théories présentées jusqu'ici. Tout d'abord, la typologie des aspects lexicaux selon Vendler (et Gosselin), Pustejovsky, Moens et Steedman, Battistelli et Kamp :

3.7. CONCLUSION

Vendler 1967	states	activities	accomplishments	achievement	
Pustejovsky 2005	states	process	transition		
Moens et Steedman 2005	states	process	culminated process	culmination	point
Battistelli 2009	état	processus		évènement	
Kamp 2011	states	event			

TABLE 3.4 – Aspect lexical des éventualités

Puis l'interaction entre aspect lexical et aspect grammatical lorsque la théorie modélise cette interface :

	Aspect lexical	Aspect grammatical	Résultat
Vendler 1967	entrée lexicale pour l'éventualité portée par le verbe muni de ses arguments	\emptyset	\emptyset
Pustejovsky 2005	ES (structure interne portée par le verbe)	LCS (forme logico-sémantique de l'énoncé)	LCS' (synthèse de l'ES et de LCS)
Moens et Steedman 2005	entrée lexicale pour l'éventualité portée par le verbe muni de ses arguments	transitions en fonction des modificateurs aspectuels	nouvelle catégorie pour l'éventualité
Battistelli 2009	entrée lexicale pour l'éventualité portée par le verbe muni de ses arguments	temps verbal et circonstants temporels	nouvelle catégorie pour l'éventualité

TABLE 3.5 – Calcul aspectuo-temporel, à l'interface entre aspect lexical et aspect grammatical

L'ensemble des travaux sur le calcul de la temporalité des éventualités représente une branche très large des théories de linguistique, et a fortiori de linguistique computationnelle, il serait impossible d'aborder la totalité des travaux qui ont été menés sur cette question, néanmoins, nous avons montré dans un premier temps divers travaux fondateurs utiles à la compréhension de la nature des objets que nous voulons manipuler, les éventualités.

Pour synthétiser notre présentation des systèmes de calcul et des modèles d'interprétation, nous dirons que certains d'entre eux ont une vision aspectuo-temporelle basée sur le temps conjugué (Reichenbach), les autres ajoutent les arguments du prédicat portant l'éventualité (Vendler). Certains ajoutent les modificateurs (Pustejovsky, Moens et Steedman, Gosselin, Battistelli). La DRT ne légitime pas l'utilité d'une catégorisation aspectuelle plus sophistiquée que la bipartition état/évènement, dans le système de calcul. La totalité des systèmes de calcul présentent l'éventualité en rapport avec le moment d'énonciation (Reichenbach, Gosselin, Battistelli, Kamp). Nous n'avons abordé

la composante modale qu’au sein des théories traitant du français mais bien d’autres, dont nous n’avons ici la place pour les introduire, l’abordent dans le calcul temporel.

Les distinctions aspectuelles, quelles soient d’ordre lexicale ou grammaticale apportent une réflexion intéressante sur la diversité des éléments en jeu pour une représentation sémantique interprétable dans un modèle.

3.7.2 Spécifications de nos travaux

Nous rappelons que notre objectif est d’interpréter correctement la temporalité des éventualités, particulièrement lorsqu’elles traitent d’une localisation spatiale ou d’un déplacement. Nous espérons traiter et extraire les informations temporelles et spatiales concernant les itinéraires décrits dans les récits de voyage du corpus ITIPY. Pour cela, nous cherchons un système de calcul de la temporalité des éventualités suffisamment élaboré pour distinguer dans un premier temps les diverses formes conjuguées des verbes, et qui puisse nous permettre de faire entrer les modificateurs temporels tels que les adverbiaux dans le calcul.

Les catégories de Vendler semblent a priori tout à fait fondées mais donnent beaucoup d’informations n’ayant pas forcément un impact immédiat sur l’interprétation d’une relation de temporalité entre un voyageur et une région. En l’occurrence, la possibilité d’utiliser une forme progressive ou non permettant de distinguer les *achievements* des *accomplishments* ne nous semble pour le moment pas pertinente, tandis que le critère d’homogénéité/hétérogénéité semble jouer un rôle prépondérant dans l’interprétation des déplacements. En revanche la distinction unique faite entre état et événement en DRT classique paraît limitée. Même si nous n’utilisons pas dans un premier temps cette typologie dans le calcul, nous argumentons en faveur d’une typologie minimale mais permettant la discrimination de différents types de déplacement.

Notre cadre théorique, la DRT, ne considère que deux types d’éventualités, les états et les événements. Nous montrons dans le chapitre suivant pourquoi nous désirons un troisième type d’éventualité et en quoi il se distingue des deux autres. Par ailleurs, nous nous inspirons de l’ensemble des travaux présentés considérant la nécessité de prendre en compte l’aspect lexical comme une composante du calcul temporel⁷. Tout comme présenté dans la section 3.2.2, notre conception de la temporalité s’inspire de Moens et Steedman, Pustejovsky, Gosselin et Battistelli, concernant l’enjeu d’un aspect grammatical dans le calcul de la temporalité, tout en ajoutant à ceci la localisation du moment de l’énonciation dans ce calcul tel que présenté par Reichenbach en premier lieu puis par Gosselin et Battistelli pour ne citer qu’eux dans ce chapitre. Nous désirons construire un lexique compositionnel permettant une représentation temporelle fine des éventualités à analyser. Nous traitons le temps verbal dans les énoncés et les limites de nos travaux sont les mêmes que les limites présentées pour la DRT. La sous-spécification et sa résolution par des procédés utilisant une granularité plus forte (de la structure du segment à la structure du discours) pose les frontières que nous ne franchissons pas dans le cadre de ce travail. Néanmoins nous avons besoin pour réaliser notre lexique d’un système de calcul temporel des éventualités distinguant tous les

7. Pour plus de précisions sur son utilité, voir la réflexion sur le lexique spatiale et le paradoxe imperfectif section 7.4.2.

temps du français, permettant de discriminer plusieurs formes de déplacement et s'intégrant à notre lexique exprimé en λ -calcul. Sur les conseils de Richard Moot, nous avons choisi comme base du lexique, un système de calcul temporel compositionnel très adapté au traitement en DRT et proposant toutes les spécificités utiles au calcul de la temporalité des éventualités, *Binary Tense* de Verkuyl (2008) [Verkuyl, 2008] que nous présentons dans le chapitre suivant.

*CHAPITRE 3. DE LA TEMPORALITÉ DES ÉVENTUALITÉS POUR LA
SÉMANTIQUE FORMELLE*

Chapitre 4

Présentation et interprétation d'un système bien adapté à la sémantique compositionnelle : *Binary Tense*

Dans ce chapitre nous présentons le système de calcul temporel des éventualités de Verkuyl tel qu'il l'a élaboré dans *Binary Tense* [Verkuyl, 2008]. Ce système, à la base de nos travaux, permet une représentation sémantique compositionnelle de la temporalité des éventualités. Nous donnons notre typologie des *éventualités*. Nous présentons le mécanisme qui, à partir d'un ensemble d'éventualités, nous permet d'obtenir une représentation sémantique temporelle de l'éventualité spécifique à l'énoncé en fonction du temps verbal qui est utilisé. A ce système, élaboré en λ -calcul, nous donnons une interprétation dans les intervalles de \mathbb{R} et détaillons, dans un langage logique typé, le traitement propre à chaque temps couvert par le système.

4.1 Introduction

Nous avons montré que l'analyse syntaxique obtenue automatiquement par *Grail* nous permettait d'associer une représentation sémantique en puisant dans ce lexique sémantique afin de réduire les termes en λ -DRT en suivant l'ordre présenté dans l'analyse syntaxique. Les termes obtenus sont donc bien formés et contiennent les informations

CHAPITRE 4. PRÉSENTATION ET INTERPRÉTATION D'UN SYSTÈME BIEN ADAPTÉ À LA SÉMANTIQUE COMPOSITIONNELLE : BINARY TENSE

essentielles à une interprétation de la relation entre individus. Nos travaux se penchent plus particulièrement sur la temporalité des éventualités de déplacement, éventualités issues de la relation entre le voyageur et l'espace. Nous limitons ici notre étude au temps verbal et proposons une interprétation du système compositionnel de Henk Verkuyl, *Binary Tense* [Verkuyl, 2008], qui sert de base à nos travaux ¹.

Dans ce chapitre, nous abordons la représentation sémantique des déplacements et des localisations dans le cadre du λ -calcul. Nous donnons tout d'abord la typologie des éventualités que nous utilisons afin de faciliter la définition même de nos objets d'étude, les éventualités, puis nous présentons le système de calcul de la temporalité des éventualités *Binary Tense* [Verkuyl, 2008]. Ce système a été créé afin de représenter les informations temporelles liées aux éventualités indépendamment d'une typologie ou de l'aspect lexical tel que nous l'avons abordé jusqu'à maintenant, l'auteur argumente en faveur d'une temporalité « neutre », sans mention d'aspect, dimension qui de notre point de vue et comme argumenté en chapitre 3 semble constitutif du calcul de la temporalité ².

Néanmoins il prend en compte une structure complexe temporelle mettant en jeu le rapport au moment d'énonciation et un phénomène de monstration ³ de l'éventualité (que nous considérons comme une forme d'aspect), ce qui fait de lui un système tout à fait adapté à nos besoins. Verkuyl dans sa théorie présente ce système de calcul avec un lexique associé, déjà exprimé en λ -termes, que nous détaillons ici.

Verkuyl présente un système de calcul de la temporalité des éventualités basé sur la composition de trois choix entre deux relations binaires (2X2X2) permettant de distinguer huit temps verbaux.

Cette approche étant inspirée des travaux de Reichenbach, les points communs et les divergences avec ces derniers seront abordés brièvement ⁴. Rappelons les unités qu'utilise Reichenbach dans [Reichenbach, 1947], sur un axe donné comme base du modèle du temps : le point d'énonciation, l'intervalle de l'évènement et le point de repère en cours (respectivement *S, E* et *R*). Les deux temps que sont le *preterit* et le *present perfect* sont représentés dans la figure 4.1 :

Les deux temps présentent une action qui s'est déroulée avant le moment d'énonciation, on a donc la relation $E < S$ dans les deux cas. La discrimination entre les deux temps se situent dans le positionnement du repère utilisé pour décrire la temporalité de l'éventualité. Dans le premier cas, le *preterit* appelé *simple past* par l'auteur propose

1. Sur le conseil de Richard Moot, que nous remercions à cette occasion.

2. Pour plus de précisions sur son utilité, voir la réflexion sur le lexique spatiale et le paradoxe imperfectif section 7.4.2.

3. terminologie que nous devons à [Gosselin, 2005]

4. Prior, quant à lui a aussi inspiré la proposition de Verkuyl avec un système logique temporel utilisant des opérateurs, qui dans le cas de Prior [Prior, 1967] sont modaux. Prior utilise plusieurs opérateurs : *P* pour *past* et *F* pour *future*, *H* pour '*has always been*' et *W* pour '*will always be*' qu'il combine. La compositionnalité du système d'opérateurs permet de proposer différentes combinaisons pour couvrir les temps conjugués principaux du discours. Néanmoins, le *preterit* selon Prior sera alors porté par le simple opérateur *P*, tout comme le *present perfect*.

Un des atouts du système de Prior réside dans la dynamicité des indices utilisés et leur rôle dans la structure temporelle du discours. A l'instar de Te Winkel, Verkuyl clot les indices par l'existentiel, ce qui permet d'avoir un référent unique dans l'univers du discours pour chaque indice utilisé.

FIGURE 4.1 – Le prétérit, appelé *Simple Past* par Reichenbach et le *Present Perfect*

un évènement qui est situé par rapport à un repère passé $R = E$ tandis que le *present perfect* localise son repère dans le présent $R = S$.

Néanmoins, quelques problèmes apparaissent, par exemple, les constructions comportant "will" en anglais ont plusieurs combinaisons attribuées : pour "will + V", on a trois valeurs possibles⁵ :

(4.1) She will write a letter.

- posterior future $S < R$ et $R < E$: she **will** go to write a letter (elle ira écrire une lettre)
- simple future $S < R$ et $E = R$: she **will** write a letter (elle écrira une lettre)
- posterior present $S = R$ et $R < E$: she is going to / **will** write a letter (elle va écrire une lettre)

Ici la spécificité commune réside dans une relation $S < E$ mais rien ne permet de situer R par rapport à aucun des deux autres points. Verkuyl résout ce problème en introduisant des relations permettant de discriminer davantage les différents temps verbaux. Par ailleurs, il propose une analyse plus précise des temps verbaux que Reichenbach, car pour chaque temps linguistique, on a une seule combinaison parmi les huit possibles et pour chaque combinaison on a un seul temps linguistique. Les temps surcomposés du français sont représentables par une composition utilisant les mêmes opérateurs, néanmoins, il propose un opérateur spécifique pour le passé simple, nommé *Anc*, que nous détaillerons en fin de la section 4.3.

4.2 Nos définitions

Nous proposons dans cette section les définitions de la terminologie que nous utilisons. Le récit met en rapport des évènements, au sens intuitif du terme, entre eux, et permet aussi de les transposer dans le temps réel. Il pourrait aussi être question du temps interne au discours lui-même ("nous verrons ci-après", "comme on l'a vu au chapitre précédent"), mais nous n'étudierons pas ce point.

L'évènement étant un objet sémantique important dans notre étude, nous en donnons ici une définition, et à ce titre nous utiliserons le terme technique d'éventualités : une éventualité est tout ce qui se déroule dans le temps. Une éventualité peut être un

5. Nous devons cette analyse à [Verkuyl, 2003].

She will write a letter.

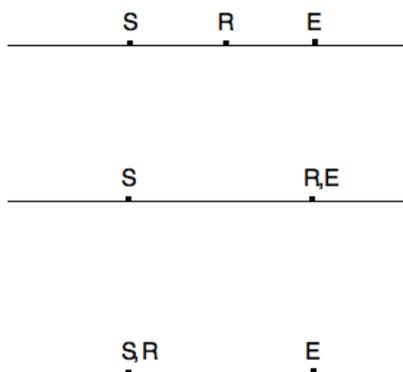


FIGURE 4.2 – Trois interprétations temporelles possibles pour *She will write a letter* dans le système de Reichenbach

événement, une activité, un état, nous donnons quelques brefs tests linguistiques pour les distinguer⁶.

Cette typologie concerne "les types de procès ou de situation que dénotent ces verbes" [Recanati et Recanati, 1999]. Notons que toute proposition, tout prédicat, renvoie à des unités de temps pendant lesquels ils sont vrais. En effet, *courir*, plutôt une activité, ne sera pas considéré de la même manière que *courir cinq kilomètres*, qui sera un événement. Nous montrons la différence entre les deux par l'exemple ci-après. L'homogénéité des unités de temps d'une éventualité, utile à sa catégorisation, est définie dans cette section.

- Etat
L'état est "la manière d'être d'une personne ou d'une chose dans ce qu'elle a de durable" [Rey et Rey-Debove, 1993].
- Activités
L'activité est une action qui est vue comme homogène.
- Evènement

6. Ces considérations ne rentrent pas dans la présentation du système *Binary Tense* qui n'est pas impacté par le type de l'éventualité, mais qui sont pertinentes dans cadre d'une analyse visant à l'extraction d'informations temporelles et spatiales.

L'évènement est une action qui est vue comme hétérogène.

– Homogénéité

Pour définir plus explicitement cette distinction nous citerons 3 éventualités distinctes :

(4.2) Je suis à Pau. Etat

(4.3) Je fais de la randonnée dans les Pyrénées. Activité

(4.4) J'atteins le sommet. Evènement

Nous dirons que les activités et les états sont des éventualités homogènes car elles peuvent être représentées par des intervalles de temps pendant lesquels l'éventualité se déroule et le prédicat concerné est vrai (dans un but illustratif, nous considérons pour cet exemple simplement l'information spatiale de localisation, le prédicat en question est donc réduit à une localisation notée *localise(région,agent,intervalle)*), tandis que pour l'évènement, on dira que l'intervalle de temps pendant lequel se déroule l'action, dans le schéma suivant, contient un intervalle plus petit pendant lequel le prédicat de localisation est vrai. Nous donnons une petite illustration de ce phénomène.

On se demande comment localiser l'intervalle de temps pendant lequel on a été en train d'atteindre le sommet, et l'intervalle de temps pour lequel le sommet est effectivement atteint. La figure 4.3 donne l'interprétation intuitive des trois prédicats. L'état *être_à(Pau, Je, [x,y])*, implique que l'intervalle de temps pendant lequel la localisation de *Je* à *Pau* est effective coïncide avec l'intervalle mobilisé par l'éventualité, pour *randonner(Pyrénées, Je, [x,y])*⁷, l'intervalle de temps pendant lequel la localisation de *Je* dans les *Pyrénées* est effective coïncide avec l'intervalle mobilisé par l'éventualité, tandis que pour *atteindre(Sommet, Je, [x,y])*, l'intervalle de temps pendant lequel la localisation de *Je* au *Sommet* est effective est située uniquement dans un sous-intervalle $[a,b]$ final de $[x,y]$. On donne les implications spatio-temporelles des trois prédicats avec des intervalles et les relations de Allen.

(4.5) $\text{être_à}(\text{Pau}, \text{Je}, [x,y]) \implies \text{localise}(\text{Pau}, \text{Je}, [a,b]) \wedge [a,b] = [x,y]$

(4.6) $\text{randonner}(\text{Pyrénées}, \text{Je}, [x,y]) \implies \text{localise}(\text{Pyrénées}, \text{Je}, [a,b]) \wedge [a,b] = [x,y]$

(4.7) $\text{atteindre}(\text{Sommet}, \text{Je}, [x,y]) \implies \text{localise}(\text{Sommet}, \text{Je}, [a,b]) \wedge [a,b] \text{ends}[x,y]$

7. Nous avons intentionnellement simplifié les prédicats et arguments, par exemple la réalisation linguistique de "faire de la randonnée" sera traduit par le prédicat *randonner*, l'argument *Sommet* est donné comme une constante alors qu'il conviendrait évidemment de le traduire différemment de *Pau* ou de *Pyrénées*. L'interprétation intuitive utilise des unités minimales plus petites que les intervalles pour faciliter la compréhension, ici on peut considérer que a, b, x et y sont des points permettant de désigner les bornes des intervalles.

*CHAPITRE 4. PRÉSENTATION ET INTERPRÉTATION D'UN SYSTÈME BIEN
ADAPTÉ À LA SÉMANTIQUE COMPOSITIONNELLE : BINARY TENSE*

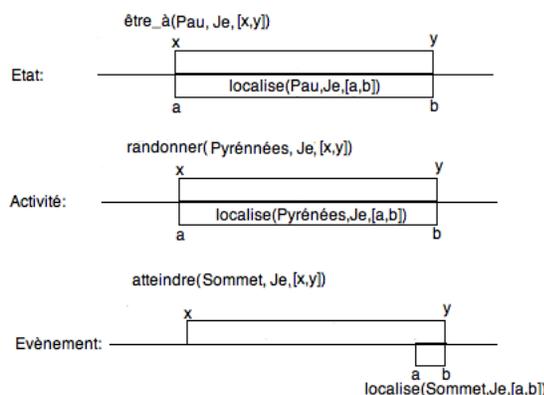


FIGURE 4.3 – L’homogénéité des états et activités et l’hétérogénéité des évènements

Pour les deux premiers exemples, la localisation effective de l’agent dans la région désignée est vraie pour tout sous intervalle de l’intervalle mobilisé par l’éventualité, tandis que pour le dernier, la localisation effective de l’agent dans la région désignée ne concerne qu’un sous-intervalle final.

Le test afin de les distinguer consiste à ajouter un complément de durée de type "pendant trois heures" par exemple.

Illustrons ce propos par des exemples de déplacement ou de localisation :

(4.8) Je suis à Pau pendant trois heures. Homogène

(4.9) Je fais de la randonnée dans les Pyrénées pendant trois heures. Homogène

(4.10) * J’atteins le sommet pendant trois heures. Hétérogène

Un autre test consiste à changer le complément par "en trois heures" et permet de mettre en évidence la finalité propre à l’évènement :

(4.11) ?Je suis à Pau en trois heures. Etat → Evènement

(4.12) *Je fais de la randonnée dans les Pyrénées en trois heures.

(4.13) J’atteins le sommet en trois heures. Evènement

Ici, on comprend bien que "Je suis à Pau en trois heures" est acceptable dans le sens où on désigne le déplacement depuis un autre lieu jusqu’à Pau qui durerait trois heures, et non l’état d’être à Pau. Ce complément fait varier le sens de "être à Pau" de l’état à l’évènement.

Nous faisons ici l’hypothèse par soucis de simplicité que les unités de temps mobilisées sont continues, mais certains cas sont bien plus complexes, par exemple⁸ :

8. Cet exemple donne ici un aperçu des phénomènes de re-catégorisation évoqués par Moens et Steedman

(4.14) Je suis souvent à Pau. Etat

– Eventualité

L'éventualité est la catégorie recouvrant les états, les activités et les événements tels que définis plus tôt. Nous citerons la définition de Saussure dans [De Saussure, 1997] qui définit l' *eventualité* comme un "terme générique pour désigner quelque chose qui se produit, qui a lieu ou qui est vrai sur une période donnée du temps". A l'instar de Bach [Bach, 1981], et comme annoncé précédemment, nous utilisons le terme *eventualité* pour désigner sans distinction état et événement et activités.

Nous introduisons ici notre proposition de typologie des éventualités en regard des théories présentées dans le chapitre précédent, *La temporalité en sémantique formelle* :

Vendler 1967	state	activity	accomplishment	achievement	
Pustejovsky 2005	state	process	transition		
Moens et Steedman 2005	state	process	culminated process	culmination	point
Battistelli 2011	état	processus		événement	
Kamp 2011	state	event			
Lefeuve 2013	état	activité	événement		

TABLE 4.1 – Aspect lexical des éventualités

Notre conception se rapproche de celle de Pustejovsky [Pustejovsky, 2005] pour laquelle trois types d'éventualités sont distinguées.

Tel que montré justement dans le chapitre précédent, les deux typologies des éventualités les plus complexes, celle de Vendler (4 types) et celle de Moens et Steedman (5 types) sont tout à fait légitimes en regard des tests qui ont permis de les mettre en évidence. Néanmoins, à ce stade de notre travail, une classification en trois types semble suffisante pour une déduction à terme des informations sur la réalisation effective d'un déplacement :

- les états, dont le statut n'est remis en cause par aucune des théories, n'impliquent pas de déplacement, mais une localisation : *être à Pau*
- les activités impliquent un déplacement homogène, autrement dit sans « cible » locative particulière : *randonner dans les Pyrénées*
- les événements qui une fois réalisés impliquent un déplacement hétérogène comprenant une « cible » : *atteindre le sommet*.

Il nous faut dire ici que Verkuyl dans *Binary Tense* défend une conception divergente de la temporalité, n'incluant pas une forme d'aspect lexical tel que nous l'avons présenté dans le chapitre sur la temporalité en sémantique formelle.

dans le chapitre précédent, l'adverbe *souvent* associé à un verbe dénotant un état et conjugué au présent de l'indicatif imposant une lecture itérative de cet état, qui par conséquent ne pourrait être lu comme mobilisant un intervalle de temps continu.

D'après l'auteur, la temporalité est neutre, et le type d'éventualité ne pouvant être définie dans la majorité des cas, qu'une fois les arguments et adverbiaux pris en compte, il propose une version non aspectuelle de l'éventualité, les regroupant toutes dans l'ensemble des éventualités sans les distinguer.

4.3 Le langage logique typé

4.3.1 Les éventualités et la λ -DRT

Dans le cadre de la DRT, on montre rapidement comment sont représentées les éventualités sans la temporalité. Les référents de discours permettent de peupler le domaine, ou univers du discours, (représenté par la première partie de la structure en DRT) sur lequel on évaluera les conditions d'interprétation de la structure (deuxième partie de la DRS).

Suivant la méthode employée par Davidson [Davidson, 1967], les éventualités sont réifiées dans la représentation sémantique des énoncés linguistiques.

(4.15) Azais quittera Cauterets à cinq heures.

On donne une variable e permettant de réifier l'éventualité exprimée par le prédicat *quitter* et à laquelle on donnera les propriétés temporelles imposées par le temps verbal, ici le futur simple, et les propriétés imposées par les adverbiaux de temps, ici à *cinq heures*.

(4.16) $\exists e. \text{quitter}(e, \text{Cauterets}, \text{Azais}) \wedge \grave{\text{a}}(e, \text{cinq_heures})$

Ici, on voit que l'on peut donner des informations sur cet évènement et on peut lier les informations temporelles des adverbiaux de temps à l'éventualité e mais on ne dit rien sur le temps verbal. La DRS sera donc construite comme suit⁹ :

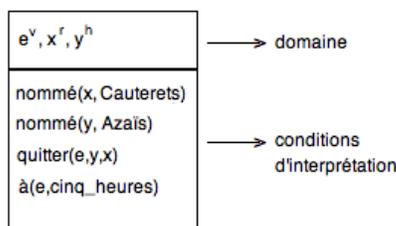


FIGURE 4.4 – La constitution d'une DRS

Dans le modèle créé¹⁰, on a quelques informations sur l'éventualité et sa temporalité, mais rien qui puisse permettre de distinguer *Azais quittera Cauterets à cinq heures*

9. La formule associée sera donc plus précisément : $\exists e \exists x \exists y. \text{nomme}(x, \text{Cauterets}) \wedge \text{nomme}(y, \text{Azais}) \wedge \text{quitter}(e, y, x) \wedge \grave{\text{a}}(e, \text{cinq_heures})$.

10. Dans une version simpliste de la DRT et ne mettant pas à profit l'ensemble des travaux menés en DRT pour traiter le temps justement.

de *Azaïs a quitté Cauterets à cinq heures*¹¹.

Tout d'abord, nous redonnons l'interprétation en intervalle distinctive entre état, activité et évènement tel que définie dans la rubrique précédente. On rappelle le critère d'homogénéité des éventualités qui permet de discriminer les activités des évènements. Ici pour l'éventualité réifiée e on donne χe l'intervalle de temps, l'extension temporelle de e .

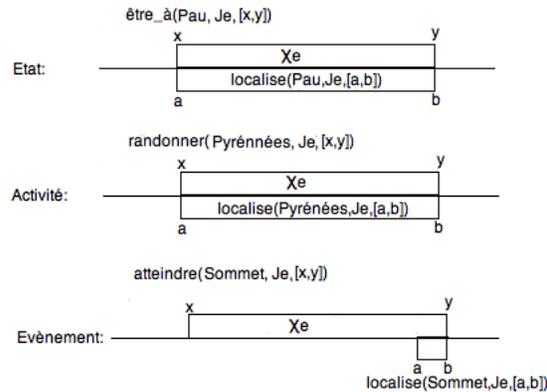


FIGURE 4.5 – L'homogénéité des états et activités et l'hétérogénéité des évènements réifiés

A ce stade de notre traitement en DRT, même si on donne aux éventualités une représentation propre, nous ne donnons pas de suite un référent du discours dans le domaine pour l'éventualité. Elle sera représentée au sein du modèle une fois que les opérateurs de la temporalité lui seront appliqués. Pour le moment, l'éventualité est liée par un λ , ce qui nous permet d'obtenir un terme exprimant un ensemble d'éventualités :

Notre DRS simplifiée¹² est donc maintenant de type $v \rightarrow t$, v étant le typé dédié aux éventualités comme nous le présenterons dans la section suivante. L'ensemble des éventualités possibles selon lesquelles la relation *quitter* est établie entre *je* et *Cauterets* est donc représenté par cette DRS. La phase de temporalisation de l'éventualité va permettre de sélectionner parmi cet ensemble d'éventualités, celle qui a les bonnes propriétés. Nous allons montrer maintenant comment donner un référent du discours à cette éventualité de type v en la temporalisant, la section suivante donne le détail du typage introduit ici.

11. De manière plus réaliste, les variables qui peuplent le domaine et représentent des entités nommées sont liées à leur dénomination par le prédicat nommé qui permet un typage correct sans avoir à réduire Cauterets à sa simple dimension géographique. On laissera de côté les autres aspects sémantiques de Cauterets par ce procédé. On pense par exemple à sa population dans l'exemple : *Cauterets est en baisse démographique depuis quelques années*.

12. Ici on ne donne pas le détail des DRS nécessaires à la résolution de nommé(x ,Cauterets) et nommé(y ,Azaïs).

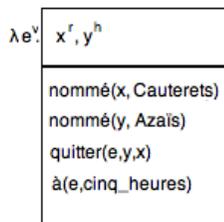


FIGURE 4.6 – La λ -DRS simplifiée pour $\lambda e. quitter(e, x, y) \wedge nommé(x, Cauterets) \wedge nommé(y, Azais)$

4.3.2 Le typage du langage

Le système de Verkuyl est représentatif des différents temps verbaux, chaque combinaison correspond à une réalisation linguistique et chaque réalisation linguistique correspond à une combinaison d'opérateurs, il y a donc une relation bi-univoque entre temps linguistique et combinaison. Ce système est binaire dans le sens où, tout d'abord dans [Verkuyl, 2003], puis dans [Verkuyl, 2008], il propose une approche combinant trois opérateurs binaires et complémentaires. Chaque opérateur est minimal et distinctif sémantiquement, et binaire en raison de l'arité de la relation qu'il porte :

- PAST / PRESENT
- SYNCHRONOUS / POSTERIOR
- PERFECT / IMPERFECT

Ce système, basé sur trois relations entre deux indices (2X2X2), contrairement à celui de Reichenbach qui combine trois relations possibles entre deux indices (3X3), permet de traiter les huit temps verbaux néerlandais à l'indicatif ainsi que les huit temps verbaux anglais correspondants. Ces indices n'ont pas dans la théorie de Verkuyl de modèle d'interprétation défini, en revanche, il est conseillé de les considérer comme des "nombres naturels". Les travaux de Verkuyl ont mis en évidence l'applicabilité de ce système au français et à bien d'autres langues. Comme présenté dans la rubrique précédente, à chaque opérateur on associe un λ -terme dans le lexique sémantique et il nous est alors possible de procéder à l'analyse temporelle des événements de manière compositionnelle et conforme à notre analyse en λ -DRT [Moot, 2010a, Amsili et Bras, 1998].

Pour les besoins de ces travaux et dans le cadre du projet ITIPY, nous avons développé une interprétation du système *Binary Tense* dans le but d'extraire automatiquement des itinéraires, ce qui nous a amené à illustrer les développements de nos travaux par des exemples de déplacements ou de localisations. Néanmoins, tout comme *Binary Tense* lors de sa conception, nos développements sont valables pour le traitement sémantique de la temporalité des éventualités, qu'elles impliquent une extension spatiale

ou non.

Avant d'expliquer plus en détail les relations utilisées et le modèle dans lequel on les interprète, il nous faut présenter le typage utile au traitement des représentations sémantiques. Nous rappelons que la variante multisorte de la logique du premier ordre que nous utilisons autorise donc les individus à être partagés en plusieurs sortes, tandis que les valeurs de vérité restent inchangées. Définissons les types utilisés pour traiter de la temporalité des déplacements :

- h sont les entités mobiles
- e sont les entités immobiles ¹³
- r sont les régions
- v sont les éventualités
- i sont les intervalles temporels
- t sont les valeurs de vérité

Nous proposons donc une distinction entre les entités : le voyageur ¹⁴, de type entité mobile h , l'espace qui est représenté en régions r , les éventualités de type v et les intervalles i qui sont des entités temporelles. Nous gardons les entités immobiles dont on ne veut rien dire si ce n'est qu'elles ne sont ni des intervalles de temps, ni des régions géographiques et qu'elles ne sont pas des voyageurs.

Nous réifions l'éventualité (e de type v) sans lui donner pour le moment de référent propre dans le domaine (λe), l'ensemble d'éventualités disponibles pour la relation prédicative propre à *quitter* sans la temporalité s'écrit donc comme ceci :

$$\lambda e^v \exists x^r \exists y^h . \text{quitter}^{r \rightarrow h \rightarrow v \rightarrow t}(e^v, x^r, y^h) \wedge \text{nomme}(x, \text{Cauterets}) \wedge \text{nomme}(y, \text{Azaïs})^{15}$$

Il nous faut maintenant montrer comment construire la temporalité qui va nous permettre d'associer un référent à l'éventualité.

- Soit ϕ l'ensemble des éventualités, de type $v \rightarrow t$, selon lesquelles la relation de *quitter* est établie entre *Azaïs* et *Cauterets*.

On peut voir ϕ comme une fonction qui prend un élément de type v en entrée et donne un élément de type t en sortie. Cette fonction est une propriété des v , et désigne dans notre cas l'ensemble des e de type v tel que *quitter*(e^v, x, y) avec x la variable pour l'entité nommée *Cauterets* et y la variable pour l'entité nommée *Azaïs*.

ϕ est donc de type $v \rightarrow t$ et s'écrit :

$$\phi^{v \rightarrow t} = (\lambda e^v \exists x^r \exists y^h . \text{quitter}^{r \rightarrow h \rightarrow v \rightarrow t}(e^v, x^r, y^h) \wedge \text{nomme}(x, \text{Cauterets}) \wedge \text{nomme}(y, \text{Azaïs}))^{v \rightarrow t}$$

13. Ici, nous aurions pu considérer les entités immobiles comme des régions, néanmoins ce type demande à être raffiné. Par exemple, "la mairie" est une entité immobile qui peut avoir plusieurs aspects, le premier, spatial peut être une région, un autre cependant pourrait être de type h en considérant le personnel de la mairie, nous pensons par exemple à : "la mairie, ils sont désagréables". Nous ne nous étendons pas sur ces questions mais pour une discussion de ces phénomènes voir : [Pustejovsky, 1998] et [Mery, 2011].

14. Nous distinguons les entités mobiles des entités immobiles afin de représenter les déplacements possibles d'un voyageur. Les pluriels impliquent des voyageurs. Il nous faudrait davantage de raffinement dans le traitement pragmatico-sémantique afin de traiter les pluriels. Nous admettons dans notre cas qu'un groupe de voyageurs porté par un "nous" contient le voyageur qui nous intéresse, de type entité mobile.

15. On ne donne pas le détail des DRS utiles à la construction de $\text{nomme}(x, \text{Cauterets})$ ni de $\text{nomme}(y, \text{Azaïs})$.

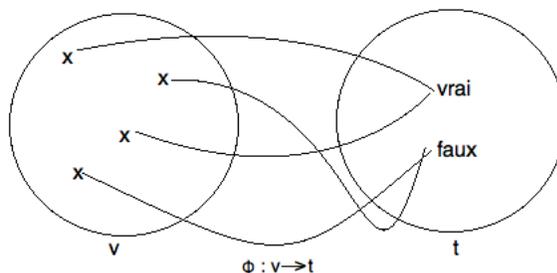


FIGURE 4.7 – Illustration ensembliste de ϕ

- Soit τ la formule exprimant la temporalité linguistique d’une éventualité.
 τ est lui aussi une fonction, il prend en entrée ϕ , soit l’ensemble des e tel que $quitter(e^v, x, y)$ ¹⁶ présentée précédemment et donne en sortie une valeur de vérité. Cette fonction sélectionne l’éventualité dans l’ensemble ϕ qui a les propriétés propres au temps verbal tel que définit dans le système, τ est donc de type $(v \rightarrow t) \rightarrow t$.
Par exemple pour le futur simple *Azais quittera Cauterets*, on aura donc $FuturSimple(\phi)$. Plus précisément, à partir de l’ensemble des e de type v et en réduisant le terme on obtiendra une formule de type t :
 $(FuturSimple)^{(v \rightarrow t) \rightarrow t} (\lambda e^v. quitter(e^v, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azais))^{v \rightarrow t}$.
- Le terme $(\tau(\phi))^t$
est donc la temporalité à laquelle on a appliqué l’ensemble des éventualités qui présentaient les bonnes propriétés sémantiques. On obtient donc une condition d’interprétation du modèle bien formée tout en donnant des informations sémantiques sur l’éventualité réifiée à partir de l’énoncé linguistique.

4.3.3 Les variables, constantes, fonctions et relations

Entrons dans le détail des opérateurs de la temporalité utiles à la construction de τ et revenons à la définition de notre langage. Pour ceci, nous donnons la correspondance de chaque fonction ou relation dans la théorie des intervalles de Allen [Allen, 1983] :

- **Variables** : des variables sont disponibles pour chaque type
- **Constante** : n est la constante du moment d’énonciation : son type est i .
- **Les fonctions et les relations** :

16. Avec toujours x la variable pour l’entité nommée *Cauterets* et y la variable pour l’entité nommée *Azais*.

4.3. LE LANGAGE LOGIQUE TYPÉ

symboles	arité	interprétation (domaine : intervalles de \mathbb{R} muni des relations de Allen [Allen, 1983])
χ	$1 : i \rightarrow i$	L'unité de temps projetée pour une éventualité x est un intervalle temporel x
a	$1 : i \rightarrow i$	pour l'intervalle x on considère l'unique intervalle ax tel que $MEETS(ax, n)$ et $STARTS(ax, x)$ ¹⁷
\diamond	$1 : i \rightarrow i$	pour l'intervalle x on considère l'unique intervalle $\diamond x$ tel que $MEETS(n, \diamond x)$ et $FINISHES(\diamond x, x)$
\circ	$2 : i \rightarrow i \rightarrow t$	$DURINGINVERSE(x, y)$
$<$	$2 : i \rightarrow i \rightarrow t$	$BEFORE(x, y)$
$=$ ¹⁸	$2 : i \rightarrow i \rightarrow t$	$EQUALS(x, y)$
\succ	$2 : i \rightarrow i \rightarrow t$	$STARTS(x, y)$ ou $DURING(x, y)$
\succcurlyeq	$2 : i \rightarrow i \rightarrow t$	$STARTS(x, y)$ ou $DURING(x, y)$ ou $FINISHES(x, y)$
$meet$	$2 : i \rightarrow i \rightarrow t$	$EQUALS(x, y)$ ou $STARTSINVERSE(x, y)$ ou $OVERLAPSINVERSE(x, y)$
		$MEETS(x, y)$

Sur les recommandations de Christian Retoré, nous désirons interpréter notre langage du temps dans un modèle standard ([Allen, 1983, Kamp et Reyle, 1993b, Gosselin, 1996, Battistelli, 2009]), une structure ordonnée, une structure homomorphe aux intervalles de \mathbb{R} .

Concernant la nature de ces intervalles, de manière analogue à Kamp [Kamp, 1981], nous assumons qu'aucun phénomène linguistique ne légitime la discrimination entre intervalle ouvert, fermé, ou semi-ouvert ou encore semi-fermé.

Pour l'énoncé *Azaïs quitte Caunterets*, on voudrait obtenir la succession de deux intervalles, le premier pour lequel $\hat{e}tre_à(Caunterets, Azaïs)$ est vraie, le second pour lequel elle est fausse. Les deux intervalles sont en relation de contact. Cette éventualité est un évènement comme défini plus tôt. L'intervalle, extension temporelle de l'éventualité pour laquelle *Azaïs quitte Caunterets*, se situe en relation de chevauchement avec le premier et le second intervalle et couvre le contact (voir figure 4.8).

Cette relation de contact entre les deux intervalles nécessite que le premier atteigne sa borne finale et les second n'atteigne pas sa borne initiale ou l'inverse, c'est à dire que les deux intervalles doivent être soit semi-ouverts soit semi-fermés, mais tous deux doivent être du même type, concernant le choix entre semi-fermé et semi-ouvert, aucune discrimination n'est justifiée linguistiquement, ce choix sera donc arbitraire mais nécessaire.

Si l'on prend l'éventualité p telle que $\hat{e}tre_à(p, Caunterets, Azaïs)$ est vraie et q telle que $\hat{e}tre_à(q, Compl(Caunterets), Azaïs)$ est vraie, pour les intervalles respectifs χp et χq , on a :

$$meet([\chi p], [\chi q])$$

ou

$$meet([\chi p[, [\chi q[$$

Soit χr l'extension temporelle de l'éventualité propre à *Azaïs quitte Caunterets*

17. Si l'intervalle considéré n'inclut pas n alors on considèrera un genre de présent dans le passé, n' qui fera office de n dans la relation.

18. Dans le système *Binary Tense*, c'est la quasi-égalité \simeq qui est utilisée. Pour des besoins de simplification, nous utilisons ici l'égalité.

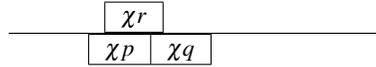


FIGURE 4.8 – Interprétation pour Azais quitte Causerets

Pour ceci, nous présentons l'opération de clôture Cl pour n'importe quel type d'intervalle X et Y .

