



HAL
open science

Étude des phénomènes itératifs en langue : Inscription discursive et Calcul aspectuo-temporel, vers un traitement automatisé

Julien Lebranchu

► **To cite this version:**

Julien Lebranchu. Étude des phénomènes itératifs en langue : Inscription discursive et Calcul aspectuo-temporel, vers un traitement automatisé. Traitement du texte et du document. Université de Caen, 2011. Français. NNT: . tel-00664788

HAL Id: tel-00664788

<https://theses.hal.science/tel-00664788v1>

Submitted on 31 Jan 2012

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.



université de Caen
Basse-Normandie



UNIVERSITÉ DE CAEN BASSE-NORMANDIE

UFR DE SCIENCES

ÉCOLE DOCTORALE

STRUCTURE, INFORMATION, MATIÈRE ET MATÉRIAUX

THÈSE

présentée par

M. JULIEN LEBRANCHU

et soutenue

le 15 décembre 2011

en vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE CAEN BASSE-NORMANDIE

Spécialité « informatique et applications »

Arrêté du 07 Août 2006

ÉTUDE DES PHÉNOMÈNES ITÉRATIFS EN LANGUE

INSCRIPTION DISCURSIVE ET CALCUL ASPECTUO-TEMPOREL,
VERS UN TRAITEMENT AUTOMATISÉ



LES MEMBRES DU JURY

M. Nicholas ASHER (<i>Directeur de recherche</i>)	IRIT - UNIVERSITÉ PAUL SABATIER	(<i>Rapporteur</i>)
Mme Delphine BATTISTELLI (<i>Maître de conférence</i>)	ISHA - UNIVERSITÉ PARIS SORBONNE	(<i>Rapporteur</i>)
M. Laurent GOSSELIN (<i>Professeur</i>)	LIDIFRA - UNIVERSITÉ DE ROUEN	
M. François LEVY (<i>Professeur</i>)	LIPN - UNIVERSITÉ PARIS XIII	
M. Yann MATHET (<i>Maître de conférence</i>)	GREYC - UNIVERSITÉ DE CAEN	(<i>co-Encadrant</i>)
M. Patrice ENJALBERT (<i>Professeur émérite</i>)	GREYC - UNIVERSITÉ DE CAEN	(<i>Directeur</i>)

Julien Lebranchu : *Étude des phénomènes itératifs en langue, Inscription discursive et Calcul aspectuo-temporel, Vers un traitement automatisé*, thèse de doctorat,
© 2011. Version finale.

Longtemps, je me suis couché de bonne heure.

— Marcel Proust (1913)

REMERCIEMENTS

Je souhaite tout d’abord exprimer à Patrice Enjalbert et Yann Mathet, mes sincères remerciements pour m’avoir fait confiance tout au long de ce projet, de leurs nombreux conseils avisés, et tout particulièrement, ces derniers mois durant ce long travail de rédaction.

Je tiens également à remercier vivement *Delphine Battistelli* et *Nicholas Asher* pour avoir accepté de m’accorder un temps précieux en tant que rapporteurs de cette thèse, ainsi que *Laurent Gosselin* et *François Lévy* de me faire l’honneur de participer au jury.

Je souhaite aussi témoigner de ma sympathie et de ma gratitude aux membres du Greyc avec qui j’ai eu le plaisir de travailler, ou simplement partager une discussion. Je souhaite notamment adresser mes remerciements et toute mon amitié à Thierry Charnois, Stéphane Ferrari, Antoine Widlöcher, Frédéric Bilhaut, et Jean-Luc Manguin.

Un remerciement tout particulier à mes deux acolytes, Julien Gosme et Adrien Lardilleux, pour leurs longues discussions que ce soit autour d’un café ou non, avec une mention spéciale pour Adrien qui a eu le privilège de relire ce manuscrit.

Mes remerciements vont également à toute ma famille, et tout particulièrement à mes parents qui m’ont toujours soutenu et encouragé.

Enfin, je tiens à remercier le rayon soleil de ma vie, mon épouse, Angélique, que je remercie tendrement pour tout l’amour qu’elle me porte, et sa patience, tout particulièrement cette dernière année. Je lui suis infiniment reconnaissant de son soutien, sans lequel rien de ce que j’ai entrepris depuis ces huit dernières années n’aurait pu se réaliser.

SOMMAIRE

INTRODUCTION	1
I LA QUESTION DE LA TEMPORALITÉ ET DES ITÉRATIONS	11
1 LES PHÉNOMÈNES ITÉRATIFS DANS LA LITTÉRATURE	15
2 DES MODÈLES DE LA TEMPORALITÉ	25
II DE LA SDT AUX REPRÉSENTATIONS DISCURSIVES	47
3 L'ITÉRATION DANS LE MODÈLE SDT	51
4 UNE FORMALISATION ALGÈBRIQUE	59
5 L'ITÉRATION SELON UNE APPROCHE OBJET	69
III LES PHÉNOMÈNES ITÉRATIFS EN CORPUS	83
6 OBSERVATION DES PHÉNOMÈNES ITÉRATIFS	87
7 REPRÉSENTATION DES STRUCTURES ITÉRATIVES	113
8 ANALYSEUR DES STRUCTURES ITÉRATIVES	135
IV LES PHÉNOMÈNES ITÉRATIFS ET LE CALCUL ASPECTUEL	157
9 LE CALCUL ASPECTUO-TEMPOREL	161
10 LE SYSTÈME <i>caste</i>	179
11 UN FORMALISME XML POUR LES STRUCTURES ITÉRATIVES	193
CONCLUSION	209
ANNEXES	215
A LEXIQUE DES TYPOLOGIES DE PROCÈS	217
B RÈGLES DE CALCUL ASPECTUO-TEMPOREL	225
C EXTRAIT DE <i>la fille aux yeux d'or</i>	245
D SYNTAXE ET SÉMANTIQUE XML	247
RÉFÉRENCES BIBLIOGRAPHIQUES	253
Bibliographie	255
Table des matières	263
Table des figures	267
Liste des tableaux	270
Liste des règles	271

INTRODUCTION

INTRODUCTION

LES PHÉNOMÈNES ITÉRATIFS

« Il est 23h30, je suis en train d'effectuer un job en ligne pour un site de petites annonces. Aujourd'hui, comme tous les jours du lundi au vendredi, j'étais en stage à RFI. Le week-end, pas de sortie, je me lève à 5h du matin pour rejoindre l'équipe de la matinale de SNCF la radio. »

Extrait d'un Blog de MediaPart (Jessica Chekroun)

Notre expérience quotidienne nous met constamment en présence de *phénomènes itératifs*. L'extrait ci-dessus est une parfaite illustration d'une structure répétitive d'une semaine type où la locutrice se rend tous les jours de la semaine sur son lieu de stage, et le samedi et dimanche sur son second lieu de travail. Cette structure répétitive est inhérente au système calendaire sur lequel reposent nos sociétés (des jours faisant des semaines, des semaines faisant des mois, eux-mêmes constituant des années. . .), mais ne s'y limite pas.

Les travaux dont nous allons faire état dans ce mémoire portent sur l'étude de ces phénomènes. Ils s'inscrivent dans une dynamique de recherche et de collaboration, déjà ancienne, entre Laurent Gosselin et Patrice Enjalbert, centrée sur la sémantique temporelle. Nous retrouvons en premier lieu les travaux sur la sémantique du temps et de l'aspect de Gosselin (1996, 2005). La sémantique de la temporalité (ci-après « *SdT* ») en français est un modèle néo-reichenbachien intégrant les différents paramètres explicatifs de la temporalité verbale en français (aspect et temps). Ce modèle a donné lieu à un travail d'implémentation dans le cadre de la thèse de Cédric Person (2004), co-dirigé par Laurent Gosselin et Patrice Enjalbert. Nous retrouvons également le projet « Ordres de Grandeur et Répétition » proposé par le TCAN¹. Le projet OGRE a concerné certaines facettes de la sémantique temporelle, à savoir la prise en compte des notions d'ordre de grandeur, et les phénomènes itératifs (Lévy *et al.*, 2005).

Dans la continuité de ces travaux, un groupe de travail autour de Laurent Gosselin, Patrice Enjalbert, Gérard Becher et Yann Mathet, et auquel nous avons également participé, s'est formé autour d'une approche pluridisciplinaire des phénomènes itératifs couvrant des orientations linguistiques, lo-

1. Traitement des Connaissances, Apprentissage et Nouvelles Technologies de l'Information et de la Communication.

giques et informatiques. Ces travaux reposent sur une double appréhension commune d'un énoncé itératif : une entité événementielle en quelque sorte générique qui est présentée comme répétée dans le temps, d'une part ; l'itération proprement dite, conçue comme ensemble des occurrences du modèle, d'autre part. Cette collaboration a donné lieu à la formalisation de trois modélisations des phénomènes itératifs : un approfondissement de cette problématique en intégrant une nouvelle définition de l'itération au sein de la *SdT* (Gosselin, 2011), une formalisation algébrique des itérations en se basant sur la notion de série (Enjalbert et Becher, 2011 – à paraître), et une modélisation du *processus* de la construction du sens des itérations selon une approche cognitive (Mathet, 2007).

Nous nous inscrivons dans la lignée de ces travaux en proposant une analyse automatique des phénomènes itératifs en langue. Notre objectif consiste en la mise en œuvre d'analyses d'ordre sémantique et discursif d'un énoncé itératif aboutissant *in fine* à un calcul des relations aspectuo-temporelles en se fondant sur les principes de la *SdT*.

L'ITÉRATION : QUELLE MANIFESTATION TEXTUELLE ?

La perspective d'une analyse automatique des phénomènes itératifs amène des questions auxquelles il est nécessaire de répondre préalablement à une opérationnalisation informatique : *De quelles manières se manifestent les phénomènes itératifs au sein d'un texte ? Quelles sont les sources de ces phénomènes ? Quels sont les marqueurs linguistiques permettant de les délimiter ?*

Comme nous le démontrerons tout au long de ce manuscrit, l'itération est un phénomène linguistique qui se révèle être complexe dans sa manifestation textuelle. Elle recoupe différentes catégories grammaticales : les substantifs qui peuvent référer à des objets (« *Marie a mangé cinq pommes* ») ou à des événements (« *Trois manifestations contre le régime iranien ont eu lieu à Téhéran* ») ; les verbes au travers de procès tels que « *Jean va à l'école à pied* » ou « *Jean va régulièrement à la piscine* ».