Soient X et Y deux intervalles :

$X \equiv Y$ si $Cl(X) = Cl(Y)$

Dans notre cas, soit X un intervalle défini par sa borne ouvrante x_1 et sa borne fermante x_2 on en déduit donc que : $[x_1, x_2) \equiv [x_1, x_2] \equiv (x_1, x_2) \equiv (x_1, x_2)$ avec les parenthèses désignant les bornes ouvertes, et les crochets, les bornes fermées. Deux intervalles sont donc équivalents s'ils ont le même intérieur au sens de la topologie usuelle. Cette opération de clôture nous permet de créer la classe d'équivalence entre intervalles ouverts, fermés, semi ouverts et semi fermés. Pour les besoins du calcul nous devons choisir un type d'intervalle, nous prendrons les intervalles fermés en leur borne initiale et ouverts en leur borne finale.

4.3.4 Les termes

Nous présentons ici quelques termes de types i et de type t :

– $(ax)^i$ et $(\diamond x)^i$

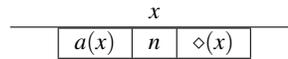


FIGURE 4.9 – Interprétation des fonctions a et \diamond

L'intervalle x est donc décomposé en trois parties, ax qui est le sous intervalle inclus dans x et qui est en relation de précédence et de contact avec n ou n' ¹⁹, et $\diamond x$ qui est le sous intervalle en relation de succession et de contact lui aussi soit avec n , soit avec n' .

– $(x \circ n)^t$

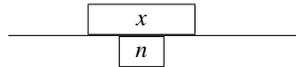


FIGURE 4.10 – Interprétation possible de la relation $x \circ n$

Ici, n est inclus dans x . Cette relation a pour particularité d'être utilisée pour mettre en relation un intervalle avec la constante n .

19. Verkuyl précise que lorsqu'un intervalle est placé antérieurement par rapport à n , le moment d'énonciation, alors systématiquement n' , un genre de "présent dans le passé" est disposé en i .

– $(x < y)^t$

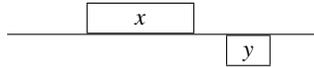


FIGURE 4.11 – Interprétation possible de la relation $x < y$

La relation d'antériorité est interprétée ici de manière standard, tout comme dans le système de Allen[Allen, 1983].

– $(x = y)^t$

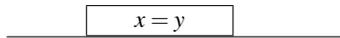


FIGURE 4.12 – Interprétation de la relation $x = y$

La relation d'égalité est elle aussi interprétée de manière standard.

– $(x \prec y)^t$

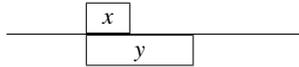


FIGURE 4.13 – Interprétations de la relation $x \prec y$

\prec doit être interprétée comme la relation que l'on donnerait entre les entiers naturels 2 et 3, si on les considère comme des ensembles, soit pour 2 on a l'ensemble $\{0,1,2\}$ et pour 3 on a l'ensemble $\{0,1,2,3\}$. 2 est à la fois précédent à 3 et inclus dans 3.

– $(x \preceq y)^t$

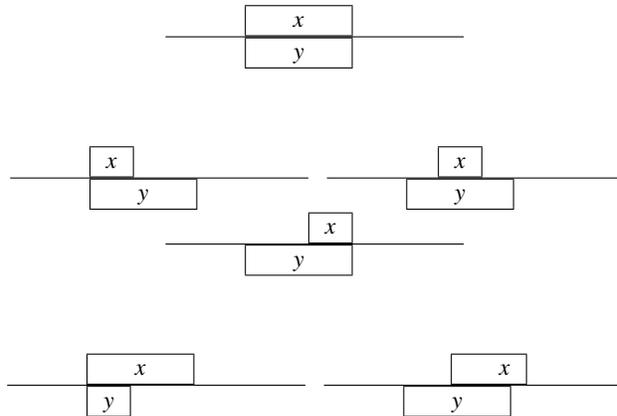


FIGURE 4.14 – Interprétations de la relation $x \preceq y$

La relation \preceq est un dual à la relation \prec , elle nous sert à déterminer que nous n'avons aucune information sur le fait que x se termine avant y .

– $(x \text{ meet } y)^t$

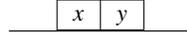


FIGURE 4.15 – Interprétation de la relation $x \text{ meet } y$

La relation *meet* quant à elle se définit de manière standard, tout comme la précédence et l'égalité.

A partir de ce langage, on construit les λ -termes des opérateurs utiles à la description de la temporalité nommés τ précédemment. Pour obtenir τ , il nous faut une combinaison des trois opérations présentées précédemment en section 4.3.2.

On applique la formule ϕ à τ , et on procède du bas vers le haut :

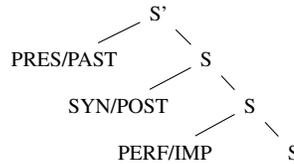


FIGURE 4.16 – Composition des opérateurs temporels, dans le système *Binary Tense*

Ce schéma permet de représenter l'ordre dans lequel on peut créer les compositions propres à chaque temps conjugué. Le système est donc compositionnel et biunivoque, c'est à dire que pour chaque composition, soit $2X2X2$ compositions, on obtient une temps conjugué, soit $2X2X2$ temps, nous présentons à la fin de cette section toutes les compositions principales du système.

Les opérateurs sont définis et s'appliquent comme suit :

– $\text{PERF} =_{def} \lambda \phi \lambda j \exists k [\phi[k] \wedge (\chi k) < (\chi j)] : (v \rightarrow t) \rightarrow v \rightarrow t$

Ici, on pose la relation entre l'intervalle pendant lequel le voyageur se déplace (χk) et l'intervalle (χj), vu comme une fenêtre depuis laquelle on regarde le déplacement.

On donne ici un référent du discours à l'éventualité choisie au sein de l'ensemble ϕ , cette éventualité est close par l'existentiel, $\exists k$. Par ailleurs, on met en relation, non pas l'éventualité elle-même, mais son extension temporelle χk .

Plus précisément, la propriété avancée est la relation entre un nouvel intervalle (χj) et l'extension temporelle de l'éventualité sélectionnée dans ϕ , comme illustré dans la figure 4.17 ci-après.

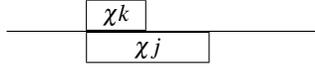


FIGURE 4.17 – Interprétation possible pour PERF

- $\text{IMP} =_{\text{def}} \lambda \phi \lambda j \exists k [\phi[k] \wedge (\chi^k) \preceq (\chi^j)] : (v \rightarrow t) \rightarrow v \rightarrow t$
 Tout comme pour PERF, on pose la relation entre l'intervalle pendant lequel le voyageur se déplace (χ^k) et l'intervalle depuis lequel on regarde le déplacement (χ^j).
 On donne ici un référent du discours à l'éventualité choisie au sein de l'ensemble ϕ , cette éventualité est close par l'existentiel, $\exists k$. Par ailleurs, on met en relation, non pas l'éventualité elle-même, mais son extension temporelle χ^k .
 Plus précisément, la propriété avancée est la relation entre χ^k et χ^j qui n'est pas déterminée quant à la borne finale de χ^k , il n'est pas spécifié si elle est située avant, en même temps ou après la borne finale de χ^j comme illustré dans la figure 4.18 ci-après.

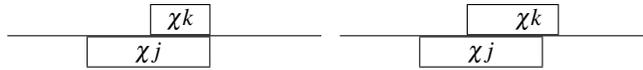


FIGURE 4.18 – Interprétations possibles pour IMP

- $\text{SYN} =_{\text{def}} \lambda \phi \lambda i \exists j [\phi[j] \wedge (\chi^j) = (\chi^i)] : (v \rightarrow t) \rightarrow v \rightarrow t$
 Ici, une fois l'intervalle propre au déplacement situé dans la fenêtre, il faut situer la fenêtre elle-même.
 La relation SYN sélectionne une éventualité dans l'ensemble d'éventualités introduit par PERF ou IMP. j est close par l'existentiel, on accorde donc un référent de discours à cette éventualité.
 Autrement dit, on pose la relation d'égalité entre l'intervalle-fenêtre positionné par PERF ou IMP (χ^j) et l'intervalle (χ^i) permettant de se positionner par rapport à la constante n comme présenté dans la figure 4.19 ci-après.

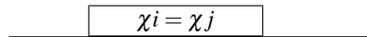


FIGURE 4.19 – Interprétation possible pour SYN

- $\text{POST} =_{def} \lambda \phi \lambda i \exists j [\phi[j] \wedge (\chi j) \preceq \diamond(\chi i)] : (v \rightarrow t) \rightarrow v \rightarrow t$

Ici, tout comme pour SYN, le positionnement de la fenêtre se fait comme suit. La relation POST sélectionne une éventualité dans l'ensemble d'éventualités introduit par PERF ou IMP. j est close par l'existentiel, on accorde donc un référent de discours à cette éventualité.

L'intervalle-fenêtre positionné par PERF ou IMP, χj , a un sous intervalle en commun avec le sous-intervalle $\diamond(\chi i)$, χi permettant de se positionner par rapport à la constante n comme présenté dans la figure 4.20 ci-après.

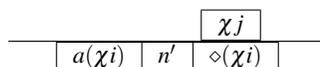


FIGURE 4.20 – Interprétation possible pour POST

- $\text{PRES} =_{def} \lambda \phi \exists i [\phi[i] \wedge (\chi i) \circ n] : (v \rightarrow t) \rightarrow t$

Tout d'abord, cet opérateur permet de clore la combinaison des opérateurs afin d'obtenir τ . On donne un référent de discours à la dernière éventualité i . Le dernier intervalle positionné par POST ou SYN (χi) est mis en relation avec la constante n , typiquement, nous sommes dans un temps du présent. Ici n est inclus dans l'intervalle χi comme présenté dans la figure 4.21 ci-après.

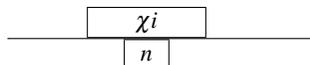


FIGURE 4.21 – Interprétation possible pour PRES

- $\text{PAST} =_{def} \lambda \phi \exists i [\phi[i] \wedge (\chi i) < n] (v \rightarrow t) \rightarrow t$

Tout comme PRES, cet opérateur permet de clore la combinaison des opérateurs afin d'obtenir τ . On donne là aussi un référent de discours à la dernière éventualité i . Le dernier intervalle positionné par POST ou SYN, (χi), est mis en relation avec la constante n , ici, nous sommes dans un temps du passé. Pour chaque intervalle χi mis en relation d'antériorité par rapport à n , on obtient un n' inclus dans i qui permet de positionner $a(\chi i)$ et $\diamond(\chi i)$, comme vu précédemment. L'intervalle χi précède n comme présenté dans la figure 4.22 ci-après.

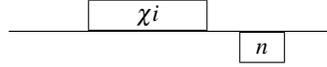


FIGURE 4.22 – Interprétation possible pour PAST

Un dernier opérateur que nous avons présenté dans l'introduction de ce chapitre et proposé par Verkuyl est l'opérateur *Anc*, utile à la représentation du passé simple en français. Le passé simple ne permet pas de faire référence au moment d'énonciation, soit n , c'est pourquoi l'opérateur *Anc* est proposé.

- $ANC =_{def} \lambda \phi \exists k [\phi[k] \wedge (\chi k) < \alpha_c] (v \rightarrow t) \rightarrow t$

En ce qui concerne cet opérateur, il permet de situer dans le passé de α_c l'intervalle propre à l'éventualité sans faire appel à aucun autre opérateur, il n'est donc pas fruit d'une combinaison mais peut être combiné par exemple à *PERF* pour former le passé antérieur. Ici α_c est un intervalle obtenu grâce à une trace dans le contexte et n'ayant pas nécessairement besoin d'être en lien avec n . Le passé simple ne fait donc pas partie du tableau des combinaisons en $2x2x2$ proposé mais est en dehors.

Une formule de ce langage est donc formée de ces variables, en relation ou non avec la constante, ce qui permet de construire les termes pour la temporalité attachée au verbe. On construira la combinaison de plusieurs opérateurs appelée précédemment τ comme PAST(POST)(IMP), PRES(POST)(PERF), etc.

On propose, pour alléger les notations, de réifier directement les extensions temporelles des éventualités au lieu des éventualités elles-mêmes à partir des prédicats de déplacement et de typer les opérateurs en conséquence :

- ϕ jusqu'à maintenant de type $v \rightarrow t$
est maintenant de type $i \rightarrow t$

$$(4.17) \quad \phi = (\lambda k^v . quitter(k', x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azais))^{v \rightarrow t}$$

sera alors

$$(4.18) \quad \phi = (\lambda k^i . quitter(k', x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azais))^{i \rightarrow t}$$

- τ jusqu'à maintenant est de type $(v \rightarrow t) \rightarrow t$
nous le remplaçons par le type $(i \rightarrow t) \rightarrow t$
 $\tau = \text{PAST(POST)(IMP)}$

$$(4.19) \quad \tau = \lambda \phi \exists i \exists j \exists k . [\phi[k] \wedge ((\chi k) \preceq (\chi j)) \wedge ((\chi j) \preceq (\chi i)) \wedge ((\chi i) < n)]^{(v \rightarrow t) \rightarrow t}$$

est alors maintenant

$$(4.20) \quad \tau = \lambda \phi \exists i \exists j \exists k . [\phi[k] \wedge (k \preceq j) \wedge (j \preceq i) \wedge (i < n)]^{(i \rightarrow t) \rightarrow t}$$

4.4 La correspondance entre combinaison d'opérateurs et temps verbal

4.4.1 Tableau des combinaisons

Ci-après, le tableaux de combinaisons des opérateurs pour six temps de l'indicatif anglais et deux conditionnels ainsi que leur correspondance en français :

		PRES	PAST
		$\lambda\phi\exists i[\phi[i] \wedge (i) \circ n]$	$\lambda\phi\exists i[\phi[i] \wedge (i) < n]$
		she writes	she wrote
		elle écrit	elle écrivait
IMP	SYN		
	$\lambda\phi\lambda i\exists j[\phi[j] \wedge (j) = (i)]$		
		she will write	she would write
	POST	elle écrira	elle écrirait
	$\lambda\phi\lambda i\exists j[\phi[j] \wedge (j) \preceq \diamond(i)]$		
		she has written	she had written
PERF	SYN	elle a écrit	elle avait écrit
	$\lambda\phi\lambda i\exists j[\phi[j] \wedge (j) = (i)]$		
		she will have written	she would have written
	POST	elle aura écrit	elle aurait écrit
	$\lambda\phi\lambda i\exists j[\phi[j] \wedge (j) \preceq \diamond(i)]$		

Le passé simple et le passé antérieur, appelés aussi aoristes²⁰ :

elle écrivit

ANC $\lambda\phi\exists k[\phi[k] \wedge k < \alpha_c]$

elle eut écrit

(ANC)(PERF) $\lambda\phi\exists k'\exists k[\phi[k] \wedge k \prec k' \wedge k' < \alpha_c]$

Tous les temps principaux du récit sont donc représentés dans les combinaisons possibles, néanmoins toutes les combinaisons ne sont pas possibles, par exemple ANC ne peut se combiner à IMP.

Nous allons maintenant donner les formules du langage pour résoudre les divers temps verbaux dans le langage, nous donnons la réduction d'un exemple, sa représentation en DRS puis son interprétation dans les intervalles de \mathbb{R} .

²⁰. Sont désignés par aoriste les passé simple et temps surcomposés à partir de celui-ci. Nous décrivons plus en détail ce qu'est un aoriste dans le chapitre 6.

4.4.2 Le présent de l'indicatif

Passons à la représentation de chaque temps du français, tout d'abord le présent de l'indicatif que l'on réduira dans l'ordre des étapes :

(4.21) Azaïs quitte Cauterets : $(PRES)(SYN)(IMP)\phi$

1. $(PRES)(SYN)(IMP)(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))$
2. $(\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j \preceq k) \wedge (i = j) \wedge (i \circ n)]^{(i \rightarrow t) \rightarrow t} (\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))^{i \rightarrow t}$
3. $\exists i \exists j \exists k [(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)[k] \wedge (j \preceq k) \wedge (i = j) \wedge (i \circ n))$
4. $(\exists i \exists j \exists k \exists x \exists y. quitter(k, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs) \wedge (j \preceq k) \wedge (i = j) \wedge (i \circ n))^t$

Le terme réduit en λ -DRT sera donc de la forme suivante :

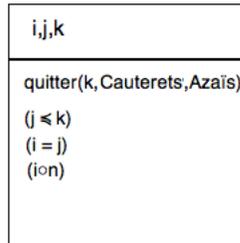


FIGURE 4.23 – DRS pour *Azaïs quitte Cauterets*

Et l'interprétation de ce terme dans les intervalles de \mathbb{R} :



FIGURE 4.24 – Interprétations possibles des variables temporelles pour le présent soit $(PRES)(SYN)(IMP)$.

4.4.3 Le futur simple

(4.22) Azaïs quittera Cauterets : $(PRES)(POST)(IMP)\phi$

1. $(PRES)(POST)(IMP)(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)$
2. $(\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j \preceq k) \wedge (j \preceq \diamond i) \wedge (i \circ n)])^{(i \rightarrow t) \rightarrow t} (\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))^{i \rightarrow t}$
3. $\exists i \exists j \exists k [(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)[k] \wedge (j \preceq k) \wedge (j \preceq \diamond i) \wedge (i \circ n))$
4. $(\exists i \exists j \exists k \exists x \exists y. quitter(k, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs) \wedge (j \preceq k) \wedge (j \preceq \diamond i) \wedge (i \circ n))^t$

Le terme réduit en λ -DRT sera donc de la forme suivante :

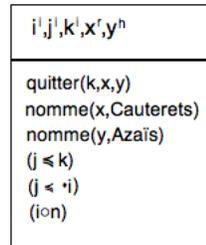


FIGURE 4.25 – DRS pour *Azaïs quittera Cauterets*

Son interprétation dans les intervalles de \mathbb{R} :



FIGURE 4.26 – Interprétations possibles des variables temporelles pour le futur soit $(PRES)(POST)(IMP)$.

4.4.4 L'imparfait

(4.23) Azaïs quittait Cauterets : (PAST)(SYN)(IMP) ϕ

1. (PAST)(SYN)(IMP)($\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)$)
2. ($\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (k \preceq j) \wedge (j = i) \wedge (i < n)]^{(i \rightarrow t) \rightarrow t} (\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))^{i \rightarrow t}$)
3. $\exists i \exists j \exists k [(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)[k] \wedge (k \preceq j) \wedge (j = i) \wedge (i < n))$
4. $(\exists i \exists j \exists k \exists x \exists y. quitter(k, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs) \wedge (k \preceq j) \wedge (j = i) \wedge (i < n))^t$

Le terme une fois réduit en λ -DRT :

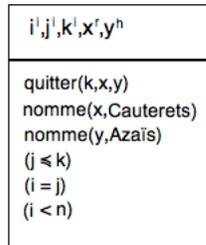


FIGURE 4.27 – DRS pour Azaïs *quittait Cauterets*

Ainsi que son interprétation dans les intervalles de \mathbb{R} :

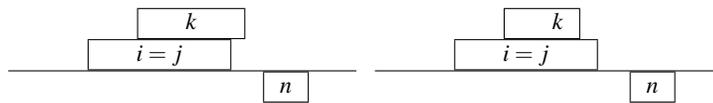


FIGURE 4.28 – Interprétations possibles des variables temporelles pour l'imparfait soit (PAST)(SYN)(IMP).

4.4.5 Le conditionnel présent

(4.24) Azaïs quitterait Cauterets : $(PAST)(POST)(IMP)\phi$

se réduit donc en :

1. $(PAST)(POST)(IMP)(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)$
2. $(\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j \preceq k) \wedge (j \preceq \diamond i) \wedge (i < n)])^{(i \rightarrow t) \rightarrow t} (\lambda c^t \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))^{i \rightarrow t}$
3. $\exists i \exists j \exists k [(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)[k] \wedge (j \preceq k) \wedge (j \preceq \diamond i) \wedge (i < n))$
4. $(\exists i \exists j \exists k \exists x \exists y. quitter(k, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs) \wedge (j \preceq k) \wedge (j \preceq \diamond i) \wedge (i < n))^t$

Voici l'équivalent en DRT :

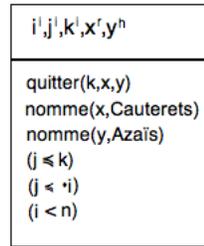


FIGURE 4.29 – DRS pour *Azaïs quitterait Cauterets*

Et son interprétation dans les intervalles de \mathbb{R} :

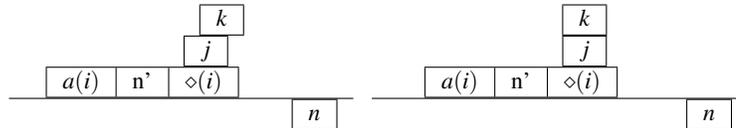


FIGURE 4.30 – Interprétations possibles des variables temporelles pour le conditionnel présent soit $(PAST)(POST)(IMP)$.

4.4.6 Le passé composé

(4.25) Azaïs a quitté Cauterets : $(PRES)(SYN)(PERF)\phi$

1. $(PRES)(SYN)(PERF)(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))$
2. $(\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j < k) \wedge (j = i) \wedge (i \circ n)]^{(i \rightarrow t) \rightarrow t} (\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))^{i \rightarrow t}$
3. $\exists i \exists j \exists k [(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)[k] \wedge (j < k) \wedge (j = i) \wedge (i \circ n))$
4. $(\exists i \exists j \exists k \exists x \exists y. quitter(k, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs) \wedge (j < k) \wedge (j = i) \wedge (i \circ n))^t$

La DRT correspondante est la suivante :

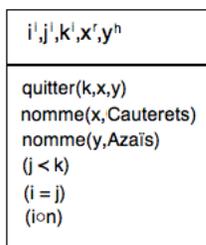


FIGURE 4.31 – DRS pour *Azaïs a quitté Cauterets*

Et son interprétation dans les intervalles de \mathbb{R} :

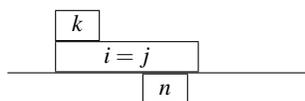


FIGURE 4.32 – Interprétation possible des variables temporelles pour le passé composé soit $(PRES)(SYN)(PERF)$.

4.4.7 Le futur antérieur

(4.26) Azaïs aura quitté Cauterets : $(PRES)(POST)(PERF)\phi$

1. $(PRES)(POST)(PERF)(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))$
2. $(\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j \prec k) \wedge (i \preceq \diamond j) \wedge (i \circ n)])^{(i \rightarrow t) \rightarrow t} (\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))^{i \rightarrow t}$
3. $\exists i \exists j \exists k [(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)[k] \wedge (j \prec k) \wedge (i \preceq \diamond j) \wedge (i \circ n))]$
4. $(\exists i \exists j \exists k \exists x \exists y. quitter(k, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs) \wedge (j \prec k) \wedge (j \preceq \diamond i) \wedge (i \circ n))^t$

Voici la réduction en λ -DRT :

i^l, j^l, k^l, x^l, y^h
$quitter(k, x, y)$ $nomme(x, Cauterets)$ $nomme(y, Azaïs)$ $(j \prec k)$ $(j \preceq \diamond i)$ $(i \circ n)$

FIGURE 4.33 – DRS pour *Azaïs aura quitté Cauterets*

Et son interprétation dans les intervalles de \mathbb{R} :

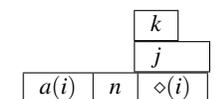


FIGURE 4.34 – Interprétation possible des variables temporelles pour le futur antérieur soit $(PRES)(POST)(PERF)$.

4.4.8 Le plus-que-parfait

(4.27) Azaïs avait quitté Cauterets : (PAST)(SYN)(PERF) ϕ

1. (PAST)(SYN)(PERF)($\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)$)
2. ($\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j < k) \wedge (i = j) \wedge (i < n)]^{(i \rightarrow t) \rightarrow t} (\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))^{i \rightarrow t}$)
3. $\exists i \exists j \exists k [(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)[k] \wedge (j < k) \wedge (i = j) \wedge (i < n))$
4. $(\exists i \exists j \exists k \exists x \exists y. quitter(k, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs) \wedge (j < k) \wedge (j = i) \wedge (i < n))^t$

Le terme équivalent en DRT :

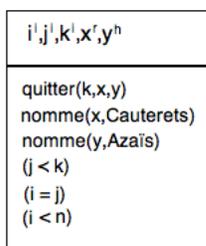


FIGURE 4.35 – DRS pour *Azaïs avait quitté Cauterets*

Et son interprétation dans les intervalles de \mathbb{R} :

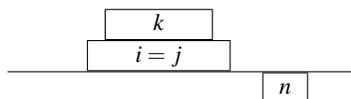


FIGURE 4.36 – Interprétation possible des variables temporelles pour le plus-que-parfait, soit (PAST)(SYN)(PERF).

4.4.9 Le conditionnel passé

(4.28) Azaïs aurait quitté Cauterets : (PAST)(POST)(PERF) ϕ

1. (PAST)(POST)(PERF)($\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)$)
2. ($\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j \prec k) \wedge (i \preceq \diamond j) \wedge (i < n)]$)^{(i → t) → t} ($\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)$)^{i → t}
3. $\exists i \exists j \exists k [(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)[k] \wedge (j \prec k) \wedge (i \preceq \diamond j) \wedge (i < n))$
4. $(\exists i \exists j \exists k \exists x \exists y. quitter(k, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs) \wedge (j \prec k) \wedge (j \preceq \diamond i) \wedge (i < n))^t$

La réduction équivalente en DRT :

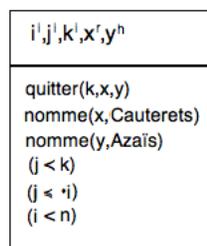


FIGURE 4.37 – DRS pour *Azaïs aurait quitté Cauterets*

Et son interprétation dans les intervalles de \mathbb{R} :

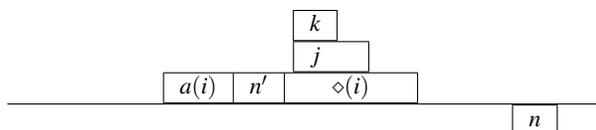


FIGURE 4.38 – Interprétation possible des variables temporelles pour le conditionnel passé première forme, soit (PAST)(POST)(PERF).

4.4.10 Le passé simple

(4.29) Azaïs quitta Cauterets : (ANC) ϕ

1. (ANC) $(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))$
2. $(\lambda \phi^{i \rightarrow t} \exists k [\phi[k] \wedge k < \alpha_c]^{(i \rightarrow t) \rightarrow t} (\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))^{i \rightarrow t})$
3. $(\exists k [(\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs))^{i \rightarrow t} [k] \wedge k < \alpha_c]^{(i \rightarrow t) \rightarrow t})$
4. $(\exists k [\exists x \exists y. quitter(k, x, y) \wedge nomme(x, Cauterets) \wedge nomme(y, Azaïs)]^{i \rightarrow t} \wedge k < \alpha_c)^{(i \rightarrow t) \rightarrow t}$

La réduction équivalente en DRT :

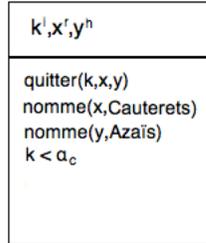


FIGURE 4.39 – DRS pour *Azaïs quitta Cauterets*

Et son interprétation dans les intervalles de \mathbb{R} :



FIGURE 4.40 – Interprétation possible des variables temporelles pour le passé simple, soit (ANC).

4.4.11 Le passé antérieur

(4.30) Azaïs eut quitté Causerets : $(ANC)\phi$

1. $(ANC)(PERF)(\lambda c \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Causerets) \wedge nomme(y, Azaïs)$
2. $(\lambda \phi^{i \rightarrow t} \exists k' \exists k [\phi[k] \wedge k \prec k' \wedge k' < \alpha_c]^{(i \rightarrow t) \rightarrow t} (\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Causerets) \wedge nomme(y, Azaïs))^{i \rightarrow t}$
3. $(\exists k [(\lambda c^i \exists x \exists y. quitter(c, x, y) \wedge nomme(x, Causerets) \wedge nomme(y, Azaïs))^{i \rightarrow t} [k] \wedge k \prec k' \wedge k' < \alpha_c]^{(i \rightarrow t) \rightarrow t}$
4. $(\exists k [\exists x \exists y. quitter(k, x, y) \wedge nomme(x, Causerets) \wedge nomme(y, Azaïs)]^{i \rightarrow t} \wedge k \prec k' \wedge k' < \alpha_c]^{(i \rightarrow t) \rightarrow t}$

La réduction équivalente en DRT :

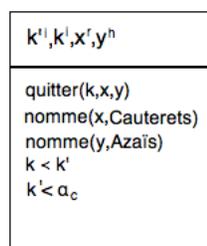


FIGURE 4.41 – DRS pour *Azaïs eut quitté Causerets*

Et son interprétation dans les intervalles de \mathbb{R} :

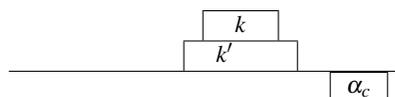


FIGURE 4.42 – Interprétation possible des variables temporelles pour le passé antérieur, soit $(ANC)(PERF)$.

4.4.12 Les adverbiaux

Une remarque importante quant à la réduction en λ -DRT des termes, une chose dont nous devons alors nous soucier est l'accessibilité des diverses variables afin de

4.4. LA CORRESPONDANCE ENTRE COMBINAISON D'OPÉRATEURS ET TEMPS VERBAL

calculer l'impact temporel des adverbiaux par exemple. Verkuyl préconise l'accès à k évidemment mais aussi à j parfois pour certains adverbiaux.

Dans le cas des perfects, on peut considérer par exemple dans la phrase *A six heures, j'ai mangé* sortie de son contexte comme le fait qu'à six heures j'ai déjà mangé (dans le cas d'une habitude par exemple) ou bien qu'à six heures je suis en train de manger (pour une situation particulière), dans la première interprétation, c'est j qui sera affecté par l'adverbial, c'est le regard sur l'éventualité accomplie qui est située temporellement, tandis que dans la seconde c'est k , on voudra faire chevaucher l'éventualité avec l'intervalle *six heures*. Nous préconisons une adaptation du système de Verkuyl afin de répartir l'influence des adverbiaux temporels sur l'indice j et k en fonction du type de l'adverbial, voire sur j et k dans certains cas, au lieu d'alterner entre j et k . Le problème que cette alternance entend résoudre est une sémantique ambiguë du perfect, très révélatrice d'un défaut du système pour traiter le français. Nous présentons dans le chapitre sur l'extension du système une solution à ce problème, néanmoins, dans le système original, on aura donc deux termes alternatifs selon l'interprétation voulue :

- (PRES)(SYN)(PERF)(ADV) ϕ
ou bien
- (PRES)(SYN)(ADV)(PERF) ϕ

Nous avons choisi ici de simplifier au maximum la représentation des DRS, en l'occurrence au sujet de la mention des entités nommées et des entités anaphorisées tel *lieu(y)* d'où *nous sommes partis* dans l'exemple suivant. Une remarque intéressante dans le cas où nous voudrions placer le calcul temporel dans une DRS imbriquée consiste en l'accessibilité des variables utiles. Nous reprenons un exemple de [Lefeuve *et al.*, 2012] dans lequel nous avons préféré la seconde interprétation entre les deux précédentes :

(4.31) Le trente et un, nous sommes partis à six heures du matin.

1. (PRES)(SYN)(PERF)(ADV)($\lambda c \exists x \exists y. \text{partir}(c, x, y) \wedge \text{lieu}(x) \wedge \text{nous}(y)$)
2. ($\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j \prec k) \wedge (j = i) \wedge (i \circ n)]$) ^{$i \rightarrow t \rightarrow t$} ($\lambda \phi' \lambda a [\phi'[a] \wedge a \prec \text{jour}(31) \wedge a \circ \text{heure}(06 : 00)]$) ^{$(i \rightarrow t) \rightarrow (i \rightarrow t)$} ($\lambda c \exists x \exists y. \text{partir}(c, x, y) \wedge \text{lieu}(x) \wedge \text{nous}(y)$) ^{$i \rightarrow t$}
3. ($\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j \prec k) \wedge (j = i) \wedge (i \circ n)]$) ^{$i \rightarrow t \rightarrow t$} ($\lambda a [(\lambda c \exists x \exists y. \text{partir}(c, x, y) \wedge \text{lieu}(x) \wedge \text{nous}(y))] [a] \wedge a \prec \text{jour}(31) \wedge a \circ \text{heure}(06 : 00)]$) ^{$(i \rightarrow t) \rightarrow (i \rightarrow t)$}
4. ($\lambda \phi^{i \rightarrow t} \exists i \exists j \exists k [\phi[k] \wedge (j \prec k) \wedge (j = i) \wedge (i \circ n)]$) ^{$i \rightarrow t \rightarrow t$} ($\lambda a (\exists x \exists y. \text{partir}(a, x, y) \wedge \text{lieu}(x) \wedge \text{nous}(y)) \wedge a \prec \text{jour}(31) \wedge a \circ \text{heure}(06 : 00)]$) ^{$i \rightarrow t$}
5. ($\exists i \exists j \exists k [(\lambda a (\exists x \exists y. \text{partir}(a, x, y) \wedge \text{lieu}(x) \wedge \text{nous}(y)) \wedge a \prec \text{jour}(31) \wedge a \circ \text{heure}(06 : 00))] [k] \wedge (j \prec k) \wedge (j = i) \wedge (i \circ n)$)
6. $\exists i \exists j \exists k (\exists x \exists y. \text{partir}(k, x, y) \wedge \text{lieu}(x) \wedge \text{nous}(y)) \wedge k \prec \text{jour}(31) \wedge k \circ \text{heure}(06 : 00) \wedge (j \prec k) \wedge (j = i) \wedge (i \circ n)$

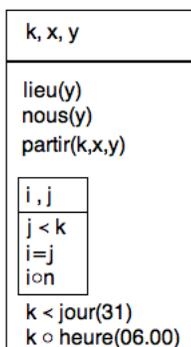


FIGURE 4.43 – DRS pour *le 31, nous sommes partis à six heures du matin*

4.5 Conclusion

Dans ce chapitre nous avons présenté le système de calcul temporel des éventualités selon Verkuyl [Verkuyl, 2008] pour lequel nous avons donné une interprétation dans les intervalles de \mathbb{R} . Néanmoins, nous avons montré que certains choix sont guidés par l'interprétation attendue, comme le choix de l'endroit où on applique (ADV) dans le terme. Nous avons aussi montré que le passé simple et le passé antérieur n'ont pas leur place au sein du tableau des combinaisons. Le passé simple est un temps central dans le récit du XIX^{ème} siècle, c'est pourquoi nous proposons une adaptation et une extension de ce système afin de proposer une version de ce système opérationnelle pour le français et permettant une représentation sémantique fidèle à cette langue.

Chapitre 5

Mise en perspective de la méthode : l'itinéraire, le récit de voyage au XIXème siècle et analyse sémantique du segment

Avant d'aller plus loin dans l'analyse sémantique du temps verbal au sein d'un énoncé, nous proposons de prendre le temps de montrer la pertinence de l'analyse du temps verbal en regard du projet au sein duquel ce travail de thèse a vu le jour. Parallèlement à notre état de l'art en sémantique formelle, nous avons mené à bien plusieurs explorations sur le corpus afin de déterminer quelle approche était la plus appropriée pour extraire les informations spatiales et temporelles ayant trait à l'itinéraire dans le récit de voyage, objectif originel du projet ITIPY. Nous présentons ici le fruit de nos recherches quant à la caractérisation du récit de voyage comme genre de discours structuré et à son analyse du point de vue de la SDRT. Ces développements justifient notre approche sémantique par segment, et a fortiori notre analyse du temps verbal, telle que l'utilisation de *Binary Tense* dans *Grail* le permet.

5.1 Introduction

Jusqu'à maintenant nous avons montré le cadre théorique dans lequel nous avons élaboré le travail présenté dans ce mémoire, nous avons présenté les méthodes que

nous utilisons en sémantique formelle pour accéder au sens des énoncés et nous avons montré en quoi nos travaux servaient l'automatisation de l'extraction d'informations sémantiques à partir d'énoncés. Néanmoins, nous ne nous sommes pas encore penchée sur l'objet précis du projet, l'itinéraire et le récit de voyage du XIXème siècle, et nous n'avons pas encore montré les spécificités linguistiques qui les régissent. Comme nous l'avons énoncé dans les chapitres précédents, notre étude se situe au niveau des énoncés, ou segments et plus exactement du verbe et de sa temporalité. Afin de situer nos travaux plus précisément par rapport aux objectifs du projet ITIPY, nous allons décrire dans ce chapitre les spécificités des productions discursives qui constituent ce genre du récit de voyage du XIXème siècle à partir d'un travail réalisé en collaboration avec Natalia Vinogradova¹.

Nous suivons plusieurs axes pour décrire au mieux notre corpus, échantillon de notre objet d'étude. Tout d'abord nous allons définir plusieurs outils qui nous permettent d'approcher cet objet, et nous allons confronter les enjeux du projet ITIPY au sein duquel nous sommes investie, aux difficultés imposées par notre corpus. Ensuite, nous montrerons comment le récit de l'itinéraire structure le récit de voyage. L'articulation entre le récit de voyage et l'itinéraire permettra par là-même d'appuyer le choix de notre sujet : le système temporel verbal du français du XIXème siècle à nos jours.

5.2 Le projet ITIPY et l'itinéraire

5.2.1 Le projet ITIPY

Nos travaux s'inscrivent dans le projet ITIPY financé par la Région Aquitaine et INRIA, dont le nom est construit sur la contraction d' « Itinéraires Pyrénéens ». Plus précisément, ce projet à son origine s'intéresse à l'extraction automatique d'itinéraires dans des récits de voyage et à son application à la recherche documentaire « intelligente » dans des fonds patrimoniaux liés au territoire aquitain. Suite à la numérisation d'un grand nombre d'oeuvres littéraires sur la région paloise par la Médiathèque de Pau, ce projet vise donc à terme l'extraction automatique d'itinéraires à partir de récits de voyage du XIXème et du début du XXème siècle.

La description d'un itinéraire est l'ensemble des segments qui se rapportent aux déplacements et localisations d'un voyageur au sein d'un discours. L'extraction quant à elle permet d'obtenir automatiquement toutes les informations spatiales et temporelles au sujet de ces déplacements. La tâche d'extraction des itinéraires peut donc être vue comme un traitement automatique d'un discours permettant d'extraire, puis d'ordonner, les différentes étapes qui le constituent.

Dans le cadre du traitement d'un corpus écrit, plusieurs méthodes peuvent être appliquées dans la discrimination des séquences et/ou segments permettant l'expression discursive de l'itinéraire :

- soit l'on cherche à distinguer les séquences textuelles propres au récit de l'itinéraire, puis l'on analyse les segments qui le constituent,
- soit on détecte directement les segments propres au récit de l'itinéraire sans prêter attention au contexte dans lequel ils apparaissent puis on les traite les uns

1. Et qui a donné lieu à une publication dans la revue Langages [Lefevre et Vinogradova, 2012].

après les autres en considérant a priori qu'ils forment un ensemble cohérent.

Nous donnons ici un aperçu des deux méthodes, en donnant une description de notre corpus comme échantillon du genre récit de voyage (désormais RV), et en montrant sa structuration en séquences. Nous éprouvons cette approche à des fins d'extraction d'information. Puis nous testons la seconde approche, en nous intéressant à la représentativité du segment narratif un déplacement au sein du corpus.

5.2.2 Les travaux antérieurs sur le corpus Itipy : l'itinéraire dans le texte

Afin d'introduire au mieux le projet, nous mentionnons brièvement une partie des travaux menés précédemment sur le corpus à l'interface entre le texte et l'itinéraire.

Parmi les travaux présentés dans la thèse de Pierre Loustau [Loustau, 2008], l'auteur propose une chaîne de traitement des segments, donc au niveau 3 de la figure 5.1. Le segment, comportant un patron lexico-syntaxique robuste mais ne donnant pas d'interprétation des données temporelles. La chaîne de traitement de Loustau, πR , s'inspire de celle de P. Enjalbert ([Enjalbert, 2005]) dans ses travaux sur la recherche et l'extraction de l'information spatiale.

Le patron lexico-syntaxique utilisé est

[verbe de déplacement, préposition ?, ES]

dans lequel la préposition peut être présente ou non, et pour lequel « ES » indique une entité spatiale. Le point d'interrogation permet d'indiquer la présence ou non d'une préposition, on notera que *je quitte Pau* ne contient pas de préposition. Par ailleurs, πR ne prend en compte que les noms propres en tant qu'ES pour le moment. Précisons, par ailleurs, la nécessité de disposer d'un lexique de toute entité pouvant être spatialisée. Par exemple, *quitter l'université*, extrait du corpus, peut aussi bien représenter l'institution dans l'explication d'un choix de carrière que le bâtiment dans la narration du quotidien d'un personnage, et donc évoquer un déplacement qui nous intéressent.

Grâce à des transducteurs, la chaîne détecte donc les segments contenant un verbe de déplacement appartenant à un lexique délimité, et une ES dans le syntagme prépositionnel ou nominal suivant ce verbe. Chaque segment est interprété comme un relais dans l'itinéraire si la polarité aspectuelle [Vandeloise, 1986] du verbe est compatible à la polarité de la préposition, ou à son absence. La polarité aspectuelle d'un verbe de déplacement est initiale pour un déplacement prenant un repère spatial comme point de départ, et finale pour un déplacement prenant un repère spatiale comme but. La polarité médiane permet de situer le repère spatial comme un repère sur le trajet mais ni comme un point de départ ni d'arrivée. Néanmoins certaines prépositions en relation avec les verbes de déplacement permettent d'exprimer une destination sans que le verbe ne spécifie si le lieu a ou non été visité [Flageul, 1997]. Regardons quelques exemples simples :

CHAPITRE 5. MISE EN PERSPECTIVE DE LA MÉTHODE : L'ITINÉRAIRE, LE
RÉCIT DE VOYAGE AU XIXÈME SIÈCLE ET ANALYSE SÉMANTIQUE DU
SEGMENT

- (5.1) Je pars de Pau.
 J'arrive à Pau.
 Je traverse Pau.
 Je quitte Pau.
 Je pars pour Pau.

SEGMENT	POLARITE ASPECTUELLE DU VERBE [Vandeloise, 1986]	EFFET DE LA POLARITE ET DE LA PREPOSITION SUR L'ES [Flageul, 1997]	RELAJ [Loustau, 2008]
Je pars de Pau.	initiale	<i>initiale</i> +de → origine	oui
J'arrive à Pau.	finale	<i>finale</i> +à → destination	oui
Je traverse Pau.	médiane	<i>mediane</i> + ∅ → intermédiaire	oui
Je quitte Pau.	initiale	<i>initiale</i> + ∅ → origine	oui
Je pars pour Pau.	initiale	<i>initiale</i> +pour → destination	non

TABLE 5.1 – Exemples d'interactions entre polarité aspectuelle du verbe de déplacement, la préposition et statut du segment dans l'itinéraire

Lorsque deux relais sont validés, on obtient un *segment* dans la terminologie de Loustau, et les *segments* permettent de construire les itinéraires. Nous avons montré rapidement l'interface entre le texte et l'extraction des informations spatiales telle que l'a présenté Loustau dans sa thèse, mais nous ne donnerons pas plus de détails ici sur le travail qu'il a mené sur les itinéraires.

Suite à ces recherches, notre travail s'est donc particulièrement focalisé sur la dimension temporelle qui a été considérée par défaut comme étant chronologique d'un segment à l'autre dans le récit, nous montrons dans ce chapitre que cet a priori est justifié et que l'approche par segment est pertinente malgré quelques nuances. Nous avons donc cherché à travers une analyse sémantique fine, et dont la granularité est le segment, à interpréter dans un modèle la temporalité des déplacements (présenté dans le chapitre précédent), puis à étendre les représentations sémantiques pour que le système de calcul de la temporalité événementielle utilisé soit adapté au français (présenté dans le chapitre suivant).

5.3 Définitions

Nous avons caractérisé le discours comme étant une unité cohérente regroupant tout ce qui est dit par un énonciateur dans un lieu et à un moment donnés. Le discours, appelé aussi production discursive, est une unité structurée et constituée d'unités

plus petites, les segments, que l'on peut rapprocher du niveau de la phrase en linguistique textuelle. Le niveau intermédiaire entre le discours et le segment est la séquence textuelle qui du point de vue purement discursif, est tout ce qui entre dans un cadre discursif [Adam, 2008]. Les séquences sont elles aussi des unités cohérentes de discours qui permettent sa structuration. Il semble difficile d'obtenir des informations sur la structuration d'une production discursive en n'observant qu'un seul de ces trois niveaux d'organisation, et l'analyse profonde du discours ne peut être obtenue qu'à la suite de l'analyse de chacun d'entre eux. Deux techniques peuvent être utilisées, « top-down » et « bottom-up ». La première consiste à retrouver les spécificités discursives du discours, en le comparant à d'autres par exemple, puis à approfondir aux niveaux de granularité inférieure par l'observation de la séquence puis du segment. La seconde consiste à observer les unités minimales comme le segment, pour ensuite à reconstruire l'architecture du discours par des unités de niveau supérieur, c'est-à-dire les séquences pour atteindre le niveau du discours.

Le genre du discours regroupe les productions discursives ayant des spécificités linguistiques, des conditions de productions et des buts communicatifs communs. Concernant ces derniers, une typologie en genres du discours est une conception du discours en contexte et régit par des phénomènes extra-linguistiques. Les spécificités quant à elle sont observables du point de vue textuel et peuvent être évaluées par des analyses quantitatives et qualitatives du matériau discursif afin de mettre en évidence ce qui rapproche ou éloigne les discours les uns des autres.

Dans une logique "top-down", nous donnons ici pour premier niveau d'analyse, une représentation schématique de la répartition des genres discursifs, sous forme de cercles dans la figure 5.1. Cette représentation pour démarquer les genres est artificielle et ne peut en aucun cas noter une limite entre les genres. Elle peut tout au plus montrer que certains discours se rapprochent du centre de cet "espace normatif" qu'est le genre de discours, tout comme ils peuvent s'en éloigner, la schématisation en cercle prenant ici tout son sens. Par ailleurs, nous nous appuyons dans le présent travail sur une typologie du discours mettant en évidence des spécificités pour le récit de voyage tout en montrant que l'approche par genre est limitée. Dans ce premier schéma, nous avons voulu montrer que la notion de genre de discours pouvait être considérée comme une forme de repérage au sein de l'ensemble des discours. Ce repérage étant basé sur l'utilisation de différents procédés linguistiques, un genre spécifique, (ici représenté donc naïvement par le cercle) peut avoir des intersections avec d'autres genres, et une production discursive peut appartenir de manière non exclusive à plusieurs genres. Par exemple, l'usage de la première personne du singulier peut être donné comme une spécificité du genre du journal intime, mais un journal intime rédigé à la première personne du pluriel appartiendra tout de même à cet espace normatif par l'utilisation d'autres procédés stylistiques, tout en étant une production discursive périphérique à ce genre dans notre représentation.

Nous montrons ensuite à un second, puis à un troisième niveau, la structure du

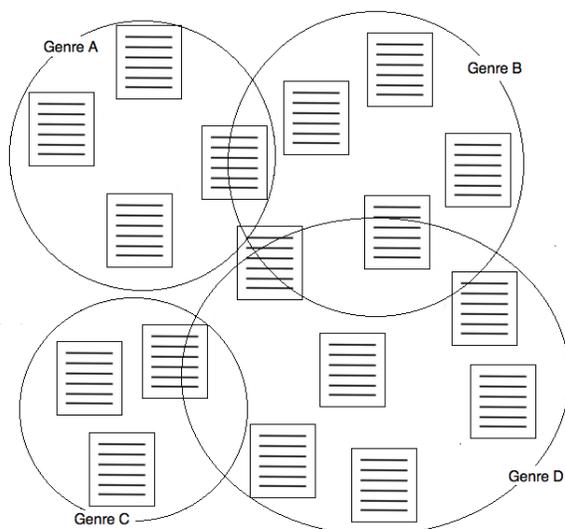


FIGURE 5.1 – Niveau 1 : Schématisation de la répartition entre genres du discours

récit de voyage dans la figure 5.2, en schématisant cette fois la répartition des unités observables, tout d'abord les séquences, puis les segments.

Nous abordons dans ce chapitre la distinction entre régime factuel et régime fictionnel, récit d'itinéraire et description du réel, tout à fait caractéristiques du récit de voyage. Ce second schéma donne une première idée de la répartition des segments dédiés à différents thèmes constitutifs du récit de voyage.

5.4 Notre corpus et le récit de voyage au XIXème siècle

5.4.1 La méthode

Comme présenté dans la rubrique précédente, nous nous sommes demandée s'il était possible de distinguer automatiquement par la structuration du discours, les séquences textuelles propres à la description de l'itinéraire. Contrairement à Lous-tau, nous nous situons donc au niveau 1, figure 5.1, afin d'accéder par la suite aux segments. En premier lieu et suite au travail de Véronique Magri-Mourgues [Magri-Mourgues, 2009], nous avons cherché à caractériser le récit de voyage du XIXème siècle, nous avons donc confronté notre corpus de RV avec des œuvres littéraires de genre similaire, puis avec des romans de la même période, et finalement avec des journaux intimes. Ces travaux ont été menés en collaboration avec Natalia Vinogradova, alors doctorante sur le projet Itipy. Quelques marqueurs utilisés dans le

5.4. NOTRE CORPUS ET LE RÉCIT DE VOYAGE AU XIXÈME SIÈCLE

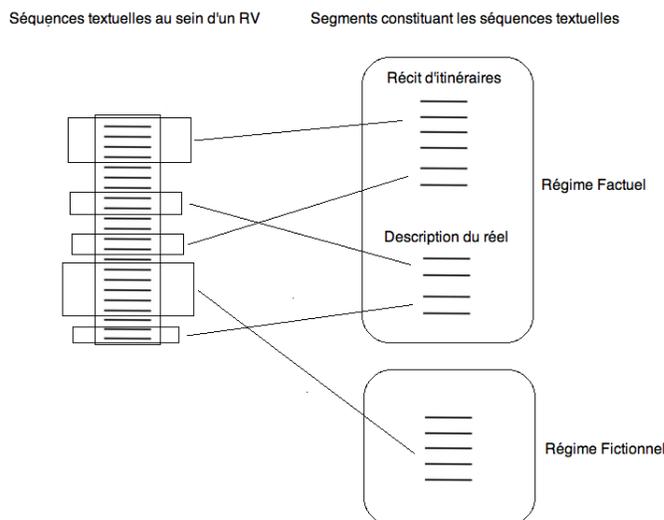


FIGURE 5.2 – Niveaux 2 et 3 : Les séquences textuelles et les segments du récit des itinéraires dans la structuration du récit de voyage

travail de V. Magri-Mourgues pour dresser un portrait du RV ont été testés afin d'en éprouver la fiabilité pour notre corpus. Nous avons supposé que l'hétérogénéité des données textuelles reconnue de ce genre devait se manifester lors de la confrontation à un autre corpus en variant les genres. Ensuite, nous avons approché le corpus sous l'angle de la théorie de A. Pasquali ([Pasquali, 1994]) selon laquelle le RV est une alternance de séquences textuelles comprenant « le récit de voyage et de découverte du réel », « le récit métaphorique », etc. (op. cit. : 26). Le RV présente une dichotomie explicite entre information factuelle, dans laquelle une place particulière est réservée aux itinéraires, et information fictionnelle, définie comme « types fonctionnels » par D. Biber, U. Connor et A. Thomas ([Vine, 2009]). Dans le cadre de la SDRT (Segmented Discourse Representation Theory, [Lascarides et Asher, 1993]), nous avons poursuivi l'analyse au niveau inférieur et examiné les unités discursives minimales, ou segments, afin d'accorder une représentation sémantique au discours et de connaître le comportement discursif des itinéraires, que nous voulons extraire. À ce niveau, nous questionnons une nouvelle fois l'hétérogénéité des données telle que nous la définissons au cours de l'analyse.

Un corpus contrastif à partir d'un corpus de référence de RV a été constitué afin de varier les conditions de production discursive et de cerner au mieux le RV. Les principaux paramètres en jeu ici sont :

- le genre,
- l'opposition factuel-fictionnel,

CHAPITRE 5. MISE EN PERSPECTIVE DE LA MÉTHODE : L'ITINÉRAIRE, LE
RÉCIT DE VOYAGE AU XIXÈME SIÈCLE ET ANALYSE SÉMANTIQUE DU
SEGMENT

- l'environnement spatial du récit,
- l'auteur
- et les dates de production.

Le corpus initial est composé de onze œuvres identifiées, par la médiathèque de Pau, comme récits de voyage pyrénéens du XIXe et début XXe siècles. Le RV est reconnu comme un genre caractérisé par des spécificités attestées dans *Le voyage à pas comptés* [Magri-Mourgues, 2009]. Nous avons utilisé quelques critères cités dans ce travail afin d'identifier le corpus de départ ITIPY comme appartenant au RV et de renforcer cette classification par notre exploration. Ainsi, le corpus IPITY a été augmenté de différents textes permettant de délimiter plusieurs corpus de travail (voir la figure 5.3) :

- deux RV se déroulant en Normandie et en Orient, qui forment avec onze textes originaux de ITIPY la partition A ;
- sept romans dont trois du même auteur et quatre d'auteurs divers regroupant différents genres, allant du « roman personnel » ([Dufiez-Sanchez, 2010]) au « roman naturaliste », ceux-ci composant la partition B ;
- trois journaux intimes datant de la fin du XIXe siècle pour les deux premiers et de la fin du siècle précédent pour le dernier, constituant la partition C.

Nous espérons a priori faire émerger les spécificités du RV par la confrontation à des œuvres ayant certains paramètres de production similaires. En effet, il ressort de la comparaison globale de ces partitions que le RV se distingue des autres genres du roman par son régime factuel. L'univers du discours n'est donc plus décroché de la réalité extralinguistique, mais se proclame au contraire « en phase » avec cette réalité pour que l'authenticité présumée du récit n'en soit que plus frappante. Cette volonté d'embrayer l'univers du discours avec une réalité extralinguistique fait bien partie des conditions de production du RV et en influence la réalisation énonciative.

En nous appuyant sur l'analyse de V. Magri-Mourgues [Magri-Mourgues, 2009], nous avons procédé à une étude lexicométrique à l'aide du logiciel Lexico3² ([Salem, 1991]) afin d'observer et de défricher notre corpus dans son ensemble. Lexico3 est destiné à opérer de manière quantitative une analyse basée sur les lois d'hypergéométrie des discours en vue d'une approche qualitative ultérieure. Sans surprise, les champs lexicaux ayant trait aux éléments naturels comme « montagnes », « vallée », « rivière », par exemple, sont principalement attachés à la partition A, tandis que les objets quotidiens et les parties du corps, « tête », « yeux », « mains », sont plutôt attachés à la partition B. Il semble assez naturel que ces deux thèmes distincts soient distribués ainsi au sein de notre corpus constitué. On notera tout de même le cas particulier de « pied(s) » qui se trouve en fréquence supérieure dans la partition A du fait des constructions syntagmatiques « au pied de la montagne » et « sous leurs pieds ». L'univers est spatialisé, et l'expérience narrée est celle de l'énonciateur face au territoire, à la nature, tandis que, dans le roman, l'aventure narrée se situe plus particulièrement dans la société, le rapport humain. L'analyse factorielle des correspondances fournie par Lexico3, figure 5.4, est construite sur l'intégralité des mots du texte, sans que l'utilisateur ne puisse sélectionner de traits particuliers. Cette AFC montre un regroupement des textes de récit de voyage (en rouge pour Itipy et rose pour les RV ajoutés), nettement distingués de la partition B (en bleu) et un rapprochement de la

2. <http://www.tal.univ-paris3.fr/lexico/>

5.4. NOTRE CORPUS ET LE RÉCIT DE VOYAGE AU XIXÈME SIÈCLE

ID	titre	auteur	Fact v/s Fict	date	genre
A0	Mémoires d'un touriste	Stendhal	Factuel	XIX	Récit de voyage
A01	Voyage en Orient	Alphonse de Lamartine	Factuel	XIX	Récit de voyage
A1	Ascension au Pic du Nethou	Platon de Tchihatcheff	Factuel	XIX	Récit de voyage
A2	Au Pays des Isards	Les frères Cadier	Factuel	Début XX	Récit de voyage
A3	Excursions autour du Vignemale	Alphonse Meillon	Factuel	Début XX	Récit de voyage
A4	La conquête du Mont-Perdu.	Louis Ramond de Carbonnières	Factuel	XIX	Récit de voyage
A5	L'aventure du Vignemale	Didier Lacaze	Factuel	XIX	Récit de voyage
A6	Les Pyrénées ou voyages pédestres (Livre I : Béarn-PaysBasque)	Vincent de Chausenque	Factuel	XIX	Récit de voyage
A7	Ann Lister, première ascension du Vignemale	traduction de Luc Maury	Factuel	XIX	Récit de voyage
A8	Voyage aux Pyrénées	Hippolyte Taine	Factuel	XIX	Récit de voyage
A9	Voyage au Mont-Perdu et Dans la partie adjacente des Hautes-Pyrénées	Louis Ramond de Carbonnières	Factuel	XIX	Récit de voyage
A10	Fragments d'un voyage sentimental et pittoresque dans les Pyrénées.	Jean Florimond Boudon de Saint-Amans	Factuel	XIX	Récit de voyage
A11	Voyages inédits dans les Pyrénées	Alain Bourneton	Factuel	XIX	Récit de voyage
B0	L'assommoir	Emil Zola	Fictionnel	XIX	Roman naturaliste
B01	L'argent	Emil Zola	Fictionnel	XIX	Roman naturaliste
B02	La curée	Emil Zola	Fictionnel	XIX	Roman naturaliste
B1	Les trois mousquetaires	Alexandre Dumas	Fictionnel	XIX	Roman historique
B2	La petite Fadette	George Sand	Fictionnel	XIX	Roman social
B3	René	François -René Chateaubriant	Fictionnel	XIX	Roman du moi
B4	Madame Bovary	Flaubert	Fictionnel	XIX	Roman réaliste
C1	Journal	Jules Renard	Témoin	Fin XIX	Roman réaliste / Journal intime
C2	Mon Journal	Léon Bloy	Témoin	Fin XVIII	Roman réaliste / Journal intime
C3	Voyage en France	Arthur Young	Témoin	Fin XVIII	Roman réaliste / Journal intime

FIGURE 5.3 – Constitution du corpus de test

CHAPITRE 5. MISE EN PERSPECTIVE DE LA MÉTHODE : L'ITINÉRAIRE, LE
RÉCIT DE VOYAGE AU XIXÈME SIÈCLE ET ANALYSE SÉMANTIQUE DU
SEGMENT

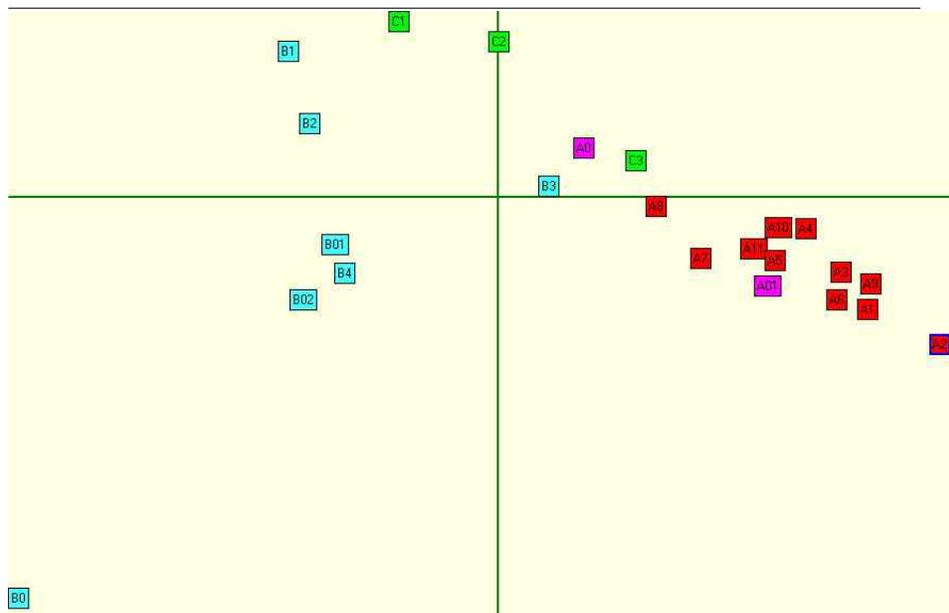


FIGURE 5.4 – L'analyse factorielles des correspondances avec Lexico3

partition C (en vert) vers le RV.

La spécificité des RV est ainsi confirmée. Elle est notamment illustrée par la répartition de certains marqueurs énonciatifs, tout particulièrement les indices de la deixis et l'expression de la personne. Chacune des partitions traitées, malgré une part d'hétérogénéité propre, peut dès lors recevoir une caractérisation linguistique.

La Figure 5.5 représente la ventilation fréquentielle des adverbes "ici" et "maintenant", afin d'illustrer le brayage³ spatial et temporel grâce à des formes morphologiquement simples et sans équivoque. Nous donnons ici ces fréquences à titre indicatif car il a été montré que ces marques pouvaient être interprétées différemment des déictiques classiques [Recanati, 2001].

Le brayage spatial, primant nettement sur le brayage temporel pour la partition A, exprime une caractérisation du RV malgré des fréquences restant très hétérogènes. Le brayage évoqué par les personnes suit la même logique. Le RV dénote un monde spatialisé et dans lequel un « je » avance dans un espace décrit, sans nécessiter systématiquement le rappel du fil temporel de la narration. Par contraste, au sein du corpus, le roman comporte généralement un narrateur hétérodiégétique qui relate les faits et

3. On définit ici la notion de brayage comme le rapprochement entre l'énonciateur et son discours, impliquant une subjectivité apparente. Les formes de brayage personnel étant principalement portées par les pronoms personnels, les formes du brayage spatial et temporel pouvant être adverbiales, adjectivales ou encore constituées de syntagmes prépositionnels complets par exemple.

5.4. NOTRE CORPUS ET LE RÉCIT DE VOYAGE AU XIXÈME SIÈCLE



FIGURE 5.5 – Fréquences des adverbes "ici" et "maintenant" avec Lexico3

CHAPITRE 5. MISE EN PERSPECTIVE DE LA MÉTHODE : L'ITINÉRAIRE, LE
RÉCIT DE VOYAGE AU XIXÈME SIÈCLE ET ANALYSE SÉMANTIQUE DU
SEGMENT

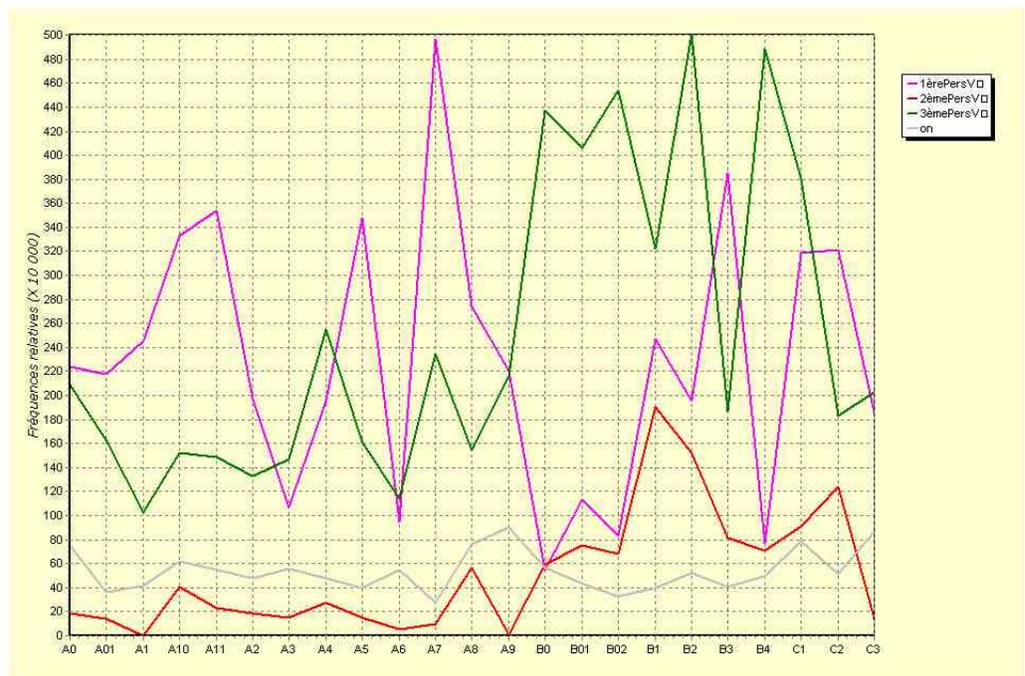


FIGURE 5.6 – Fréquences des pronoms personnels employés avec Lexico3

actions autour du protagoniste de la fiction. La spatialisation est moins prégnante et le rapport entre la deixis temporelle et spatiale est plus équilibré. On observe des comportements semblables entre tous les RV et B3, C2, C3 (visible dans la Figure 5.4). D'après les analystes du roman du XIXe s. [Gasparini, 2004], René de Châteaubriant (B3) est catégorisé comme un « roman personnel » [Gasparini, 2004] dont la portée déictique est franchement contrastée par rapport à ses contemporains. Nous prendrons en compte dans notre analyse le statut particulier de ce genre. C2 et C3, quant à eux, sont nos témoins. Leur comparaison avec les RV permet de les rapprocher de la partition dite factuelle sans que, pour autant, ils ne fassent partie des récits de voyage : ils partagent avec les RV un aspect de « découverte du réel », le réel étant un territoire, et la volonté de relater des faits authentiques. Nous constatons que le RV est un genre spécifique qui possède des caractéristiques lexicales et déictiques attestées. Nous remarquons, néanmoins, les similarités avec les textes des autres genres et des fréquences hétérogènes d'un texte à l'autre. Ce fait nous invite à explorer notre corpus d'origine de plus près, en envisageant notamment le niveau d'analyse intra-textuel afin de mieux comprendre son organisation. Ainsi, nous nous attachons à l'étape suivante de notre analyse, à observer au sein du RV, c'est-à-dire de la partition A, les manifestations de l'alternance des séquences textuelles. Malgré les spécificités observées caractérisant la partition A, nous avons déjà remarqué une forme d'hétérogénéité dans la construction de l'univers discursif de cette partition, nous nous intéressons maintenant à la construction interne au RV, représentée par le niveau 2 de la Figure 5.2.

5.4.2 Les types fonctionnels

Selon J.-M. Adam ([Adam, 1992]), tout texte est « une configuration réglée par divers modules ou sous-systèmes en constante interaction », l'un de ces modules étant son organisation séquentielle. La séquence est définie comme une entité textuelle « constituée de paquets de propositions (les macro-propositions), elles-mêmes constituées de *n* propositions » (op. cit. : § 41). Le RV, tout comme les autres productions discursives, est composé de séquences textuelles, que l'on peut discriminer les unes des autres par divers critères. Suite à la théorie d'Affergan dans *Exotisme et altérité* en 1987, A. Pasquali ([Pasquali, 1994]) souligne que :

le récit de voyage et de découverte du réel [alterne avec] le récit métaphorique, le récit métonymique et le récit synecdochique.

Il existe donc une structuration propre au RV dans laquelle on trouve différents procédés narratifs. Il semble que cette hétérogénéité des séquences textuelles est intrinsèque au RV. Pour traiter ce phénomène, nous adoptons la terminologie de D. Biber ([Biber *et al.*, 1998]). Il évoque les types fonctionnels pour désigner l'ensemble des séquences textuelles ayant le même but communicatif. Le mécanisme de production du discours, qu'il soit oral ou écrit, est régi par un but communicatif : donner une information, demander un renseignement, formuler un ordre, etc. qui influence le type fonctionnel. Dans le discours, ces buts communicatifs alternent, ce qui se traduit par une alternance des types fonctionnels. Nous remarquons ce même phénomène au sein du RV qui se prête tout à fait à une approche thématique factuelle/fictionnelle. D. Biber a construit cette catégorisation de manière interne au texte, en croisant les marqueurs

CHAPITRE 5. MISE EN PERSPECTIVE DE LA MÉTHODE : L'ITINÉRAIRE, LE
RÉCIT DE VOYAGE AU XIXÈME SIÈCLE ET ANALYSE SÉMANTIQUE DU
SEGMENT

linguistiques émergents. Ces marqueurs sont des traits lexicaux et grammaticaux qui permettent de caractériser les différents types fonctionnels et de les distinguer les uns des autres. Nous citerons en guise d'exemple la dimension « informative » versus « impliquée » regroupant plusieurs traits tels que les pronoms personnels employés, par exemple : l'utilisation de je proposant plutôt un discours impliqué et l'absence du locuteur présentant plutôt un discours informatif ([Pery-Woodley et Scott, 2006]). Dans le RV, ces deux dimensions se présentent par le récit factuel et le récit fictionnel, respectivement. En effet, dans le récit factuel – plutôt informatif – l'énonciateur prétend décrire le monde réel comme il est, tandis que, dans le récit fictionnel – plutôt brayé (impliqué) – il cherche à montrer des émotions et une forme de subjectivité. Puis, au sein du récit factuel, un découpage plus fin nous intéresse, à savoir le type fonctionnel propre au récit de l'itinéraire qu'il faudra séparer du récit de la vie quotidienne ou encore des descriptions naturalistes.

Dans notre corpus, certaines séquences textuelles régies par un type fonctionnel sont explicitement délimitées par le découpage en paragraphe. Prenons un exemple extrait de *Fragments d'un voyage sentimental et pittoresque dans les Pyrénées*⁴ :

(5.2) Mais quittons ce lieu désolé. M. Mercère mon protecteur, part pour Gavarnies. Je vais l'accompagner à travers les hautes montagnes qui séparent Héas de ce dernier district. Quelle route, grands Dieux ! Mais que dis-je ? Il n'y a point de route ici : le voyageur monte, descend, traverse les prairies & les gaves, sans chemins, sans traces, sans autre renseignement, que la position respective du lieu d'où il vient, & du lieu où il va. M. Pasumot, M. Dusaulx, & vous curieux, amateurs, ou promeneurs de Barège, je commence à vous féliciter de n'être point du voyage. Ne croyez point cependant que je forme le moindre regret de l'avoir entrepris ; les montagnards & les montagnes me seront plus connus désormais, que je n'aurais pu me flatter de les connaître ; dans cinq cents de nos courses ordinaires. Mais revenons, partons de Héas. [A10 : *Fragments d'un voyage sentimental et pittoresque dans les Pyrénées*]

Le contenu factuel des trois premières phrases, suivies par une expression forte de subjectivité, introduit un récit plutôt fictionnel. Nous notons une rupture fonctionnelle, marquée par des commentaires subjectifs sur la difficulté du voyage, puis la clôture de cet aparté par *Mais revenons*. Ce passage indique clairement que nous nous situons sur deux plans énonciatifs liés : le premier décrivant le voyage réalisé dans le passé, ancré dans une réalité révolue, et le second incluant le lecteur dans le voyage qui se déroule au fil du récit. Ce passage relève du même phénomène que nous avons décrit avec l'exemple des pronoms avec Lexico3. Le *nous* inclusif permet de délivrer des informations factuelles sur l'itinéraire, tout en brayant l'univers discursif et en intégrant le lecteur. L'aparté central dissocie le protagoniste du lecteur par l'opposition *je/vous* et laisse place à des informations d'ordre fictionnel.

Le repérage de ces types fonctionnels laisse apparaître les traces de l'hétérogénéité structurelle et énonciative du RV. Trois types fonctionnels sont pertinents pour notre recherche : le récit fictionnel teinté d'une forte subjectivité de l'énonciateur, le type factuel de l'itinéraire, incluant une deixis personnelle et spatiale plus faible en

4. Quelques coquilles sont présentes dans l'extrait et nous le livrons ici sans corrections, tel qu'il a été intégré dans le corpus ITIPY.

subjectivèmes lexicaux, et le récit factuel « autre », englobant les descriptions de l'environnement et le récit de la vie quotidienne. Nous rappelons que nous nous situons à la seconde étape de l'analyse (allant du niveau 2, les séquences, au niveau 3, les segments, de la Figure 5.1) au sein de laquelle nous voulons dégager les séquences textuelles pertinentes à l'extraction des itinéraires. Ainsi, nous pouvons conclure que l'hétérogénéité des données doit être prise en compte dans l'extraction d'information. Une distinction entre les différentes séquences semble donc pertinente pour toute tâche d'extraction d'informations à partir de discours dont on connaît les types fonctionnels qui le composent. C'est pourquoi nous avons analysé l'extrait choisi 5.2 sous l'angle de la SDRT, afin d'en observer les segments et de leur articulation.

5.5 La structure du discours au sein du récit de voyage

5.5.1 La segmentation du discours dans le cadre de la SDRT

Nous rappelons brièvement⁵ que dans le cadre théorique de la SDRT (*Segmented Discourse Representation Theory*), version enrichie de la DRT, nous cherchons à traiter automatiquement les informations factuelles ayant trait aux itinéraires.

La SDRT permet d'intégrer les articulations entre segments au sein du discours sous forme de relation rhétorique. L'unité utilisée est l'EDU *elementary discourse unit* l'EDU et correspond à un segment textuel minimal du discours représentant soit un événement, soit une période temporelle.

Les CDU, *complex discourse unit* désignent un bloc comprenant plusieurs EDU, en préservant la cohérence thématique et rhétorique. La définition formelle de ce concept est proposée dans le travail de ([Asher, 2011]). Nous redonnons ici l'exemple de la section 3.6.2 inspiré de notre corpus et analysé en SDRT :

- (5.3) Notre arrivée à Bagnères ne fût pas facile. [a]
- (5.4) Nous avançons à un rythme irrégulier depuis un moment [b]
- (5.5) et nous entrâmes tard dans la rue Saint-Blaise. [c]
- (5.6) Depuis là-bas, nous fîmes une promenade à Superbagnères. [d]
- (5.7) Nous suivions d'abord un chemin d'une pente peu rapide [e]
- (5.8) et ensuite nous nous élevâmes presque sans nous en apercevoir jusqu'à Superbagnères. [f]

π_A et π_D sont les labels de CDU, contenant π_b et π_c pour le premier, et π_e et π_f pour le second. Ici π_a est élaboré par π_A et π_d est élaboré par π_D . L'intégration des relations discursives à la DRT a été motivée par l'impossibilité d'exprimer la temporalité d'une suite d'événements en s'appuyant simplement sur la syntaxe et sur le lexique, et par la nécessité de faire appel à des connaissances du langage et du monde pour une représentation sémantique du discours juste.

Les relations discursives *elaboration* et *narration* sont deux relations incompatibles, deux événements ne pouvant par ailleurs pas être successifs et inclus l'un dans l'autre. Nous présentons quelques unes des relations disponibles dans la section suivante.

5. Une introduction à la SDRT est donnée en section 3 du chapitre 3.6.2.

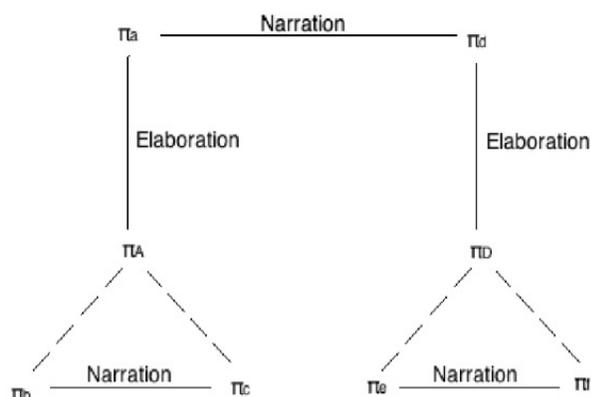


FIGURE 5.7 – une représentation en SDRS de l'exemple 5.3

5.5.2 Segmentation et relations discursives dans le RV

Regardons de plus près notre corpus et reprenons l'exemple 5.2 segmenté :

- (5.9) [Mais quittons ce lieu désolé.]πa [M. Mercère mon protecteur, part pour Gavarnies.]πb [Je vais l'accompagner à travers les hautes montagnes]πc [qui séparent Héas de ce dernier district.]πd [Quelle route, grands Dieux !]πe [Mais que dis-je ?]πf [Il n'y a point de route ici :]πg [le voyageur monte, descend, traverse les prairies& les gaves, sans chemins, sans traces, sans autre renseignement, que la position respective du lieu]πh [d'où il vient]πi [,& du lieu] πh [où il va.]πi [M. Pasumot, M. Dusaulx,& vous curieux, amateurs, ou promeneurs de Barège, je commence à vous féliciter de n'être point du voyage.]πk [Ne croyez point cependant]πl [que je forme le moindre regret de l'avoir entrepris :]πm [les montagnards& les montagnes me seront plus connus désormais, que je n'aurais pu me flatter de les connaître ; dans cinq cents de nos courses ordinaires.]πn [Mais revenons,]πo [partons de Héas.]πp

Nous analysons cet exemple en EDU et CDU et attribuons les relations entre les segments. Néanmoins, comme montré dans la section précédente, cet extrait est hétérogène, et contient des segments de deux types fonctionnels différents. A l'origine, la SDRT propose des relations entre les segments de même type fonctionnel. La question

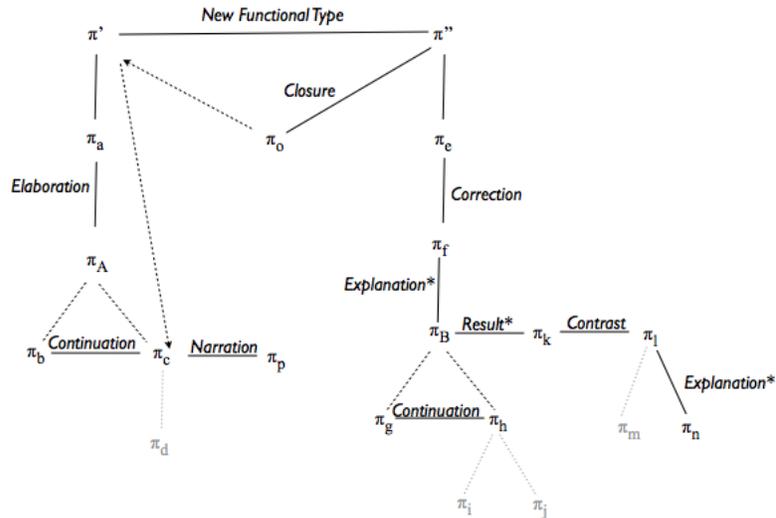


FIGURE 5.8 – une représentation en SDRS de l'exemple précédent

se pose alors de définir comment élargir le cadre théorique pour pouvoir analyser correctement cet extrait en particulier et le RV de manière plus générale. Nous présentons en figure 5.8 l'analyse des relations discursives entre les EDU et CDU de l'exemple précédent.

La séquence textuelle ayant trait à l'itinéraire semble caractérisée, dans ce passage, par les relations de discours telles que *elaboration* et *continuation* (π_a , π_b , π_c). Nous avons déjà introduit la relation *elaboration*. Quant à la relation *continuation*, relation coordonnante, elle implique une forme de narration sans domination thématique de l'une sur l'autre et sans succession temporelle entre les événements portés par les deux EDU. Les constituants de la seconde séquence sont liés par les relations de *correction*, permettant de remettre en question le contenu sémantique de son premier argument par le second, *continuation* (expliquée précédemment), *explanation** et *result** qui introduisent respectivement l'explication et le résultat meta-discursifs, et enfin *contrast*, permettant d'exprimer par le second segment la violation d'une attente induite par le

premier. Nous précisons que nous ne nous intéressons pas pour le moment aux segments imbriqués, en gris sur le graphe. Il faut maintenant établir la relation entre les deux séquences qui se suivent dans notre exemple. Afin de conserver la cohésion énonciative et la cohérence temporelle, il faut considérer que la seconde séquence est une forme de cadre⁶, le cadre étant un regroupement de segments homogènes sous la portée d'un critère sémantique qui permet de structurer le discours. Notre critère pragmatico-sémantique portant sur l'itinéraire conserve une cohérence énonciative et temporelle. La SDRT permet d'établir les relations entre les segments discursifs simples et complexes, elle est parfaitement adaptée à l'analyse des relations entre cadres discursifs. Pourtant, ce comportement nous semble ne pas avoir été traité au sein de la SDRT : quelle relation discursive utiliser entre nos deux séquences ? Aucune ne convient, il faut donc introduire une nouvelle relation permettant de manipuler le phénomène rencontré.