L'itération peut également revêtir un aspect discursif en corpus. Si nous pouvons retrouver des phrases simples comme les différents exemples que nous venons d'évoquer, il existe également des propositions itératives au sein de phrases complexes comme dans « *Chaque lundi matin, le brocanteur qui logeait sous l'allée étalait par terre ses ferrailles.* ». Ces énoncés ne sont pas nécessairement limités à une phrase comme dans l'extrait introductif. Nous retrouvons ce type d'énoncé fréquemment dans les œuvres littéraires telles que « *À la recherche du temps perdu* » de Marcel Proust dont l'incipit, « *Longtemps je me suis couché*

de bonne heure. », est une phrase itérative qui intègre une longue description d'événements décrivant sa nuit sur plusieurs paragraphes, ou encore l'extrait suivant tiré de la même œuvre :

« *Nous rentrions toujours de bonne heure de nos promenades pour pouvoir faire une visite à ma tante Léonie avant le dîner. Au commencement de la saison, où le jour finit tôt, quand nous arrivions rue du Saint-Esprit, il y avait encore un reflet du couchant sur les vitres de la maison et un bandeau de pourpre au fond des bois du Calvaire, qui se reflétait plus loin dans l'étang (...). Dans l'été au contraire, quand nous rentrions, le soleil ne se couchait pas encore ; et pendant la visite que nous faisons chez ma tante Léonie, sa lumière qui s'abaissait et touchait la fenêtre était arrêtée entre les grands rideaux et les embrasses, divisée, ramifiée, filtrée, et incrustant de petits morceaux d'or le bois de citronnier de la commode, illuminait obliquement la chambre (...). Mais certains jours fort rares, quand nous rentrions, il y avait bien longtemps que la commode avait perdu ses incrustations momentanées, il n'y avait plus quand nous arrivions rue du Saint-Esprit nul reflet de couchant étendu sur les vitres et l'étang au pied du calvaire avait perdu sa rougeur, quelquefois il était déjà couleur d'opale (...). Alors, en arrivant près de la maison, nous apercevions une forme sur le pas de la porte et maman me disait :*

"Mon Dieu ! voilà Françoise qui nous guette, ta tante est inquiète ; aussi nous rentrons trop tard." Et sans avoir pris le temps d'enlever nos affaires, nous montions vite chez ma tante Léonie pour la rassurer (...). »

Du côté de chez Swann, Marcel Proust

Cet extrait illustre la complexité, notamment textuelle, que nous pouvons rencontrer. Nous avons une première proposition, « *Nous rentrions toujours de bonne heure de nos promenades* », qui est itérative et situe le contexte pour les événements qui suivent. Et elle est suivie d'une première série d'événements, marqué par « *Au commencement de la saison* » qui viennent enrichir les informations liées à la promenade (*arriver rue du Saint-Esprit, avoir encore un reflet, ...*)². Une seconde série d'événements (*le soleil se coucher, s'abaïsser, ...*) vient enrichir les promenades sur une autre période temporelle, marquée par « *Dans l'été au contraire* ». Nous avons donc un ensemble de propositions qui font partie d'une même itération. Ainsi est construite une série d'occurrences de procès emboîtant des sous-séries (cf. figure 1) dont certaines contiennent des procès différents simultanés ou successifs. Nous remarquerons que ces sous-séries sont temporellement disjointes.

2. Nous reprenons la terminologie établie dans (Mathet, 2007) où il distingue deux structures : l'itération et la sélection. Une sélection, via un sélecteur (adverbe de fréquence, ...), vient sélectionner certains itérés d'une itération afin d'enrichir ou de modifier ces derniers.



FIGURE 1 – Représentation schématique d’une itération.

Il existe également une complexité liée à l’interprétation de l’énoncé. D’une part, il existe des exemples pouvant exprimer, selon le contexte, une habitude ou non, comme « *Paul fume* », que nous pouvons interpréter par « *Paul est un fumeur* » ou « *Paul est en train de fumer* ». D’autre part, la notion de présupposition intervient dans l’interprétation des itérations. En effet, en fonction de la position de l’itérateur, nous avons une interprétation stricte comme dans « *Jean va à la piscine le lundi* », où il n’y va que le lundi contrairement à l’exemple « *Le lundi, Jean va à la piscine* », qui n’exclut pas qu’il y aille en plus d’autres jours de la semaine. Toutefois, dans le cadre de cette thèse, nous n’aborderons que modestement la facette liée à l’interprétation en nous focalisant principalement sur sa dimension textuelle.

Afin de mieux caractériser les phénomènes itératifs, nous avons mis en place une étude descriptive des itérations en corpus qui nous a permis de mettre en évidence des propriétés structurelles (position des indices dans la phrase, discontinuité des procès au sein d’une itération, portée des déclencheurs...) et sémantiques (reprise anaphorique, temps morphologique, connecteurs temporels...). De cette étude, nous avons abouti à une automatisation de la découverte des phénomènes itératifs au sein d’un énoncé, en nous reposant sur l’hypothèse forte qu’une itération est fortement contrainte dans sa disposition textuelle par la cohésion des temps morphologiques de la dite itération.

LA TEMPORALITÉ DES PHÉNOMÈNES ITÉRATIFS

La seconde facette de notre travail a consisté à mettre en œuvre un système de calcul des relations aspectuo-temporelles en se fondant sur le modèle de la sémantique de la temporalité en français de Gosselin (1996). Le modèle repose sur une représentation du temps linguistique au moyen de deux lignes temporelles à orientations opposées, sur quatre intervalles temporels. De type hypothético-déductif, le modèle s’inscrit dans le cadre d’une sémantique instructionnelle holiste qui postule le dépassement possible des conflits linguistiques entre instructions. La puissance explicative de la *SdT* est en partie imputable aux différents modes de résolution des conflits qui rendent compte de la multiplicité des effets de sens auxquels participent les temps en discours.

Gosselin propose un ensemble de règles permettant de rendre compte des phénomènes linguistiques liés à la temporalité, en se basant sur deux concepts linguistiques : le temps qui situe le procès dans le passé, le présent ou le futur, et l'aspect qui précise la façon dont est présenté le procès. Les travaux de (Person, 2004) proposent une opérationnalisation de la *SdT*, qui à partir des marqueurs linguistiques contenus dans un énoncé en construit une représentation aspectuo-temporelle.

Les récents travaux de Laurent Gosselin sur les phénomènes itératifs reposent sur les mêmes principes que pour la *SdT*. Il introduit cependant un nouveau concept pour pouvoir représenter les itérations : la série itérative. Elle est conçue comme un macro-procès englobant une série d'occurrences d'un même élément itéré, ce dernier étant considéré comme « modèle » (ou prototype) de chacun des itérés. Un aspect important de ce dispositif est qu'il permet de rendre compte du fait que l'itération peut faire l'objet, simultanément, de deux visées aspectuelles éventuellement différentes, comme dans l'exemple 1. Dans cet exemple, la visée est aoristique sur le procès modèle, et inaccomplie sur la série (d'où la compatibilité de circonstanciels apparemment contradictoires : *en dix minutes* évalue la durée du procès modèle, tandis que *depuis deux mois*, porte sur le début de la série et le moment considéré).

« Depuis deux mois, il mangeait en dix minutes. »

EXEMPLE 1 – Exemple extrait de Gosselin (2011 – à paraître)

L'introduction de ce nouveau concept amène à s'interroger de nouveau sur l'opérationnalisation de la *SdT*. L'exemple ci-dessus (cf. exemple 1) illustre cette interrogation au travers de la question du rattachement des circonstanciels au niveau de la série et du procès modèle. Plus généralement, cette question se pose pour l'ensemble des marqueurs temporels et aspectuels caractérisés au sein de la *SdT* tel que les verbes et leurs compléments, les temps morphologiques, les circonstanciels temporels (*hier, pendant 3 jours, le jour où ...*), les connecteurs temporels (*souvent, puis*), les adverbes aspectuels ou encore la construction syntaxique des propositions. Pour conforter le lecteur de la réalité de cette question, nous reprenons l'exemple précédent en étudiant le temps morphologique. Classiquement, l'imparfait dans un énoncé sémelfactif porte une valeur d'inaccompli, et cette valeur reste identique dans les énoncés itératifs. Ici, cette valeur est affectée à la série, mais dans certains exemples tel que *Il avait fait ses devoirs à chaque fois qu'elle rentrait du travail*, cette valeur est affectée au procès modèle (c'est-à-dire à la version prototypique de l'ensemble

des itérés). Il est donc nécessaire de caractériser l'ensemble des configurations possibles.

En nous fondant sur ces travaux, nous proposons une solution autonome pour le traitement des itérations. Notre contribution consiste à adapter les règles existantes aux phénomènes itératifs, tout en respectant les principes de la *SdT*. Nous nous sommes focalisé sur deux facettes : les règles d'affectation des valeurs aspectuelles des différents marqueurs soit sur la série soit sur le procès contenus au sein d'une proposition. La seconde facette correspond au traitement des structures itératives dans une dimension textuelle, en proposant notamment une modélisation de la notion de sélection comme nous avons pu le voir avec l'extrait de Proust.

PLAN DE LA THÈSE

Dans ce contexte, la problématique posée s'inscrit donc dans une double perspective : d'une part, nous nous intéressons aux phénomènes itératifs dans le cadre d'une analyse automatique, consistant à repérer textuellement lesdites itérations et à les circonscrire. D'autre part, nous souhaitons réaliser le calcul des représentations aspectuo-temporelles d'un énoncé itératif. Notre sujet se positionne à la croisée de trois domaines :

- la linguistique descriptive, au travers d'approches caractérisant la source des itérations ;
- la linguistique de corpus, au travers d'une étude sur corpus des indices et structures discursives relatifs aux phénomènes itératifs ;
- l'informatique, dans la perspective d'un traitement automatique se fondant sur les deux aspects évoqués, et dont l'objectif est de proposer un calcul de la temporalité d'un énoncé itératif.

Dans la première partie de ce document, nous esquissons une présentation des principaux travaux sur les phénomènes itératifs en français. Puis nous mettons en perspective le modèle *SdT* par rapport à d'autres travaux sur la temporalité. La seconde partie s'intéresse au modèle théorique dans lequel nous nous inscrivons dans notre travail. Nous nous sommes fondée sur deux approches différentes d'une part, une approche aspectuelle avec la *SdT*, et d'autre part, une approche cognitive avec MOPi. La troisième partie expose le cœur de notre travail de thèse en présentant les résultats de notre étude de corpus qui nous a amené à mettre en place un modèle d'annotation pour la phase automatique, et la mise en œuvre d'un analyseur automatique des structures itératives : ADSI. Nous présentons dans une quatrième partie notre opération-

nalisation de la *SdT* dans sa version itérative en exposant les principes et les contraintes du calcul des structures aspectuo-temporelles, en comparaison des travaux existants sur les énoncés semelfactifs. Nous abordons également une formalisation des structures itératives tant du point de vue de leur inscription textuelle que de l'aspect au sein de la norme TIMEML. Nous concluons enfin en envisageant les perspectives possibles qui ressortent de ces travaux.



PUBLICATIONS

Julien Lebranchu : Vers une analyse automatique des phénomènes itératifs en corpus. In G Williams, éditeur : *Actes des 6èmes Journées de la Linguistique de Corpus*, Lorient, France, septembre 2009. Université de Bretagne-Sud. URL <http://jlebranc.perso.info.unicaen.fr/publication/JLC2009-analyse-automatique-phenomenes-iteratifs-corpus.pdf>.

Julien Lebranchu et Yann Mathet : Vers une prise en charge approfondie des phénomènes itératifs par TimeML. In *Actes de TALN 2011*, Montpellier, France, 2011. URL <http://jlebranc.perso.info.unicaen.fr/publication/taln2011-timeml-phenomenes-iteratifs.pdf>.

Première partie

LA QUESTION DE LA TEMPORALITÉ
ET DES PHÉNOMÈNES ITÉRATIFS

PRÉAMBULE

L'objectif de cette première partie est de présenter le cheminement au sein des travaux existants qui nous a conduit à proposer un type de traitement pour appréhender les phénomènes itératifs. La somme des travaux linguistiques consacrés à l'expression de la temporalité et au calcul des relations temporelles dans la (ou les) langue(s) est très importante. Par conséquent, nous ne proposons pas ici un état de l'art comparatif des travaux en linguistique de la temporalité, un tel projet dépassant largement le cadre de notre étude.

Nous proposons un état de l'art restreint aux théories, et plus largement aux travaux, mettant directement en perspective notre problématique, à savoir les phénomènes itératifs. Nous nous intéressons dans un premier temps aux travaux linguistiques traitant des itérations dans une perspective de traitement automatique. Dans un second temps, nous présentons les modèles traitant plus largement de la temporalité avec comme objectif d'exposer les principes de la *SdT* (Sémantique de la Temporalité) qui est la base linguistique sur laquelle reposent nos travaux.



LES PHÉNOMÈNES ITÉRATIFS DANS LA LITTÉRATURE

Ce chapitre est consacré aux travaux portant sur les phénomènes itératifs présents dans la littérature. Cette question y est traitée de manière abondante. Pour les présenter, nous reprenons la trichotomie proposée par Gosselin (2011 – à paraître) distinguant les travaux selon une certaine typologie de travaux linguistiques : la linguistique d’« inspiration » formelle, la linguistique formelle et la linguistique descriptive.

SOMMAIRE

1.1	Linguistique d’« inspiration » formelle	15
1.1.1	Les phrases habituelles	15
1.1.2	La pluriactionnalité	16
1.2	Linguistique formelle	18
1.3	Linguistique descriptive	20
1.3.1	La fréquence et son expression	20
1.3.2	Les phrases à subordonnées temporelles	21

1.1 LINGUISTIQUE D’« INSPIRATION » FORMELLE

1.1.1 *Les phrases habituelles*

Les travaux de Kleiber (1987) proposent une étude critique des travaux contemporains sur l’habitualité, notamment ceux de Carlson (1981). Kleiber présente la question de l’habitude sous différents angles, à savoir la généricité et la quantification. Il distingue tout d’abord les phrases itératives et fréquentatives. La phrase itérative est une phrase qui présente une situation comme étant vérifiée à plusieurs reprises à l’intérieur d’un intervalle temporel (*cf.* exemple 2a). La phrase fréquentative présente l’itération comme s’étendant sur tout l’intervalle temporel. L’itération est vraie pour l’intervalle temporel, et non dans l’intervalle, ce qui implique une notion de distribution des occurrences sur une période (*cf.* exemples 2b, 2c et 2d).

Kleiber constate qu’une des origines du sens de l’habitude, en plus des marqueurs extrinsèques, provient d’une valeur d’habitualité portée par cer-

- (a) Paul est allé 10 fois à l'école à pied le mois dernier.
- (b) Paul va à l'école à pied.
- (c) Généralement, Paul va à l'école à pied.
- (d) Paul va souvent à l'école à pied.

EXEMPLE 2

tains prédicats. Il distingue deux types de prédicats. D'une part, des prédicats comme *fumer*, qui dénotent une situation *a priori* déjà considérée comme une habitude, c'est-à-dire une manière de se comporter fréquemment répétée. D'autre part, certains prédicats dénotent une situation qui est par avance structurante, et donc susceptible de fournir la division régulière nécessaire à une interprétation fréquentative ou habituelle, comme *Paul chaptalise son vin*.

1.1.2 *La pluriactionnalité*

Les études menées sur la pluriactionnalité ont pour but de décrire la morphologie de certaines langues d'Afrique, et plus récemment les langues indo-européennes. On citera pour l'exemple les travaux de Van Geenhoven (2004, 2005) sur le groenlandais occidental ou encore ceux de Laca (2005) sur l'espagnol. La pluriactionnalité consiste à étudier les différents marqueurs de la pluralité des événements marquant la distribution des sous-événements d'un événement global sur plusieurs localisations spatiales, la répétition d'un événement, ou les lectures distributives d'un argument. Ces marqueurs peuvent être, selon les langues étudiées, des particules, des affixes, des radicaux spécifiant une lecture fréquentative, etc.

Certaines extensions de la notion de pluriactionnalité visent à rendre compte de phénomènes aspectuels, comme ceux liés aux marqueurs fréquentiels ou habituels, ou ceux liés à des conflits pragmatico-référentiels (*cf.* exemple 3).

Pierre s'est baigné tout l'été.

EXEMPLE 3

Il existe, selon Laca, deux types de pluriactionnalité : une dite temporelle exigeant que l'intervalle pluriactionnel comprenne plusieurs sous-intervalles successifs disjoints, et une seconde dans laquelle ils ne sont pas nécessaire-

ment disjoints. Cette disjonction des sous-intervalles est vérifiable par l'incompatibilité avec des expressions niant l'existence d'intervalles disjoints telles que *simultanément, en même temps* ou par des expressions comme *d'un coup, d'une traite* assertant l'indivisibilité de l'intervalle (cf. exemple 4¹). C'est notamment cette dernière approche qui est utilisée dans (Laca, 2005) en analysant les périphrases aspectuelles comme des opérateurs pluriactionnels (par exemple, *occasionnellement* est un opérateur de fréquence).

- (a) Maria a lu *Guerre et Paix* d'une traite.
- (b) ★Nous paierons progressivement notre dette d'une traite.

EXEMPLE 4 – Exemple extrait de (Laca, 2005, page 5).

Dans Van Geenhoven (2004, 2005), l'auteur propose également de saisir la pluriactionnalité temporelle en définissant des opérateurs qui s'attachent au verbe et déterminent comment les intervalles doivent être distribués sur l'intervalle du pluriactionnel. Dans son analyse, elle distingue les lectures fréquentatives produites par la combinaison d'un adverbial de mesure temporelle (*pendant X temps*), avec un télique, de celles issues des adverbes de fréquences par exemple. Pour l'auteur, le premier cas résulte de l'insertion optionnelle d'un opérateur pluriactionnel silencieux (c'est-à-dire sans un ancrage textuel), *FREQ*, au niveau du verbe uniquement. Dans le second cas par contre, c'est l'adverbe de fréquence qui est responsable de la lecture fréquentative de la phrase portant soit sur le verbe, soit sur le groupe verbal. L'exemple 5a n'est interprétable qu'en supposant que la balle en question a été renvoyée dans le lac à plusieurs reprises, contrairement à l'exemple 5b.

1. Nous précédonos nos exemples non valides de la marque ★.

- (a) Jim hit a golf ball into the lake for an hour.
1. There is a golf ball and Jim hit into the lake repeatedly for an hour.
 2. ★For an hour, Jim hit each time another golf ball into the lake.
- (b) Jim hit a golf ball *every five minutes* into the lake for an hour.
1. There is a golf ball and for an hour Jim hit into the lake every five minutes.
 2. For an hour, Jim hit every five minutes another golf ball into the lake.

EXEMPLE 5 – Exemple extrait de (Van Geenhoven, 2005), page 119 (ex 40-41).

En conclusion de cette section, et en anticipant sur la suite du manuscrit, nous remarquons que l'opérateur fréquentiel silencieux, introduit par Van Geenhoven comme explication de la lecture fréquentative de l'exemple 5a, fait référence au conflit d'ordre pragmatico-référentiel étudié par L. Gosselin dans le cadre de la *SdT* (*cf.* chapitre 2, section 2.2.2).

1.2 LINGUISTIQUE FORMELLE

Dans le cadre de la linguistique formelle, la question de l'itération est traditionnellement présentée et traitée comme un problème de quantification. La quantification est généralement associée à des utilisations très restrictives faites notamment dans le cadre de la logique du premier ordre. Une phrase, comme celle de l'exemple 6 extrait de (Enjalbert et Becher, 2011 – à paraître), peut être paraphrasée par « pour toute période temporelle étiquetée lundi, il existe un événement ayant pour type *Jean aller à la piscine* et comme inscription temporelle une sous-période de ce lundi ».

Tous les lundis, Jean allait à la piscine.

EXEMPLE 6

Une représentation simple de la formalisation en logique du premier ordre pourrait être :

$$(1.1) \quad \forall t[(\text{lundi}(t) \wedge t < t_0) \\ \rightarrow \exists e(e \subseteq t \wedge \text{aller à la piscine}(e) \wedge \text{agent}(e, \text{Jean}))]$$

où t , t_0 sont des variables de type « période », t_0 désigne l'instant d'énonciation et e une variable de type « événement ». Il existe toutefois des limites, notamment dans l'expression des quantificateurs de la langue naturelle tels que *beaucoup* ou *au moins deux*.