5.5.3 Proposition d'une nouvelle relation discursive

La structure du RV repose principalement sur l'avancée de l'itinéraire et les autres types d'informations dépendent de cette structure. Ainsi, nous avançons que les informations intégrées dans le RV, sont rattachées au fil narratif de l'itinéraire sous la forme de cadres discursifs. Chacun contient des informations fictionnelles ou factuelles spécifiques qui n'ont pas forcément de lien avec le cadre précédent. À la fin d'une séquence fictionnelle par exemple, le cadre se ferme et l'auteur poursuit avec la description de l'itinéraire. Ce retour à l'itinéraire, après un ou plusieurs cadres, est donc systématique, il garantit la cohérence au sein du RV. Dans l'exemple étudié, l'ouverture et la fermeture du cadre sont très explicites du point de vue énonciatif et des relations discursives. Nous proposons deux relations rhétoriques généralisées propres à ce type de discours : *New Functional Type* et son dual *Closure*. En effet, les informations fictionnelles étant variées, nous supposons plusieurs types fonctionnels correspondants. Dans ces conditions, chaque nouveau cadre discursif est régi par le type fonctionnel des informations qu'il contient. Lorsque le cadre se ferme, la relation de clôture renvoie au moment où la description de l'itinéraire s'est arrêtée. Dans la SDRT, il est reconnu que la liste des relations n'est pas complète et que les relations sont souvent ambiguës. Nous cherchons à enrichir le cadre théorique en ajoutant deux nouvelles relations à cette liste, ce qui nous permettra de traiter le phénomène décrit. Notre proposition est le fruit d'une nécessité, elle se restreint à un phénomène spécifique à un seul genre discursif et beaucoup de problèmes restent encore à résoudre. Néanmoins, on ne peut généraliser totalement le fonctionnement textuel de l'itinéraire, car nous pouvons prévoir de nombreux cas où ce schéma ne s'applique pas. Prenons l'exemple suivant :

- (5.10) En faisant ces réflexions désolantes pour moi qui désirerais bien vivement partager avec mes amis tout le plaisir que je prends, toutes les sensations que j'éprouve, nous arrivâmes dans le district de Grip où M. l'abbé P**** ne place qu'une seule maison, où nous ne vîmes qu'un seul village : cela saute aux yeux. Comment se peut-il qu'on n'indique en cet endroit qu'une seule, qu'une unique

6. Nous remercions Patrice Enjalbert pour la précision suivante : il faut noter une distinction entre les cadres de [Charolles, 1997] et les cadres étudiés ici, qui concernent la structure narrative.

maison, lorsque la route & la vallée entières sont couvertes d'habitations, depuis un quart de lieue au-dessus de l'auberge jusqu'au village de Sainte-Marie ?
[A10 : Fragments d'un voyage sentimental et pittoresque dans les Pyrénées]

On voit que le segment appartenant au type fonctionnel de l'itinéraire est isolé dans une séquence dédiée à l'expression des sentiments du protagoniste, puis à la description de son environnement. Le changement de type fonctionnel n'est pas explicite et, comme précédemment, aucun marqueur ne le réalise. Pour autant, il nous faut relier la représentation de ce segment à la suite de la représentation dédiée à l'itinéraire. La conclusion que nous pouvons tirer de l'analyse de cet exemple est que nos séquences textuelles sont effectivement hétérogènes. Notre analyse du RV montre que l'itinéraire ne peut être traité exclusivement en terme de cadres. Les données textuelles du RV étant hétérogènes, les segments sont parfois isolés et ceci nous informe que la structuration du discours en séquences ne peut être limitée à des cadres s'ouvrant et se fermant. En complément de ces observations et de la proposition de relation discursive, nous avons voulu analyser les diverses formes que pouvait revêtir le segment dédié à l'itinéraire. En restant au niveau 3 de la figure 5.2, nous avons testé une technique traditionnelle en extraction d'information pour être plus précise sur notre analyse de l'hétérogénéité dans le RV, le patron lexico-syntaxique proposé par Loustau.

5.6 Le segment type : quelle approche ?

Nous avons cherché à dresser le portrait énonciatif du RV pour comprendre en corpus le comportement des séquences présentant l'itinéraire. Le corpus d'origine a été confronté à d'autres textes ayant des traits similaires. Nous avons démontré qu'en dépit de cette similarité la partition des RV se distingue des autres partitions en révélant ainsi une forme d'homogénéité et d'hétérogénéité à différents niveaux. L'analyse de l'hétérogénéité interne du RV a été abordée sous l'angle de sa structuration : elle laisse une place à l'existence de nombreux cadres discursifs de types fonctionnels différents, mais les organise de manière bien spécifique, en les greffant à la narration de l'itinéraire. La SDRT dispose d'outils appropriés pour traiter le discours, néanmoins nous proposons une relation discursive supplémentaire afin de traiter ce phénomène. Les relations ouvrant et fermant un cadre fonctionnel nécessitent des définitions formelles que nous n'avons pas l'opportunité de développer ici. Finalement, cette analyse rend compte des cadres de narration de l'itinéraire au sein du RV. Sachant que les productions discursives sont de fait hétérogènes dans leur structure, nous avons cherché et trouvé des exemples où la structuration du RV, comme nous l'avons décrite, n'est pas respectée : le cas où un segment d'un type fonctionnel est isolé au sein d'une séquence d'un autre type. Nous avons alors éprouvé le patron lexico-syntaxique rigide de Pierre Loustau qui permet de détecter les segments de l'itinéraire dans n'importe quel cadre discursif :

[verbe de déplacement, préposition ?, ES]

Néanmoins, certains problèmes apparaissent dans l'analyse. Selon l'évaluation de Loustau (2008), le patron ne représente que de 63 % à 85 % de ces expressions au sein du corpus et l'algorithme d'extraction automatique utilisant le patron permet d'extraire

CHAPITRE 5. MISE EN PERSPECTIVE DE LA MÉTHODE : L'ITINÉRAIRE, LE
RÉCIT DE VOYAGE AU XIXÈME SIÈCLE ET ANALYSE SÉMANTIQUE DU
SEGMENT

autour de 80% d'entre elles. Tout d'abord, le patron ne couvre pas tous les segments de l'itinéraire. Nous dirons ensuite que l'hétérogénéité des données discursives fait partie des spécificités du RV au niveau de la structuration fonctionnelle du discours, de sa construction énonciative et de la construction des segments appartenant au récit de l'itinéraire, malgré la convergence de plusieurs conditions de production laissant penser que ces segments formeraient un ensemble plutôt homogène. On peut dire que l'hétérogénéité du RV à tous les niveaux – du discours dans son ensemble à la composition lexico-syntaxique des segments le structurant – légitime une mobilisation de plusieurs techniques afin d'extraire efficacement les formes linguistiques que peut revêtir l'itinéraire.

Ce patron présente l'avantage d'être robuste. Nous reprendrons en exemple les segments *partons de Héas* et *nous arrivâmes dans le district de Grip* qui obéissent au patron décrit, que le cadre discursif précédant soit fermé ou non. Pour éprouver la validité du patron, nous testons sa distribution au sein du corpus afin de le confronter dans les différentes productions discursives représentées. Nous avons choisi d'extraire les concordances de quatre verbes. Plus précisément, nous avons choisi deux verbes dont l'emploi est polysémique au sein du corpus, "arriver" et "passer", et deux verbes dont la sémantique propose clairement un déplacement dans l'espace, "parcourir" et "descendre"⁷. Concernant les verbes non polysémiques "parcourir" et "descendre", on note une franche diminution des emplois entre la partition A et la partition B dans le graphique en figure 5.10.

Les comportements de ces 4 verbes sont très différents les uns des autres, "passer" garde une fréquence très normée d'un texte à l'autre sans distinction de partition. "Arriver" quant à lui diminue sensiblement entre la partition A et les deux autres, quant à "descendre" et "parcourir", leur fréquence diminue franchement dans les partitions B et C. Après examen minutieux des concordanciers du verbe "descendre"⁸, nous pouvons dire qu'il ne présente que peu ou pas d'emploi métaphorique en A tandis qu'en B et C la fréquence augmente clairement, et en B3, l'usage métaphorique dépasse même les autres usages. "Parcourir" ne montre pas de différence d'utilisation entre les partitions, on trouve dans chacune autant d'emplois métaphoriques ("parcourir des yeux").

Cette utilisation métaphorique est prise en compte par la présence de l'entité spatiale dans les arguments du verbe de déplacement.

Par ailleurs, ces deux derniers verbes se distinguent l'un de l'autre par l'emploi métaphorique du second, comme par exemple dans *Madame Bovary* :

(5.11) elle l'avait descendu tout au fond de son cœur.

Nous remarquons la présence intéressante dans notre corpus de récit de voyage de la pronominalisation de l'objet quand il est spatial.

(5.12) Cette route que je viens de parcourir [...] a souvent sur son côté gauche de fort jolies petites maisons [...]. [A0 : Mémoires d'un touriste]

Ici, la position du constituant portant le déplacement dans la relative marque une cohésion forte entre les différents segments du discours. La distinction entre segment du

7. D'après étude du concordancier sur le corpus il apparaît que les deux premiers verbes acceptent bien des emplois "métaphoriques" plus fréquemment que les deux autres.

8. 1178 occurrences du verbe dans la totalité du corpus.

5.6. LE SEGMENT TYPE : QUELLE APPROCHE ?

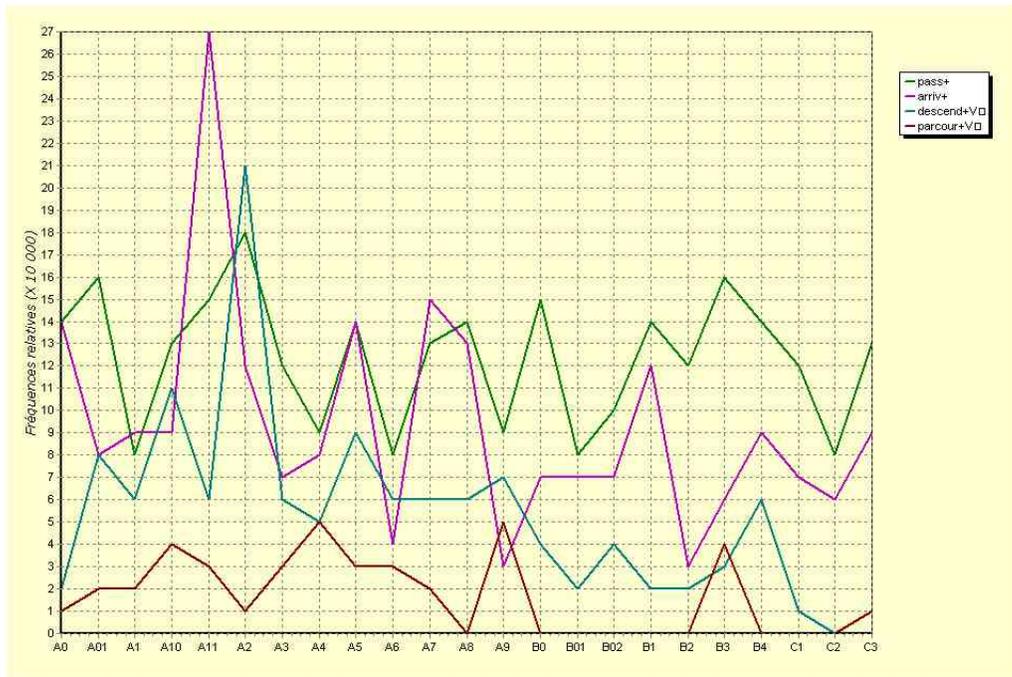


FIGURE 5.9 – Fréquence de répartition des verbes observés dans le corpus

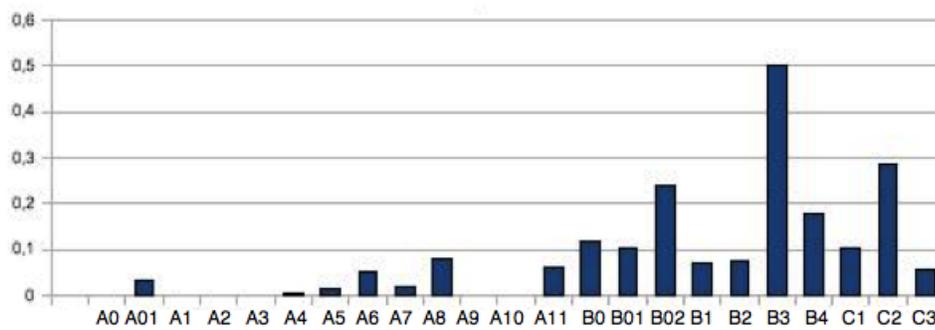


FIGURE 5.10 – Fréquence de répartition de l'emploi métaphorique du verbe "descendre" dans le corpus

type fonctionnel itinéraire et du type fonctionnel description est donc brouillée syntaxiquement au sein du même segment.

On trouvera donc une légitimité à représenter les segments imbriqués.

L'hétérogénéité des réalisations lexico-syntaxiques est à prendre en compte en vue d'une application efficace qui capturerait toutes les étapes de l'itinéraire d'un RV. L'emploi du verbe n'est donc pas restreint au seul sens locatif et l'expression de l'itinéraire par le verbe ne se réalise pas selon un seul patron lexico-syntaxique. Néanmoins, le verbe de déplacement semble un indice robuste quant à la détection des segments propres à la description des itinéraires. La granularité segmentale construite autour du verbe de déplacement semble donc être la méthode la plus robuste quant à l'extraction d'une information ayant trait à l'itinéraire.

Nous remarquons que l'hétérogénéité des données discursives préconise une approche où les deux approches sont complémentaires. Nous avons tout d'abord observé la structure du genre du récit de voyage qui nous a renseigné sur la présence de récits d'itinéraire et de description du réel ou encore de récits fictionnels. Nous avons pu observer le fonctionnement des séquences textuelles propres aux types fonctionnels, néanmoins, certains segments relatant une étape de l'itinéraire semblent isolés au sein de séquences différentes, ce qui complique une extraction d'information s'appuyant intégralement sur la structure du récit. D'autre part un patron lexico-syntaxique semble trop rigide pour capturer l'ensemble des segments propres à l'itinéraire. Nos investigations quant à ce genre discursif nous portent à penser qu'un pré-traitement extrayant en premier lieu les segments comportant un verbe de déplacement est pertinent. Ensuite, l'analyse sémantique segment par segment est appropriée, il reste cependant une dernière question quant au respect supposé de la linéarité chronologique entre segment une fois les segments pertinents regroupés.

5.7 Narration, description et segment dédié à l'itinéraire dans récit de voyage du XIXème siècle

Nous avons montré que le discours du RV est hétérogène, structuré sur une alternance entre segments (et/ou séquences textuelles dans le meilleur des cas) ayant pour type fonctionnel le récit de l'itinéraire, la description du réel et le récit fictionnel. Nous citons à ce propos Magri-Mourgues dans *La description dans le récit de voyage*[Magri-Mourgues, 1996] :

Le voyageur veut reconstituer un parcours, celui de son itinéraire réel. Comme le voyage a été jalonné par des paysages ou des scènes orientales, le parcours discursif est balisé par des séquences descriptives. Les conséquences de ce parti pris descriptif sont doubles :

- Le récit de voyage suit pour l'essentiel l'ordre chronologique qui reproduit la succession aléatoire et contingente des étapes du voyage référentiel. Les descriptions sont alors confrontées au danger de l'émiettement, du fragmentaire non motivé et simplement juxtaposé. A ce règne du hasard s'oppose le principe logique du texte de fiction où toute séquence apparaît comme maillon nécessaire d'un schéma narratif d'ensemble, comme serti dans une chaîne de causalité. De fait, une histoire au sens narratif est susceptible d'être résumée puisqu'elle com-

5.7. NARRATION, DESCRIPTION ET SEGMENT DÉDIÉ À L'ITINÉRAIRE DANS RÉCIT DE VOYAGE DU XIXÈME SIÈCLE

porte une situation initiale et une succession d'évènements hiérarchisés qui font aboutir logiquement à une situation finale. Chaque évènement, agencé en vue d'une finalité bien précise et préconçue, a dans ce contexte une fonction principale ou secondaire.

[...] Le résumé d'un récit de voyage ne peut être en définitive que l'énumération des lieux traversés accompagnés des dates qui les inscrivent dans une durée, et il peut se réduire simplement à une carte géographique sur laquelle on aurait retracé l'itinéraire réel. Les séquences descriptives qui se succèdent ont une égale importance et doivent être placées de front, pareilles aux grains d'un long chapelet qui relierait le début à la fin du voyage.

- De nouvelles relations espace-temps sont instaurées : dans ce type de récit en effet, espace et temps deviennent interchangeable. A chaque date, correspond un lieu, à chaque lieu une date, si bien que n'importe quel fait peut être identifié soit par la mention de la date à laquelle il s'est produit, soit par celle du lieu où il s'est produit. La contiguïté spatiale se mue en proximité temporelle lors du voyage réel et se trouve restituée par la juxtaposition textuelle dans le récit de voyage. C'est là une façon de (re)structurer le réel en lui imposant d'abord l'arbitraire du parcours réel et en lui surimposant ensuite la grille du texte écrit.
[...]

On peut en effet déceler des analogies et des récurrences de motifs, ce qui donne une structure profonde au récit de voyage comme pour compenser l'absence de structure apparente inhérente au genre même. Le désert, par exemple, motif-clé du voyage en Orient, se trouve décrit par touches successives, par bribes qui assurent la cohérence du récit. Mais la seule motivation de l'ordre des séquences descriptives reste celle de l'itinéraire réel du voyageur, promu seul héros de l'aventure et dont les déplacements relient les lieux décrits ; dès lors, puisque les descriptions écrites coïncident avec des étapes de son voyage, on pourrait parler, plutôt que d'une subordination de l'une à l'autre catégorie, d'une équivalence entre description et narration. La description n'est nullement une pause dans le récit[Genette, 1972].

On peut déduire de ces observations ainsi que de notre étude que le récit est majoritairement une narration séquentielle des visites des lieux effectuées par le voyageur. La narration permet donc de décrire un itinéraire qui a eu lieu dans le réel. Au sein de cette structure fondamentale, s'incrument des passages descriptifs, eux aussi cruciaux dans le déroulement de cet itinéraire, et dans le travail littéraire visant à rendre le plus authentique possible le voyage. On peut donc dire que le récit de voyage a une structure fortement contrainte par une chronologie reflétant le réel qui permet de décrire un itinéraire "référentiel", et c'est sur cette état de faits que l'on peut s'appuyer pour extraire l'itinéraire automatiquement d'après le corpus. On peut en déduire aussi une alternance entre narration et description. Mais cela reste à explorer plus en détail. Nous avons vu que très peu d'usages métaphoriques des verbes de déplacement sont présents dans le corpus, ce qui nous permet de nous appuyer sur le verbe en premier lieu pour analyser sémantiquement les déplacements énoncés. Le niveau 3 de la figure 5.2 sera donc privilégié, donnant la granularité nécessaire à la détection des itinéraires, mettant en relief les déplacements et localisations au sein de séquences textuelles diverses.

5.8 Conclusion

Notre portrait du RV a permis de mettre en relief différents points :

- le récit est structuré sur la chronologie globale de l'itinéraire "référentiel"
- les séquences textuelles sont donc des récits de l'itinéraire auquel on a greffé des descriptions du réel et des récits fictifs
- au sein des ces séquences il nous faut ordonner temporellement les segments relatant des déplacements et localisations
- dans la représentation sémantique de ces segments on veut obtenir les informations spatiales
- pour obtenir ces informations sémantiques, une analyse s'appuyant sur la détection des segments correspondant à un ou plusieurs patrons lexico-syntaxiques semble judicieuse dans une partie des cas (entre 63 et 85 % des cas pour le patron de Pierre Loustau [Loustau, 2008] évalué sur une partie du corpus ITIPY), néanmoins, une analyse sémantique à partir de l'analyse syntaxique profonde semble judicieuse pour les expressions des déplacements n'obéissant pas aux patrons proposés.

Par ailleurs, nous avons montré qu'une approche hybride allant de la séquence textuelle au segment et du segment à la séquence textuelle trouverait toute légitimité ici, d'une part pour ne traiter que le récit propre à l'itinéraire, et de l'autre pour repérer les segments isolés.

Ces différents raisons nous ont poussé à approfondir le système de Verkuyl, *Binary Tense* [Verkuyl, 2008], qui propose une représentation sémantique compositionnelle de la temporalité des éventualités. Ce système a été adapté à la temporalité verbale du français du XIXème siècle et par conséquent du français contemporain comme nous le montrerons dans le chapitre suivant. Cette adaptation a pour être objectif d'être intégrée à GRAIL, notre analyseur syntaxique [Moot, 1999] afin de donner une représentation sémantique fine des éventualités de déplacement et de localisation comme une première étape pour l'extraction d'itinéraires.

Chapitre 6

Adaptation du système *Binary Tense*

Nous abordons maintenant l'adaptation du système *Binary Tense* [Verkuyl, 2008] pour le français du XIX^{ème} siècle à nos jours. Afin de cerner au mieux les adaptations nécessaires, nous nous appuyons sur une étude diachronique fine menée sur le français du XI^{ème} au XX^{ème} siècles par Caudal et Veters [Caudal et Veters, 2007]. Nous nous penchons particulièrement sur le cas du passé simple et sa valeur sémantique aoriste, partagée avec le passé composé à ce moment particulier de l'évolution du paradigme sémantique des temps verbaux. Nous introduisons en premier lieu une réflexion sur les valeurs sémantiques des opérateurs PERF et IMP (pour *perfect* et *imperfect*) définis comme étant des opérateurs purement temporels et non-aspectuels dans *Binary Tense*, système à destination de l'analyse de langues d'origine germanique. Nous donnons notre conception de l'interaction entre temps, aspect et énonciation au sein du système et confrontons les analyses produites avec des exemples en français. Puis nous proposons une adaptation éclairée du système et de son interprétation pour couvrir les usages de ces deux temps centraux dans le récit de voyage.

6.1 Introduction

Suite à nos travaux sur la structuration du récit de voyage dans le chapitre précédent, nous proposons une analyse sémantique temporelle intégrée à *Grail* en adaptant

le système *Binary Tense*. En conservant l'atout majeur de ce système, la compositionnalité, nous l'adaptions afin de l'appliquer au français du XIX^{ème} siècle, l'objectif étant d'automatiser l'analyse sémantique temporelle des segments constituant les récits d'itinéraires dans des récits de voyage. Comme nous l'avons montré dans le chapitre précédent, le récit de voyage privilégie une structure chronologique du récit de l'itinéraire sur laquelle il semble pertinent de s'appuyer en vue de l'extraction d'informations temporelles. Néanmoins, il nous faut décrire précisément les différents temps du français utilisés afin d'évaluer la proposition de lexique sémantique de la temporalité verbale, *Binary Tense* au discours auquel on désire l'appliquer. Deux facteurs sont limitant dans ce cas, l'adaptation au français d'un système créé pour des langues d'ascendance germanique, et d'un point de vue diachronique, la description du paradigme des temps verbaux en français du XIX^{ème} siècle au sein de l'histoire de ses évolutions sémantiques.

Nous montrons que l'enjeu majeur de ce paradigme est un usage concurrent du passé composé et du passé simple d'un point de vue aspectuel, qui influence grandement la sémantique temporelle. Le point sur lequel nous nous attardons est la distinction entre perfect et aoriste.

Nous donnons en premier lieu une définition de l'aoriste, puis nous confrontons le système à des exemples en français afin de mieux décrire notre point de vue sur l'articulation entre temps, aspect et énonciation. Puis nous synthétisons les résultats d'une étude diachronique du français de Caudal et Veters [Caudal et Veters, 2007] afin de cerner au sein des évolutions du paradigme temporel verbal les enjeux du XIX^{ème} siècle. Par la même occasion, nous montrons que l'état des temps au XIX^{ème} siècle a gardé globalement la même sémantique jusqu'à nos jours, nos résultats au niveau de l'analyse du segment sont alors valables aussi pour le français contemporain.

Ensuite nous proposons une adaptation de *Binary Tense* permettant de respecter les contraintes présentées, c'est-à-dire la sémantique spécifique au français du perfect et de l'aoriste. Nous abordons tout d'abord l'impact de l'introduction d'un opérateur aoriste sur le système puis nous décrivons cet opérateur et son interprétation au sein du lexique proposé.

6.2 Définition de l'aoriste

L'origine de l'aoriste est grecque et en grec ancien on distingue deux formes, toutes deux présentant l'éventualité comme s'étant déroulée dans le passé. Le Trésor de la Langue Française Informatisé¹ dit à son sujet :

AORISTE, subst. masc. LING. Temps de la conjugaison grecque correspondant approximativement au passé simple et au passé antérieur français :

1. § 181. La théorie stoïcienne des temps, à laquelle l'aoriste doit son nom, a justement mis en lumière ce qu'il y a de négatif dans l'aoriste, surtout si on l'oppose au présent. Les stoïciens distinguaient deux types de temps déterminés (...) et indéterminés (...). Ils considéraient comme « déterminés » le « duratif » (...) c.-à-d. le présent et l'imparfait ainsi que l'« achevé » (...) c.-à-d. le parfait et le plus-que-parfait ; au contraire sont « indéterminés » l'aoriste et le futur. L'aoriste est effectivement ce

1. <http://atilf.atilf.fr/dendien/scripts/tlfiv5/advanced.exe?8;s=2189290260> ;

qui est dépouillé des valeurs subjectives de durée ou d'achèvement qu'expriment présent et parfait, et il est mis sur le même plan que le futur qui, on le sait, est dépourvu d'aspect. J. HUMBERT, *Syntaxe gr.*, Paris, Klincksieck, 1945, p. 120. (...) [En fr.], p. ext., auj. inus. Le passé simple (cf. Ac. 1835, LITTRÉ, Ac. 1878, DG ; BESCH. 1845, Lar. 19e et Nouv. Lar. ill. enregistrent également ce sens mais en attribuant l'usage aux „anciens grammairiens français“).

On retiendra que l'aoriste est "dépouillé des valeurs subjectives de durée ou d'achèvement" et qu'il est, tout comme le futur, "dépourvu d'aspect". Il convient de nous demander pourquoi notre usage aujourd'hui du terme aoriste pour le passé simple se veut héritier des "anciens grammairiens français". Il n'est plus appelé aoriste en langue courante, nous montrons en section 6.3.3 que le passé simple a gardé jusqu'aujourd'hui la sémantique qui lui a été attachée en français à partir du XVIIIème siècle.

Plus précisément, nous citons la définition donnée par Caudal et Veters dans [Caudal et Veters, 2007] :

Le terme d'aoriste désigne ici un temps verbal permettant l'inscription d'avènements dans une succession temporelle en rupture avec le plan du présent. Ce terme recouvre par exemple le *passé simple* en français et les usages non-résultatifs du *simple past*.

En d'autres termes, un aoriste est illustré par l'emploi du passé simple en français dans :

(6.1) Azaïs quitta Cautelets.

tandis qu'en anglais, l'exemple donné par Caudal dans [Caudal, 2006b] concerne *covered* dans :

(6.2) We reached Dorset at last. A thick blanket of snow covered the fields.

que nous traduirons ainsi :

(6.3) Nous atteignîmes enfin le Dorset. Une couche épaisse de neige couvrit les champs.

Dans cet exemple l'usage du *simple past* ou *preterit* est aoriste si la neige a commencé à tomber lorsque nous sommes arrivés dans le Dorset, qui se traduit par l'usage du passé simple en français. En revanche, une autre acception du *simple past* dont l'usage est résultatif force une interprétation selon laquelle lorsque nous arrivâmes, nous avons découvert les champs recouverts de neige.

Pour éclaircir la définition du Trésor de la Langue Française Informatisé, il nous faut revenir sur ce qu'est l'aspect. Il peut être lexical et permettre de classer les éventualités dans une typologie, cette typologie distinguant les états des activités ou encore des événements tel que nous l'avons montré dans les chapitres 3 et 4. Dans le cas de l'aoriste, il s'agit de la seconde forme d'aspect reconnue, l'aspect grammatical puisque ce n'est pas le type d'éventualité qui est questionné mais bien sa réalisation en contexte par le temps verbal utilisé. Les notions centrales de l'expression aspectuelle par les temps verbaux sont la perfectivité et l'imperfectivité, qu'il faut bien distinguer de *perfect* et *imperfect* chez Verkuyl², nous citons Caudal pour préciser ce qu'est l'aspect grammatical [Caudal, 2006a] :

2. Parfait et imparfait en français, mais nous gardons délibérément la terminologie anglaise, les concepts ayant des réalisations différentes d'une langue à une autre.

Ils [les termes perfectivité et imperfectivité] servent en effet à caractériser le contenu aspectuel des temps verbaux décrivant des éventualités de manière transitionnelle ou non, c'est-à-dire en rapport à la présence ou à l'absence d'un changement d'état. Un changement d'état est ordinairement conçu comme impliquant l'altération, la destruction ou la création d'un ou plusieurs entités objet du modèle (cf. la notion d' "affectation" (affectedness) de Roberts 1985 [Roberts, 1985]).

L'aspect grammatical de l'aoriste "indéterminé" s'oppose à l'"achevé", et le "duratif" dans la première définition, le passé simple a pour spécificité son aspect, ou plutôt son absence d'aspect. En effet, il peut être considéré comme une fonction objectivante, si l'on peut s'exprimer ainsi, permettant de présenter l'éventualité comme un bloc, sans phase préparatoire, interne ou encore consécutive. Notre acception de la sémantique de l'aspect grammatical pour les temps du français s'oppose à la conception de Verkuyl : l'opposition *PERF/IMP* telle qu'elle est utilisée dans le système constitue de notre point de vue une part aspectuelle dans la construction temporelle du temps verbal. Ainsi l'angle avec lequel on regarde l'éventualité, ou la fenêtre à travers laquelle on la regarde est selon nous une partie de ce qui compose l'aspect. Nous montrons dans les développements notre définition de l'aspect qui justifie une telle position.

Concernant l'énonciation maintenant, Benveniste donne une approche pertinente de l'aoriste [Benveniste, 1966], il le définit comme "décroché de la situation d'énonciation, propre au discours historique", Culioli parlerait ici de rupture énonciative, évoquée par Caudal par une "rupture avec le plan du présent". Aoristes et perfectifs sont, d'après lui en distribution complémentaire sur le plan sémantico-pragmatique : dans le récit historique on trouvera l'aoriste, l'imparfait, le plus que parfait et le prospectif (en forme de 3ème personne) et dans le discours on trouvera toutes les autres formes exceptés l'aoriste simple et composé. Il nous faut pourtant nous défendre d'une mauvaise interprétation de ce que nous posons comme étant l'aoriste dans notre système. Une erreur serait de vouloir construire la temporalité de l'éventualité aoriste comme hors du temps, ou encore sur un autre référentiel que celui construit par l'acte d'énonciation. Nous allons montrer comment ré-introduire l'aoriste dans le tableau des combinaisons et pourquoi il est légitime de le situer par rapport à l'acte d'énonciation porté par la constante *n* chez Verkuyl.

6.3 Rappel du système *Binary Tense* et évaluation pour le français

6.3.1 *Binary Tense*, l'aspect et l'énonciation

Tout d'abord le système est construit sur le concept d'opposition, les opérateurs proposés permettent de mettre en évidence de manière binaire un trait particulier à chaque niveaux :

Nous rappelons brièvement la sémantique des différents opérateurs. Nous avons renommé les variables d'intervalles pour une lecture plus fluide des différents opérateurs :

6.3. RAPPEL DU SYSTÈME BINARY TENSE ET ÉVALUATION POUR LE FRANÇAIS

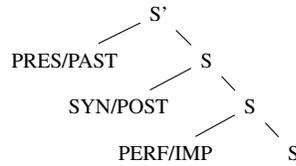


FIGURE 6.1 – Composition des opérateurs temporels, dans le système *Binary Tense*

- i , en relation avec n ou n' et avec j , et qui permet de distinguer par son positionnement si le récit appartient à une temporalité du présent ou du passé sera désormais nommé p pour position, ou encore présent/passé.
- j , originellement en relation avec i et k détermine un point de vue ou encore une fenêtre à travers laquelle on expose l'éventualité, nous la nommons f pour fenêtre ou *frame*.
- k présente l'éventualité et son positionnement dans le cadre précédemment introduit, nous la nommons e .

Nous rappelons la sémantique des différents opérateurs en donnant quelques exemples non exhaustifs d'interprétation :

- $\text{PERF} =_{def} \lambda \phi \lambda f \exists e [\phi[e] \wedge (e) < (f)]$
- $\text{IMP} =_{def} \lambda \phi \lambda f \exists e [\phi[e] \wedge (e) \preceq (f)]$

Ici, on pose la relation entre l'intervalle pendant lequel le voyageur se déplace (e) et l'intervalle (f), vu comme une fenêtre depuis laquelle on regarde le déplacement. Plus précisément, la propriété avancée est la relation entre un nouvel intervalle (f) et l'extension temporelle de l'éventualité sélectionnée dans ϕ , comme illustré dans les figures 6.2 et 6.3 ci-après.

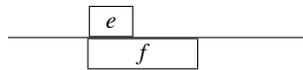


FIGURE 6.2 – Interprétation possible pour PERF

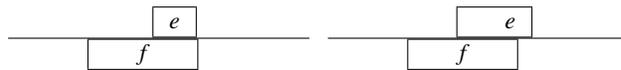


FIGURE 6.3 – Interprétations possibles pour IMP

- SYN =_{def} $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) = (p)]$
- POST =_{def} $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) \preceq \diamond(p)]$

Ici, une fois l'intervalle propre au déplacement situé dans la fenêtre, il faut situer la fenêtre elle-même.

Autrement dit, on pose la relation d'égalité entre l'intervalle-fenêtre positionné par PERF ou IMP (f) et l'intervalle (p) permettant de se positionner par rapport à la constante n comme présenté dans les figures 6.4 et 6.5 ci-après.

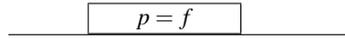


FIGURE 6.4 – Interprétation possible pour SYN

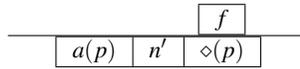


FIGURE 6.5 – Interprétation possible pour POST

- PRES =_{def} $\lambda\phi\exists p[\phi[p] \wedge (p) \circ n]$
- PAST =_{def} $\lambda\phi\exists p[\phi[p] \wedge (p) < n]$

On donne un référent de discours à l'éventualité p . Le dernier intervalle positionné par POST ou SYN (p) est mis en relation avec la constante n , typiquement, nous sommes dans un temps du présent. Ici n est inclus dans l'intervalle p comme présenté dans les figures 6.6 et 6.7 ci-après.

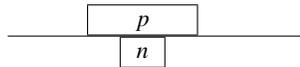


FIGURE 6.6 – Interprétation possible pour PRES

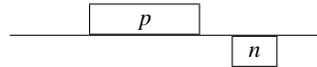


FIGURE 6.7 – Interprétation possible pour PAST

Nous nous intéressons maintenant plus en détail à l'opposition PERF/IMP. Ces deux opérateurs sont réalisés respectivement par les relations $<$ et \preceq entre deux intervalles e et f :

6.3. RAPPEL DU SYSTÈME BINARY TENSE ET ÉVALUATION POUR LE FRANÇAIS

- $e < f$

Selon Verkuyl, $<$ doit être interprétée comme la relation que l'on donnerait entre les entiers naturels 2 et 3, si on les considère comme des ensembles, soit pour 2 on a l'ensemble $\{0,1,2\}$ et pour 3 on a l'ensemble $\{0,1,2,3\}$. 2 est à la fois précédent à 3 et inclus dans 3. Sa traduction dans les relations entre intervalles de Allen est illustrée en

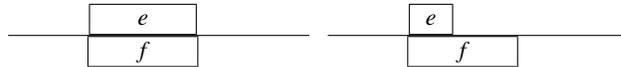


FIGURE 6.8 – Interprétations de la relation $e < f$

figure 6.9 ci-après.

Relations	Interprétations
$x < y$	
$x m y$	
$x o y$	
$x d y$	
$x s y$	
$x f y$	
$x = y$	
$x > y$	
$x mi y$	
$x oi y$	
$x di y$	
$x si y$	
$x fi y$	

FIGURE 6.9 – Les relations de Allen et $<$

- $e \preceq f$

Selon Verkuyl, la relation \preceq est un dual à la relation \prec , elle nous sert à exprimer l'absence d'information sur le fait que e se termine avant f ou non. Autrement dit, ce que nous devons retenir des formes *perfect* et *imperfect* c'est qu'elle nous permet de définir si la fin de l'éventualité est visible ou non en fonction de l'angle avec lequel on perçoit cette éventualité. Outre le fait que les temps ayant le trait PERF sont des formes composées et que les temps portant le trait IMP sont réalisés par des formes synthétiques, une distinction sémantique est en jeu concernant la partie de l'éventualité mise en lumière par cette opposition (indépendamment de l'aspect lexical). Nous retrouvons pour les temps utilisant PERF un changement d'état, voir une mise en lumière d'un état résultant, ce qui est radicalement opposé à l'apport d'IMP dans la composition temporelle, qui au contraire ne met en lumière qu'une phase interne de l'éventualité sans spécifier son issue.

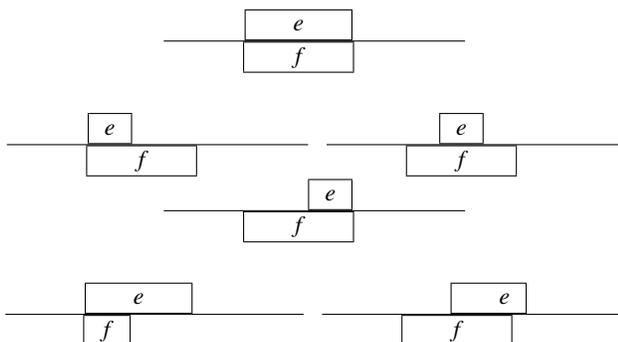


FIGURE 6.10 – Interprétations de la relation $e \preceq f$

Sa traduction dans les relations entre intervalles de Allen sont illustrées en figure 6.11 ci-après.

Pour préciser notre pensée, observons ces deux exemples construits :

(6.4) Elle quittera Caunterets à seize heures, je pourrai donc partir.

(6.5) Elle aura quitté Caunterets à seize heures, je pourrai donc partir.

Les opérateurs temporels pour *elle quittera Caunterets* et *elle aura quitté Caunterets* sont respectivement (PRES POST IMP) et (PRES POST PERF). Ces deux exemples sont distincts par la localisation temporelle de l'éventualité que l'on peut traduire par la considération de l'existence d'un changement d'état ou non succédant à l'éventualité. *Je pourrai donc partir* se situe effectivement à *seize heures*, et toutes deux sont des unités temporelles, équivalentes dans les deux exemples, par contre les formes *elle quittera* et *elle aura quitté*, seule la "phase interne" ou "coeur" l'éventualité est considérée et montrée comme pertinente temporellement tandis que pour le second, un second "objet" succède à l'éventualité et sert de support à l'éventualité suivante, c'est l'état conséquent qui est signifiant dans la sémantique temporelle de l'énoncé. Cette

6.3. RAPPEL DU SYSTÈME BINARY TENSE ET ÉVALUATION POUR LE FRANÇAIS

Relations	Interprétations
$x < y$	
$x m y$	
$x o y$	
$x d y$	
$x s y$	
$x f y$	
$x = y$	
$x > y$	
$x mi y$	
$x oi y$	
$x di y$	
$x si y$	
$x fi y$	

FIGURE 6.11 – Les relations de Allen et \preceq

opposition correspond à notre définition de l'aspect et nous défendons une conception aspectuo-temporelle de la représentation sémantique via *Binary Tense*.

L'opposition entre PRES et PAST constitue une relation de nature énonciative selon nous. L'éventualité est positionnée par rapport au moment d'énonciation selon ces deux relations que sont le chevauchement (\circ) ou l'antériorité ($<$). D'un point de vue énonciativiste, même s'il n'y a pas de rupture dans la prise en charge de l'énoncé³, on peut dire qu'il y a une mise à distance de l'éventualité posée par l'énonciateur quant à l'objet qu'il évoque dans l'utilisation de PAST. C'est d'ailleurs tout l'enjeu d'un *n'* utile à évoquer le "présent de l'éventualité", proposé par Verkuyl pour pouvoir obtenir deux sous-extensions saisies par la fonction *a* (initiale) et \diamond (finale).

Prenant en compte les deux dimensions citées, aspectuelle et énonciative, nous

3. Comme on l'observe dans le cas du discours indirecte par exemple, où les propos d'une personne seconde sont énoncés par l'énonciateur sans qu'il ne les prenne en charge ou n'accepte de les assumer.

définissons notre conception du calcul temporel comme mettant en jeu deux interfaces, l'une énonciativo-temporelle et l'autre aspectuo-temporelle. Nous donnons à titre d'exemple l'imparfait :

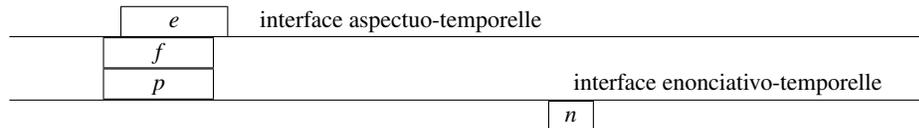


FIGURE 6.12 – Interprétation possible des variables temporelles pour l'imparfait soit (PAST)(SYN)(IMP).

On peut décomposer les trois relations entre indices en deux articulations de nature différente pour chaque temps conjugué, la première, ici portée par l'opérateur PAST pour l'imparfait permet de mettre en relation un premier indice p avec le moment d'énonciation n , cette articulation est d'ordre *énonciativo-temporelle*. La seconde qui nous intéresse ici, place l'éventualité dans une perspective, l'objet (l'indice e) par rapport au regard (l'indice f), porté ici par l'opérateur IMP, sera quand à elle une interface dite *aspectuo-temporelle*. On utilisera alors le système comme un calcul de la temporalité basé sur les articulations entre l'énonciation et l'aspect.

Ayant abordé d'un point de vue conceptuel notre appropriation du système, nous abordons maintenant une évaluation de ce système pour le français.

6.3.2 Évaluation du système pour le français contemporain

Observons de plus près l'opérateur PERF en français avec un exemple du corpus, extrait de *Voyage aux Pyrénées* de George Sand :

(6.6) Dans dix minutes, j'aurai quitté Nohant.

Selon le lexique issu de *Binary Tense*, la représentation temporelle de cette éventualité suit donc l'ordre des applications suivant :

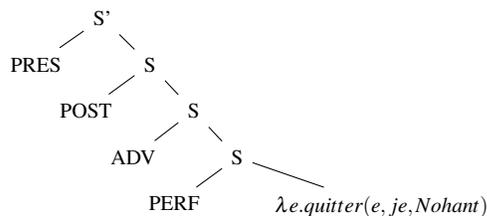


FIGURE 6.13 – Structure sémantique de l'analyse temporelle dans l'exemple 6.6

Le terme suivant

$$(PRES)(POST)(ADV)(PERF)(\lambda e.quitter(e, je, Nohant))$$

se réduit donc ainsi :

$$\exists p \exists f \exists e. \text{quitter}(e, je, \text{Nohant}) \wedge (e) \prec (f) \wedge (f) \prec \diamond(p) \wedge (p) \circ n \wedge \text{distance}_{\min}(f, n) = 10$$

Ce qui doit être interprété de la manière suivante :

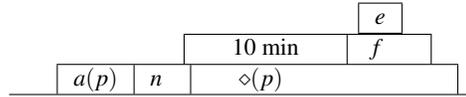


FIGURE 6.14 – Interprétation de *Dans dix minutes, j'aurai quitté Nohant.* dans le système original *Binary Tense*

L'extension temporelle de l'évènement *quitter* démarre donc *dix minutes* après le moment d'énonciation n , pourtant, cet évènement devrait avoir lieu avant le terme des *10 min*, la relation \prec n'est pas appropriée selon nous, ici il faut une relation d'antériorité stricte entre e et f , relation originellement proposée dans une version antérieure du système [Verkuyl, 2003] :

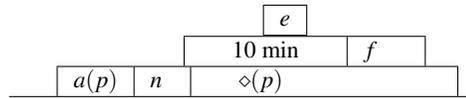


FIGURE 6.15 – Interprétation de *Dans dix minutes, j'aurai quitté Nohant.* attendue

La position de l'application de l'adverbe dans *Binary Tense* peut changer, tel que nous l'avons montré dans le chapitre 4. Dans cet exemple, on peut se demander si ce n'est pas justement la position de l'adverbe qui influence une mauvaise interprétation de la temporalité de l'éventualité. Ici, *dans dix minutes* pourrait être appliqué de la manière suivante dans le système :

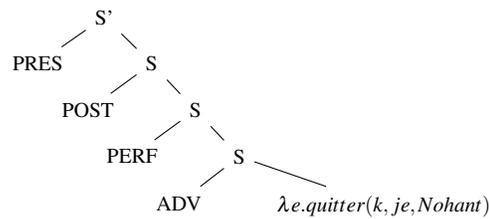


FIGURE 6.16 – Position possible de l'adverbe temporel dans le système *Binary Tense* pour l'exemple 6.6

Le terme suivant

$$(\text{PRES})(\text{POST})(\text{PERF})(\text{ADV})(\lambda e. \text{quitter}(e, je, \text{Nohant}))$$

se réduit donc ainsi :

$$\exists p \exists f \exists e. \text{quitter}(e, je, \text{Nohant}) \wedge (e) \prec (f) \wedge (f) \prec \diamond(p) \wedge (p) \circ n \wedge \text{distance}_{\min}(e, n) = 10$$

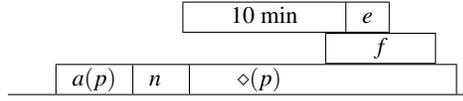


FIGURE 6.17 – Interprétation de *Dans dix minutes, j'aurai quitté Nohant.* dans le système original *Binary Tense*

Le problème n'est pas résolu et l'issue de l'extension temporelle de l'éventualité réifiée par le verbe *quitter* dans l'exemple 6.6 n'est toujours pas atteinte une fois les dix minutes écoulées. C'est pourquoi nous recommandons l'antériorité stricte. Afin d'isoler l'effet de PERF, nous changeons le temps de l'exemple pour le plus-que-parfait :

(6.7) Dix minutes plus tard, j'avais quitté Nohant.

Le terme suivant

$$(PAST)(SYN)(ADV)(PERF)(\lambda e.quitter(e, je, Nohant))$$

se réduit donc tel que :

$$\exists p \exists f \exists e.quitter(e, je, Nohant) \wedge (e) \prec (f) \wedge (f) = (p) \wedge (p) < n \wedge \mathbf{distance}_{\min}(\mathbf{e}, \mathbf{c}) = \mathbf{10}$$

ou encore le terme

$$(PAST)(SYN)(PERF)(ADV)(\lambda e.quitter(e, je, Nohant))$$

se réduit en

$$\exists p \exists f \exists e.quitter(e, je, Nohant) \wedge (e) \prec (f) \wedge (f) = (p) \wedge (p) < n \wedge \mathbf{distance}_{\min}(\mathbf{f}, \mathbf{c}) = \mathbf{10}$$

Ce qui peut être interprété de la manière suivante⁴ :

4. c étant un repère déjà présent dans le contexte et par rapport auquel l'adverbial temporel est relatif. Par exemple on peut imaginer que la phrase précédente soit : *A six heures je faisais mon sac. Dix minutes plus tard, j'avais quitté Nohant.* Ici, c serait l'intervalle dédié à à six heures.

6.3. RAPPEL DU SYSTÈME BINARY TENSE ET ÉVALUATION POUR LE FRANÇAIS

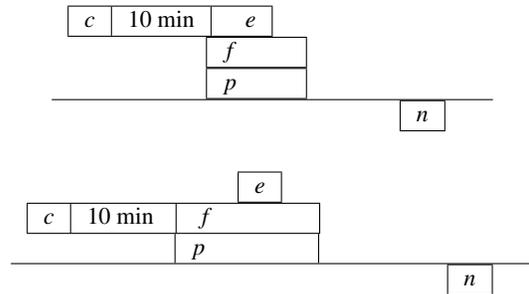


FIGURE 6.18 – Interprétations selon les deux positions possibles de l’adverbe de *Dix minutes plus tard*, *j’avais quitté Nohant*. dans le système original *Binary Tense*

Une fois de plus, la borne finale de l’extension temporelle de l’éventualité e doit se placer avant la fin des *10 min*, mais ce n’est pas le cas, nous recommandons donc l’infériorité stricte comme relation pour PERF.

En ce qui concerne la composition avec l’opérateur POST, on obtient un futur dans le passé dont l’aspect serait aoriste, on peut interroger l’interprétation de cet opérateur, dans les temps comportant l’opérateur PRES, on obtient alors les formes de futur *perfect* et *imperfect*, tandis que dans les formes du passé, c’est à dire dans les formes comportant l’opérateur PAST, on obtient en français des conditionnels. Observons les exemples suivant :

(6.8) Azaïs quitterait Cauterets (si Jean y restait).

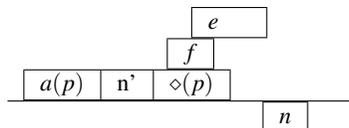


FIGURE 6.19 – Interprétations possibles des variables temporelles pour le conditionnel présent soit (PAST)(POST)(IMP).

(6.9) Azaïs aurait quitté Cauterets (si Jean y était resté).

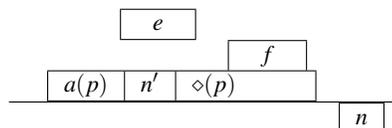


FIGURE 6.20 – Interprétation possible des variables temporelles pour le conditionnel passé, soit (PAST)(POST)(PERF).

Les deux interprétations sont justes si l'on interprète e étendue largement en sa borne finale dans le cas du conditionnel présent 6.19. On peut comprendre ce qui pousserait un angliciste et a fortiori un néerlandiste à créer un opérateur POST pour les formes verbales utilisant *will*, puis à considérer *would* comme sa forme passée (on aura donc respectivement PAST POST IMP et PAST POST PERF) et malgré une étymologie différente, on observe que les deux interprétations sont correctes. Nous évaluons donc l'opérateur POST lorsqu'il est combiné à PAST comme pertinent pour traiter les conditionnels français.

Il nous faut signaler un point sémantique important qui distingue les conditionnels et les temps de l'indicatif. Dans le cas des conditionnels, la sémantique du déplacement énoncé a bien une extension temporelle réelle que nous désirons représenter, même si le mode utilisé se situe dans le domaine de l'irréel. Dans une partie des cas l'utilisation du conditionnel implique la négation du déplacement en question. Ainsi dans *Azaïs quitterait Causerets si Jean y restait*, on peut en déduire l'information selon laquelle Azaïs n'a pas effectivement quitter Causerets au moment où est prononcée cette phrase. Dans la phrase suivante :

(6.10) Azaïs aurait quitté Causerets à huit heures si Jean y était resté. (Supposant qu'en réalité Azaïs a quitté Causerets à sept heures)

On ne peut inférer qu'Azaïs est toujours à Causerets, passé huit heures. L'implication suivante est donc juste si l'on représente le fait que le voyageur x est localisé en la région y dans l'intervalle z , on utilisera le prédicat $localise(z,x,y)$ et pour simplifier, nous écrivons $8 : 00$ pour l'intervalle de temps représentant *huit heures*. Cet intervalle est en relation meets avec z pour "*passé huit heures*".

D'après $(PAST)(POST)(ADV)(PERF)(quitter(e,x,y) \wedge nomme(x, Causerets) \wedge nomme(y, Azaïs))$ on peut déduire ceci :

$$\begin{aligned} \exists p \exists f \exists e \exists x \exists y. quitter(e, Causerets, Azaïs) \wedge (e < f) \wedge (f \preceq \diamond p) \wedge (p < n) \wedge (f \preceq 8heures) \\ \implies \\ \exists z. 8 : 00 \text{ meets } z \wedge \neg localise(z, x, y) \end{aligned}$$

en revanche, on ne peut rien déduire sur la localisation de Azaïs d'après la proposition suivante :

$$\begin{aligned} \exists p \exists f \exists e \exists x \exists y. \neg (quitter(e, Causerets, Azaïs) \wedge (e < f) \wedge (f \preceq \diamond p) \wedge (p < n) \wedge (f \preceq 8heures)) \\ \not\Rightarrow \\ \exists z. \neg (8 : 00 \text{ meets } z \wedge localise(z, x, y)) \end{aligned}$$

Remarque sur la légitimité à représenter des éventualités qui n'auront ou n'ont pas eu lieu mais dont la temporalité nécessite une représentation dans l'univers du discours :

Nous rappelons ici que nos développements ne se préoccupent pas du mode mais bien du calcul de la temporalité d'une éventualité. L'actualisation de cette éventualité

n'étant pas abordée dans ce système, on remarquera simplement qu'une éventualité exprimée au futur ou dans un temps de l'indicatif et sous la portée d'une négation peut tout aussi bien mobiliser une extension temporelle sans que l'actualisation ou non de l'éventualité n'intervienne dans la représentation sémantique temporelle :

(6.11) Azaïs escladera le versant nord du Vignemale demain.

(6.12) Azaïs n'escaldera pas le versant nord du Vignemale demain.

Azaïs escaladera-t-il vraiment demain ? Azaïs escaladera-t-il le versant sud ? Escaladera-t-il le versant nord le jour suivant ? Toujours est-il que dans ces deux exemples, une information est donnée à propos de la relation entre *demain*, *Azaïs* et *Cauterets*, que l'éventualité ait ou non eu lieu, sa sémantique temporelle reste représentable. Quant à la confrontation du système au français, nous retiendrons l'interprétation problématique de PERF et la légitimité des conditionnels dans l'utilisation de *Binary Tense* pour le français. Nous nous intéressons dorénavant à l'objet de ce projet, le récit de voyage du XIX^{ème} siècle.

6.3.3 Le récit de voyage du XIX^{ème} siècle et le français contemporain

Il est reconnu par plusieurs études diachroniques [Treikelder, 2006, Caudal et Vettters, 2007] que le passé composé a revêtu au fil du temps différentes valeurs aspectuelles dont l'aoriste, ce qui est concomitant avec l'effacement progressif du passé simple dans les récits. Le passé composé a doucement recouvert les usages qui étaient relégués au passé simple jusqu'alors. Nous nous appuyerons dans un premier temps sur l'étude diachronique des temps du français réalisées par Patrick Caudal et Carl Vettters [Caudal et Vettters, 2007] afin de montrer à quel moment du glissement sémantico-pragmatique des temps verbaux le français du XIX^{ème} siècle se situe. Nous décrivons de manière cohérente ce que recouvre l'aoriste dans notre tableau des combinaisons, et quelles interprétations il faut lui accorder dans notre modèle, les intervalles de \mathbb{R} .

Caudal et Vettters proposent une analyse sémantique par les usages en contexte du passé composé et du passé simple pour trois étapes de l'évolution du français, l'ancien et le moyen français : du XI^{ème} au XV^{ème} siècle, la période classique du XVI^{ème} au XVII^{ème} siècle et la période post-classique, du XVIII^{ème} au XX^{ème} siècles. Ils utilisent pour décrire les évolutions une structure des évènements en *stages* vus comme des phases, on distingue deux phases particulièrement, la phase interne, *Inner Stage* et la phase résultante, *Result Stage*.

Cette étude présente la sémantique accordée aux trois *points de vue aspectuels* qui permet de "décrire la contribution aspectuelle des temps verbaux, qui consiste pour l'énonciateur à adopter une "focale" particulière pour présenter une éventualité donnée"[Caudal, 2006a], mis en évidence et que l'on peut résumer ainsi :

- Le *point de vue imperfectif* concerne la phase interne de l'éventualité :

Sa sémantique est imperfective

Exemple en français actuel : Jean mangeait

Ici, on remarquera qu'on ne cherche pas à donner d'information sur la fin de cette éventualité, ni ses conséquences, ce qui est traduit par l'opérateur IMP chez Verkuyl.

- Le *point de vue perfectif* concerne le cœur de l'éventualité et un changement d'état :

Sa sémantique est aoriste

Exemple en français actuel : Jean mangea

Ici, on comprendra que l'éventualité est considérée comme une mise à jour de l'état du monde, c'est pourquoi on voudrait restreindre les positionnements de la borne finale de l'éventualité avec Anc. L'antériorité stricte exprimée par cet opérateur permet de dire que la borne finale de l'éventualité est située avant la borne initiale du repère α_c .

- Le *point de vue résultatif* concerne l'état résultant :

Sa sémantique est perfective

Exemple en français actuel : Jean a mangé (maintenant)

Ici, on remarque la volonté de marquer l'antériorité ou l'égalité de la borne finale de l'éventualité par rapport à la borne finale de la fenêtre-point de vue pour PERF.

Pour résumer les évolutions du passé composé et du passé simple du XI^{ème} siècle au début du XX^{ème} siècle, nous présentons un tableau réalisé par nos soins résumant très brièvement en trois étapes les conclusions de [Caudal et Veters, 2007]. On ne donne ici aucune information sur l'interface entre sémantique et pragmatique, nous ne montrons que les conclusions sur les *points de vue* et leur **sémantique** entre parenthèses adoptés en fonction des contextes dans lesquels on trouve ces temps. Tout d'abord l'évolution du passé composé puis le passé simple :

6.3. RAPPEL DU SYSTÈME BINARY TENSE ET ÉVALUATION POUR LE FRANÇAIS

	Passé composé	Passé simple	partage des aspects	partage du co(n)-texte
ancien et moyen français XI - XV Sémantique :	<i>résultatif</i> perfective	-quelques <i>imperfectifs</i> et quelques <i>résultatifs</i> imperfective perfective	<i>resultatif</i> perfective	oui
période classique XVI - XVII Sémantique :	- <i>résultatif</i> et quelques <i>perfectifs</i> -lien entre la situation d'énonciation et l'éventualité peut être flou perfective aoriste	- <i>imperfectif</i> , et preterits perdus -lien occasionnel entre la situation d'énonciation et l'éventualité imperfective	aucun	oui avec <i>depuis</i>
post-classique XVIII - XX Sémantique :	- <i>résultatif</i> - <i>perfectif</i> perfective aoriste	- <i>perfectif</i> aoriste	<i>perfectif</i> aoriste	non soit l'un soit l'autre

TABLE 6.1 – Le passé simple et le passé composé en diachronie selon [Caudal et Vettters, 2007] (en italique : les points de vue aspectuels, et en gras : la sémantique)

Le **passé simple** et le **passé composé** en *ancien et moyen français* partageaient le même point de vue *résultatif* dont la sémantique est donc **perfective**.

Durant la *période classique*, tous deux sont en distribution complémentaire, se partageant les points de vue aspectuels *résultatif* et *aoriste* pour le **passé composé**, c'est à dire une sémantique **perfective** et **aoriste** ; et le point de vue aspectuel *imperfectif* pour le **passé simple**, c'est à dire une sémantique **imperfective**.

Cette distribution complémentaire s'effondre à nouveau à partir du XVIIIème siècle où tous deux revêtent un point de vue et une sémantique **aoristes**.

En regard de cette étude, on peut tirer quelques conclusion de *Binary Tense* quant aux opérateurs PERF et IMP. Nous présentons les correspondances entre points de vue aspectuels et sémantique chez Caudal et Vettters et opérateurs temporels chez Verkuyl pour le passé composé et le passé simple :

Temps concernés	Points de vue aspectuels	Sémantique	Opérateurs
imparfait	Imperfectif	Imperfective	(PAST)(SYN)(IMP)
passé composé	Résultatif Perfectif	Perfective Aoriste	(PRES)(SYN)(PERF) ∅
passé simple	Perfectif	Aoriste	(ANC)

TABLE 6.2 – Imparfait, passé composé et passé simple en français (période post-classique) : Points de vue aspectuels et sémantiques selon Caudal et Veters 2007, et opérateurs temporels chez Verkuyl 2008

Cette correspondance fait clairement apparaître l'absence d'une représentation satisfaisante pour le second usage du passé composé, lorsque sa sémantique est aoriste, c'est pourquoi nous proposons maintenant une adaptation du système pour les besoins du français de la période post-classique.

6.4 Adaptation du système : l'aoriste

6.4.1 Aoriste et énonciation

L'exemple donné dans [Verkuyl, 2008] pour justifier une sémantique de l'aoriste ne mettant pas en jeu la constante n pour le moment d'énonciation est le suivant :

(6.13) *Je vois que Pierre abattit le vieux chêne.

Il est vrai que cette phrase semble incorrecte du point de vue énonciatif, néanmoins, ce n'est pas tant la concordance avec le présent qui bloque cette composition mais plutôt le contenu sémantique du verbe *voir*. Ce verbe force en effet une simultanéité entre sa propre éventualité et la chose vue, que ce soit l'état d'un individu ou bien l'activité ou processus mettant en scène des individus. De la même manière on a :

(6.14) *Je vois que Pierre abattait le vieux chêne.

En revanche un autre verbe au présent peut très bien introduire une proposition subordonnée contenant un passé simple :

(6.15) Je crois que nous y réussîmes car jamais nous ne nous étions sentis plus frais et plus dispos, et c'est dans les meilleures conditions que, le 11 août 1838, à 8 heures du matin, nous nous présentions au poteau du départ.

Dès lors, il est bien question de plans énonciatifs différents, ici deux récits se distinguent, le premier donne accès à l'analyse du protagoniste quant à ses exploits passés, qui constituent un second plan énonciatif. Le verbe *croire* introduit une forme de modalité que nous ne traitons pas dans le système, néanmoins, les deux plans sont fortement liés dans l'énoncé (proposition subordonnée) et doivent être localisés temporellement l'un par rapport à l'autre. Dans un calcul temporel n'introduisant aucune forme de modalité, rien de notre point de vue ne retient le passé simple de figurer parmi

6.4. ADAPTATION DU SYSTÈME : L'AORISTE

les autres temps et d'obtenir dans la composition de sa sémantique une relation entre l'éventualité et le moment d'énonciation.

En revanche, si cela ne concerne un changement de référentiel, nous remarquons une distance sur le même référentiel imposée par l'aoriste au niveau énonciatif, entre l'éventualité et le moment d'énonciation. La conséquence d'une telle considération est l'impossibilité d'une quelconque intersection entre notre constante n et l'intervalle k réservé à l'éventualité elle-même. En d'autres termes l'opérateur AOR doit être autorisé pour les temps dont la sémantique contient la relation PAST mais ne peut pas être utilisé avec un temps contenant l'opérateur PRES.

Si l'on introduit alors l'aoriste au même niveau que PERF et IMP on obtient alors ce tableau :

		PRES	PAST
IMP	SYN	elle écrit	elle écrivait
	POST	elle écrira	elle écrirait
PERF	SYN	elle a écrit	elle avait écrit
	POST	elle aura écrit	elle aurait écrit
AOR	SYN	∅	elle écrivit
	POST	∅	?

Qu'en est-il de PAST POST AOR ? Si l'on observe précisément les autres temps formés avec PAST et POST, on obtient des conditionnels comme montré précédemment et le conditionnel aoriste ne trouve pas d'expression en français.

Afin de donner une représentation sémantique du passé composée recouvrant aussi son usage aoriste, on donne le tableau suivant :

Temps verbal	Sémantique	Opérateurs	Exemple
Passé composé	perfective aoriste	PRES(PERF)	(A l'heure qu'il est,) Azaïs a quitté Cauterets.
		PAST(AOR)	(A seize heures quinze,) Azaïs a quitté Cauterets.
Passé simple	aoriste	PAST(AOR)	Azaïs quitta Cauterets.
Imparfait	imperfective	PAST(IMP)	Azaïs quittait Cauterets.

TABLE 6.3 – Point de vue aspectuel et sémantique selon Caudal et Veters 2007 et représentation sémantique du passé composé, passé simple et imparfait au XIX^{ème} siècle

Afin de donner une sémantique appropriée à l'opérateur AOR, rappelons l'opérateur proposé par Verkuyl. Il parle d'ancrage ("*anchorage*") pour aborder l'aoriste, avec l'opérateur $Anc =_{def} \lambda \phi \exists e. [\phi[e] \wedge e < \alpha_c]$. Cet opérateur ne nous semble pas approprié car l'ancrage ne doit pas forcer une relation d'antériorité par essence qui serait plutôt la relation imposée par le *Perfect* tel que nous l'avons discuté jusqu'ici. La volonté de mettre en relation une éventualité, plutôt son extension temporelle, avec un indice

plutôt qu'avec la constante n semble motivée par une volonté de décrocher l'aoriste du moment d'énonciation plus que par une volonté de le raccrocher à quelque chose d'autre. Mais comme nous l'avons défini plus tôt, l'aoriste pour nous ne saurait être purement énonciatif, il est tout d'abord un aspect absolu ou encore une absence d'aspect, c'est pourquoi nous discutons ce sujet dans la section suivante.

6.4.2 Aoriste et aspect

C'est en premier lieu à l'interface aspectuo-temporelle que l'on doit trouver une spécificité pour l'aoriste, cette spécificité ayant effectivement une incidence sur le rapport au moment d'énonciation. L'aspect est une information sur la manière dont on regarde une éventualité, il est légitime de projeter temporellement un lien entre éventualité et moment d'énonciation pour un temps "décroché". Nous rappelons ici que notre visée est une représentation sémantique temporelle avec le moins de modalité possible, aucun opérateur de rupture énonciative ne trouverait donc sa place, a contrario, l'aoriste trouve place au sein du système.

Comme l'a argumenté Gosselin [Gosselin, 2005] l'aoriste est avant tout une manière compacte de voir l'éventualité, elle est envisagée comme absolue, "globale" sans information sur la fin de cette éventualité. L'aspect global chez Gosselin sera aussi appelé "aoristique", il impose une coïncidence exacte entre l'intervalle de l'éventualité et l'intervalle "fenêtre" (f dans nos calculs). Nous choisissons ici de garder la synchronicité, l'égalité entre intervalles, comme interprétation privilégiée de l'aoriste, qui devient dès lors une valeur aspectuelle à part entière.

Dans les faits Verkuyl a désiré construire un système compositionnel et biunivoque dans lequel pour chaque temps verbal, une combinaison correspond. Or on sait qu'en français l'aoriste est une valeur aussi portée par le passé composé, c'est pourquoi la relation \preceq a été utilisée, c'est pour inclure dans l'interprétation de PERF, une dimension aoriste. Chose dont nous avons besoin, à moins que l'on réduise l'interprétation de \preceq . Ainsi nous avons deux possibilités d'étendre le système afin de couvrir les temps verbaux du XIXème siècle en français. Soit on donne une relation qui est incluse dans les deux autres, soit on restreint toutes les interprétations.

6.4.3 L'opérateur AOR et le système

Comme nous l'avons montré précédemment, nous proposons d'ajouter premièrement l'aoriste au niveau PERF/IMP.

La sémantique de chaque opérateur de décline donc ainsi :

IMP : $\lambda \phi \lambda f \exists e [\phi[k] \wedge (e) \preceq (f)]$

PERF : $\lambda \phi \lambda f \exists e [\phi[k] \wedge (e) < (f)]$

AOR : $\lambda \phi \lambda f \exists e [\phi[k] \wedge (e) = (f)]$

La relation de synchronicité = établie entre notre indice j et l'éventualité k , est un tronc commun entre l'imperfect et l'aoriste mais exclut dorénavant le perfect. Les

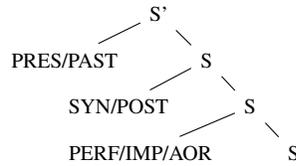


FIGURE 6.21 – Composition des opérateurs temporels, dans notre adaptation

adaptations sémantiques pour les opérateurs PERF et AOR sont les suivantes :

Opérateurs	Relations dans <i>Binary Tense</i>	Propositions de modification
IMP	$e \preceq f$	$e \preceq f$
PERF	$e \prec f$	$e < f$
Anc	$e < \alpha_c$ ⁵	$e = f$

TABLE 6.4 – Répartition des opérateurs selon Verkuyl 2008 et proposition d’adaptation

L’antériorité de e par rapport à α_c est due manifestement au fait que le passé simple est un temps du récit, situant les éventualités dans un passé lointain, sans conséquence sur le présent⁶. Néanmoins, nous avons montré que rien n’interdisait de calculer la valeur sémantique temporelle du passé simple et du passé composé par rapport au moment d’énonciation.

A la lumière de l’étude diachronique de ces deux temps, nous proposons les deux interprétations dans le système de Verkuyl :

Temps verbal	Points de vue	Sémantique	Opérateurs
Passé composé	Résultatif	Perfective	PRES(PERF)
	Perfectif	Aoriste	PAST(AOR)

TABLE 6.5 – Point de vue aspectuel et sémantique selon Caudal, et représentation sémantique du passé composé au XIXème siècle

- (A l’heure qu’il est,) Azaïs a quitté Cauterets.
 (PRES)(ADV)(PERF) ($\lambda e. \text{quitter}(e, \text{Cauterets}, \text{Azaïs})$)
 $(\exists p \exists f \exists e \exists x \exists y. \text{quitter}(e, \text{Cauterets}, \text{Azaïs}) \wedge (e \prec f) \wedge (fRn) \wedge (f = p) \wedge (p \circ n))$

Et son interprétation dans les intervalles de \mathbb{R} :

Ici à l’heure qu’il est modifie les contraintes sur f , ce que nous avons voulu représenter par fRn .

- (A seize heures quinze,) Azaïs a quitté Cauterets.
 (PRES)(PERF)(ADV) ($\lambda e. \text{quitter}(e, \text{Cauterets}, \text{Azaïs})$)

6. Qu’il soit un présent du récit ou un présent d’énonciation.

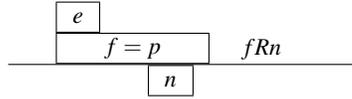


FIGURE 6.22 – Première interprétation possible des variables temporelles pour le passé composé soit (PRES)(SYN)(PERF) dans la version originale de Verkuyl.

$(\exists p \exists f \exists e \exists x \exists y. quitter(e, Caeterets, Azaïs) \wedge (eR16 : 15) \wedge (e \prec f) \wedge (f = p) \wedge (p < n))$
 Et son interprétation dans les intervalles de \mathbb{R} :

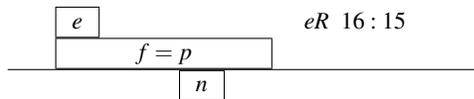


FIGURE 6.23 – Deuxième interprétation possible des variables temporelles pour le passé composé soit (PRES)(SYN)(PERF) dans la version originale de Verkuyl.

Ici à *seize heure quinze* modifie les contraintes sur e , ce que nous avons voulu représenter par $eR 16 : 15$.

Nous obtenons donc les deux représentations en accord avec les analyses linguistiques du passé composé, mais la première interprétation perfective semble peu informative, il faut forcer e à ne pas avoir d'intersection avec n , chose ici plutôt artificielle. De plus, nous proposons que le contexte, lorsqu'il apporte des informations temporelles, affecte l'indice f , ainsi dans le premier exemple si f est inclus dans l'intervalle correspondant "*à l'heure qu'il est*", l'interprétation est mauvaise, tandis que pour la seconde, si f est inclus dans l'extension temporelle de "*à seize heures quinze*", l'interprétation est juste. Le passé composé a donc deux valeurs au XIX^{ème} siècle, valeurs qu'il a gardées jusqu'à nos jours.

Dans notre adaptation du système, les deux usages du passé composé peuvent être discriminés et représentés sémantiquement :

(6.16) (A l'heure qu'il est,) Azaïs a quitté Caeterets.

(PRES)(ADV)(PERF) ($\lambda e. quitter(e, Caeterets, Azaïs)$)

$(\exists p \exists f \exists e \exists x \exists y. quitter(e, Caeterets, Azaïs) \wedge (e < f) \wedge (fRn) \wedge (f = p) \wedge (p \circ n))'$

Et son interprétation dans les intervalles de \mathbb{R} :

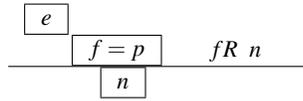


FIGURE 6.24 – Première interprétation possible des variables temporelles pour le passé composé soit (PRES)(SYN)(PERF) dans notre adaptation.

(6.17) (A seize heures quinze,) Azaïs a quitté Cauterets.

(PAST)(ADV)(AOR) ($\lambda e.$ quitter(e ,Cauterets,Azaïs))

$(\exists p \exists f \exists e \exists x \exists y. \text{quitter}(e, \text{Cauterets}, \text{Azaïs}) \wedge (e = f) \wedge (fR\ 16 : 15) \wedge (f = p) \wedge (p < n))$ ^t

Et son interprétation dans les intervalles de \mathbb{R} :



FIGURE 6.25 – Deuxième interprétation possible des variables temporelles pour le passé composé soit (PAST)(SYN)(AOR) dans notre adaptation.

Les deux interprétations sont plus fidèles à leur expression linguistique, le passé composé pouvant avoir valeur d'aoriste dans certains contextes et valeur *perfect* dans d'autres. Le choix de l'interprétation temporelle ne dépend plus de la position à laquelle on applique l'adverbial mais du choix de la sémantique temporelle de l'éventualité exprimée, mettant en jeu l'ensemble du système à savoir l'interface énonciativo-temporelle, avec la relation PRES ou PAST et l'interface aspectuo-temporelle, avec la relation PERF ou AOR.

6.4.4 La relation PERF, la relation AOR et l'enjeu adverbial

Les deux interprétations sont donc maintenant justes avec l'indice j affecté par le contexte temporel, ce qui force la place de l'adverbial temporel à être appliqué en second et non en premier :

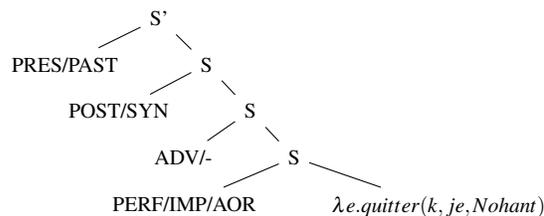


FIGURE 6.26 – Structure sémantique de l'analyse temporelle

Cette structure sémantique ne peut être définitive, et pour attribuer une place fixe à l'adverbial, il convient de se demander si tout type d'adverbial temporel doit être appliqué au même niveau ou bien si la place attribuée à ce modifieur temporelle dépend d'autre chose. Par exemple, *j'avais marché pendant quatre heures* propose une sémantique de *perfect* tout en réclamant une modification adverbiale temporelle sur l'éventualité nue, l'indice e . On ne saurait considérer que c'est le point de vue qui est altéré par le modifieur, mais bien l'éventualité. Nous abordons cette question dans la section suivante et plus particulièrement dans le chapitre suivant.

6.5 Interprétation

Nous donnons maintenant toutes les formes dans le système adapté ainsi que des exemples du corpus illustrant les temps conjugués et leurs différentes valeurs. Pour représenter l'incidence des adverbiaux temporels, les relations ne sont pas spécifiées, simplement R indique une relation et (" ") encadre l'information pseudo-normalisée du contenu sémantique contenu dans l'adverbial. Nous donnons un lexique adverbial précis dans le chapitre suivant, nous cherchons ici à confronter notre adaptation à des exemples du corpus ITIPY.

		PRES
IMP	SYN	nous atteignons, en trente-cinq minutes, le port de Vénasque
	POST	nous partirons le 25 (octobre)
PERF	SYN	(Venez,) j'ai découvert un dortoir
	POST	dans dix minutes, j'aurai quitté Nohant

TABLE 6.6 – Tableau des formes conjuguées du présent et leur interprétation

– Commençons par un exemple au présent :

(6.18) (...) **nous atteignons, en trente-cinq minutes, le port de Vénasque** (...)
 (PRES) (SYN) (ADV) (IMP) (λe . atteindre(e , V, N))

\rightarrow_{β}

$\exists p \exists f \exists e$.atteindre(e , V, N) $\wedge e \preceq f \wedge fR$ "35 min" $\wedge f = p \wedge p \circ n$

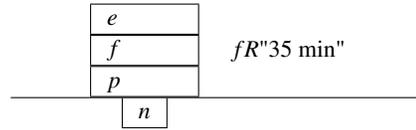


FIGURE 6.27 – Interprétation possible des variables temporelles pour *nous atteignons, en trente-cinq minutes, le port de Vénasque (...)*

– On observe maintenant le futur simple :

(6.19) (...) **nous partirons le 25 (octobre)**(...)

(PRES) (POST) (ADV) (IMP) ($\lambda e. \text{partir}(e, r, N)$)

$\rightarrow\beta$

$\exists p \exists f \exists e. \text{partir}(e, r, N) \wedge e \preceq f \wedge fR''25 \text{ octobre}'' \wedge f \preceq \diamond p \wedge p \circ n$

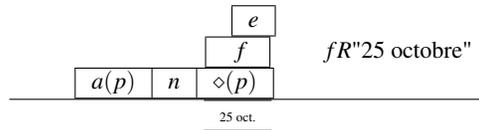


FIGURE 6.28 – Interprétation possible des variables temporelles pour *nous partirons le 25 (octobre)*

– Ensuite, on illustre le passé composé perfectif :

(6.20) (...) - Pourrons-nous, dans cette Sibérie, réparer l'insomnie de la nuit dernière ? - Oui, oui ! Venez, **j'ai découvert un dortoir**. (...)

(PRES) (SYN) (PERF) ($\lambda e. \text{découvrir}(e, D, J)$)

$\rightarrow\beta$

$\exists p \exists f \exists e. \text{découvrir}(e, D, J) \wedge e < f \wedge f = p \wedge p \circ n$

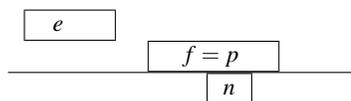


FIGURE 6.29 – Interprétation possible des variables temporelles pour *(Venez,) j'ai découvert un dortoir*

– Et enfin nous montrons l'interprétation du futur antérieur :

(6.21) **Dans dix minutes, j'aurai quitté Nohant.**

(PRES) (POST)(ADV) (PERF) (λe . quitter(e , N, J))

$\rightarrow\beta$

$\exists p\exists f\exists e$.quitter(e , N, J) $\wedge e < f \wedge fR$ "10 minutes" $\wedge f \preceq \diamond p \wedge p \circ n$

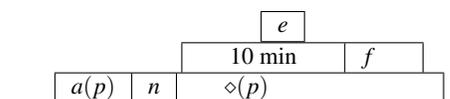


FIGURE 6.30 – Interprétation possible des variables temporelles pour *dans dix minutes, j'aurai quitté Nohant*

Maintenant nous abordons les temps du passé dans le système :

		PAST
IMP	SYN POST	(Enfin) à une heure et quart, je quittais la banque on monterait pendant 2 h 30
PERF	SYN POST	nous avons parcouru également un espace de douze à quinze mille mètres Sur cette confortable terrasse si bien appelée Belle-Vue, nous aurions volontiers prolongé notre séjour (si nous n'avions voulu faire un détour et redescendre vers la Prade Saint-Jean)
AOR	SYN	Nous quittâmes Gavarnie à sept heures moins le quart / je les ai traversées en 1828 à Figuières
AOR	PERF (surcomposé)	Lorsque nous eûmes assisté au repas de l'annélide Carnivore [...] (nous continuâmes d'avancer vers le village de Sarp)

TABLE 6.7 – Tableau des formes conjuguées du passé et leur interprétation

– Nous débutons par l'imparfait :

(6.22) (Enfin) à une heure et quart, je quittais la banque (...)

(PAST) (SYN) (ADV) (IMP) ($\lambda e. \text{quitter}(e, B, J)$)

$\rightarrow\beta$

$\exists p \exists f \exists e. \text{quitter}(k - e, B, J) \wedge e \preceq f \wedge fR"1h15" \wedge f = p \wedge p < n$

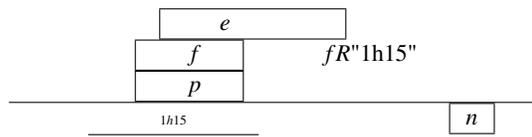


FIGURE 6.31 – Interprétations possibles des variables temporelles pour *(Enfin) à une heure et quart, je quittais la banque*

– Puis arrive le conditionnel présent :

(6.23) (...)on monterait pendant 2 h 30, (...)

(PAST) (POST) (ADV) (IMP) ($\lambda e. \text{monter}(e, r, O)$)

$\rightarrow\beta$

$\exists p \exists f \exists e. \text{monter}(e, r, O) \wedge e \preceq f \wedge fR"2h30" \wedge f \preceq \diamond p \wedge p < n$

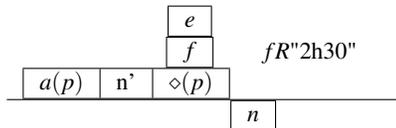


FIGURE 6.32 – Interprétation possible des variables temporelles pour *on monterait pendant 2 h 30*

– Observons maintenant le plus-que-parfait :

(6.24) (...) **nous avons parcouru** (également) **un espace de douze à quinze mille mètres**, (...)

(PAST) (SYN) (PERF) ($\lambda e. \text{parcourir}(e, E, N)$)

$\rightarrow\beta$

$\exists p \exists f \exists e. \text{parcourir}(e, E, N) \wedge e < f \wedge f = p \wedge p < n$

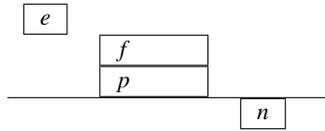


FIGURE 6.33 – Interprétation possible des variables temporelles pour *nous avons parcouru (également) un espace de douze à quinze mille mètres*

– Ensuite on montre le conditionnel passé :

(6.25) Sur cette confortable terrasse si bien appelée Belle-Vue, **nous aurions volontiers prolongé notre séjour** (si nous n’avions voulu faire un détour et redescendre vers la Prade Saint-Jean) (...)