À ce problème du trop faible nombre de quantificateurs, Partee (1987) et Barwise et Cooper (1981) ont proposé la théorie des quantifieurs généralisés. Cette théorie permet de classer les déterminants selon les propriétés sémantiques qu'ils manifestent. On distingue souvent deux approches de la TQG. La première approche est dite fonctionnelle, elle associe à tout syntagme nominal une fonction ayant comme argument un ensemble et retourne une valeur de vérité.

La seconde approche est dite relationnelle. Elle considère que le déterminant dénote une relation entre deux ensembles. C'est l'approche choisie dans Keenan (1996), qui définit un ensemble de fonctions associées au déterminant, comme pour *MOST* dans l'exemple 7 où il définit un opérateur prenant deux prédicats. C'est également celle privilégiée dans les travaux de Kamp et Reyle (1993), que nous abordons plus longuement au chapitre 2.

Most students work hard. MOST(student)(work hard)
--

EXEMPLE 7 – Exemple extrait de Keenan (1996).

Pour conclure, il existe d'autres approches plus formelles qui se sont intéressées aux relations convexes d'Allen; citons par exemple les travaux de Cukierman et Delgrande (1996, 2000) qui proposent une approche basée sur les intervalles convexes pour la représentation des occurrences.

1.3 LINGUISTIQUE DESCRIPTIVE

1.3.1 *La fréquence et son expression*

Les travaux de Lim (2002) constituent une étude linguistique de la notion de fréquence en français. Elle présente les différents mécanismes mis en place lors de ce qu'elle appelle une « lecture répétitive » d'un procès. Elle caractérise l'aspect itératif par les deux propriétés suivantes : d'abord, la répétition est une notion aspectuelle qui peut se surajouter à la ponctualité et à la durativité ; ensuite, elle s'exprime le plus souvent à l'aide de moyens qualifiés par l'auteur d'*extrinsèques*, c'est-à-dire non portés par la sémantique du verbe. Elle se fixe comme objectif le recensement des moyens lexicaux qui permettent de traduire l'aspect itératif ainsi que l'aspect semelfactif dans une phrase. Elle se focalise sur la phrase simple en la représentant sous la forme d'un schéma d'arguments, suivant le modèle prédicat(*arg1, arg2, arg3*), le prédicat étant un verbe, un substantif ou un adjectif.

Pour cela, elle décrit les différents marqueurs explicites tels que les affixes (le préfixe *re-* et des suffixes comme *-iller, -oter, ...*), les adjectifs comme *fréquemment, rarement, ...*, les adverbes de fréquence (*souvent, parfois, ...*) ou de durée (*tout au long du séminaire, pendant des années, ...*) ainsi que des adverbes de progression (*au fur et à mesure, au fil des jours*). Elle décrit en outre la nature compositionnelle des itérations au travers de l'interaction entre les éléments de la phrase qui peuvent entraîner un changement dans sa lecture, par exemple dans le cas où le genre singulier du complément est changé en pluriel.

- | |
|---|
| (a) Il a fumé son cigare. ⇒ non répétitif
(b) Il a fumé ses cigares. ⇒ répétitif |
|---|

EXEMPLE 8

Lim expose aussi les différents mécanismes compositionnels existant en langue française en fonction de ces marqueurs explicites tels que les adverbes de progression. Deux types sont définis : l'un continu et l'autre discontinu. Si on associe la locution prépositive *au fur et à mesure de* avec un prédicat ponctuel, ce dernier ne peut qu'être au pluriel et on ne peut que dénoter une progression discontinue :

- (a) Au fur et à mesure de ses besoins (*son besoin), Paul a vendu ses livres.
- (b) Luc a compté les gens au fur et mesure de leur arrivée (*son arrivée).

EXEMPLE 9

Elle montre également l'opposition entre un état permanent et un état passager. Ce dernier est caractérisé par le fait qu'il peut être soumis à la notion d'occurrence, contrairement à un état permanent.

- (a) Marc est souvent malade.
- (b) *Marc est souvent mort.

EXEMPLE 10

L'étude menée a permis de mettre en évidence les grandes catégories (unicité des occurrences, seconde occurrence, pluralité des occurrences) dans lesquelles sont décrits les différents mécanismes, et étudié les principaux marqueurs de chaque catégorie. Par exemple, l'auteur a étudié pour les adverbes *souvent* et *rarement* à la fois la pluralité des occurrences et le degré de répétition où *souvent* exprime une fréquence élevée tandis que *rarement* exprime une fréquence faible.

1.3.2 Les phrases à subordonnées temporelles

Les travaux de Condamines (1990, 1992) portent sur la répétition dans la phrase à subordonnée temporelle. L'auteur présente dans un premier temps les caractéristiques des répétitions au sein d'une phrase simple pour les confronter à celles de la phrase complexe.

Dans la phrase simple, la production de la répétition est classée en trois catégories : les phrases habituelles, où la phrase peut être paraphrasée en ajoutant un adverbe d'habitude (tel que *généralement* ou *habituellement*), les phrases itératives, et les fréquentatives.

Les phrases simples et complexes ont un élément commun qui joue un rôle dans l'interprétation itérative de la phrase. Cet élément, appelé *rupteur*, se décline de trois manières :

- les adverbes de répétition itératifs (*trois fois, tous les dimanches*) et les fréquentatifs (*souvent, parfois*);
- les rupteurs menant à une répétition déduite, c'est-à-dire lorsqu'il y a une incompatibilité entre la durée exprimée par le procès et le rupteur (*cf. exemple 11a*);
- les rupteurs menant à une répétition héritée (*cf. exemple 11b*).

- | |
|---|
| <p>(a) Longtemps je me suis couché de bonne heure.
 (b) Les jours de pluie, Paul va à l'école en bus.</p> |
|---|

EXEMPLE 11

Un autre critère rentre en compte dans la répétition : la nature du prédicat. Tout comme dans le modèle *SdT*, elle décrit l'aspect lexical :

- Les activités ont la capacité de se répéter.
- Les achèvements pourront engendrer des répétitions si le sujet n'a pas un référent unique. L'exemple 12a présente un procès de type achèvement dont le référent n'est pas unique et dont la lecture itérative est possible, contrairement à l'exemple 12b sémantiquement invalide.
- Seuls les états transitionnels peuvent se répéter.
- Les accomplissements dépendent de l'objet sur lequel porte la valeur du prédicat : la répétition se fera seulement si l'objet n'est pas modifiable.

- | |
|---|
| <p>(a) Un chien meurt tous les jours.
 (b) *Le chien meurt tous les 15 ans.</p> |
|---|

EXEMPLE 12

Tous ces éléments jouent aussi un rôle dans le cas de la phrase complexe à subordonnée temporelle. D'autres paramètres vont également contribuer à ceux présents dans le cas des phrases simples, notamment la relation temporelle entre deux propositions et la nature de ce rapport. Cette relation est influencée par les trois mêmes critères que dans une phrase simple, à savoir le temps, le connecteur et les prédicats.

L'INFLUENCE DU TEMPS MORPHOLOGIQUE : La combinaison de certains temps, entre la proposition principale et la subordonnée, favorise une interprétation itérative comme le présent et le passé composé (*cf. exemple 13a*).

- (a) Quand il a bu du vin, il est gai.
(b) Je vous contacte dès qu'il se manifeste.

EXEMPLE 13

Cependant, la combinaison n'entraîne pas systématiquement une interprétation itérative, par exemple dans les cas où le présent prend une des valeurs aspectuelles suivantes : présent historique, à valeur de futur (cf. exemple 13b) ou à valeur déictique.

L'INFLUENCE DU CONNECTEUR : Les connecteurs peuvent être de deux formes :

- Préposition/Déterminant/Nom de référence temporelle/(où + que), par exemple *Le jour où*, *À l'époque où* ;
- (Préposition + Adverbe + verbe)/que, par exemple *maintenant que*, *dès que*.

Indifféremment de la forme du connecteur, certains connecteurs vont bloquer une interprétation ou la favoriser, et d'autres vont dépendre du contexte.

L'INFLUENCE DES PRÉDICATS : Les prédicats fonctionnent de la même manière que dans le cas d'une phrase simple, à l'exception du prédicat à valeur d'achèvement. Ce dernier permet de combiner un adverbe de fréquence avec certains procès du type achèvement, à la condition que le sujet soit générique.

Quand le Président de la République meurt, c'est le Président du Sénat qui assure l'intérim.

EXEMPLE 14

EN RÉSUMÉ

Dans ce chapitre, nous avons présenté qu'il existait différentes approches pour traiter les itérations, allant de la représentation formelle des itérations à une description exhaustive du mécanisme de leur création. Nous remarquons que ces études se limitent à une analyse au niveau phrastique, laissant de côté les relations pouvant exister entre les procès.

Dans le cadre de cette thèse, nous nous sommes appuyé sur les travaux relevant de la linguistique descriptive. En effet, les travaux de J.-H. Lim et de A. Condamines nous ont permis respectivement de créer des ressources et des règles de calcul sur la subordonnée dans le cadre de l'analyse automatique que nous avons mise en œuvre, et que nous exposerons au chapitre 8.



DES MODÈLES DE LA TEMPORALITÉ

Ce chapitre est consacré à la description des modèles existants traitant de la temporalité. Nous distinguons trois grandes orientations :

- des modèles essentiellement référentiels, basés sur la logique ;
- des modèles essentiellement aspectuels, basés sur l’exploitation des marqueurs linguistiques ;
- des modèles essentiellement discursifs, qui visent à rendre compte de l’organisation du discours assurant la cohérence d’un texte.

Nous présentons successivement ces trois types de modélisation.

SOMMAIRE

2.1	Modèles référentiels	25
2.1.1	DRT : Discourse Representation Theory	25
2.1.2	SDRT : Segmented Discourse Representation Theory	29
2.2	Modèles aspecto-temporels	30
2.2.1	Reichenbach : un modèle temporel de la référence	30
2.2.2	Le modèle de la Sémantique de la Temporalité de L. Gosselin	32
2.2.3	Le modèle de J.-P. Desclés et de Z. Guentchéva	40
2.3	Modèles discursifs	42
2.3.1	Les cadres du discours	42
2.3.2	Les cadres temporels	43
2.3.3	La dynamique énonciative de surface	44

2.1 MODÈLES RÉFÉRENTIELS

2.1.1 DRT : *Discourse Representation Theory*

La *Discourse Representation Theory*, élaborée par Kamp et Reyle (1993), est une théorie de représentation du discours ayant pour objectif d’interpréter sémantiquement le discours, c’est-à-dire des suites cohérentes de phrases. Pour cela, elle propose un niveau intermédiaire entre la syntaxe et son interprétation, représenté sous une forme logique par les structures de représentation discursive, ou DRS (*Discourse Representation Structure*), qui sont les « unités de base » de la théorie. L’une des motivations de la DRT est de surmonter certaines difficultés de la logique classique concernant l’interaction entre quan-

tification et anaphores. Une DRS est une description partielle de ce que devrait être le monde pour que le discours soit vrai. Il en est attribué une à chacun des énoncés et représente la contribution sémantique qu'ils apportent au discours. Les DRS sont construites à partir d'un algorithme descendant que nous présentons plus loin. Une DRS se compose de deux parties, la première contenant les référents de discours qui représentent les entités présentes dans les énoncés du textes, et la seconde contenant les conditions sur les référents. Prenons l'exemple suivant que nous allons décrire :

Hier, John a appelé Mary. Elle dormait.

EXEMPLE 15

L'interprétation d'un discours en DRT commence par une DRS vide. Nous l'étendons, lors de l'interprétation de la première phrase, par les cinq référents présents dans le discours et six conditions.

- x et y représentent *John* et *Mary* ;
- e représente l'événement *appeler* ;
- n correspond au temps du discours ;
- t correspond à la localisation temporelle.

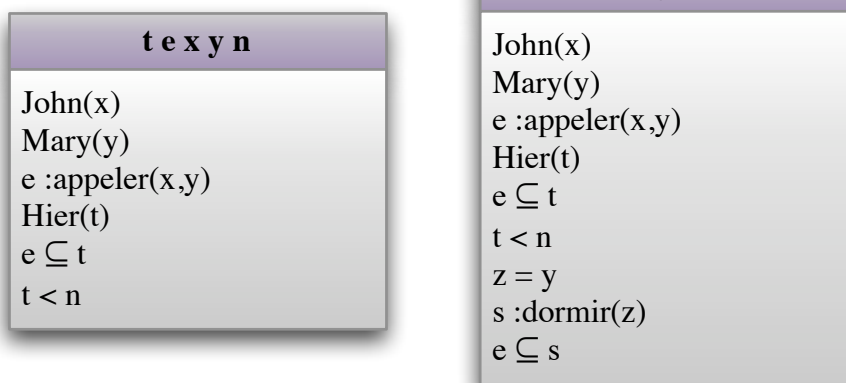


FIGURE 2 – Représentation de la construction d'une DRS.

La seconde phrase exprime un état, représenté par le référent du discours s . Nous étendons notre DRS avec le référent z qui représente le pronom *elle*. L'interprétation de cette seconde phrase nous permet de résoudre deux anaphores. La première est pronominale et est résolue en faisant correspondre le référent du pronom, ici z , avec un autre référent qui est accessible (la notion

d'accessibilité est détaillée dans (Kamp et Reyle, 1993). Dans notre exemple, le seul référent candidat est *y*.

Fondamentalement, la DRT se base sur deux notions : le « concept de vérité » et le dynamisme de l'interprétation. Le concept de vérité de la DRT repose sur la notion d'enchâssement d'une DRS dans un modèle complet de la réalité. L'interprétation d'une DRS est donc l'ensemble des modèles dans lesquels elle peut s'enchâsser. La signification est la relation établie entre les situations ou contextes dans lesquels on réalise un discours et les situations décrites par l'énonciation de ce discours. Une DRS est une description partielle de l'état du monde afin que le discours soit vrai. L'interprétation d'une DRS est donc l'ensemble des modèles dans lesquels elle peut s'enchâsser.

La seconde notion fondamentale de la DRT est le dynamisme de l'interprétation. Certaines expressions, comme les expressions quantifiées existentielle ou les descriptions indéfinies, ont deux contributions :

- une contribution statique à la signification de la phrase, c'est-à-dire leurs conditions de vérité ;
- une contribution dynamique par l'introduction d'un référent ou marqueur de discours comme antécédent potentiel pour les pronoms anaphoriques ultérieurs.

D'autre part, la DRT a cherché à incorporer la vision dynamique du sens, issue des approches cognitivistes, dans celles de la logique et de la philosophie du langage. Du point de vue dynamique, la signification d'une phrase n'est plus une fonction entre mondes possibles et valeurs de vérité, mais une fonction de contextes à contextes. Lorsque l'on aborde les phénomènes discursifs dans cette perspective, la notion de vérité dans la DRT ne s'identifie pas complètement à celle de la tradition logico-philosophique, c'est-à-dire que la DRT tient compte non seulement des conditions de vérité d'une proposition, mais aussi de son interprétation en contexte. Cette interprétation en contexte se matérialise dans l'exemple 15 par la reprise anaphorique marquée par le pronom *elle*.

Du point de vue des phénomènes itératifs, la DRT conçoit les itérations en termes quantificationnels, et se détache en cela de l'approche aspectuelle de l'itération que nous avons esquissée dans l'introduction de ce manuscrit. Les auteurs abordent les itérations à travers deux aspects, à savoir la pluralité nominale et les adverbes de quantification temporelle.

La pluralité

La pluralité est donc traitée à l'aide de quantifieurs généralisés qui sont définis comme une relation entre deux ensembles. Soit pour l'exemple 16, repris

de (Kamp et Reyle, 1993, p. 310), nous avons deux sous-ensembles de livres issus des expressions *avoir besoin de livres* et *avoir trouvé les livres* qui se trouvent reliés par le quantifieur *la plupart*. En terme de représentation boxologique, nous obtenons la figure 3.

Susan a trouvé la plupart des livres dont Bill a besoin.

EXEMPLE 16

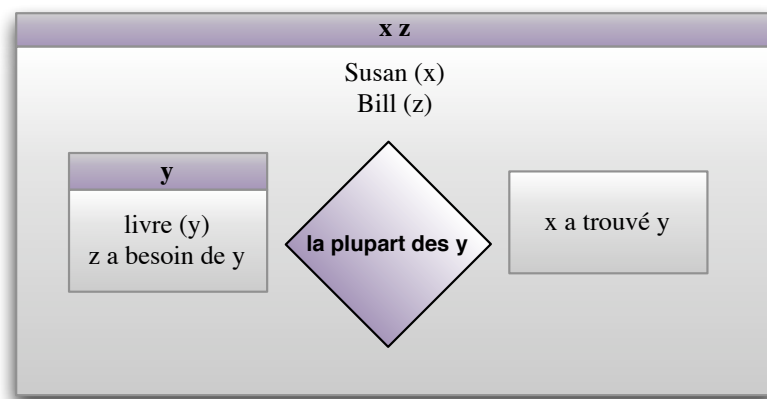


FIGURE 3 – Illustration d’une DRS sur la pluralité.

La quantification

Pour la quantification temporelle, les auteurs utilisent également les quantifieurs généralisés mais appliqués à des ensembles d’occurrences d’événements ou de localisation temporelle. Dans l’exemple 17, nous avons l’ensemble *des matins* mis en relation avec l’ensemble des *aller à la piscine*, et ce de manière bijective grâce au quantifieur *tous les*. Nous pouvons remarquer que les occurrences sont contraintes par ce que les auteurs appellent un localiseur simple, c’est-à-dire un circonstanciel de temps, ici *l’année 1985* (cf. figure 4).

En 1985, Mary allait à la piscine tous les matins.

EXEMPLE 17

Kamp et Reyle (1993) soulèvent plusieurs difficultés dans la représentation des phénomènes itératifs au sein de la DRT. Elles ne relèvent pas de la sémantique mais de l’évaluation d’une DRS. Ils mentionnent en particulier la question de la représentation des adverbes de fréquence comme *souvent*, *rarement*.

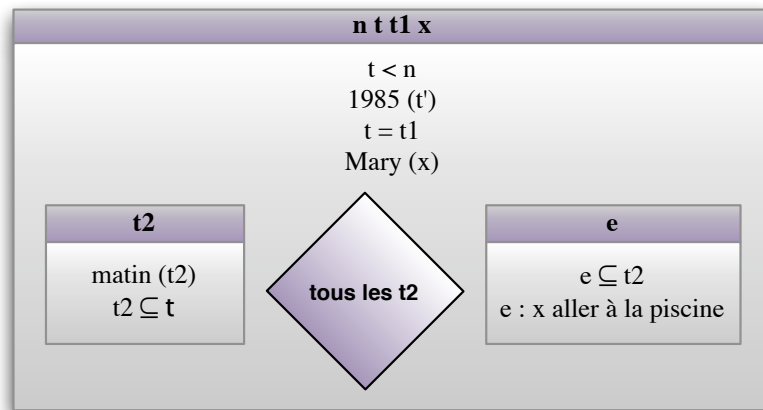


FIGURE 4 – Illustration d’une DRS sur la quantification.

Ces diverses difficultés ont amené à laisser la question de la représentation boxologique ouverte.

D’un point de vue calculatoire, il existe à notre connaissance peu de travaux ayant amené à une opérationnalisation de la DRT. Nous pouvons citer toutefois le travail de Curran *et al.* (2007) sur l’outil *C&C* fondé sur une grammaire extraite de la *CCG* (*Combinatory Categorical Grammar*), une version du Penn Treebank, et l’utilisation d’un tagger spécifique permettant de calculer les DRS d’un énoncé, fonctionnant sur la langue anglaise¹. Nous avons testé la manière dont il traite la temporalité mais nos premiers tests furent peu concluants sur des exemples simples. Son application nous a donc paru peu envisageable sur des énoncés en français.

2.1.2 SDRT : *Segmented Discourse Representation Theory*

La SDRT, présentée dans (Asher, 1993; Asher et Lascarides, 2005), est une théorie représentationnelle dynamique du discours qui prend en compte la segmentation et l’organisation structurelle du discours. Elle étend la DRT en s’inspirant des différentes approches de l’analyse du discours, pour expliciter dans un cadre vériconditionnel les interactions entre le contenu sémantique des segments et la structure globale, pragmatique, du discours : on peut légitimement décrire la SDRT comme une théorie de l’interface sémantique-pragmatique. Comme la DRT, la SDRT est une théorie opératoire, en ce sens qu’elle vise à décrire une méthode déterministe de construction des structures de représentation du discours segmentées, ou SDRS (*Segmented Discourse Representation Structure*). Comme les DRS, les SDRS représentent le contenu

1. <http://svn.ask.it.usyd.edu.au/trac/candc/wiki/Demo>

propositionnel du discours, ainsi que sa macro-structure. La notion de structure retenue est fondée sur les propositions et les relations de discours du point de vue de leur nature sémantique (*élaboration, narration, ...*). La figure 5 représente la structure du discours de l'exemple 15, qui reprend le contenu des DRS, à laquelle est ajoutée la relation de discours entre les deux « boîtes ».