(PAST) (POST) (PERF) ($\lambda e. \text{prolonger}(e, T, N)$)

$\rightarrow\beta$

$\exists p \exists f \exists e. \text{prolonger}(e, T, N) \wedge e < f \wedge f \preceq \diamond p \wedge p < n$

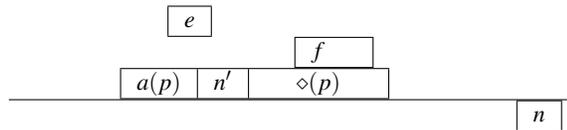


FIGURE 6.34 – Interprétation possible des variables temporelles pour *Sur cette confortable terrasse si bien appelée Belle-Vue, nous aurions volontiers prolongé notre séjour*

– Et voici le passé composé aoriste et son interprétation :

(6.26) (...) **je les ai traversées en 1828** à Figuières (...)

(PAST) (SYN) (ADV) (AOR) ($\lambda e. \text{traverser}(e, I, J)$)

$\rightarrow \beta$

$\exists p \exists f \exists e. \text{traverser}(e, I, J) \wedge e = f \wedge fR"1828" \wedge f = p \wedge p < n$

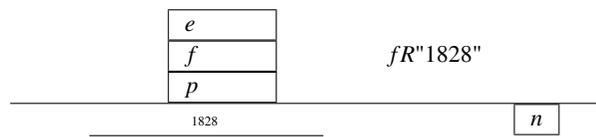


FIGURE 6.35 – Interprétation possible des variables temporelles pour *je les ai traversées en 1828 à Figuières*

– Puis le passé simple et son interprétation :

(6.27) **Nous quittâmes Gavarnie à sept heures moins le quart**,(...)

(PAST) (SYN) (ADV) (AOR) ($\lambda e. \text{quitter}(e, G, N)$)

$\rightarrow \beta$

$\exists p \exists f \exists e. \text{quitter}(e, G, N) \wedge e = f \wedge fR"6h45" \wedge f = p \wedge p < n$

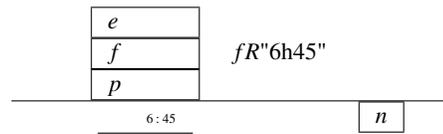


FIGURE 6.36 – Interprétation possible des variables temporelles pour *nous quittâmes Gavarnie à sept heures moins le quart*

– Et nous donnons un cas de surcomposition, le passé antérieur :

(6.28) (Lorsque) **nous eûmes assisté au repas de l’annélide Carnivore**, (...)
(nous continuâmes d’avancer vers le village de Sarp)

(PAST) (SYN) (AOR) (PERF) ($\lambda e. assister(e, R, N)$)

$\rightarrow\beta$

$\exists i \exists j \exists j' \exists k. assister(e, R, N) \wedge e < f' \wedge f' = f \wedge f = p \wedge p < n$

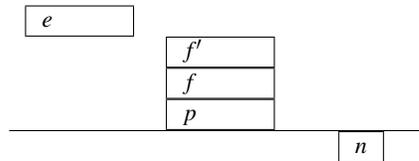


FIGURE 6.37 – Interprétation possible des variables temporelles pour *nous eûmes assisté au repas de l’annélide Carnivore*

Cette section nous a permis de mettre en évidence une interprétation correcte pour chacun des temps conjugués à partir d’exemples authentiques du corpus ITIPY. Il a par ailleurs permis de montrer une distinction entre PERF et AOR tel que nous proposons de les utiliser dans le cas du passé composé. Concernant la place de l’adverbe, nos exemples ne posent pas de problème particulier si, lorsqu’un adverbial de durée est appliqué à une éventualité portée par un verbe dont le temps conjugué possède une interface aspectuo-temporelle réalisée par l’opérateur IMP, alors le choix de son interprétation est limité à une forme de synchronicité, qui est une des interprétation de la relation \preceq . Nous abordons néanmoins ce problème dans le chapitre suivant, notre objectif suivant étant d’uniformiser l’application des adverbiaux temporels dans le système. Nous avons proposé, dans nos exemples, une surcomposition avec *nous eûmes assisté au repas de l’annélide Carnivore* qui, comme dans le système originel, est le fruit d’un ancrage propre au passé simple et d’un aspect ajouté *perfect*, il se réalise dans notre adaptation par les opérateurs PAST AOR PERF.

6.6 Conclusion

Nous avons présenté une adaptation du système de Verkuyl *Binary Tense* nous permettant de traiter les temps du français du XIXème siècle à nos jours. Nous avons apportée quelques modifications que nous résumons ici :

Nous avons modifié la relation portée par l’opérateur PERF afin qu’elle se restreigne uniquement aux emplois dont la sémantique est perfective. Nous avons intégré l’aoriste dans l’interface aspectuo-temporelle, ce qui nous permet d’obtenir une représentation du passé composé aoriste distincte du passé composé perfectif, de représenter le passé simple et le passé antérieur. Par ailleurs, pour ce qui est de l’utilisation

de l'opérateur AOR, nous avons montré son incompatibilité avec les temps utilisant l'opérateur PRES pour des raisons d'ordre énonciatif.

Et pour finir nous avons légitimé les conditionnels sauf PAST POST AOR qui ne trouve pas d'expression en français. Nous avons régularisé l'influence des adverbiaux temporels sur l'indice f en distinguant les deux usages du passé composé, néanmoins quelques questions persistent quant à l'application des adverbiaux de temps dans le système, nous avons donné les interprétations de divers exemples issus du corpus sans analyser le types de ces adverbiaux, nous proposons donc dans le chapitre suivant une analyse des phénomènes qui régissent la position de leur application dans le système.

Chapitre 7

Extension du système

Nous nous intéressons dans cette section à l'impact de l'introduction de l'opérateur AOR dans la composition des opérateurs temporels. Verkuyl aborde deux types d'ambiguïtés soulevées par le calcul temporel de *Binary Tense*. La première concerne les temps verbaux construits avec l'opérateur PERF et la position à laquelle on applique les adverbiaux temporels suivant l'interprétation attendue. Nous montrons qu'en donnant une représentation sémantique pour la notion aspectuo-temporelle aoriste, nous permettons une application homogène des expressions temporelles adverbiales. L'introduction de l'opérateur AOR à cette fin implique la sous-spécification d'une partie du paradigme temporel verbal, ce qui déplace l'ambiguïté sur l'aspect au lieu d'être liée à la position de l'adverbe dans le calcul. La seconde ambiguïté liée à l'utilisation de l'opérateur POST est analysée et permet de cerner les limites du calcul de la temporalité verbale. Nous montrons ensuite la pertinence d'une typologie des éventualités en fonction de leur aspect lexical en vue de l'extraction simultanée des informations temporelles et spatiales au sein des discours.

7.1 Introduction

Nous nous intéressons maintenant à l'impact de l'introduction de l'opérateur AOR dans la composition des opérateurs temporels. Le premier enjeu identifié est le lieu d'application de l'adverbial temporel en fonction de la composition temporelle choisie. Nous définissons dans un premier temps ce qu'est un adverbial temporel, puis nous mettons en relation les différents adverbiaux temporels identifiés et les formes temporelles verbales. Nous montrons par ce procédé que l'ambiguïté attachée à l'opérateur

PERF fait émerger de nouvelles analyses et la discrimination entre PERF et AOR résout une partie du problème. Nous interrogeons la seconde ambiguïté soulevée par Verkuyl [Verkuyl, 2008] : l'impact de l'opérateur POST sur la sélection de l'indice altéré par un éventuel adverbial temporel. Dans ce chapitre nous nous intéressons ensuite à la manière de représenter la relation entre le voyageur et le lieu, que nous appelons région (et dont c'est le type dans notre système), dans la représentation sémantique d'un segment comportant une éventualité de déplacement ou de localisation. Afin de résoudre cette question, à l'aide d'un prédicat minimal de localisation, nous discutons une interprétation possible de la sémantique spatiale des éventualités de déplacement et localisation en fonction des informations qu'elles véhiculent sur la relation entre le voyageur et le lieu en question.

7.2 Définitions

Par adverbial temporel, nous entendons toute expression temporelle ayant en surface une fonction adverbiale, c'est-à-dire qui modifie le verbe ¹, et dans une analyse plus profonde, affectant la temporalité de l'éventualité (l'éventualité, telle que nous la traitons, étant réifiée depuis un verbe conjugué). L'éventualité ayant le type ν dans notre système, nous rappelons que pour simplifier les calculs, nous réifions directement l'extension temporelle de l'éventualité de type i , pour intervalle temporel.

$$\lambda e^i . \text{quitter}(e, \text{Cauterets}, \text{Azaïis})^{i \rightarrow t}$$

Les indices créés par l'application des opérateurs temporels sont au nombre de 2, f et p tels que présentés dans *Binary Tense* originellement et dont nous gardons les noms pour simplifier les explications. Ils sont eux aussi de type i :

PRES	$\lambda \phi^{i \rightarrow t} \exists p^i [\phi[p] \wedge (p) \circ n]^{(i \rightarrow t) \rightarrow t}$
PAST	$\lambda \phi^{i \rightarrow t} \exists p^i [\phi[p] \wedge (p) < n]^{(i \rightarrow t) \rightarrow t}$
SYN	$\lambda \phi^{i \rightarrow t} \lambda p^i \exists f^i [\phi[f] \wedge (f) = (p)]^{(i \rightarrow t) \rightarrow i \rightarrow t}$
POST	$\lambda \phi^{i \rightarrow t} \lambda p^i \exists f^i [\phi[f] \wedge (f) \preceq \diamond(p)]^{(i \rightarrow t) \rightarrow i \rightarrow t}$
IMP	$\lambda \phi^{i \rightarrow t} \lambda f^i \exists e^i [\phi[e] \wedge (e) \preceq (f)]^{(i \rightarrow t) \rightarrow i \rightarrow t}$
PERF	$\lambda \phi^{i \rightarrow t} \lambda f^i \exists e^i [\phi[e] \wedge (e) < (f)]^{(i \rightarrow t) \rightarrow i \rightarrow t}$
AOR	$\lambda \phi^{i \rightarrow t} \lambda f^i \exists e^i [\phi[e] \wedge (e) = (f)]^{(i \rightarrow t) \rightarrow i \rightarrow t}$

FIGURE 7.1 – Lexique temporel dans notre adaptation du système *Binary Tense*

L'adverbial est une expression temporelle qui modifie la temporalité de l'éventualité en agissant directement sur les intervalles qui la constituent. Ils peuvent aussi être connecteurs entre deux éventualités (par exemple *ensuite* ou encore des adverbiaux "qualitatifs" (tels que *bientôt*, *auparavant*)²) mais nous ne les traitons pas dans cette

1. D'un point de vue syntagmatique génératif.

2. Nous remercions Patrice Enjalbert pour la remarque.

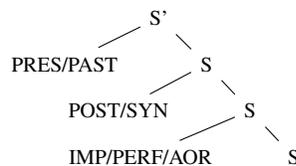


FIGURE 7.2 – Rappel de la structure du système

étude. Tout comme les opérateurs temporels mentionnés jusqu'ici, ils nécessitent un lexique qui leur est propre et qui doit entrer dans le système de calcul de la temporalité des éventualités. Les exemples classiques d'expression temporelle sont les syntagmes prépositionnels introduits par la préposition *à*, *en* ou *pendant* par exemple.

(7.1) Azaïs quittera Cauterets *à cinq heures*.

(7.2) Azaïs avait randonné *pendant cinq heures*.

(7.3) Azaïs escalada le Vignemale *en 1886*.

Les syntagmes prépositionnels *à cinq heures*, *pendant cinq heures* et *en 1886* sont autant d'adverbiaux temporels. Notre objectif ici n'est pas de donner une liste exhaustive des adverbiaux temporels mais d'approcher cette notion du point de vue du calcul de la représentation sémantique.

7.3 Les adverbiaux temporels

7.3.1 Spécification dans *Binary Tense*

Deux ambiguïtés ont été soulevées lors de la conception de *Binary Tense* :

- la place de l'adverbial dans le calcul de la temporalité d'une éventualité comprenant l'opérateur PERF
- la place de l'adverbial dans le calcul de la temporalité d'une éventualité comprenant l'opérateur POST

L'ambiguïté s'exprime dans la structure sémantique par deux positions possibles pour l'adverbial :

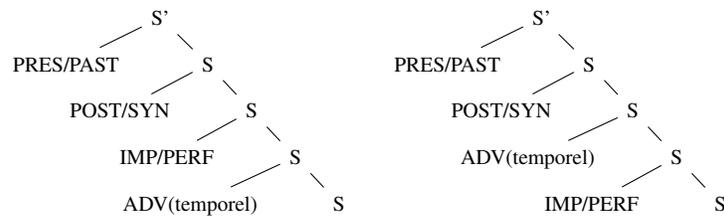


FIGURE 7.3 – Structure sémantique de la temporalité avec l’adverbiale temporel en position 1 (à gauche) et en position 2 (à droite)

Les deux positions montrées en figure 7.3, révèlent une ambiguïté temporelle, stipulant que la modification adverbiale temporelle agit :

- soit sur l’indice e , l’extension temporelle de l’éventualité réifiée à partir du prédicat verbal que l’on peut se représenter intuitivement comme "objectivée", c’est à dire non temporalisée, ce qui implique qu’elle n’ait pas été modifiée par la prise en considération du phénomène de *monstration* d’ordre aspectuel et sans distance énonciative, et à vrai dire sans temporalité ;
- soit sur l’indice f , introduit dès l’interface aspectuo-temporelle par les opérateurs IMP, PERF et AOR dans notre adaptation et permettant de positionner le cadre, la fenêtre à travers laquelle on regarde l’éventualité dans le processus de *monstration* précédemment cité d’après [Gosselin, 2005].

L’exemple typique donné par Verkuyl concernant l’ambiguïté du *perfect* est le suivant :

(7.4) Mary had left at five o’clock. (departure time = 5.00PM) : *position 1*

(7.5) Mary had left at five o’clock (departure time < 5.00PM) : *position 2*

Nous traduisons l’exemple :

(7.6) ?Marie était partie à cinq heures. (heure de départ = 5h00) : *position 1*

(7.7) Marie était partie à cinq heures (heure de départ < 5h00) : *position 2*

Notre adaptation du système privilégie le second contexte (heure de départ < 5h00), avec la composition (PAST) (SYN) (PERF). Ce qui permet d’appliquer préférentiellement l’adverbiale en position 2. La première interprétation selon nous est peu probable, particulièrement par l’utilisation de l’auxiliaire être au lieu d’avoir pour former le plus-que-parfait. Pour plus de précision, prenons un verbe dont la forme plus que parfait se réalise avec l’auxiliaire avoir, *quitter*, l’utilisation de l’auxiliaire *être* pouvant avoir un impact sur l’interprétation de l’éventualité en français.

(7.8) Azaïs avait quitté Causerets à cinq heures. (heure de départ = 5h00) : *position 1*

(7.9) Azaïs avait quitté Cauterets à cinq heures (heure de départ < 5h00) : *position 2*

Dans ce cas les deux contextes sont plausibles, la seconde étant réalisée correctement par (PAST) (SYN) (PERF) et l'adverbial en position 2 et la première ayant selon nous une interprétation aoriste tout comme le passé composé. L'ambiguïté ici soulevée semble liée à notre discussion sur l'ambiguïté entre les deux possibilités d'interprétations du passé composé. Dans la première interprétation 7.8, on ne considère pas un état du monde conséquent mais bien l'éventualité en elle-même que l'on modifie par l'adverbe. Dans ce cas précis, la valeur sémantique du plus-que-parfait suppose la localisation précise de l'éventualité dans le passé par rapport au moment d'énonciation et montre une équivalence entre le passé composé et le passé simple : (PAST)(SYN)(AOR). Très justement, Verkuyl suppose que l'ambiguïté provient de l'utilisation de PERF et nous remarquons que cette distinction est équivalente à la distinction présentée dans le chapitre précédent avec les exemples suivants :

(7.10) (A seize heures quinze,) Azaïs a quitté Cauterets.

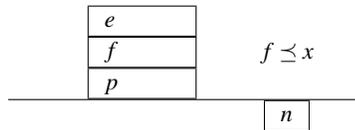


FIGURE 7.4 – Interpretation pour (PAST)(SYN)(ADV)(AOR)(λ e. quitter(e,Cauterets,Azaïs)

(7.11) (A l'heure qu'il est,) Azaïs a quitté Cauterets.

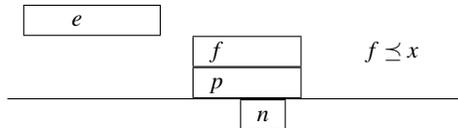


FIGURE 7.5 – Interpretation pour (PRES)(SYN)(ADV)(PERF)(λ e. quitter(e,Cauterets,Azaïs)

Ainsi pour le plus-que parfait, de la même manière que pour le passé composé, deux interprétations sont acceptées, l'une *perfect* et l'autre aoriste :

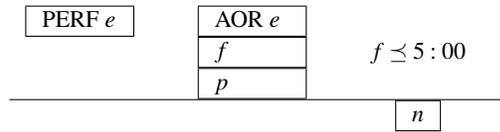


FIGURE 7.6 – Synthèse des deux interprétations, *perfect* et aoriste d’*Azaïs avait quitté Caunterets à cinq heures*

Cette ambiguïté peut être traitée par cette distinction aspectuo-temporelle pour le cas du passé composé et pour le plus-que-parfait mais qu’en est-il des autres formes utilisant PERF ? Deux autres formes doivent être analysées pour mettre en évidence le fonctionnement de PERF :

		PRES	PAST
<p>PERF</p> $\lambda \phi \lambda f \exists e [\phi[e] \wedge (e) \prec (f)]$	<p>SYN</p> $\lambda \phi \lambda p \exists f [\phi[f] \wedge (f) = (p)]$	<p>elle a écrit</p>	<p>elle avait écrit</p>
	<p>POST</p> $\lambda \phi \lambda p \exists f [\phi[f] \wedge (f) \preceq \diamond(p)]$	<p>elle aura écrit</p>	<p>elle aurait écrit</p>

– (PRES) (POST) (PERF) est la combinaison correspondant au futur antérieur. Nous proposons un contexte plausible pour saisir la subtilité sémantique présente ici :

(7.12) (Il était cinq heures et quart et personne ne vint aider Ann Lister a descendre de cheval, comme elle le redoutait,) Azaïs aura (donc) quitté Caunterets à cinq heures. (heure de départ = 5h00)

(7.13) (Si Ramond décide de faire l’ascension,) Azaïs aura quitté Caunterets à cinq heures(, il ne pourra venir au comice). (heure de départ < 5h00)

Ces deux interprétations sont effectivement valides si l’on rajoute un contexte. On remarque une interprétation facilitée par le contexte au passé pour la forme aoriste (heure de départ = 5h00), et un contexte préférablement au présent pour l’interprétation *perfect* (heure de départ < 5h00).

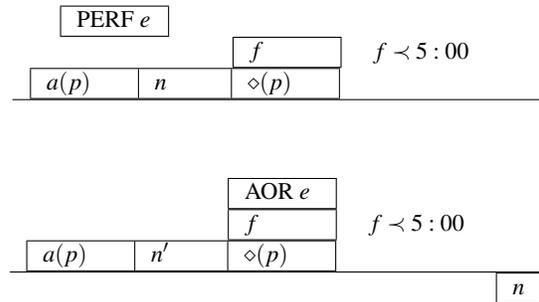


FIGURE 7.7 – Synthèse des deux interprétations, *perfect* et aoriste d’*Azaïs aura quitté Cauterets à cinq heures*

– (PAST) (POST) (PERF) est la combinaison correspondant au conditionnel passé :

(7.14) (Il était cinq heures et quart et personne ne vînt aider Ann Lister a descendre de cheval, comme elle le redoutait,) Azaïs aurait (donc) bien quitté Cauterets à cinq heures. (heure de départ = 5h00)

(7.15) (Si Ramond décidait de faire l’ascension,) Azaïs aurait quitté Cauterets à cinq heures(, il ne pourrait venir au comice) (heure de départ < 5h00)

Ces deux interprétations sont effectivement valides ici aussi si l’on rajoute un peu de contexte. On remarque là aussi une interprétation facilitée par le contexte au passé pour la forme aoriste et un contexte qui reste au passé pour l’interprétation *perfect*.

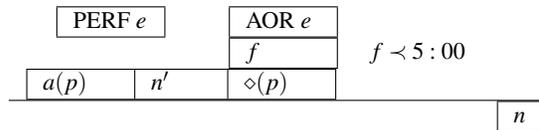


FIGURE 7.8 – Synthèse des deux interprétations, *perfect* et aoriste d’*Azaïs aurait quitté Cauterets à cinq heures*

La distinction majeure entre la forme *perfect* et la forme aoriste dans ces deux exemples est orientée par l’emploi conditionnel ou hypothétique de chacune des formes, mais nous ne rentrons pas dans ces considérations ici. Nous cherchons simplement à donner une représentation temporelle en adéquation avec la sémantique de la langue, sans prendre en compte aucune forme de modalité et a fortiori aucune distinction ayant trait au mode. L’aoriste peut donc s’étendre à toutes les formes *perfect* en ayant pour influence de les basculer dans une sémantique du passé si elle font partie des formes au présent.

Les quatre interprétations sont possibles, avec pour chacune d’entre elle une interprétation aoriste pour la version dans laquelle l’heure du départ est 5h00, et une interprétation *perfect* si le départ est antécédent à 5h00, ce qui nous permet d’étendre

le tableau en respectant la contrainte sur l'interface énonciativo-temporelle pour AOR : AOR ne peut être combiné à PRES.

		PRES $\lambda\phi\exists p[\phi[p] \wedge (p) \circ n]$	PAST $\lambda\phi\exists p[\phi[p] \wedge (p) < n]$
IMP $\lambda\phi\lambda f\exists e[\phi[e] \wedge (e) \preceq (f)]$	SYN $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) = (p)]$	elle écrit	elle écrivait
	POST $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) \preceq \circ(p)]$	elle écrira	elle écrirait
PERF $\lambda\phi\lambda f\exists e[\phi[e] \wedge (e) < (f)]$	SYN $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) = (p)]$	elle a écrit	elle avait écrit
	POST $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) \preceq \circ(p)]$	elle aura écrit	elle aurait écrit
AOR $\lambda\phi\lambda f\exists e[\phi[e] \wedge (e) = (f)]$	SYN $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) = (p)]$	∅	elle a écrit elle avait écrit elle écrivait
	POST $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) \preceq \circ(p)]$	∅	elle aura écrit elle aurait écrit

FIGURE 7.9 – Représentation des emplois aoristes dans le système

Dans les deux cas, l'adverbial temporel peut être appliqué en position 2, sans qu'il n'y ait besoin de modifier cette place. En revanche tous les temps *perfect* ont un pendant aoriste qui fait d'eux des temps sous-spécifiés et déplace l'ambiguïté qui résidait dans le choix de la position d'application de l'adverbial. L'ambiguïté qui reposait sur le choix entre position 2 et position 1 pour l'application de l'adverbial a été levée et repose maintenant sur le choix entre la sémantique *perfect* et aoriste du temps conjugué.

Les temps *imperfect* ne présentent pas de telles ambiguïtés. Le système peut être divisé en deux parties :

- PRES présente deux choix entre deux opérateurs : SYN/POST et IMP/PERF dans une version fidèle au système originel
- PAST présente un choix entre deux opérateurs SYN/POST et un choix entre trois opérateurs IMP/PERF/AOR

Nous obtenons un système à deux branches, la première ayant 1x2x2 feuilles et la seconde 1x2x3 feuilles :

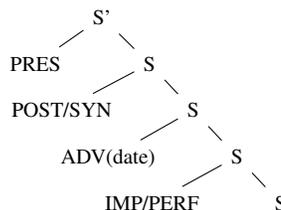


FIGURE 7.10 – Structure sémantique de la temporalité du présent

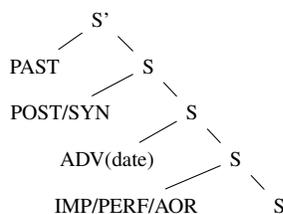


FIGURE 7.11 – Structure sémantique de la temporalité du passé

Chacune des branches présente une place fixe en position 2 pour l'application de l'adverbial lorsqu'il exprime une date. Nous présentons une classe des adverbiaux de durée et à ancrage variable dans la section suivante.

L'opérateur POST quant à lui présente une ambiguïté quant au degré de certitude selon lequel s'est réalisée l'éventualité lorsqu'elle est localisée par rapport au sous-intervalle final de p , c'est-à-dire $\diamond p$, pour le néerlandais et l'anglais. Ceci est vrai aussi pour le français. L'exemple donné dans [Verkuyl, 2008] est :

(7.16) Mary would leave at five o'clock.

(7.17) At five o'clock (it was clear that) Mary would leave.

que nous traduirons par

(7.18) Marie partirait (bien) à cinq heures.

(7.19) (Il était clair qu') A cinq heures, Marie partirait.

Pour le premier exemple, Marie aimerait partir au moment où "cinq heures" sonne tandis que pour le second exemple, lorsque "cinq heures" a sonné, l'énonciateur était convaincu que Marie partirait à un moment indéterminé ultérieurement. Cette ambiguïté n'est pas partagée avec le futur simple, si l'on retire le contexte, la seconde interprétation n'est pas plausible :

(7.20) Marie partira (bien) à cinq heures.

(7.21) *(Il est clair qu') A cinq heures, Marie partira.

En ce qui concerne le conditionnel passé, les deux exemples suivants ne peuvent recevoir deux interprétations divergentes dans le même sens.

(7.22) Azaïs aurait (bien) quitté Cauterets à cinq heures.

(7.23) (Il était clair qu') A cinq heures, Azaïs aurait quitté Cauterets.

Ici le premier exemple impose une lecture aoriste tandis que le second est, de notre point de vue, un usage élidé mettant en scène une autre éventualité qui serait portée par *il était clair qu'*. Cette élosion présente une limite à notre système qui ne prend pas en compte les présupposés.

Verkuyl appelle la particularité du conditionnel présent une ambiguïté "supratemporelle" car elle dépasse effectivement les enjeux du calcul d'une temporalité et entre

tout à fait dans le calcul de la modalité, proposant une lecture de l'éventualité ciblée comme étant soumise au "possible" contrairement à une temporalité attestée. En d'autres termes, n , le moment d'énonciation, ou n' , un "présent dans le passé" qui permettent de distinguer le sous-intervalle initial (ap) ou final ($\diamond p$) de l'intervalle p , est exactement le lieu de la coupure modale entre l'irrévocable et le possible évoqué par Gosselin [Gosselin, 2005] et Battistelli [Battistelli, 2009], présentés dans le chapitre 3. L'utilisation de l'opérateur POST impose une indétermination de la temporalité de l'éventualité. Nous ne pouvons résoudre cet aspect de la sous-spécification temporelle du conditionnel présent. Par ailleurs dans l'exemple *A cinq heures (il était clair que) Marie partirait.* montre bien que l'adverbial ne s'applique pas à l'éventualité dénotée par *partir* mais à l'éventualité dénoté par *il était clair*. C'est pourquoi nous pensons que l'ambiguïté supratemporelle est un problème modal plus large que le calcul du temps verbal, et nous confirmons la légitimité des conditionnels à figurer dans le système du fait de leur localisation temporelle, que l'éventualité soit ou non actualisée, comme "feuilles" des arbres de composition de la sémantique temporelle.

7.3.2 Classification des expressions temporelles adverbiales

Nous avons classé ici les adverbiaux afin de mettre en relief les indices qu'ils modifient en Table 7.1

- (7.24) Dans dix minutes je quitterai Nohant
- (7.25) Dans dix minutes j'aurai quitté Nohant (je serai alors hors de Nohant)
- (7.26) Dix minutes plus tard, je quitterai Nohant
- (7.27) Dix minutes plus tard, j'avais quitté Nohant (je n'étais alors plus à Nohant)
- (7.28) Nous quittâmes Gavarnie à sept heures moins le quart
- (7.29) Nous avons quitté Gavarnie à sept heures moins le quart (nous n'étions alors plus à Nohant)
- (7.30) Azaïs atteignit le sommet du Vignemale au mois de juin
- (7.31) Azaïs avait atteint le sommet du Vignemale au mois de juin (il était alors sur au sommet)
- (7.32) Azaïs randonnait depuis cinq mois
- (7.33) Azaïs avait quitté Cauterets depuis cinq mois (il était alors hors de Nohant)
- (7.34) On monterait pendant deux heures trente
- (7.35) On avait monté pendant deux heures trente (nous n'étions alors plus en train de monter passées les deux heures trente)

Pour les exemples 7.24, 7.25, 7.26, 7.27, 7.28, 7.29, 7.30, 7.31, 7.32 et 7.33, l'intervalle affecté par le modifieur est f , ce qui justifie la position 2 pour l'application de l'adverbial. Tandis que pour les exemple 7.34 et 7.35, c'est e qui est modifié. Par exemple,

- (7.36) A sept heures moins le quart, on monterait pendant deux heures trente.

7.3. LES ADVERBIAUX TEMPORELS

Ancrage	Type d'unité	Opérateur	Exemple	Enoncés	Intervalle affecté
Relatif	Date	ADVRelDate	dans dix minutes/ dix minutes plus tard	7.24, 7.25 7.26, 7.27	<i>f</i>
Absolu	Date	ADVAbsDate	le six juin 1886/ au mois de juin	7.28, 7.29 7.30, 7.31	<i>f</i>
Relatif	DateDurée	ADVRelDateDurée	depuis cinq mois	7.32, 7.33	<i>f</i>
Absolu	Durée	ADVAbsDurée	pendant quatre heures	7.34, 7.35	<i>e</i>

TABLE 7.1 – Les classes d’expressions temporelles adverbiales

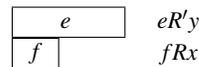


FIGURE 7.12 – Interaction entre les adverbiaux de date et de durée et les représentations des temps verbaux pour les *imperfect* (les compositions contenant IMP)

fRx signifie que l’intervalle *f* est en relation *R* avec l’intervalle temporel *x*, réifié à partir de l’adverbial de date. Dans le cas d’IMP, l’intervalle de l’éventualité *e* est affecté par l’adverbial par transitivité car *e* et *f* sont en relation \preceq .

eR'y exprime la relation *R'* entre l’extension temporelle de l’éventualité *e* et une mesure *y* exprimée par l’adverbial de durée. C’est *e* qui est altéré cette fois-ci.

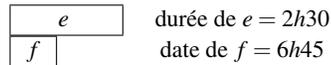


FIGURE 7.13 – Interaction entre les adverbiaux de date et de durée et la représentation de l’exemple 7.36

Uniformément, pour l’exemple 7.37, les mêmes postions sont conservées en ce qui concerne la modification adverbiale temporelle :

(7.37) A sept heures moins le quart, on avait monté pendant deux heures trente.



FIGURE 7.14 – Interaction entre les adverbiaux de date et de durée et les représentations des temps verbaux pour les *perfect* (les compositions contenant PERF)

Pour les verbes dont la valeur aspectuo-temporelle est PERF, l’intervalle de l’éventualité *e* n’est pas affecté directement par la date mais c’est la phase résultante à laquelle on accède par *f* qui est modifiée, ce qui correspond à la sémantique présentée

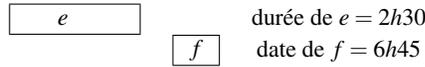


FIGURE 7.15 – Interaction entre les adverbiaux de date et de durée et la représentation de l'exemple 7.37

jusqu'alors. Cependant pour les exemples 7.34 et 7.35, l'adverbial demande systématiquement à modifier l'intervalle e , extension de l'éventualité, pour lesquels e est en relation R' avec y , mesure dénotée par *pendant deux heures trente*. L'application des adverbes de durée s'applique en position 1 tandis que les dates même si elles impliquent une forme de durée (*depuis cinq mois*) s'appliquent en position 2.

7.3.3 Le lexique des adverbiaux

Afin d'homogénéiser l'application des adverbiaux temporels, nous proposons d'appliquer les syntagmes prépositionnels en position 2 pour les dates et ce que nous appelons les dates-durées tels que *depuis cinq mois*, tandis que les durées sont appliquées systématiquement en position 1 tel que montré en figure 7.16.

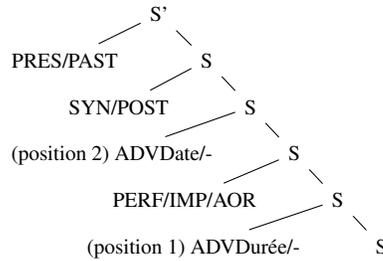


FIGURE 7.16 – Composition des opérateurs temporels et adverbiaux dans le système

Maintenant que nous avons montré dans quel ordre appliquer les différents opérateurs temporels, nous proposons le lexique adapté en λ -calcul simplement typé. Nous proposons de réifier non plus un intervalle mais un ensemble d'intervalles depuis le prédicat verbal. Nous proposons cette approche afin de donner la possibilité à terme d'intégrer les éventualités complexes composées de plusieurs intervalles, qu'ils soient multiples par itération ou par leur caractère discontinu tels que *Pendant un an j'ai fait du jogging* ou encore *J'ai hoqueté pendant une heure*. Nous désirons à ce stade donner une représentation sémantique des éventualités la plus proche possible de ce que peut être une éventualité, par conséquent, c'est l'ensemble fini d'intervalle qui nous a semblé le plus approprié.

Les types :

- le type complexe $i \rightarrow t$ est un ensemble d'intervalles

Les fonctions :

- la fonction $\mathbb{C}(un_ensemble^{i \rightarrow t})^i$ est la fonction qui prend en entrée un ensemble fini d'intervalles $\{i_1, \dots, i_n\}$ et renvoie un intervalle, enveloppe convexe de la totalité des intervalles de l'ensemble tel que $\mathbb{C}(\{i_1, \dots, i_n\}) \supseteq \{i_1, \dots, i_n\}$ et qu'il soit le plus petit intervalle tel que pour tout intervalle appartenant à l'ensemble $\{i_1, \dots, i_n\}$, il est inclus dans $\mathbb{C}(\{i_1, \dots, i_n\})$.
- la fonction $\mu_{unit}^{i \rightarrow \mathbb{R}}$ mesure la durée d'un intervalle et l'exprime en un réel. Ici on exprime une durée³. On a donc l'intervalle $\mu_{minutes}(x) = 5$ par exemple qui exprime que l'intervalle x a une durée de 5 minutes.
- la fonction $loc_{unit}(nat^{nat})^i$ renvoie l'intervalle temporel correspondant à l'entier nat dans son unité $unit$, soit heure ou minute, etc. Ici on exprime une date. On a donc l'intervalle $loc_{annee}(2013)$ par exemple qui renvoie l'intervalle associé à l'année 2013.
- la fonction $\delta_{unit}^{i \rightarrow i \rightarrow \mathbb{R}}$ désigne la mesure orientée entre deux intervalles et l'exprime en un réel. Ici on exprime une distance. On a donc un intervalle $\delta_{jours}(x, y) = 10$ par exemple qui exprime que la distance entre la borne finale de x et la borne initiale de y est de 10 jours.
- la fonction $\gamma_{unit}^{i \rightarrow i \rightarrow \mathbb{R}}$ désigne la mesure orientée entre la borne initiale d'un intervalle et la borne initiale d'un second intervalle et l'exprime en un réel. Ici on exprime une distance "recouverte" en quelque sorte. On a donc un intervalle $\gamma_{jours}(x, y) = 10$ par exemple qui exprime que la distance entre la borne initiale de x et la borne initiale de y est de 10 jours.

Les prédicats dorénavant nous permettront de réifier un ensemble d'intervalles et non plus ni une éventualité ni un intervalle simple, ϕ est donc de type $i \rightarrow t$ tandis que les entrées lexicales pour les verbes seront donc par de la forme :

$$(\lambda y^r \lambda x^h \lambda e^{i \rightarrow t} .quitter(e, x, y))^{r \rightarrow h \rightarrow (i \rightarrow t) \rightarrow t}$$

Nous modifions les types utilisés pour les intervalles e et f déjà introduits. f et e deviennent des ensembles finis d'intervalles :

- f^i et e^i de type i deviennent $f^{i \rightarrow t}$ et $e^{i \rightarrow t}$ de type $i \rightarrow t$ tels que $f^{i \rightarrow t} = \{f_1^i, \dots, f_a^i\}$ et $e^{i \rightarrow t} = \{e_1^i, \dots, e_a^i\}$.
- Respectivement pour PERF, IMP et AOR :
 $e < f$ devient $\mathbb{C}(e) < \mathbb{C}(f)$
 $e \preceq f$ devient $\mathbb{C}(e) \preceq \mathbb{C}(f)$
 $e = f$ devient $\mathbb{C}(e) = \mathbb{C}(f)$

Remarque : Pour un couple d'ensembles (e, f) on définit les relations $<_{ens}$, \preceq_{ens} et $=_{ens}$ comme étant des relations de eXf totales et bijectives (tout élément de e est en relation avec un et un seul élément de f) comme suit :

- $e <_{ens} f$ si et seulement si $\mathbb{C}(e) < \mathbb{C}(f)$ et chaque élément de e est en relation $<$ avec un élément de f ,
- $e \preceq_{ens} f$ si et seulement si $\mathbb{C}(e) \preceq \mathbb{C}(f)$ et chaque élément de e est en relation \preceq avec un élément de f ,

3. Nous avons uniformisé nos opérateurs a posteriori avec ceux présentés dans les travaux de Moot [?].

- $e =_{ens} f$ si et seulement si $\mathbb{C}(e) = \mathbb{C}(f)$ et chaque élément de e est en relation = avec un élément de f .

PERF, AOR, IMP sont donc modifiés comme suit :

$$\text{PERF} =_{def} \lambda \phi^{i \rightarrow t \rightarrow t} \lambda f^{i \rightarrow t} \exists e^{i \rightarrow t} . [\phi[e] \wedge \mathbb{C}(e) < \mathbb{C}(f)]$$

Le type de PERF est donc $(i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$,

$$\text{IMP} =_{def} \lambda \phi^{i \rightarrow t \rightarrow t} \lambda f^{i \rightarrow t} \exists e^{i \rightarrow t} . [\phi[e] \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)]$$

Le type de IMP est donc $(i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$,

$$\text{AOR} =_{def} \lambda \phi^{i \rightarrow t \rightarrow t} \lambda f^{i \rightarrow t} \exists e^{i \rightarrow t} . [\phi[e] \wedge \mathbb{C}(e) = \mathbb{C}(f)]$$

Le type d'AOR est donc $(i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$.

Cette modification induit une modification de SYN et POST qui prennent en entrée dorénavant un ensemble d'intervalles pour f comme suit :

$$\text{SYN} =_{def} \lambda \phi^{i \rightarrow t \rightarrow t} \lambda p^i \exists f^{i \rightarrow t} . [\phi[f] \wedge \mathbb{C}(f) = p] : (i \rightarrow t \rightarrow t) \rightarrow i \rightarrow t$$

$$\text{POST} =_{def} \lambda \phi^{i \rightarrow t \rightarrow t} \lambda p^i \exists f^{i \rightarrow t} . [\phi[f] \wedge \mathbb{C}(f) \preceq \diamond p] : (i \rightarrow t \rightarrow t) \rightarrow i \rightarrow t$$

Le premier opérateur optionnel concerne donc la durée de l'éventualité et affecte l'indice e dans le système :

$$\text{ADVAbsoluDuree} =_{def} \lambda \phi^{i \rightarrow t \rightarrow t} \lambda e^{i \rightarrow t} . [\phi[e] \wedge \mu_{unit}(\mathbb{C}(e)) = y] : (i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$$

Pour l'adverbial *pendant quatre heures* par exemple, c'est bien de l'éventualité dont on parle, on donne donc la valeur 4 à y qui doit être interprété comme un intervalle de 4 heures $\mu_{heures}(\mathbb{C}(e)) = 4$, l'intervalle $\mathbb{C}(e)$ incluant tous intervalles de l'ensemble e .

Maintenant nous présentons les adverbiaux affectant f :

$$\text{ADVAbsoluDate} =_{def} \lambda \phi^{i \rightarrow t \rightarrow t} \lambda f^{i \rightarrow t} . [\phi[f] \wedge loc_{unit}(\mathbb{C}(f)) = y] : (i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$$

Pour l'adverbial *en 2013*, on situe l'enveloppe convexe de f dans l'intervalle obtenu pour l'année 2013, on a donc $loc_{annee}(\mathbb{C}(j)) = 2013$.

$$\text{ADVRelatifDate} =_{def} \lambda \phi^{i \rightarrow t \rightarrow t} \lambda f^{i \rightarrow t} . [\phi[f] \wedge \delta_{unit}(\mathbb{C}(f), n) = y] : (i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$$

Pour l'adverbial *dix minutes plus tard*, on donne une distance de 10 minutes entre l'enveloppe convexe de f et n dans le cas d'un temps du présent ou n' dans le cas d'un temps du passé, on a donc $\mu_{minutes}(\mathbb{C}(f), n) = 10$ ou $\mu_{minutes}(\mathbb{C}(f), n') = 10$.

$$\text{ADVRelatifDateDurée} =_{def} \lambda \phi^{i \rightarrow t \rightarrow t} \lambda f^{i \rightarrow t} . [\phi[f] \wedge \gamma_{unit}(\mathbb{C}(f), n) = y] : (i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$$

Pour l'adverbial *depuis 5 mois*, par exemple, j doit recouvrir les cinq mois, il nous faut donner une distance entre la borne initiale de $\mathbb{C}(f)$ et la borne initiale de n ou n' en mois et au nombre de 5, soit $\gamma_{\text{mois}}(\mathbb{C}(f), n) = 5$.