À notre connaissance, les phénomènes itératifs sont traités de la même façon qu'au sein de la DRT.

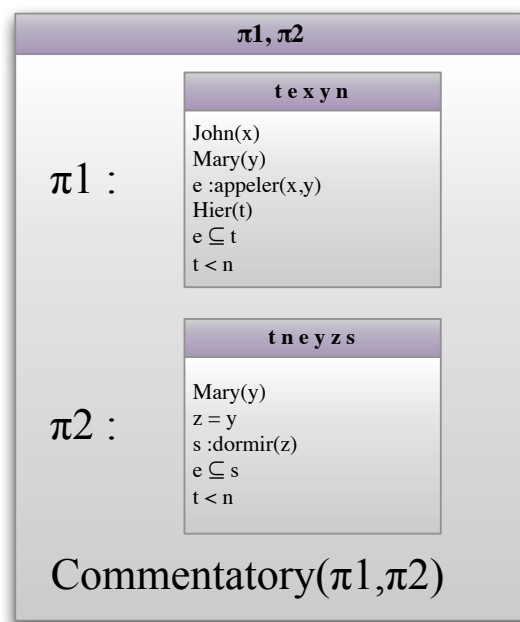


FIGURE 5 – Exemple de SDRS avec les relations de discours.

2.2 MODÈLES ASPECTO-TEMPORELS

Nous exposons ici les travaux de Laurent Gosselin, dans lesquels s'inscrit pour partie notre travail de thèse. En préliminaire, nous rappelons le modèle d'Hans Reichenbach, lequel a proposé dans (Reichenbach, 1947) un modèle de traitement des temps verbaux dont se sont inspirés de nombreux linguistes dans leurs approches de la sémantique temporelle, en particulier L. Gosselin.

2.2.1 Reichenbach : un modèle temporel de la référence

H. Reichenbach propose un système où le moment de la parole est le point central de la structure du temps verbal, soulignant ainsi la nature essentiellement déictique du temps (*cf.* (Vet, 1980)). Il propose un modèle temporel fondé

sur la représentation du temps en deux fragments articulés autour d'un repère symbolisant le présent, et distingue trois points :

- un point de l'énonciation (*point of speech*) noté S ;
- un point de l'événement (*point of the event*) noté E ;
- un point de référence (*point of reference*) noté R.

Cette articulation autour du moment de référence permet de délimiter trois périodes temporelles pouvant se confondre ou se dissocier sur l'axe temporel (antérieure, postérieure ou simultanée au moment de l'énonciation). Il existe donc des jeux de relations entre S et R d'une part et E et R d'autre part. La relation entre S et E n'est pas jugée pertinente. Les relations entre R et S représentent le temps et sont désignées par :

- R-S : passé (R est avant S) ;
- R,S : présent (R et S sont confondus) ;
- S-R : futur (R est après S).

Les relations entre E et R représentent l'aspect et sont désignées par :

- E-R : antérieur (E est avant R) ;
- E,R : simple (E et R sont confondus) ;
- R-E : postérieur (E est après R).

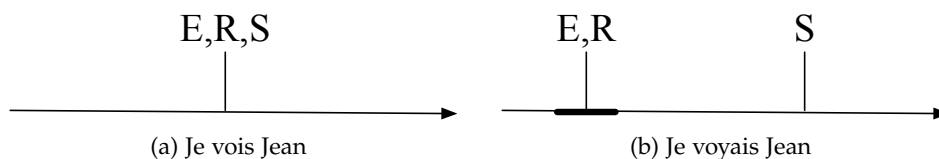


FIGURE 6 – Représentation du temps dans le modèle de Reichenbach.

Ces relations permettent d'assigner une représentation à chaque temps verbal. Ainsi, les valeurs pour les temps de l'indicatif les plus courants en français sont :

- présent \Rightarrow S,R,E ;
- imparfait \Rightarrow E,R-S ;
- passé simple \Rightarrow E,R-S ;
- passé composé \Rightarrow E-S,R ;
- futur simple \Rightarrow S-R,E.

Reichenbach propose également une représentation formelle de R, dit « étendue », en le présentant comme un laps de temps et non plus comme un point. Selon l'auteur, cette représentation étendue de R permet d'exprimer aussi bien la durée d'un procès que sa répétition au cours du temps.

Il est possible d'avancer un certain nombre de critiques sur ce modèle, en particulier :

- la représentation en points ne permet pas de distinguer l'imparfait du passé simple, ni de représenter le présent et l'imparfait d'habitude ;
- dès lors que la notion d'intervalle est considérée, les trois relations, antériorité, simultanété et postériorité, adoptées pour le traitement des points, sont insuffisantes. Il serait en effet nécessaire de pouvoir traiter des relations d'inclusion et de chevauchement. Allen (1983) propose un ensemble des relations possibles entre intervalles, que reprend Laurent Gosselin dans son modèle.

En conclusion, le modèle d'Hans Reichenbach n'est pas un dispositif suffisant pour la description de systèmes temporels dans les langues. Il a toutefois été à la source de nombreux travaux qui ont adopté le modèle pour en résoudre les limites. Certains de ces travaux substituent des intervalles aux points. On peut citer par exemple, en sémantique formelle, Dowty (1986), qui utilise des valeurs de vérité sur des intervalles de temps, et, en sémantique descriptive, Gosselin (1996), que nous présentons à la section suivante, et qui réinterprète la représentation de Reichenbach en postulant que les relations entre intervalles servent à représenter le temps et l'aspect grammatical.

2.2.2 *Le modèle de la Sémantique de la Temporalité*

Le modèle nommé *Sémantique de la Temporalité* (ci-après *SdT*), élaboré par Laurent Gosselin (1996, 2005), repose sur la *compositionnalité holiste* des valeurs aspectuo-temporelles, intégrant aussi bien des éléments purement sémantiques que des connaissances encyclopédiques et pragmatiques.

2.2.2.1 *Principes généraux*

Les structures d'intervalles

Au sein du modèle, les catégories et les relations temporelles et aspectuelles sont représentées au moyen de structures d'intervalles disposés sur l'axe temporel. Quatre grands types d'intervalles sont mis en œuvre :

- l'intervalle d'énonciation, $IE = [IE_1, IE_2]$, renvoie à l'énonciation effective de l'énoncé ;
- l'intervalle de procès, $IP = [IP_1, IP_2]$, correspond à la subsomption d'une série de changements ou de situations sous la détermination d'un procès ;
- l'intervalle de référence, $IR = [IR_1, IR_2]$, représente ce qui est montré par le procès² ;

2. L'intervalle de référence est parfois appelé intervalle de monstration.

- l'intervalle circonstanciel, $IC = [IC_1, IC_2]$, délimite des portions de l'axe temporel permettant de localiser l'intervalle du procès ou de référence. Les intervalles circonstanciels sont marqués par des compléments de localisation temporelle (par ex. : *mardi dernier, le jour où Pierre est venu*) et de durée (par ex. : *pendant trois heures*).

À chaque énoncé est associé un unique intervalle d'énonciation. À chaque proposition est associée au moins un intervalle de procès et au moins un de référence. À chaque complément circonstanciel de temps est associé au moins un intervalle circonstanciel. Nous obtenons donc pour l'énoncé de l'exemple 18 la structure de la figure 7.

Hier, son fils chercha son jouet.

EXEMPLE 18

Le locuteur de l'énoncé parle d'un certain moment de référence (noté $[IR_1, IR_2]$) situé dans le passé (il est antérieur au moment de l'énonciation), et fait aussi référence au procès dans sa globalité. De plus, le procès est borné par le circonstanciel « hier » (noté $[IC_1, IC_2]$).

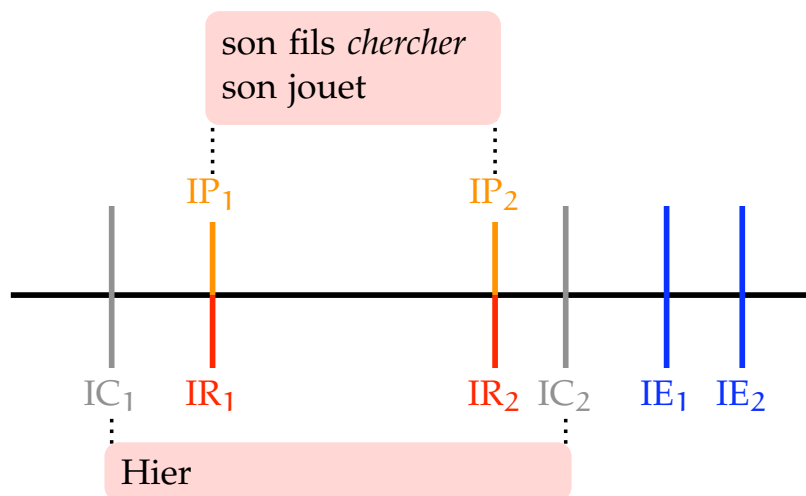


FIGURE 7 – Représentation des intervalles de l'exemple 18 disposés sur l'axe du temps.

Afin de rendre compte des phrases complexes de façon plus visible, le modèle prévoit de dupliquer l'axe temporel pour chaque proposition subordonnée. Ainsi l'énoncé de l'exemple 19 se verra associer la structure de la figure 8.

Hier, les journalistes ont annoncé que la police recherchait le coupable depuis trois jours.

EXEMPLE 19

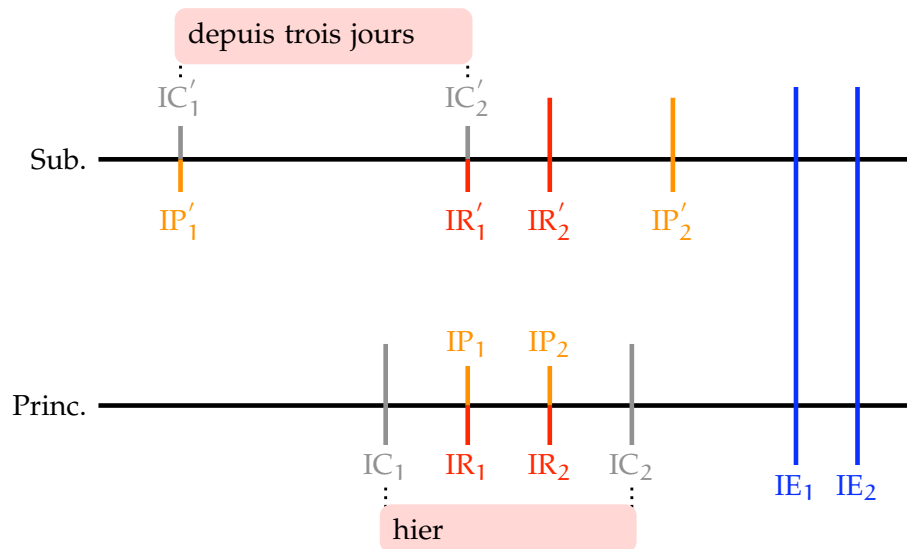


FIGURE 8 – Représentation des intervalles de l'exemple 19 disposés sur deux axes du temps pour en faciliter la lecture.

Les relations

Afin de pouvoir relier les intervalles entre eux, le modèle précise tout d'abord des relations entre bornes. Soient i et j les bornes d'un intervalle :

- $i = j \Rightarrow i$ coïncide exactement avec j
- $i \propto j \Rightarrow (i$ précède $j) \wedge (i$ est infiniment proche de $j)$
- $i \prec j \Rightarrow (i$ précède $j) \wedge (i$ n'est pas dans le voisinage de $j)$
- $i < j \Rightarrow (i \propto j) \vee (i \prec j)$
- $i \leq j \Rightarrow (i < j) \vee (i = j)$

Les relations entre intervalles se laissent exprimer à partir des relations entre bornes. Soient $I = [I_1, I_2]$ et $J = [J_1, J_2]$ deux intervalles ; leurs relations temporelles sont formalisées comme suit (figure 9) :

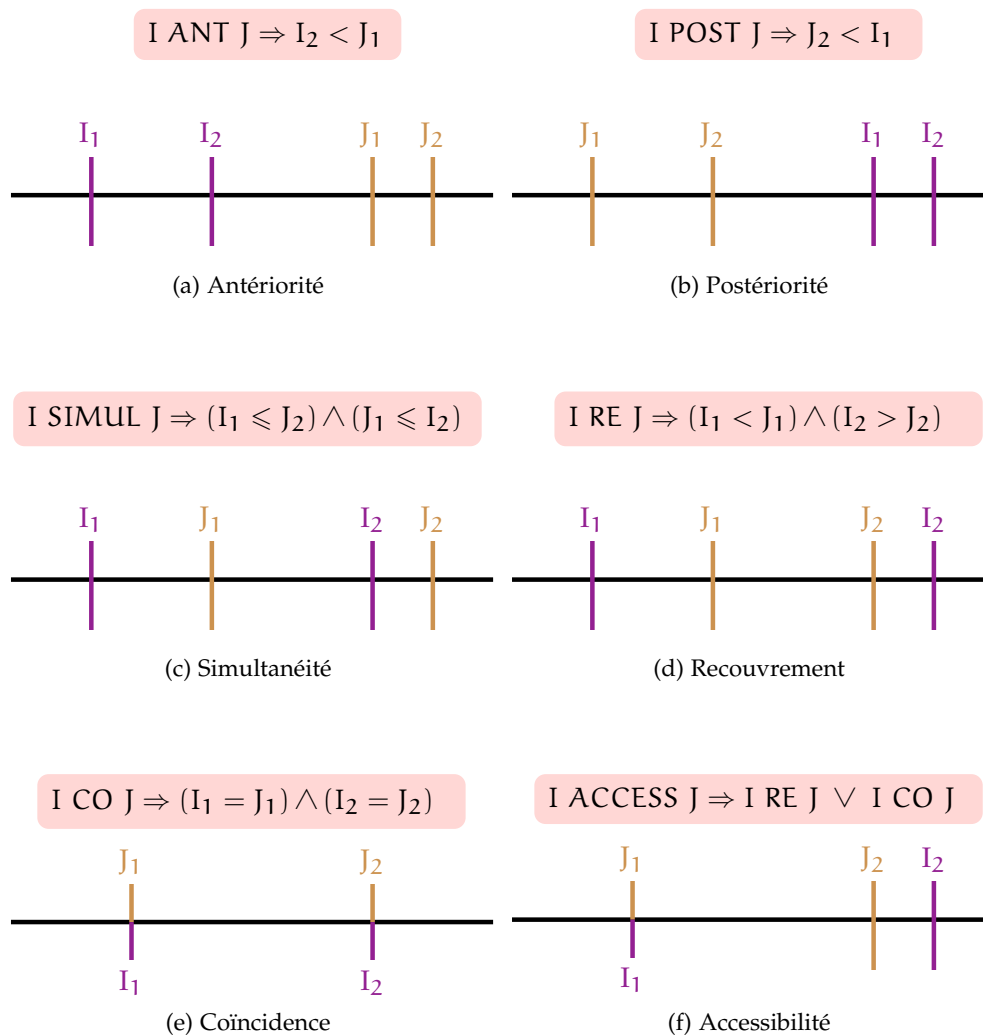


FIGURE 9 – Liste des relations temporelles au sein de la *SdT*.

2.2.2.2 Les marqueurs aspectuo-temporels

Pour réaliser les calculs, le modèle s'appuie sur les « marqueurs » suivants :

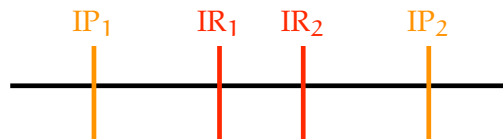
Les types de procès sont classés selon les quatre grandes catégories de la typologie de Vendler (1967)³ :

- les états sont non dynamiques, atéliques et non ponctuels (*habiter une maison, être en vacances*) : ils sont compatibles avec « pendant + durée » ;
- les activités sont dynamiques, atéliques et non ponctuelles (*marcher, manger des frites, dormir*) : elles sont compatibles avec « être en train de verbe à l'infinitif » et « pendant + durée » ;

3. Les tests permettant de classer les types de procès sont à appliquer avec certaines précautions. Nous renvoyons le lecteur à Gosselin (1996, p.41-72), pour plus de détails.

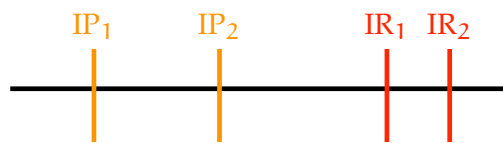
- les accomplissements sont dynamiques, téliques et non ponctuels (*manger une pomme, terminer un travail*) : ils sont compatibles avec « être en train de *verbe à l'infinitif* » et « en + durée » ;
- les achèvements sont dynamiques, téliques et ponctuels (*apercevoir un avion, atteindre un sommet, trouver une solution*) : leur ponctualité les rend incompatibles avec les compléments de durée, d'où la possibilité de paraphraser l'expression « mettre *n temps à verbe à l'infinitif*. » par « mettre *n temps avant de verbe à l'infinitif* ».

La visée aspectuelle est marquée par la conjugaison ou par des auxiliaires. Elle correspond à « l'aspect grammatical » dans les langues romanes, et se trouve définie par la relation entre l'intervalle de référence et celui du procès. On distingue quatre types de visées aspectuelles de base en français :



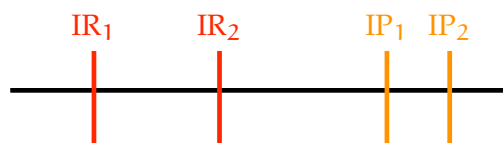
La visée *inaccomplie* (imperfectif) présente une vue partielle du procès : IP RE IR.

Exemple : *Il traversait le carrefour*



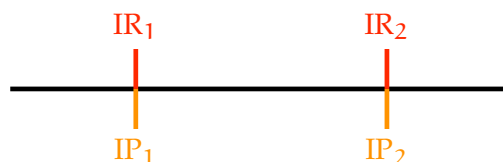
La visée *accomplie* montre l'état résultant du procès : IP ANT IR.

Exemple : *Il a traversé le carrefour*



La visée *prospective* présente la phase préparatoire : IP POST IR.

Exemple : *Il allait traverser le carrefour.*



La visée *aoristique* (perfectif) montre le procès dans sa globalité : IP CO IR.

Exemple : *Il traversa le carrefour.*

Le temps absolu est défini comme la relation entre l'intervalle de référence et le moment de l'énonciation :

1. temps présent : les deux intervalles coïncident ou se chevauchent, IR SIMUL IE ;
2. temps passé : IR ANT IE ;
3. temps futur : IR POST IE.

Le temps relatif se trouve défini comme la relation entre deux intervalles de référence (notés respectivement $[IR_1, IR_2]$ et $[IR_1', IR_2']$). La figure 8 page 34 illustre la relation de *simultanéité* entre les intervalles de référence de la subordonnée et de la principale.

1. simultanéité : IR CO IR' ;
2. antériorité : IR ANT IR' ;
3. postériorité : IR POST IR'.

2.2.2.3 Propriétés des intervalles

L'intervalle de référence n'est pas autonome et a besoin d'un intervalle antécédent. Il est intrinsèquement anaphorique. Cet antécédent doit satisfaire la relation de coïncidence sans contrevenir aux instructions codées par les autres marqueurs de l'énoncé, ni aux contraintes générales sur les structures aspectuo-temporelles. L'intervalle antécédent doit disposer d'un certain ancrage circonstanciel, c'est-à-dire qu'il doit pouvoir correspondre à une réponse possible à la question *Quand?* L'intervalle retenu est le plus saillant dans le contexte, en suivant une échelle de saillance relative (*cf.* figure 10).