Ce lexique pose différents problèmes qui constituent les limites de nos travaux. Premièrement, le choix entre n et n' dans la sémantique des adverbiaux relatifs dépend de la forme conjuguées du verbe utilisé, soit c'est un temps du passé contenant l'opérateur PAST, n' sera donc sélectionné, soit c'est un temps du présent, et n sera alors sélectionné.

Ensuite, certains usages des relatifs nécessitent de prendre en compte une variable contextuelle qui relève de l'anaphoricité du récit que nous ne traitons pas ici. Par exemple, *je partirai le 5 juillet, trois heures plus tard je serai à Bagnères* nécessite la formulation d'une propriété de distance entre l'issue de l'enveloppe convexe de l'ensemble d'intervalles e énoncée par *je partirai le 5 juillet* (dont nous n'avons pas la mesure en heures) et ce que désigne *trois heures plus tard*. Nous pouvons traiter correctement *Je partirai le 5 juillet*, tandis que *trois heures plus tard* requiert une entrée lexicale tels que $\mu_{\text{heures}}(\mathbb{C}(f), x) = 3$ où x est $\mathbb{C}(e_{\text{partir}})$. Une piste qui pourrait être explorée consiste à repositionner le n suivant dans un récit au présent ou le n' suivant, dans un récit au passé, systématiquement à l'intersection entre $\diamond p$ et f en cours. Ainsi le fil du récit avance, tout comme le temps extra-linguistique duquel l'énonciation émerge.

Nous proposons de montrer la réduction d'un exemple permettant d'apprécier le positionnement des adverbiaux de datation relatifs (au moment d'énonciation pour cet exemple) et des adverbiaux de durée.

(7.38) Dans une semaine, Ramond séjournera à Bagnères pendant cinq jours.

Lexique		
Dans une semaine	ADVRelDate	$\lambda \phi \lambda f. [\phi[f] \wedge \delta_{\text{jours}}(\mathbb{C}(f), n) = 7]$
Ramond	Ramond	R
séjourner _à	$\lambda x \lambda y \lambda z. se\ journeyer(z, x, y)$	$\lambda x \lambda y \lambda z. se\ journeyer(z, x, y)$
futur simple (-a)	PRES	$\lambda \phi \exists p. [\phi[p] \wedge (p) \circ n]$
	POST	$\lambda \phi \lambda p \exists f. [\phi[f] \wedge \mathbb{C}(f) \preceq \diamond p]$
	IMP	$\lambda \phi \lambda f \exists e. [\phi[e] \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)]$
Bagnères	Bagnères	B
pendant cinq jours	ADVAbsDurée	$\lambda \phi \lambda e. [\phi[e] \wedge \mu_{\text{jours}}(\mathbb{C}(e)) = 5]$

PRES POST ADVRelatifDate IMP ADVAbsoluDuree ($\lambda x. se\ journeyer(x, B, R)$)

PRES POST ADVRelatifDate IMP($\lambda \phi \lambda e. [\phi[e] \wedge \mu_{\text{jours}}(\mathbb{C}(e)) = 5]$)($\lambda x. se\ journeyer(x, B, R)$)

PRES POST ADVRelatifDate IMP($\lambda e. [(\lambda x. se\ journeyer(x, B, R))[e] \wedge \mu_{\text{jours}}(\mathbb{C}(e)) = 5]$)

PRES POST ADVRelatifDate IMP($\lambda e. [se\ journeyer(e, B, R) \wedge \mu_{\text{jours}}(\mathbb{C}(e)) = 5]$)

Nous sommes alors rendu au niveau 2 de la figure 7.16

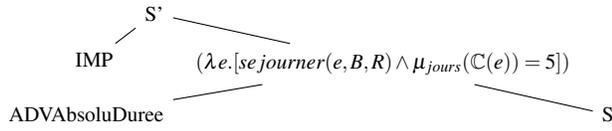


FIGURE 7.17 – Première étape de la réduction

PRES POST ADVRelatifDate $(\lambda \phi \lambda f \exists e. [\phi[e] \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)])(\lambda z. [se_journer(z, B, R) \wedge \mu_{jours}(\mathbb{C}(z)) = 5])$

PRES POST ADVRelatifDate $(\lambda f \exists e. [(\lambda z. [se_journer(z, B, R) \wedge \mu_{jours}(\mathbb{C}(z)) = 5])[e] \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)])$

PRES POST ADVRelatifDate $(\lambda f \exists e. [([se_journer(e, B, R) \wedge \mu_{jours}(\mathbb{C}(e)) = 5]) \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)])$

Nous sommes alors rendu au niveau 3 de la figure 7.16

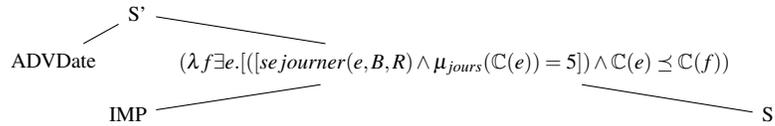


FIGURE 7.18 – Deuxième étape de la réduction

PRES POST $(\lambda \phi \lambda f. [\phi[f] \wedge \delta_{jours}(\mathbb{C}(f), n) = 7])$
 $(\lambda f \exists e. [([se_journer(e, B, R) \wedge \mu_{jours}(\mathbb{C}(e)) = 5]) \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)])$

PRES POST $(\lambda \phi \lambda f. [\phi[f] \wedge \delta_{jours}(\mathbb{C}(f), n) = 7])$
 $(\lambda z \exists e. [([se_journer(e, B, R) \wedge \mu_{jours}(\mathbb{C}(e)) = 5]) \wedge \mathbb{C}(e) \preceq \mathbb{C}(z)])$

PRES POST $(\lambda f.([\lambda z \exists e.([\textit{se\ journaler}(e, B, R) \wedge \mu_{\textit{jours}}(\mathbb{C}(e)) = 5] \wedge \mathbb{C}(e) \preceq \mathbb{C}(z))][f]$
 $\wedge \delta_{\textit{jours}}(\mathbb{C}(f), n) = 7])$

PRES POST $(\lambda z.([\exists e.([\textit{se\ journaler}(e, B, R) \wedge \mu_{\textit{jours}}(\mathbb{C}(e)) = 5] \wedge \mathbb{C}(e) \preceq \mathbb{C}(z)) \wedge \delta_{\textit{jours}}(\mathbb{C}(z), n) = 7])$

Nous sommes alors rendu au niveau 4 de la figure 7.16

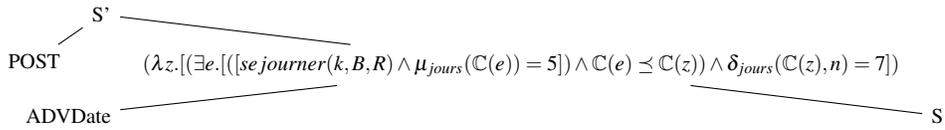


FIGURE 7.19 – Troisième étape de la réduction

PRES $(\lambda \phi \lambda p \exists f.[\phi[f] \wedge \mathbb{C}(f) \preceq \diamond p])(\lambda z.([\exists e.([\textit{se\ journaler}(e, B, R) \wedge \mu_{\textit{jours}}(\mathbb{C}(e)) = 5] \wedge \mathbb{C}(e) \preceq \mathbb{C}(z))$
 $\wedge \delta_{\textit{jours}}(\mathbb{C}(z), n) = 7])$

PRES $(\lambda p \exists f.([\lambda z.([\exists e.([\textit{se\ journaler}(e, B, R) \wedge \mu_{\textit{jours}}(\mathbb{C}(e)) = 5] \wedge \mathbb{C}(e) \preceq \mathbb{C}(z)) \wedge \delta_{\textit{jours}}(\mathbb{C}(z), n) = 7])][f]$
 $\wedge \mathbb{C}(f) \preceq \diamond p])$

PRES $(\lambda p \exists f.([\exists e.([\textit{se\ journaler}(e, B, R) \wedge \mu_{\textit{jours}}(\mathbb{C}(e)) = 5] \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)) \wedge \delta_{\textit{jours}}(\mathbb{C}(f), n) = 7])$
 $\wedge \mathbb{C}(f) \preceq \diamond p])$

Nous sommes alors rendu au niveau 5 de la figure 7.16

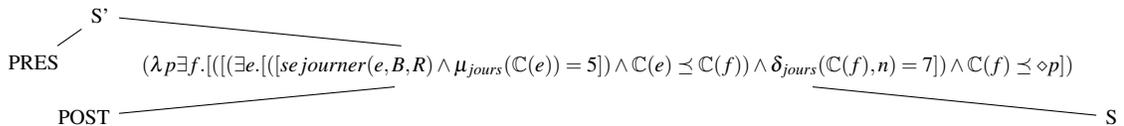


FIGURE 7.20 – Quatrième étape de la réduction

$$\begin{aligned}
 & (\lambda\phi\exists p[\phi[p] \wedge (p) \circ n]) (\lambda p\exists f. [([\exists e. [([se\ journey(e, B, R) \wedge \mu_{jours}(\mathbb{C}(e)) = 5]) \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)) \\
 & \quad \wedge \delta_{jours}(\mathbb{C}(f), n) = 7]) \wedge \mathbb{C}(f) \preceq \diamond f]) \\
 & (\exists p [(\lambda z\exists f. [([\exists e. [([se\ journey(e, B, R) \wedge \mu_{jours}(\mathbb{C}(e)) = 5]) \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)) \wedge \delta_{jours}(\mathbb{C}(f), n) = 7]) \\
 & \quad \wedge \mathbb{C}(f) \preceq \diamond z]) [p] \wedge (p) \circ n]) \\
 & (\exists p [(\exists f. [([\exists e. [([se\ journey(e, B, R) \wedge \mu_{jours}(\mathbb{C}(e)) = 5]) \wedge \mathbb{C}(e) \preceq \mathbb{C}(f)) \wedge \delta_{jours}(\mathbb{C}(f), n) = 7]) \\
 & \quad \wedge \mathbb{C}(f) \preceq \diamond p]) \wedge (p) \circ n]) \\
 & \exists p\exists f\exists e. se\ journey(e, B, R) \wedge \mu_{jours}(\mathbb{C}(e)) = 5 \wedge \mathbb{C}(e) \preceq \mathbb{C}(f) \wedge \delta_{jours}(\mathbb{C}(f), n) = 7 \wedge \mathbb{C}(f) \preceq \diamond p \wedge (p) \circ n
 \end{aligned}$$

Nous sommes alors rendu au niveau 6 de la figure 7.16

$$\begin{array}{ccc}
 \exists p\exists f\exists e. se\ journey(e, B, R) \wedge \mu_{jours}(\mathbb{C}(e)) = 5 \wedge \mathbb{C}(e) \preceq \mathbb{C}(f) \wedge \delta_{jours}(\mathbb{C}(f), n) = 7 \wedge \mathbb{C}(f) \preceq \diamond p \wedge (p) \circ n & & \\
 \text{PRES} \text{-----} & & \text{-----} \text{S}
 \end{array}$$

FIGURE 7.21 – Cinquième étape de la réduction

Ici, *dans une semaine* indique la distance temporelle entre le moment d'énonciation et l'intervalle $\mathbb{C}(f)$ pour la seule raison que l'adverbial temporel est relatif et le verbe conjugué dans un temps du présent (qui invoque donc l'opérateur PRES, en opposition aux temps du passé et qui invoquent l'opérateur PAST). Pour l'exemple *Une semaine plus tard, Ramond séjournera à Bagnères pendant cinq jours*, ce n'est pas n mais une variable contextuelle x présentée dans le lexique, nous ne nous soucions pas ici de l'anaphoricité du récit et considérons cette question comme le travail qui pourrait poursuivre notre recherche.

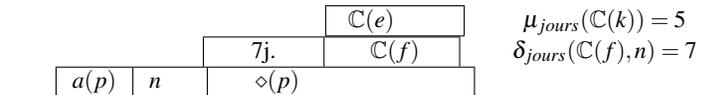


FIGURE 7.22 – Interprétation de l'exemple *Dans une semaine, Ramond séjournera à Bagnères pendant cinq jours*.

7.4 Aspect lexical et information spatiale

7.4.1 Pertinence du calcul sémantique de la temporalité et information spatiale des verbes de déplacement et localisation

Nous montrons dans cette partie en quoi les informations temporelles telles qu'elles ont été analysées jusqu'à maintenant sont pertinentes dans le repérage de la localisation d'un voyageur, pour extraire l'itinéraire décrit.

A cette fin, nous présentons dans cette partie une typologie spatiale standard des éventualités de déplacement et de localisation. Cette typologie nous mène à constater la pertinence du calcul de la temporalité tel qu'il est effectué jusqu'ici lorsqu'il est confronté à des données spatiales. Nous ne nous intéressons qu'aux verbes impliquant la ou les localisations d'un voyageur en position sujet et ne faisons pas de distinction sur la manière de réaliser le déplacement ou la localisation⁴. Nous concevons que ces verbes expriment des informations spatiales ainsi que temporelles. Notre système fonctionne pour tout type de verbe et nous rappelons que c'est dans l'objectif d'une extraction d'itinéraires effective à partir de textes bruts que nous désirons confronter le système de calcul de la temporalité aux données spatiales des verbes concernés. Nous ne distinguons pas par exemple *sortir* de *partir* qui indiquent deux positionnements différents par rapport à une entité spatiale. Ici, *sortir* et *partir* sont équivalents et représentent les localisations successives d'un voyageur dans, puis hors d'une région. Dans la typologie de [Asher *et al.*, 2008], nous reprenons typiquement les verbes de changement de lieu (ou "change of location"), auxquels nous ajoutons les localisations, tels que *rester*, *séjourner*, etc.

Nous définissons le prédicat élémentaire :

$$localise(x^r, y^h)$$

qui se lit "en la région x , est localisé le voyageur y ". Nous rappelons que le voyageur est de type h , appelée aussi entité mobile, tandis que la région est représentée par le type r . L'essence d'un déplacement réside dans le changement de localisation du voyageur, afin de représenter la dynamique du déplacement grâce au prédicat statique $localise(x, y)$, nous proposons de représenter canoniquement le déplacement par un intervalle e composé de trois sous-intervalles e_1 , e_2 et e_3 qui vérifient les relations de contact entre e_1 et e_2 et entre e_2 et e_3 .

Nous cherchons ici à représenter le caractère continu du déplacement au sein des intervalles, ainsi on cherche à donner une valuation positive ou négative au prédicat $localise(x, y)$ pour les trois intervalles considérées. La granularité de notre description s'arrête à l'ordonnancement des intervalles pendant lesquels on peut dire qu'en

4. Par conséquent, nous n'observons pas les verbes pouvant comporter une forme de localisation spatiale tels que *poser au dessus*. Même si beaucoup de verbes peuvent donner des informations indirectement sur la localisation d'un voyageur tels que *se lever* qui indique plutôt un changement de position en réactualisant la localisation du voyageur, ou encore *manger* qui pourrait indiquer qu'un déplacement en cours a été arrêté, par exemple dans l'extrait : *Nous disons adieu à ce sommet enchanteur, pour effectuer la descente malaisée de la paroi s.-o., où l'on avance lentement. En bas de ce fastidieux précipice, et dans la rimaye d'un névé qui nous procure eau et fraîcheur, nous mangeons une salade de chicorée que nous venons de cueillir : manger est le verbe depuis lequel on réifie l'ensemble d'intervalle pendant lequel les voyageurs sont localisés en bas du précipice.*

mière colonne et $\langle 0, (1/0), 1 \rangle$ pour la seconde. 0 indiquant que x n'est pas localisé en y , 1 que x est effectivement localisé en y et 0/1 que nous ne pouvons accéder à cette information. Nous nous inspirons ici des théories [Vandeloise, 1986, Boons, 1987] sur la polarité aspectuelle permettant de créer trois classes de verbes. Nous proposons trois exemples naïfs pour illustrer rapidement ces trois classes :

– Polarité initiale :

(7.39) Azaïs quitte Cauterets.

Quitter présente un déplacement dans lequel le voyageur part d'une région dans lequel il se trouve pour aller ailleurs. Le lieu considéré dans l'énoncé est donc le point de départ de la trajectoire, on appelle la polarité de *quitter* **initiale**. La valuation de *localise*(Cauterets, Azaïs) pour e sera donc $\langle 1, 1/0, 0 \rangle$, l'information pertinente étant qu'au début de l'éventualité, le voyageur est localisé *Cauterets*, tandis qu'à l'issue de celle-ci, il ne le sera plus.

– Polarité médiane :

(7.40) Azaïs traverse le Gave. Azaïs a traversé la France (pour rejoindre Ramond).

Traverser présente un déplacement dans lequel le voyageur utilise le repère spatial énoncé comme lieu par rapport auquel se déroule le déplacement. Le lieu considéré dans l'énoncé n'est donc ni le point de départ ni le point d'arrivée, il est médian, on appelle la polarité de *traverser* **médiane**. La valuation de *localise*(Gave, Azaïs) pour e sera donc $\langle 0, 1, 0 \rangle$, mais peut être aussi $\langle 1, 1, 1 \rangle$ pour *localise*(France, Azaïs) s'il est parti de Lille pour aller à Bagnères.

– Polarité finale :

(7.41) Azaïs atteint Cauterets.

Atteindre présente un déplacement dans lequel le voyageur utilise le repère spatial énoncé comme destination. Le lieu considéré dans l'énoncé est donc le point d'arrivée, il est final, on appelle la polarité de *atteindre* **finale**. La valuation de *localise*(Cauterets, Azaïs) pour e sera donc $\langle 0, (1/0), 1 \rangle$, tout comme pour *quitter*.

Typiquement on voudrait à partir d'une telle distinction proposer les valuations possibles suivantes :

polarité	Initiale	Mediane	Finale
valuations	$\langle 1, (1/0), 0 \rangle$	$\langle 1, 1, 1 \rangle$ $\langle 1, 0, 1 \rangle$ $\langle 0, 0, 0 \rangle$ $\langle 0, 1, 0 \rangle$	$\langle 0, (1/0), 1 \rangle$

TABLE 7.2 – Valuation de la relation *localise*(x, y) pour chaque polarité aspectuelle

Nous donnons quelques exemples de verbes de déplacement ou de localisation avec leur valuations associées.

Polarité	e_1	e_2	e_3	
Initiale	1	(1/0)	0	$quitter(e, x, y) / partir(e, x, y)$
Mediane	1	1	1	$setrouver(e, x, y)$
	1	0	1	$s'eclipser(e, x, y)$
	0	0	0	$éviter(e, x, y)$
	0	1	0	$traverser(e, x, y)$
Finale	0	(1/0)	1	$atteindre(e, x, y) / arriver(e, x, y)$

TABLE 7.3 – Propositions de verbes pour chaque valuations du prédicat $localise(x, y)$ dans la structure interne d'une déplacement ou d'une localisation

En revanche *parcourir* et *traverser* peuvent revêtir différentes valuation, c'est pourquoi nous désirons garder les trois intervalles. Observons les exemple suivants :

(7.42) Je suis née à Bordeaux et je m'installe à Hendaye, j'ai traversé l'Aquitaine.

(7.43) Je suis partie du Puy en Velay pour rejoindre San Sébastian, j'ai traversé l'Aquitaine.

Traverser devrait donc revêtir la valuation $\langle (1/0), 1, (1/0) \rangle$.

Donnons un exemple qui ne présente pas d'ambiguïté :

(7.44) Azaïs quitte Cauterets.

On a la λ -réduction suivante :

$$\exists p \exists f \exists e. quitter(e, Cauterets, Azaïs) \wedge (\mathbb{C}(e) \preceq \mathbb{C}(f)) \wedge (p = \mathbb{C}(f)) \wedge (p \circ n)$$

Pour illustrer notre propos, on propose le cas où l'ensemble e et l'ensemble f sont constitués chacun d'un seul élément, l'enveloppe convexe de e est l'intervalle e_u unique et l'enveloppe convexe de f est l'intervalle unique f_u :

L'interprétation de *Azaïs quitte Cauterets* dans les intervalles de \mathbb{R} sera donc figure 7.25 :

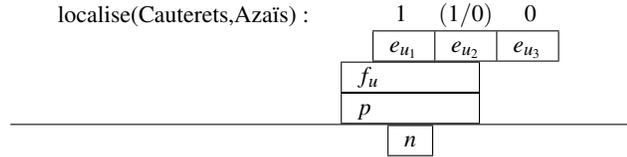


FIGURE 7.25 – Représentation spatio-temporelle minimale de *Azaïs quitte Cauterets*

On remarque qu'avec un temps *imperfect*, il est possible d'obtenir un sous-intervalle "final" de $\mathbb{C}(e)$ (ou e_u dans notre cas), qui n'a pas d'intersection avec $\mathbb{C}(f)$ ou f_u , ce sous-intervalle est un bon candidat pour être e_{u_3} . L'ensemble e représente un ensemble d'un seul élément, Néanmoins, il faut imaginer que cet ensemble puisse contenir plusieurs intervalles par exemple *Deux fois Azaïs quittait Cauterets quand Ramond le retint*. On suppose l'ensemble e comprenant deux intervalles e_a et e_b ainsi que l'ensemble f comprenant deux intervalles f_a et f_b . Dans ce cas e_{a_3} et e_{b_3} ne sont pas atteints :

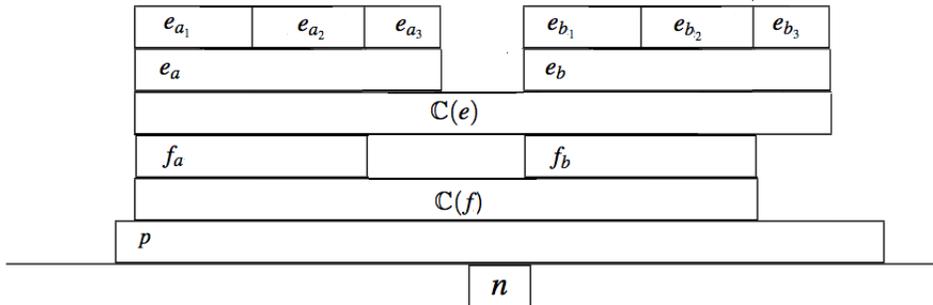


FIGURE 7.26 – Exemple d’interprétation avec les ensembles $e = \{e_a, e_b\}$ et $f = \{f_a, f_b\}$

7.4.2 Eventualités, aspect lexical et structure sémantique spatiale

Nous nous demandons désormais s’il nous est possible de définir automatiquement quand un sous-intervalle est atteint à partir de notre système temporel et des informations spatiales accordées aux classes de valuations définies plus tôt. Pour faciliter la présentation de notre propos, on considère une fois de plus que nos ensembles d’intervalles ne comprennent qu’un intervalle unique. Nous rappelons que notre typologie aspectuelle des eventualités compte trois types :

- Etat

L’état est "la manière d’être d’une personne ou d’une chose dans ce qu’elle a de durable" [Rey et Rey-Debove, 1993].

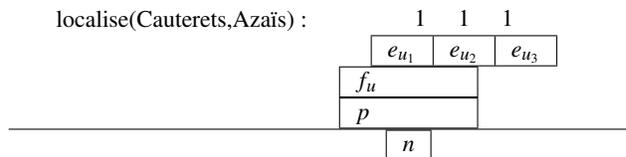


FIGURE 7.27 – Représentation spatio-temporelle minimale de *Azais reste à Cauterets*

- Activités

L’activité est une action qui est vue comme homogène.

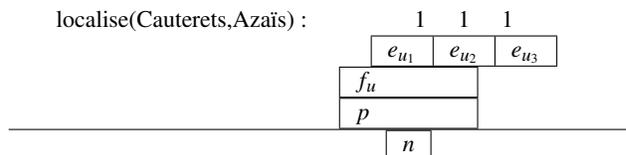


FIGURE 7.28 – Représentation spatio-temporelle minimale de *Azais visite Cauterets*

- Evènement
L'évènement est une action qui est vue comme hétérogène.

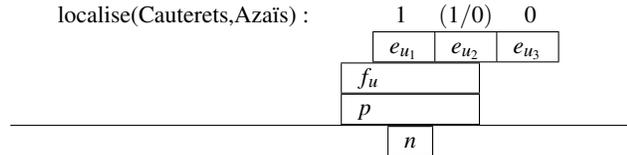


FIGURE 7.29 – Représentation spatio-temporelle minimale de *Azaïs quitte Cauterets*

L'hétérogénéité en ce qui concerne la localisation dans les éventualités de déplacement se traduit par la caractérisation d'un sous intervalle "final" qui rend vrai la localisation visée par le déplacement.

Qu'en est-il de la distinction état/non-état ? Si l'on observe le développement de [Kamp et Reyle, 1993a], la distinction réside dans la possibilité pour un état de n'avoir qu'une intersection avec l'extension temporelle de l'adverbial qui lui est appliqué, par exemple :

- (7.45) Azaïs est resté à Cauterets samedi dernier. En réalité, il est resté à Cauterets toute la semaine. Etat
- (7.46) Azaïs a visité Cauterets samedi dernier. En réalité, il a visité Cauterets toute la semaine. Activité
- (7.47) Azaïs a quitté Cauterets samedi dernier. #En réalité, il a quitté Cauterets toute la semaine. Evènement

Un second test concerne l'intolérance des états à la forme progressive :

- (7.48) #Azaïs était en train de rester à Cauterets samedi dernier. Etat
- (7.49) Azaïs était en train de visiter Cauterets samedi dernier. Activité
- (7.50) Azaïs était en train de quitter Cauterets samedi dernier. Evènement

C'est ce second test qui nous semble le plus pertinent pour discriminer les états des non-états si l'on veut conserver une distinction entre état et activité. Néanmoins, d'un point de vue spatial, la seule distinction entre éventualité homogène et éventualité hétérogène suffit. Ainsi, les états et activités peuvent être traités de la même manière, c'est-à-dire que les éventualités homogènes revêtent les valuations $\langle 0,0,0 \rangle$ ou $\langle 1,1,1 \rangle$, les évènements quant à eux, ou éventualités hétérogènes peuvent avoir toutes les autres valuations tant qu'au moins deux sous-intervalles de l'éventualité ne présentent pas d'égalité.

Nous avons pu montrer l'intérêt de la relation IMP en regard de la sémantique spatiale et nous considérons l'ajout de l'interprétation sémantique spatiale comme un gage de l'utilité de la distinction homogène/hétérogène au sein des éventualités. En voici un exemple avec la valuation : $v(\text{quitter}(e,x,y)) = \langle 1, (1/0), 0 \rangle$.

L'évènement peut ne jamais arriver à son terme comme c'est le cas dans cet exemple. Ce second exemple nous montre un paradoxe de l'imperfectif pour lequel on ne peut dire si le sous-intervalle final de l'éventualité est atteint. Ce qui empêche de considérer le voyageur comme ayant véritablement quitté les Pyrénées. Nous obtenons une bonne

7.4. ASPECT LEXICAL ET INFORMATION SPATIALE

Polarité	e_1	e_2	e_3	prédicat	Hétérogénéité/homogénéité
Initiale	1	(1/0)	0	$quitter(e, x, y) / partir(e, x, y)$	hétérogène
Mediane	1	1	1	$setrouver(e, x, y)$	homogène
	1	0	1	$s'eclipser(e, x, y)$	hétérogène
	0	0	0	$eviter(e, x, y)$	homogène
	0	1	0	$traverser(e, x, y)$	hétérogène
Finale	0	(1/0)	1	$atteindre(e, x, y) / arriver(e, x, y)$	hétérogène

TABLE 7.4 – Hétérogénéité des éventualités et sémantique spatiale

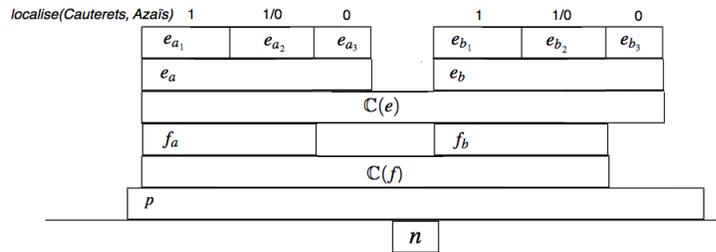


FIGURE 7.30 – Interprétation possible de *Deux fois, Azais quittait Cauterets ... (quand Ramond le retint)*

interprétation de *Deux fois Azais quittait Cauterets quand Ramond le retint*. En effet, les deux fois, on ne peut déduire qu’Azais a quitté Cauterets puisque e_{a_3} pour lequel la valuation $localise(Cauterets, Azais)$ est 0 n’a aucune intersection avec f_a . De la même manière, e_{b_3} pour lequel la valuation $localise(Cauterets, Azais)$ est 0 n’a aucune intersection avec f_b .

Ce paradoxe, étudié par beaucoup d’auteurs est interprété par Gosselin et Battistelli comme un coupure entre l’irrévocable et le possible, doit être interprété dans notre système par le fait que pour tout intervalle e_n de l’ensemble e , e_{n_3} doit être inclus en un intervalle f_n de l’ensemble f pour attester la localisation du voyageur en la région nommée et ce pour chaque couple (e_n, f_n) .

Ces considérations dépassent le propos de nos travaux actuels mais constituent une suite logique à nos recherches. Pour traiter correctement l’actualisation et le lien entre les éventualités au fil du discours il nous faut traiter les relations entre ces éventualités, régies par bien d’autres mécanismes et d’autres marqueurs que les connecteurs adverbiaux comme montré dans l’exemple :

(7.51) Mon guide faisait un pas, regardait son chemin, se retournait vers moi, et nous avançons ainsi fort lentement, nous accrochant sur une immense muraille de rochers à toutes les saillies qui s’offraient à nous ;

La relation de granularité entre les éventualités *faire un pas* et *avancer* est nécessaire pour interpréter correctement l'itération, ainsi que la relation discursive entre chacun des imparfaits utilisés ici.

7.5 Conclusion

Nous avons discuté la conception aspectuelle des opérateurs PERF et IMP et l'impact de l'introduction de l'interprétation aoriste dans le système compositionnel de calcul de la temporalité *Binary Tense*. Nous avons mis en évidence les calculs nécessaires à la capture des usages aspectuels des temps du paradigme verbal en français du XVIIIème siècle à nos jours. Notre adaptation et extension du système *Binary Tense* compte maintenant un opérateur aspectuo-temporel supplémentaire et quatre types d'adverbiaux. L'introduction de l'opérateur AOR a permis d'uniformiser l'application des adverbiaux, ce qui engendre pour cinq des temps du français une sous-spécification.

Nous avons présenté ici notre lexique afin de traiter les adverbiaux temporels le plus systématiquement possible. Les adverbiaux de durée s'appliquent en premier tandis que les autres adverbiaux temporels s'appliquent en seconde position (voire troisième dans le cas où un adverbial de durée aurait déjà été appliqué). Nous avons montré la pertinence d'une typologie des éventualités en fonction de leur homogénéité ou leur hétérogénéité. Cette typologie trouve légitimité dans l'extraction d'informations spatiales en adéquation avec les informations temporelles déjà traitées. Nous avons proposé un typage incluant des ensembles d'intervalles qui permettraient à terme de traiter les phénomènes d'éventualités complexes telles que l'itération, ou l'es adverbiaux de fréquence qui dénotent une forme de duplication l'éventualité.

Chapitre 8

Evaluation et Limites

Cette partie du présent manuscrit propose de confronter le lexique construit à l'analyse syntaxique de *Grail*, afin d'exposer l'automatisation totale de la chaîne de traitement à partir du texte brut jusqu'à la représentation sémantique des éventualités temporalisées. Nous proposons une synthèse de ce qui est couvert par le système de calcul et d'en montrer les limites.

8.1 Introduction

Dans notre cadre théorique et pour une intégration compositionnelle de la temporalité dans la sémantique du discours, nous nous situons à l'interface syntaxe-sémantique pour déterminer où seront intégrés les termes nécessaires à la représentation temporelle des déplacements et localisations. En effet *Grail* [Moot, 2001] permet d'obtenir complètement automatiquement une analyse syntaxique en catégories de Lambek de la phrase. A partir de cette analyse, et du lexique sémantique, *Grail* donne la représentation sémantique de cette même phrase. La représentation sémantique est résolue en puisant pour chaque terme de la phrase son entrée dans le lexique et en avançant dans la réduction de la preuve dans l'ordre de la structure syntaxique. En surface, les morphèmes propres à la temporalité verbale sont attachés au verbe en français. Traditionnellement, en grammaire syntagmatique on considère que le VP (verbal phrase, ou SV pour syntagme verbal) est contenu dans le TP (temporal phrase ou ST pour syntagme temporel). Dans le cadre des grammaires catégorielles, l'analyse syntaxique de la temporalité n'est pas à notre connaissance explorée. Il semble donc pertinent de confronter notre système de calcul sémantique de la temporalité à l'analyse syntaxique catégorielle ce qui permet d'observer le caractère opérationnel du lexique proposé.

Nous montrons ensuite les limites de notre analyse et proposons quelques perspectives intéressantes à ce travail. En particulier nous abordons la résolution de la sous-spécification aspectuo-temporelle provoquée par l'introduction de l'opérateur AOR, la couverture des réalisations linguistiques ayant trait à l'expression des éventualités de déplacement et localisation temporalisés, puis nous aborderons quelques phénomènes sémantiques temporels qu'il reste à traiter.

8.2 Synthèse du système

8.2.1 Syntaxe, sémantique et compositionnalité

En premier lieu il convient d'évaluer l'opérationnalité de notre système de calcul sémantique de la temporalité en regard de l'analyse syntaxique sur laquelle il repose dans *Grail*. Nous reprenons une dernière fois l'exemple canonique suivant :

(8.1) Azaïs quittera Cauterets.

Cet énoncé est analysé traditionnellement dans le cadre des grammaires syntagmatiques issue de la grammaire générative [Chomsky, 1963] de la manière suivante :

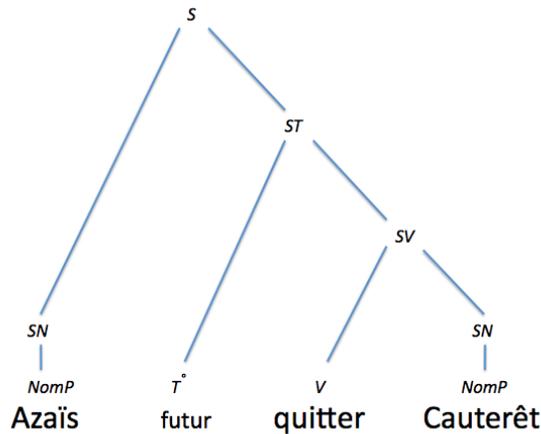


FIGURE 8.1 – Arbre syntaxique de "Azaïs quittera Cauterets" en grammaire générative

La phrase est correctement formée par un noeud *SN* sujet et un noeud *ST* (pour syntagme temporel, traditionnellement *TP* pour *temporal phrase*) qui contient le syntagme verbal. Le temps conjugué modifie l'ensemble des éventualités qui ont les propriétés sémantiques introduites par l'analyse du prédicat verbal muni de ses arguments sauf le sujet.

Tout d'abord d'un point de vue technique, les deux interprétations donnent à la temporalité un rôle différent. Il nous faut alors bien saisir ce que les opérateurs de la temporalité affectent, ce qu'ils modifient. Observons les entrée et sortie d'un opérateur temporel.

Le système de calcul de la temporalité dans le système *Binary Tense* suppose de traiter la temporalité comme prenant en entrée le prédicat verbal muni de ses arguments sujet et objet puis de l'appliquer aux opérateurs temporels, provenant du temps conjugué choisi et des adverbiaux temporels. Passer d'un système à l'autre nécessite un mouvement qui permet de faire intervenir l'opérateur temporel plus haut dans la hiérarchie syntaxique. Le mouvement inverse est lui aussi tout à fait possible.



FIGURE 8.2 – Comparaison schématique des structures syntaxiques et de l'application du temps : en grammaire générative (à gauche) et dans notre système (à droite)

Néanmoins, cette représentation est inexacte si l'on regarde de plus près les adverbiaux temporels qui devraient être intégrés au VP, nous détaillons tout ceci ensuite. Observons de plus près la forme logique de chacune des interprétations :

(8.2) (futur ((quitter)(Cauterets)))(Azaïs) : structure syntagmatique

(8.3) (futur (((quitter)(Cauterets)))(Azaïs)) : structure catégorielle

Nous allons montrer quatre exemples d'analyses syntaxiques nécessaires à l'utilisation correcte du lexique temporel, et pour cela nous rappelons les structures du système de calcul. Par ailleurs, les adverbiaux posent un problème crucial dans notre cadre, ils prennent la forme d'un syntagme prépositionnel ou d'un syntagme nominal, situé auprès ou non du verbe. Un adverbial par définition modifie un verbe et dans notre cas, doit faire plus que se déplacer auprès des autres modifieurs temporels, en ce qui concerne les adverbiaux de date par exemple, il doit se mêler à eux.

Nous rappelons la structure de composition des modifieurs :

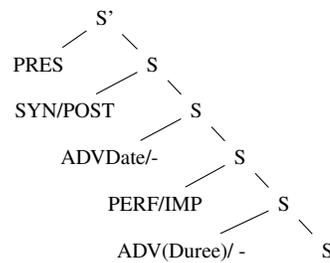


FIGURE 8.3 – Composition des opérateurs temporels, aspectuels et adverbiaux pour les temps du "présent"

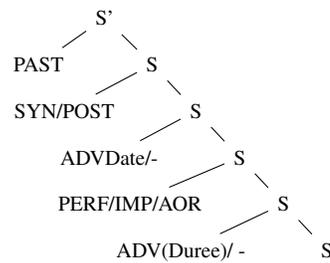


FIGURE 8.4 – Composition des opérateurs temporels, aspectuels et adverbiaux pour les temps du "passé"

Comparativement à la structure syntaxique dans le cadre des grammaires génératives, le système de calcul de la temporalité ne pose pas de problème majeur structurel. Néanmoins, les adverbiaux ayant la même catégorie que les opérateurs PERF, IMP ou AOR, il nous faut distinguer les adverbiaux de date qui ne doivent être acceptés qu'en seconde place de la structure du calcul temporel, nous apposons * à certaines catégories. Les types sont formés par *eI* signifiant ensemble d'intervalles, n'étant pas représentés au niveau linguistique par des unités visibles, *I* désignant intervalle. Nous illustrons l'analyse syntaxique et donnons les réductions à partir du lexique suivant :

Lexique sémantique	Catégorie	Type
PRES/PAST	$S/(I \setminus S)$	$(i \rightarrow t) \rightarrow i \rightarrow t$
POST/SYN	$(I \setminus S)/(eI \setminus S)_*$	$(i \rightarrow t \rightarrow t) \rightarrow i \rightarrow t$
IMP/PERF/AOR	$(eI \setminus S)_*/(eI \setminus S)$	$(i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$
ADVAbsoluDuree	$(eI \setminus S) \setminus (eI \setminus S)$	$(i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$
ADVAbsoluDate	$(eI \setminus S)_* \setminus (eI \setminus S)_*$	$(i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$
ADVRelatifDate	$(eI \setminus S)_* \setminus (eI \setminus S)_*$	$(i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$
ADVRelatifDateDurée	$(eI \setminus S)_* \setminus (eI \setminus S)_*$	$(i \rightarrow t \rightarrow t) \rightarrow (i \rightarrow t) \rightarrow t$

TABLE 8.2 – Catégories et types du lexique temporel

		PRES	PAST
		$\lambda\phi\exists p[\phi[p] \wedge (p) \circ n]$	$\lambda\phi\exists p[\phi[p] \wedge (p) < n]$
IMP	SYN $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) = (p)]$	elle écrit	elle écrivait
$\lambda\phi\lambda f\exists e[\phi[e] \wedge (e) \preceq (f)]$	POST $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) \preceq \circ(p)]$	elle écrira	elle écrirait
PERF	SYN $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) = (p)]$	elle a écrit	elle avait écrit
$\lambda\phi\lambda f\exists e[\phi[e] \wedge (e) < (f)]$	POST $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) \preceq \circ(p)]$	elle aura écrit	elle aurait écrit
AOR	SYN $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) = (p)]$	\emptyset	elle a écrit elle avait écrit elle écrivit
$\lambda\phi\lambda f\exists e[\phi[e] \wedge (e) = (f)]$	POST $\lambda\phi\lambda p\exists f[\phi[f] \wedge (f) \preceq \circ(p)]$	\emptyset	elle aura écrit elle aurait écrit

FIGURE 8.5 – Représentation des emplois aoristes dans le système

emplois aoristes sont représentés, ce qui permet d'appliquer de manière systématique les adverbiaux temporels de date et de durée.

Ce lexique permet de procéder aux créations automatiques des représentations sémantiques des segments utilisant un de ces temps tout en laissant une forme d'ambiguïté sur le choix de la forme *perfect* ou aoriste. Nous avons montré par l'ajout d'une représentation spatiale que la discrimination entre éventualités hétérogènes et homogènes pouvait avoir un impact sur l'actualisation d'une localisation dans la sémantique d'un déplacement. Sont donc pris en charge les paradoxes imperfectifs utiles à l'extraction des informations spatio-temporelle des éventualités hétérogènes.

D'un point de vue lexical, nous avons traité les adverbiaux portés pas des syntagmes prépositionnels et construits de façon standard, qu'il présente une sémantique ayant trait à la durée ou à la datation de l'éventualité.

Néanmoins, les cadratifs ne sont pas traités. Leur acception au sein du lexique demande une autre catégorie syntaxique du fait de leur rejet en début de phrase et une autre structure sémantique si l'on veut les introduire de la même manière que nous avons procédé jusqu'à maintenant.

(8.8) En 1828, Azaïs quittait Causerets.

obtiendra la structure suivante :

ADVDate(en1828)Azaïs (PRES)(SYN)(IMP) quitter Causerets

qui impose une structure que l'on ne peut résoudre sans dépendance, l'adverbe ne pouvant intervenir dans la preuve qu'entre l'opérateur IMP et l'opérateur SYN, on ne peut résoudre cet énoncé.

Prenons maintenant deux exemples du corpus Itipy :

(8.9) Les sapins s'arrêtent à dix heures.

(8.10) Le chemin descend pendant deux heures.

Le premier exemple indique que le voyageur ne voit plus de sapins à partir de dix heures, ici le sujet syntaxique ne peut être interprété comme étant le sujet logique et thématique du prédicat s'arrêter, néanmoins, il en faut pas non plus interpréter cette phrase par le prédicat *s_arrêter*(voyageur). Ici, tout comme l'ont proposé Moot et Rétoré dans [Moot *et al.*, 2011] le glissement sémantique nécessite l'utilisation d'une implication entre prédicats dont l'argument est un voyageur dans la forme sémantique, ce qui fait apparaître un voyageur fictif pour réaliser le déplacement évoqué.

Par ailleurs, nous n'abordons pas dans ce système d'opérateur de la modalité et ne prenons pas en compte dans notre modèle d'interprétation ce qui est attesté de ce qui ne l'est pas, outre le paradoxe imperfectif. Cette dimension pourtant cruciale en extraction d'information nécessite une couche supplémentaire dans la représentation sémantique de la temporalité événementielle, ce qui constitue une perspective intéressante à ce travail.

8.3 Limites du système

8.3.1 Structure discursive et sous-spécification

Dans notre proposition d'adaptation du système *Binary Tense*, un problème persiste sur la sélection d'une sémantique ou d'une autre pour les temps bénéficiant d'un emploi aoriste : le passé composé, le plus-que-parfait, le futur antérieur et le conditionnel passé 1ère forme. Effectivement, dans notre tableau des combinaisons possibles, ces temps peuvent revêtir deux sémantiques possibles :

(8.11) - Pourrons-nous, dans cette Sibérie, réparer l'insomnie de la nuit dernière ?
 - Oui, oui ! Venez, j'ai découvert un dortoir.

(8.12) je les ai traversées en 1828 à Figuières

Dans ces deux exemples, le passé composé revêt deux valeurs distinctes, la première étant une forme perfective du présent tandis que la seconde est une forme aoriste du passé.

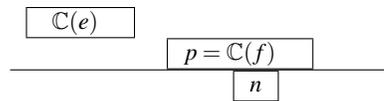


FIGURE 8.6 – Interprétation possible des variables temporelles pour *(Venez,) j'ai découvert un dortoir*

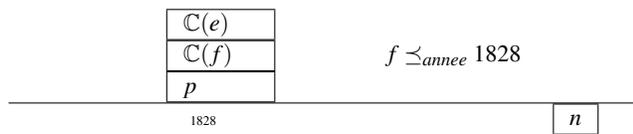


FIGURE 8.7 – Interprétation possible des variables temporelles pour *je les ai traversées en 1828 à Figuières*

Dans ce cas précis, l'utilisation du présent d'énonciation *Venez* et d'une datation *en 1828* permettrait de résoudre l'ambiguïté aspectuelle et temporelle du passé composé, cette ambiguïté nous force à garder les deux interprétations, on parle alors de sous-spécification.

Malgré des marqueurs linguistiques permettant d'analyser la sémantique de l'emploi, ce type de calcul sort du cadre de nos travaux et il semble judicieux d'opter pour des analyses discursives prenant en compte une unité plus grande que le segment pour le résoudre. Nous pensons notamment à la SDRT et aux séquences textuelles présentées dans le chapitre *L'itinéraire et le récit de voyage au XIXème siècle* de ce présent manuscrit.

Nous n'avons pas pu dans le cadre de ce travail présenter des opérateurs pour les subjunctifs, qui demande une prise en compte du mode et de la modalité de manière générale.

Dans un registre plus accessible, nous n'avons eu le temps de traiter les participes/gérondifs, les impératifs et les infinitifs qui présentent une temporalité intéressante dans le discours.

Comme nous l'avons annoncé concernant les conditionnels, nous désirons représenter la temporalité des éventualités même si elle n'ont pas été attestées, nous pensons aux impératifs, aux éventualités logées dans des questions et à celles sous la portée d'une négation. Pour ces trois cas de figure, nous pensons qu'une extension temporelle peut être manipulée, que l'on soit face à une éventualité ou une absence d'éventualité. Par exemple :

(8.13) Pour raccourcir, pourquoi n'irions-nous pas tout droit au Marmurè ?

L'éventualité portée par le verbe *aller* ici est sous la portée d'une négation et d'une interrogation, néanmoins, il semble bien là qu'un objet doit être représenté, plus précisément :

(8.14) -Mais nous grimperons donc toujours ? disais-je à mon guide ; j'ai les mains et les pieds déchirés ; il me semble qu'il est temps que cela finisse !

(8.15) - Courage ! messieurs, répondait-il, mais ne marchez pas si près du bord ; dans quelques instants, nous y serons.

Il est évident ici que l'extension temporelle de *grimper* (dans une interrogation) est en relation avec l'extension temporelle de *marcher* (sous la portée d'une négation et de forme impérative). Ce sont des objets dont on peut discuter les propriétés et avec dont la référence jonche le discours (*ceci*). Faire abstraction de ceux-ci semble aberrant dans une logique du discours.

Une autre catégorie retient aujourd'hui notre attention, c'est la classe des substantifs, qu'ils soient déverbaux ou non.

(8.16) A quatre heures, nous commençâmes à gravir le premier escarpement du Néthou ; une heure et demie d'ascension pénible nous amena à une espèce de moraine (...)

En effet, les nominaux évènementiels sont des objets dont on peut interroger l'extension temporelle et la manipuler comme c'est le cas pour *une heure et demie d'ascension pénible*, jusqu'à la mettre en relation avec d'autres objets temporels.

8.3.2 Perspectives : l'inchoatif et le terminatif et l'itératif

La première perspective évidente de ce travail est le traitement de l'aspect tel que l'inchoatif, dans l'acception suivante : qui montre le début d'une éventualité ; ou terminatif : qui en montre plutôt la fin. Nous argumentons pour un traitement temporel équivalent des verbes portant en suffixe ou en préfixe cette information (nous pensons à *s'élançer*, *s'enfuir*) et des verbes modaux permettant de mettre en lumière de tels aspects comme *se mettre à*, *commencer à*, *s'arrêter de* ou *terminer*. Si l'aspect inchoatif/terminatif est porté par le temps lui même et son contexte discursif, tout comme pour la sous-spécification du passé composé nous supposons que la résolution de cette ambiguïté relève d'un traitement à l'échelle du discours et non du segment tel que nous le réalisons ici. Par exemple :

(8.17) Azaïs gravît la montagne, les guides l'abandonnèrent.

(8.18) Azaïs gravît la montagne, redescendit à Cauterets, il était fatigué.

Ces deux exemples montrent la difficulté à obtenir une interprétation automatiquement de l'aspect inchoatif présenté dans le second exemple. Cette interprétation est due au fait que la seconde partie du segment est soit en relation d'*elaboration* qui implique une imbrication temporelle de la seconde éventualité dans la première, ou pour le premier cas, une relation de *narration* simple, qui implique cette fois-ci une succession des éventualités. En effet, sont en jeu dans la résolution de l'ambiguïté d'une part une forme de raisonnement spatial, et d'une autre, une forme de connaissance du monde.

Les périphrases telles que *venir de* et *aller* placées avant un verbe de déplacement à l'infinitif sont toutes deux portées par des verbes soit *aller* et *venir* qui forcent une interprétation.

(8.19) Azaïs vient de quitter Cauterets.

(8.20) Azaïs va quitter Cauterets.

Les deux présents utilisés ont une influence aspectuelle sur l'éventualité portée par *quitter Cauterets*. En effet, pour le premier exemple, on observe un usage concurrent au passé composé, dans son interprétation perfective du présent (PRES SYN PERF), et on observe que la perfectivité est forcée, c'est par ailleurs le cas pour *Azaïs venait de quitter Cauterets* qui sera interprété comme un PAST SYN PERF au lieu d'un imparfait classique (PAST SYN IMP). L'interprétation temporelle de cet exemple est équivalente à celle de *Azaïs a (juste) quitté Cauterets*.

Pour *va quitter Cauterets*, de notre point de vue, on a à faire à un futur simple, ou dans le système un PRES POST IMP, ici c'est la postériorité qui est forcée, on observe le même phénomène dans *Azaïs allait quitté Cauterets*, cette fois ci c'est un futur antérieur, PAST POST IMP, qui sera choisit ¹.

Dans le cas de l'ambiguïté sémantique du passé composé, la sélection de la sémantique appropriée est déclenchée en partie par des marqueurs dans le contexte (*Venez* ou encore *en 1828*), c'est aussi le cas des aspects inchoatif et terminatif imposés par les périphrases *venir de* ou *aller* utilisés comme des modificateurs temporels de l'éventualité portée par le verbe à l'infinitif auquel ils sont apposés. Nous ne donnons pas ici de représentation sémantique à ces deux mécanismes mais nous pouvons dire qu'ils dominent l'interprétation des éventualités, soit dans le choix de la représentation soit dans la sélection d'une partie de l'éventualité. La représentation sémantique en ensembles finis d'intervalles propose alors une représentation fidèle à ce que nous désirons représenter, l'itératif étant alors un ensemble de plus de un élément, tandis que l'inchoatif et le terminatif devraient sélectionner des sous-ensembles d'intervalles ayant des propriétés particulières. Nous donnons un aperçu plus détaillé des limites de nos travaux dans le chapitre traitant de l'évaluation du lexique.

1. Quoi qu'il en soit les deux formes indiquent une notion quantitative restreinte, une notion de "proche" respectivement dans le passé et le futur. En parallèle avec la sélection forcée d'une composition sémantique temporelle pour ces formes il y aurait à prendre en compte la notion sémantique de ce "proche" qui aurait pour but de restreindre la distance entre *n* et *e* pour *va arriver* si on le considère comme un aspect à part entière ou entre *n* et *f* si c'est une localisation de la *fenêtre* qui est en jeu.

Notre interprétation sémantique du temps et de l'aspect nous permet d'attribuer un intervalle temporel à une éventualité assez finement mais certains phénomènes linguistiques ne sont pas traités jusqu'alors, nous pensons notamment à l'itératif.

Observons cet exemple du corpus :

- (8.21) Mon guide faisait un pas, regardait son chemin, se retournait vers moi, et nous avançons ainsi fort lentement, nous accrochant sur une immense muraille de rochers à toutes les saillies qui s'offraient à nous ;

La quatre premières éventualités de cet exemple sont exprimées à l'imparfait, pourtant les trois premières présentent des éventualités itératives, tandis que la quatrième englobe les trois précédentes. Ici les modifieurs ne peuvent nous aider à résoudre cette sous-spécification, et le type des éventualités est soit un évènement pour *faire un pas*, *se retourner vers quelqu'un*, soit une activité pour *regarder le chemin*. Le facteur supplémentaire permettant de déduire l'aspect de l'éventualité est la relation discursive. *Ainsi* provoque l'englobement temporel des trois éventualités précédant celle qu'il modifie. En effet la relation de discours ici utilisée est ce qu'on pourrait interpréter comme une élaboration inversée. Ces considérations sortent du cadre de nos travaux mais l'exemple suivant montre une autre forme d'itération plus précise et portée par l'adverbial *deux fois*. Cet exemple exemples marquent la nécessité de traiter l'itératif.

- (8.22) Deux fois nous parcourûmes ainsi tout le bord de l'effrayant balcon qui domine la vallée de Béousse.

Nous voudrions obtenir l'interprétation présentée en figure 8.8.

Nous voyons ici que la structure proposée dans le modèle d'interprétation et dans

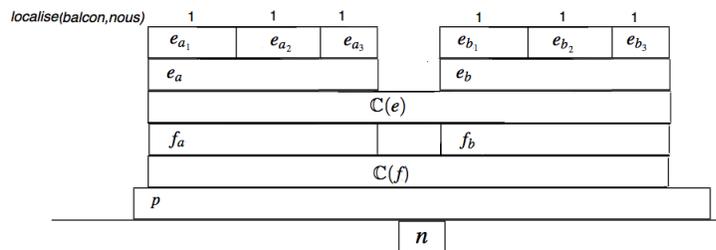


FIGURE 8.8 – Interprétation possible de *Deux fois nous parcourûmes ainsi tout le bord de l'effrayant balcon qui domine la vallée de Béousse*

le lexique laisse la possibilité de traiter ensuite l'itératif, chose que nous aurions aimé avoir le temps de mettre en oeuvre.

Nous montrons pour l'exemple *Deux fois nous parcourûmes ainsi tout le bord de l'effrayant balcon qui domine la vallée de Béousse* que l'itération donne aux ensembles f et e deux intervalles chacun f_a , e_a , f_b et e_b et conserve la relation entre eux pour chaque couple. L'éventualité et la perspective posées sur elle sont dupliquées tandis

que l'interface énonciativo-temporelle n'est pas affectée, nous ne conservons qu'un seul p en relation avec n . Nous désirons interpréter les intervalles multiples imposés par l'itératif comme formant une seule éventualité complexe composée de deux couples à l'interface aspectuo-temporelle (e_a et f_a ainsi e_b et f_b), et d'un seul couple à l'interface énonciativo-temporelle (p et n).

8.4 Conclusion

Nous avons montré dans ce dernier chapitre l'ensemble des énoncés couverts par le lexique et les cas qu'il reste à traiter pour obtenir une analyse automatique efficace de la temporalité dans le discours du français. Beaucoup de choses restent à faire telles que la résolution de l'ambiguïté entre *perfect* et aoriste ou encore le traitement des aspects inchoatif ou terminatif. Ce travail trouve des perspectives dans l'analyse discursive offrant des outils pour traiter les unités plus larges que les segments et qui permettent de saisir l'impact de la structure discursive sur la structure temporelle.

D'un point de vue surfacique, les représentants de la temporalité en langue française sont divers et variés, et même si elle est centrale, la sémantique verbale ne suffit pas à traiter la temporalité au sein d'un discours. Les adverbiaux quant à eux présentent une diversité équivalente que nous n'avons pas abordée, et ce travail constitue un bon point de départ pour le traitement d'autres formes de temporalité, modalisée ou non, substantivée ou non.

Conclusion

Le présent manuscrit de thèse relate une partie des travaux menés en sémantique formelle dans le cadre du projet ITIPY. Suite à la numérisation d'un grand nombre d'oeuvres littéraires sur la région paloise par la Médiathèque de Pau, la volonté de valoriser cette ressource a permis de mettre en place le projet ITIPY, qui vise l'extraction automatique d'itinéraires à partir de récits de voyage du XIX^{ème} et du début du XX^{ème} siècle. Nos travaux de thèse se rapprochent de la compréhension de texte en extraction d'information et nous cherchons à construire une représentation du sens des énoncés, en nous appuyant sur le principe de compositionnalité de la sémantique formelle. Plus précisément, l'extraction d'itinéraire revient à ordonner temporellement les événements (que nous nommons éventualités) de déplacement et de localisation d'un voyageur tel qu'on peut se le représenter à travers le discours.

Cette thèse s'intéresse à la construction d'un lexique sémantique de la temporalité de l'éventualité en français et à leur interprétation. Nous avons par ailleurs caractérisé notre corpus comme échantillon du français, ce qui a nécessité une étude quantitative et qualitative dans laquelle les onze textes constituant ce corpus ont été confrontés à d'autres genres discursifs, ainsi qu'une approche de la structure interne de ce genre de discours afin d'observer le comportement du récit de l'itinéraire au sein de celui-ci. Cette étude a permis de mettre en avant le récit de l'itinéraire comme la colonne vertébrale du récit de voyage du XIX^{ème} siècle. Après avoir exploré plusieurs pistes afin de traiter ce type de discours au mieux, nous avons orienté notre choix sur l'analyse sémantique du segment, obtenue suite à l'analyse syntaxique profonde de celui-ci, le segment étant une unité linguistique cohérente et minimale d'un point de vue propositionnel. Au sein de ce segment se loge le verbe, forme d'expression majeure de l'éventualité de déplacement. Nous avons abordé dans la majeure partie de ce manuscrit comment traiter la temporalité des éventualités exprimées par le langage et au sein du discours.

Notre approche en sémantique formelle permet d'obtenir automatiquement une représentation sémantique fine. L'analyse syntaxique sur laquelle nous nous appuyons est réalisée dans le cadre des grammaires catégorielles, telle que proposée par *Grail*. Grâce à la transparence de l'interface-sémantique entre grammaire catégorielle et λ -

calcul, nous obtenons immédiatement une représentation sémantique en DRT de la phrase. L'élaboration d'un lexique attaché au temps verbal utilisé et à quelques adverbiaux temporels permet donc d'accéder à une interprétation de la temporalité des éventualités portées par les verbes conjugués. Ce traitement entre totalement dans les objectifs du projet ITPY pour le traitement plus spécifique des segments ayant trait à l'expression du déplacement et de la localisation géographique dans le cadre du récit d'itinéraires dans des récits de voyage. L'extraction d'informations en sémantique formelle permet d'obtenir des informations automatiquement sans restriction sur la forme des énoncés, le typage garantissant des expressions bien formées et permettant d'enrichir les représentations obtenues, nous pensons par exemple à [Moot *et al.*, 2011] sur l'introduction d'un voyageur fictif utile à l'interprétation d'un énoncé tel que *la route descend pendant deux heures* dans le cadre de la théorie des types et adapté au traitement dans *Grail*, puis [Retoré *et al.*, 2013] sur le traitement des pluriels en sémantique formelle, ou encore [Real-Coelho et Retoré, 2013] sur les déverbaux.

La contribution centrale de ces travaux s'inscrit dans le traitement sémantique du temps et de l'aspect de l'éventualité portée par le verbe conjugué et à leur modélisation, proposant une adaptation et une extension de la théorie *Binary Tense* de Henk Verkuyl (2008) au français du XIX^{ème} siècle. Notre proposition est un modeste maillon dans la lignée de nombreux travaux sur la question dont nous tentons de présenter quelques uns des plus pertinents.

En adaptant le système au français, il nous a été permis de montrer sa pertinence tant pour traiter du français du XIX^{ème} siècle que le français contemporain. Nous avons étendu ce système afin de traiter certains phénomènes tels que l'aspect aoriste du passé simple et les usages aoristes des temps du parfait représenté en français par les formes composées. Nous avons ajouté au lexique étendu quatre classes d'adverbiaux essentiels à l'expression de la temporalité événementielle. Néanmoins, cette proposition n'est pas exhaustive et réclame certains approfondissements au sujet des adverbiaux particulièrement, puis au sujet d'autres formes d'aspect tels que l'itératif, l'inchoatif ou le terminatif par exemple. Nous ajoutons que l'étude des relations temporelles entre les segments ainsi que la résolution de la sous-spécification de certains temps est une suite naturelle à cette première étape dans l'analyse profonde de la sémantique temporelle du français. Par ailleurs, le passage de l'énoncé comportant une éventualité seule au discours qui présente un maillage référentiel complexe est discuté et permet de définir la limite de ce présent travail.

Bibliographie

- [Abeillé *et al.*, 2003] ABEILLÉ, A., CLÉMENT, L. et TOUSSENEL, F. (2003). Building a treebank for french. *In Treebanks*. Kluwer.
- [Adam, 1992] ADAM, J.-M. (1992). *Les textes : types et prototypes*. Nathan-Université.
- [Adam, 2008] ADAM, J.-M. (2008). Note de cadrage sur la linguistique textuelle. *Congrès Mondial de Linguistique Française 2008*.
- [Adjukiewicz, 1935] ADJUKIEWICZ, K. (1935). Die syntaktische Konnexität. *Studia Philosophica*, 1:1–27.
- [Allen, 1983] ALLEN, J. F. (1983). Maintaining knowledge about temporal intervals. *In Communications of the ACM*, numéro 26(11), pages 832–843.
- [Amsili et Bras, 1998] AMSILI, P. et BRAS, M. (1998). DRT et compositionnalité. *T.A.L.*, 39(1):131–160.
- [Asher, 2011] ASHER, N. (2011). *Lexical Meaning in context – a web of words*. Cambridge University Press.
- [Asher *et al.*, 2008] ASHER, N., MULLER, P. et GAIO, M. (2008). Spatial entities are temporal entities too : the case of motion verbs. *In Methodologies and Resources for Processing Spatial Language*.
- [Aurnague, 2008] AURNAGUE, M. (2008). Qu’est-ce qu’un verbe de déplacement ? critère spatiaux pour une classification des verbes de déplacement intransitifs du français. *In Actes du Congrès mondial de Linguistique Française, CMLF’08*.
- [Bach, 1981] BACH, E. W. (1981). Time, Tense, and Aspect : An Essay in English Metaphysics . *In Radical Pragmatics*, pages 63–81.
- [Bachtine, 1970] BACHTINE, M. M. (1970). *La poétique de Dostoïevski*. Seuil.
- [Bar-Hillel, 1953] BAR-HILLEL, Y. (1953). A quasi arithmetical notation for syntactic description. *Language*, 29:47–58.
- [Battistelli, 2009] BATTISTELLI, D. (2009). *La temporalité linguistique : Circonscrire un objet d’analyse ainsi que les finalités à cette analyse*. Mémoire d’habilitation à diriger des recherches, Paris Ouest Nanterre La Défense.
- [Benveniste, 1966] BENVENISTE, E. (1966). *Problèmes de linguistique générale*, volume 1. Gallimard.

- [Biber *et al.*, 1998] BIBER, D., CONRAD, S. et REPPEN, R. (1998). *Corpus Linguistics. Investigating language structure and use*. Cambridge University Press.
- [Black, 1960] BLACK, M. (1960). *Translations from the Philosophical Writings of Gottlob Frege*, chapitre On Sense and Reference (4), pages 56–78. Basil Blackwell.
- [Boons, 1987] BOONS, J.-P. (1987). La notion sémantique de déplacement dans une classification syntaxique des verbes locatifs. *Langue française*, 76:5–40.
- [Caudal, 2006a] CAUDAL, P. (2006a). Aspect. In GODARD, D., ROUSSARIE, L. et CORBLIN, F., éditeurs : *Sémanticopédie : dictionnaire de sémantique*. GDR Sémantique & Modélisation, CNRS, <http://www.semantique-gdr.net/dico/>.
- [Caudal, 2006b] CAUDAL, P. (2006b). The aspectual contribution of tenses and the semantics/pragmatics interface. In *UT Discourse Workshop*.
- [Caudal et Veters, 2007] CAUDAL, P. et VETTERS, C. (2007). Passé composé et passé simple : Sémantique diachronique et formelle. *Cahiers Chronos*, 16:121–151.
- [Charolles, 1995] CHAROLLES, M. (1995). Cohésion, Cohérence et pertinence du discours. *Travaux de Linguistique*, 29:125–151.
- [Charolles, 1997] CHAROLLES, M. (1997). L'encadrement du discours, UNIVERS, CHAMPS, DOMAINES ET ESPACES. *Cahier de Recherche Linguistique*, pages 1–60.
- [Chomsky, 1963] CHOMSKY, N. (1963). Formal properties of grammars. In *Handbook of Mathematical Psychology*, volume 2, pages 323 – 418. Wiley, New-York.
- [Cori et Léon, 2002] CORI, M. et LÉON, J. (2002). La constitution du TAL. *Traitement Automatique des Langues*, 43(3):21–55.
- [Culioli, 1999] CULIOLI, A. (1999). *Pour une linguistique de l'énonciation*. Orphys.
- [Davidson, 1967] DAVIDSON, D. (1967). The logical form of action sentences. In RESCHER, N., éditeur : *The Logic of Decision and Action*, pages 81–120. Univ. of Pittsburgh Press.
- [De Saussure, 1997] DE SAUSSURE, L. (1997). « le temps chez beauzée : algorithmes de repérage et comparaison avec reichenbach ». *Cahiers Ferdinand de Saussure*, (49):171–195.
- [Desclés et Guentchéva, 2000] DESCLÉS, J.-P. et GUENTCHÉVA, Z. (2000). Énonciateur, locuteur, médiateur. In *Les Rituels du dialogue*. Société d'ethnologie.
- [Ducrot, 1994] DUCROT, O. (1994). *Le dit et le dire*. Editions de minuit.
- [Dufiez-Sanchez, 2010] DUFIEZ-SANCHEZ, V. (2010). *Philosophie du roman personnel de Chateaubriand à Fromentin 1802-1863*. Genève : Droz.
- [Enjalbert, 2005] ENJALBERT, P. (2005). *Sémantique et traitement automatique du langage naturel*. Lavoisier, Paris.
- [Flageul, 1997] FLAGEUL, V. (1997). *Description sémantico-cognitive des prépositions spatiales du français*. Thèse de doctorat, Paris-Sorbonne (Paris-IV).
- [Gasparini, 2004] GASPARINI, P. (2004). *Est-il Je ? Roman autobiographique et autofiction*. Le Seuil.
- [Genette, 1972] GENETTE, G. (1972). *Figures III*. Seuil, Paris.

- [Gosselin, 1996] GOSSELIN, L. (1996). *Sémantique de la temporalité en français : un modèle calculatoire et cognitif du temps et de l'aspect*. Duculot.
- [Gosselin, 2005] GOSSELIN, L. (2005). *Temporalité et modalité*. De Boeck supérieur.
- [Gosselin, 2010] GOSSELIN, L. (2010). *Interpréter les temps verbaux*, chapitre Les relations entre périphrases aspectuelles et conjugaisons en français. Peter Lang.
- [Howard, 1980] HOWARD, W. A. (1980). The formulae-as-types notion of construction. In HINDLEY, J. et SELDIN, J., éditeurs : *To H.B. Curry : Essays on Combinatory Logic, λ -calculus and Formalism*, pages 479–490. Academic Press.
- [Kamp, 1981] KAMP, H. (1981). Événements, représentations discursives et référence temporelle. *Langages*, 15(64):39–64.
- [Kamp et Reyle, 1993a] KAMP, H. et REYLE, U. (1993a). *From Discourse to Logic*. Kluwer Academic Publishers.
- [Kamp et Reyle, 1993b] KAMP, H. et REYLE, U. (1993b). *From Discourse to Logic*. D. Reidel, Dordrecht.
- [Kamp et al., 2011] KAMP, H., van GENABITH, J. et REYLE, U. (2011). Discourse representation theory. *Handbook of Philosophical Logic*, 15.
- [Kerbrat-Orecchioni, 1999] KERBRAT-ORECCHIONI, C. (1999). *L'énonciation*. Armand Colin.
- [Kripke, 1999] KRIPKE, S. (1999). *La logique des noms propres (Naming and Necessity)*. Les Editions de Minuit.
- [Lambek, 1958] LAMBEK, J. (1958). The mathematics of sentence structure. *American mathematical monthly*, pages 154–170.
- [Lascarides et Asher, 1993] LASCARIDES, A. et ASHER, N. (1993). Temporal Interpretation, Discourse Relations, and Commonsense Entailment. *Linguistics and Philosophy*, 16:437–493.
- [Lefevre et al., 2012] LEFEUVRE, A., MOOT, R., RETORÉ, C. et SANDILLON-REZER, N.-F. (2012). Traitement automatique sur corpus de récits de voyages pyrénéens : Une analyse syntaxique, sémantique et temporelle. In *TALN'2012*.
- [Lefevre et Vinogradova, 2012] LEFEUVRE, A. et VINOGRADOVA, N. (2012). Hétérogénéité et extraction d'information factuelle dans un corpus de récits de voyage. *Langages*, (187):127–144.
- [Loustau, 2008] LOUSTAU, P. (2008). *Interprétation automatique d'itinéraires dans des récits de voyages . D' une information géographique du syntagme du syntagme à une information géographique du discours*. Thèse de doctorat, Université de Pau et des Pays de l'Adour.
- [Magri-Mourgues, 1996] MAGRI-MOURGUES, V. (1996). La description dans le récit de voyage. *Cahier de narratologie*, 7:35–48.
- [Magri-Mourgues, 2009] MAGRI-MOURGUES, V. (2009). *Le voyage à pas comptés. Pour une poétique du récit de voyage au XIXème siècle*. Numéro 9 de Lettres numériques. Honoré Champion.
- [Maingueneau, 1991] MAINGUENEAU, D. (1991). *L'analyse du discours*. Hachette.

- [Mery, 2011] MERY, B. (2011). *Modélisation de la sémantique lexicale dans le cadre de la théorie des types*. Thèse de doctorat, Université de Bordeaux 1.
- [Moens et Steedman, 2005] MOENS, M. et STEEDMAN, M. (2005). Temporal ontology and temporal reference. In MANI, I., PUSTEJOVSKY, J. et GAIZAUSKAS, R., éditeurs : *The Language of Time : A Reader*, chapitre 6. Oxford Linguistics.
- [Moot, 1998] MOOT, R. (1998). Grail : An automated proof assistant for categorial grammar logics. In BACKHOUSE, R., éditeur : *Proceedings of the 1998 User Interfaces for Theorem Provers Conference*, pages 120–129.
- [Moot, 1999] MOOT, R. (1999). Grail : an interactive parser for categorial grammars. In DELMONTE, R., éditeur : *Proceedings of VEXTAL'99*, pages 255–261, Venice. University Cà Foscari.
- [Moot, 2001] MOOT, R. (2001). A short introduction to Grail. In ARECES, C. et de RIJKE, M., éditeurs : *Proceedings of Methods for Modalities 2*.
- [Moot, 2010a] MOOT, R. (2010a). Semi-automated extraction of a wide-coverage type-logical grammar for French. In *Proceedings of Traitement Automatique des Langues Naturelles (TALN)*, Montreal.
- [Moot, 2010b] MOOT, R. (2010b). Wide-coverage French syntax and semantics using Grail. In *Proceedings of Traitement Automatique des Langues Naturelles (TALN)*, Montreal.
- [Moot et al., 2011] MOOT, R., PRÉVOT, L. et RETORÉ, C. (2011). Un calcul de termes typés pour la pragmatique lexicale. In *Traitement Automatique du Langage Naturel, TALN 2011*, pages 161–166, Montpellier.
- [Moot et Retoré, 2012] MOOT, R. et RETORÉ, C. (2012). *The Logic of Categorial Grammars : A deductive Account of Natural Language Syntax and Semantics*. Springer.
- [Nicolas, 2006] NICOLAS, D. (2006). Compositionnalité : questions philosophiques. In GODARD, D., ROUSSARIE, L. et CORBLIN, F., éditeurs : *Sémantico-pédagogie : dictionnaire de sémantique*. GDR Sémantique & Modélisation, CNRS, <http://www.semantique-gdr.net/dico/>.
- [Partee, 1996] PARTEE, B. H. (1996). Montague grammar. In van BENTHEM, J. et ter MEULEN, A., éditeurs : *Handbook of Logic and Language*, chapitre 1, pages 5–91. North-Holland Elsevier, Amsterdam.
- [Pasquali, 1994] PASQUALI, A. (1994). *Le Tour des horizons : critique et récit de voyage*. Klincksieck.
- [Pery-Woodley et Scott, 2006] PERY-WOODLEY, M.-P. et SCOTT, D. (2006). Computational approaches to discourse and document processing. *Traitement Automatique des Langues*, 47(2):7–19.
- [Prior, 1967] PRIOR, A. (1967). *Past, Present and Future*. Oxford University Press.
- [Pustejovsky, 1998] PUSTEJOVSKY, J. (1998). *The Generative Lexicon*. The MIT Press, Cambridge, Massachusetts.
- [Pustejovsky, 2005] PUSTEJOVSKY, J. (2005). The syntax of event structure. In MANI, I., PUSTEJOVSKY, J. et GAIZAUSKAS, R., éditeurs : *The Language of Time : A Reader*, chapitre 2. Oxford Linguistics.

- [Ramsay, 2005] RAMSAY, A. (2005). *Discourse*, chapitre 6, pages 112–135. Oxford University Press, USA, Oxford.
- [Real-Coelho et Retoré, 2013] REAL-COELHO, L.-M. et RETORÉ, C. (2013). A Generative Montagovian Lexicon for Polysemous Deverbal Nouns. *In 4th world congress on Universal Logic UNILOG*, pages <http://www.uni-log.org/>, Rio de Janeiro, Brésil. Accepted for publication in the proceedings of 4th congress on Universal Logic — Logic and linguistics workshop. Rio de Janeiro April 4-7 2013.
- [Recanati et Recanati, 1999] RECANATI, C. et RECANATI, F. (1999). La classification de Vendler revue et corrigée. *Chronos*, 4:167–184.
- [Recanati, 2001] RECANATI, F. (2001). Are 'here' and 'now' indexicals? *Texte*, 127/8:115–127.
- [Reichenbach, 1947] REICHENBACH, H. (1947). *Elements of Symbolic Logic*. The Mac millan Company.
- [Retoré, 2000] RETORÉ, C. (2000). Systèmes déductifs et traitement des langues : un panorama des grammaires catégorielles. *Technique et Science Informatiques*, 20(3):301–336. Numéro spécial *Traitement Automatique du Langage Naturel* sous la direction de D. Kayser et B. Levrat. Version préliminaire RR-3917 <http://www.inria.fr/>.
- [Retoré et al., 2013] RETORÉ, C., MOOT, R. et MERY, B. (2013). Plurals : individuals and sets in a richly typed semantics. *In Logic and Engineering of Natural Language Semantics*, numéro 10. LENLS 10.
- [Rey et Rey-Debove, 1993] REY, A. et REY-DEBOVE, J., éditeurs (1993). *Le Nouveau Petit Robert*. Dictionnaire Le Robert.
- [Roberts, 1985] ROBERTS, I. (1985). *The Representation of Implicit and Dethematized Subjects*. Thèse de doctorat, Los Angeles : University of Southern California.
- [Roussarie, 2006] ROUSSARIE, L. (2006). Sens et dénotation. *In GODARD, D., ROUSSARIE, L. et CORBLIN, F., éditeurs : Sémanticlopédie : dictionnaire de sémantique*. GDR Sémantique & Modélisation, CNRS, <http://www.semantique-gdr.net/dico/>.
- [Saint-Dizier, 2006] SAINT-DIZIER, P. (2006). Rôles thématiques. *In GODARD, D., ROUSSARIE, L. et CORBLIN, F., éditeurs : Sémanticlopédie : dictionnaire de sémantique*. GDR Sémantique & Modélisation, CNRS, <http://www.semantique-gdr.net/dico/>.
- [Salem, 1991] SALEM, A. (1991). Lexico1-logiciel pour l'analyse lexicométrique des textes. Rapport technique, E.N.S. de Fontenay-Saint-Cloud.
- [Treikelder, 2006] TREIKELDER, A. (2006). Remarques sur les valeurs du passé composé en ancien français. *In Skandinaavia romanstide*, numéro 16.
- [Vandeloise, 1986] VANDELOISE, C. (1986). *L'espace en français*. Seuil.
- [Vendler, 1967] VENDLER, Z. (1967). *Linguistics in philosophy*. Cornell University Press.
- [Verkuyl, 2003] VERKUYL (2003). On the compositionality of tense : Merging Reichenbach and Prior. Utrecht University.

BIBLIOGRAPHIE

[Verkuyl, 2008] VERKUYL, H. (2008). *Binary Tense*. CSLI Publications.

[Vine, 2009] VINE, E. W. (2009). Discourse on the move : Using corpus analysis to describe discourse structure. *Discourse Studies*, 11(1):123 – 125.

Résumé et mots-clefs

Titre

Sémantique des temps du français : une formalisation compositionnelle

Résumé

Cette thèse s'inscrit dans le cadre du projet Région Aquitaine - INRIA : ITIPY : Extraction automatique d'itinéraires dans des récits de voyage et application à la recherche documentaire « intelligente » dans des fonds patrimoniaux liés au territoire aquitain. Suite à la numérisation d'un grand nombre d'oeuvres littéraires sur la région paloise par la Médiathèque de Pau, ce projet vise à terme l'extraction automatique d'itinéraires à partir de récits de voyage du XIXème et du début du XXème siècle. Notre premier travail fut de caractériser le corpus comme échantillon du français, par une étude contrastive d'une part de données quantitatives et d'autre part de la structure des récits de voyage. Nous nous sommes ensuite consacrée à l'étude du temps, et plus particulièrement à l'analyse automatique de la sémantique des temps verbaux du français. Disposant d'un analyseur syntaxique et sémantique à large échelle du français, basé sur les grammaires catégorielles et la sémantique compositionnelle (λ -calcul ou λ -DRT), notre tâche a été de prendre en compte les temps des verbes pour reconstituer la temporalité des événements et des états, notions que nous regroupons sous le termes d'éventualité. Cette thèse se concentre sur la construction d'un lexique sémantique traitant des temps verbaux du français. Nous proposons une extension et une adaptation d'un système d'opérateurs compositionnels conçu pour les temps du verbe anglais, aux temps et à l'aspect du verbe français du XIXème siècle à nos jours. Cette formalisation est *de facto* opérationnelle, car elle est définie en terme d'opérateurs du λ -calcul dont la composition et la réduction, déjà programmées calculent automatiquement les représentations sémantique souhaitées, des formules multisortes de la logique d'ordre supérieur. Le passage de l'énoncé comportant une éventualité seule au discours, dont le maillage référentiel est plus complexe, est discuté dans le cadre de la λ -DRT, et nous concluons notre travail par les perspectives qu'ouvre notre formalisation pour l'analyse du discours.

Mots-Clefs

Sémantique formelle - Temporalité des éventualités - Lexique compositionnel - Récits de voyage - Lambda-calcul simplement typé

Title

French Tenses Semantics : a compositionnal formalisation

Abstract

This manuscript is a part of the work that have been done in formal semantics in the frame of the ITIPY project. After the digitalization of numerous literary artworks about the region around Pau by the Médiathèque de Pau, the will to valorize this resource allowed to set up the project ITIPY for Itinéraires Pyrénénes (Pyrenean Itineraries), which goal was to automatically extract itineraries from travel novels from the XIXth century and from the beginning of the XXth. Our thesis work are close to text understanding task in information retrieval and we aim at building a representation of meaning of linguistic utterances, leaning on the compositionnality principle in formal semantics. More precisely, the itinerary extraction supposes to temporally represent displacement and localization events (that we actually call eventuality) of a traveler as far as we understand it through discourse.

This work is focused on the building of an events temporality semantic lexicon in French. We actually characterized our corpora as a sample of French language, by the means of a quantitative and qualitative analysis in which its eleven texts have been confronted to other discursive genres, as well as a study of the internal structure of this genre in order to observe how the itinerary behaves through it. The main contribution of this thesis deals with tense and aspect semantic processing of the event expressed by tensed verbs and with its modeling. In this respect, we propose an adaptation and an extension for French from XIXth century to nowadays of the lexicon proposed in Henk Verkuyl's *Binary Tense* (2008). Our contribution is a modest link in numerous works on this question, among which we try to present some of the most relevant ones. Transition from a single event uterrance to a whole discourse which contains a complex referential network is discussed and allows us to define the limits of this hereby work.

Keywords

Formal Semantics - Eventualities temporality - Compositionnal Lexicon - Travel novels - Simply typed λ -calcul