L'intervalle circonstanciel est marqué par les compléments circonstanciels temporels, y compris les subordonnées temporelles. Les circonstanciels temporels se distinguent comme suit :

- les circonstanciels de durée, qui définissent la taille de l'intervalle circonstanciel sans le localiser autrement que par rapport au procès ou à l'intervalle de référence ;
- les circonstanciels de localisation temporelle, qui situent l'intervalle circonstanciel de façon plus ou moins précise et plus ou moins déterminée

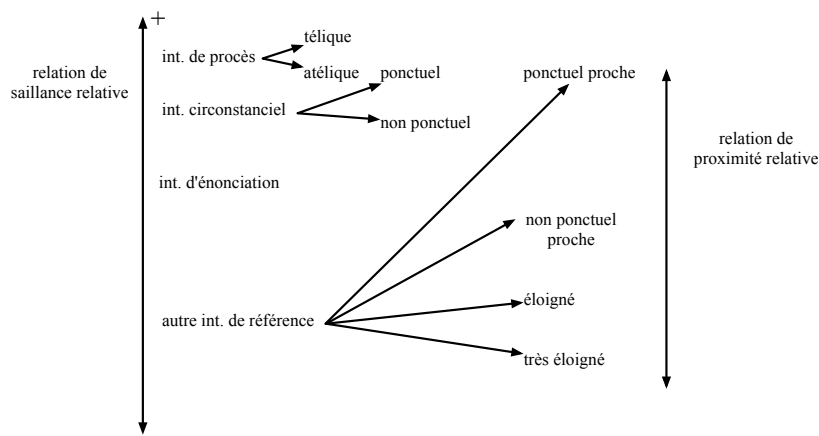


FIGURE 10 – Échelle de saillance relative pour le calcul du degré de saillance des intervalles.

par rapport au calendrier (localisation absolue), à l'intervalle de l'énonciation (localisation déictique), ou à un autre procès (localisation relative). Ce second type de localisation est caractéristique des subordonnées circonstancielles, dans lesquelles l'intervalle circonstanciel est situé par rapport au procès exprimé par la subordonnée.

Dans le cas d'une subordonnée, le mécanisme doit aussi spécifier la relation entre l'intervalle circonstanciel et l'un des intervalles IP/IR du procès de la subordonnée (relation circonstancielle subordonnée ou seconde). On distingue deux modes de subordination temporelle : la subordination directe et indirecte.

La subordination directe est marquée par une conjonction ou une locution conjonctive (*cf.* exemple 20a). La relation circonstancielle se fait entre l'intervalle circonstanciel et l'intervalle de procès de la subordonnée.

La subordination indirecte est construite au moyen d'un syntagme prépositionnel jouant le rôle de locution conjonctive, et désignant une période de temps servant à identifier ou à caractériser la proposition principale, par exemple *le jeudi où, après le jour où* (*cf.* exemple 20b). La relation circonstancielle se fait entre l'intervalle circonstanciel et l'intervalle de référence de la subordonnée.

- (a) Dès que je suis arrivé, nous avons commencé une partie de poker.
- (b) Un jour où il pleuvait, j'ai fait une rencontre surprenante.

EXEMPLE 20

2.2.2.4 *Principes de calcul*

Le modèle pose le principe que tous les marqueurs aspectuo-temporels codent une ou plusieurs instruction(s) pour la construction d'intervalles ou de relations entre bornes sur l'axe temporel. Cet assemblage d'instructions a pour but d'obtenir une structure globale cohérente et compatible avec les contraintes pragmatico-référentielles du contexte, comme nous avons pu le voir dans les figures 7 et 8. Toutefois, il arrive que différentes instructions codées pour un même énoncé soient contradictoires ou que la structure ne soit pas compatible avec le contexte.

Le modèle *SdT* prévoit donc la mise en œuvre de modes de résolution de conflits, qui consistent à déformer — le moins possible — les structures globales qui seraient mutuellement incohérentes. Cette déformation peut prendre les formes suivantes :

- le déplacement ;
- la contraction ;
- la dilatation ;
- la duplication, c'est-à-dire l'*itération*, que nous aborderons au chapitre 3.

Les règles de constructions des structures sont définies pour chacune des marques pertinentes. Elles sont décrites dans (Gosselin, 1996, 2005). Un certain nombre de propriétés sur les intervalles y sont décrites, notamment sur l'anaphoricité des intervalles de référence ou sur la portée des intervalles circonstanciels. Les règles furent également l'objet d'une implémentation dans le cadre des travaux de Person (2004).

[Il a dit qu']_{P₁} [il était venu la veille]_{P₂}.

EXEMPLE 21

Pour illustrer le modèle, les différentes instructions codées par les marqueurs présent dans l'énoncé de l'exemple 21 sont :

IR_{P_1} coïncide avec IP_{P_1} (aspect aoristique);

IR_{P_1} est antérieur à IE (temps passé);

IR_{P_2} coïncide avec IP_{P_2} (aspect aoristique);

IC_{P_2} recouvre IP_{P_2} ;

IR_{P_2} est antérieur à IE (temps passé);

donc IP_{P_2} est antérieur à IP_{P_1} .

Nous obtenons la structure globale illustrée par la figure 11.

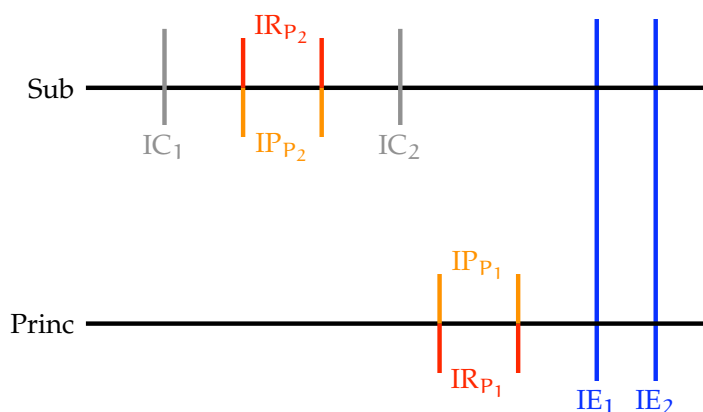


FIGURE 11 – Structure globale résultant des instructions codées par les marqueurs de l'énoncé.

De par l'importance qu'occupe le modèle *SdT* dans nos travaux, nous consacrons le chapitre 3 au traitement de l'itération au sein de ce modèle.

2.2.3 Le modèle de J.-P. Desclés et de Z. Guentchéva

Nous terminons cette section par le modèle aspectuel développé par Desclés et Guentchéva (1980), qui par certains aspects possède des ressemblances avec le modèle *SdT*. Il relève de la théorie de l'énonciation et prend, comme le modèle de L. Gosselin, l'énonciateur et son acte de parole comme point de départ du système. Il constitue l'un des modules de la Grammaire Applicative et Cognitive de Desclés (1990), qui propose un principe général d'analyse du langage et des langues, dans une perspective cognitive qui s'articule autour de différents niveaux de représentation. Il repose notamment sur la description sémantique des verbes sous forme de schèmes. Le module sur le traitement du temps et de l'aspect peut fonctionner de manière autonome, si bien que nous nous limiterons à la seule présentation de ce module.

Le modèle repose sur la notion d'intervalles topologiques, issue du concept des bornes ouvertes ou fermées des intervalles proposé par A. Culioli. Ces intervalles sont associés aux trois valeurs aspectuelles du modèle :

- état : intervalle à bornes ouvertes à gauche et à droite
(*elle est grande*);
- événement : intervalle à bornes fermées à gauche et à droite
(*il a mangé une pomme au goûter*);
- processus : intervalle à bornes fermée à gauche et ouverte à droite (*il est en train de manger une pomme*);

Le traitement de l'itération a fait l'objet de travaux plus récents (*cf.* (Chagnoux, 2006)). Chagnoux propose une représentation des phénomènes itératifs s'appuyant entre autres sur les travaux de Gosselin (1996). Elle définit l'itération comme une série de procès identiques. Elle associe à la série de procès un intervalle global dans lequel sont enchâssés des intervalles globaux représentant les occurrences du procès. Nous illustrons les choix de représentation à l'aide des exemples 22 et 23. Dans les termes de ce modèle, l'exemple 22 se décompose en :

- un intervalle global : un événement;
- deux intervalles globaux : deux événements, c'est-à-dire les deux occurrences de *boire une bière*;
- un intervalle temporel : un état.

Hier, il a bu deux bières.

EXEMPLE 22 – Exemple extrait de Chagnoux (2006, p. 125).

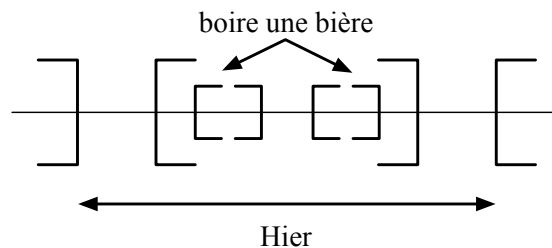


FIGURE 12 – Représentation de l'exemple 22.

L'exemple 23 présente une itération issue d'un intervalle temporel où l'on peut voir la répétition du procès sur une période de deux ans⁴.

4. Dans la *SdT*, mais également dans la suite ce document, il y est fait référence comme marqueur calendaire.

Il est venu *tous les jours* pendant deux ans.

EXEMPLE 23 – Exemple extrait de Chagnoux (2006, p. 126).

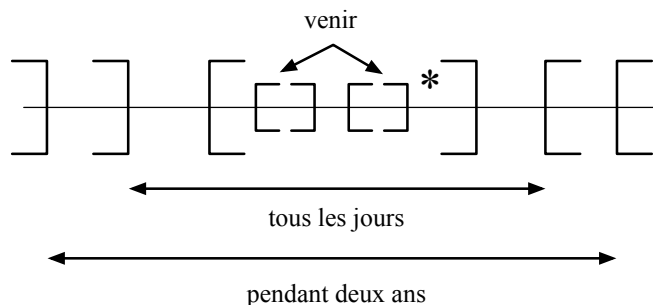


FIGURE 13 – Représentation de l'itération de l'exemple 23.

En conclusion, nous relevons qu'il existe une limite importante à ce modèle en intervalles topologiques. En effet, il n'y a pas de distinction entre les intervalles représentant le procès et ceux représentant le circonstanciel, à savoir *venir* et *jour* sur la figure 13. Il semble donc que le modèle manque d'une certaine finesse dans la représentation des structures : *comment représenter les jours où il est venu deux fois ?*

2.3 MODÈLES DISCURSIFS

Cette dernière section traite de travaux dont le paradigme diffère des modèles précédents. En effet, nous nous intéressons ici aux modèles qui analysent la temporalité au niveau discursif. À notre connaissance, aucun travail ne traite des phénomènes itératifs dans leur dimension textuelle. Nous présentons toutefois trois modélisations traitant la temporalité au sens large : les cadres de discours de Charolles (1997), les cadres temporels de Terran (2002) et la dynamique énonciative de Battistelli et Chagnoux (2008).

2.3.1 Les cadres du discours

Le modèle des cadres de discours de Michel Charolles propose de s'intéresser non plus aux types de relations binaires entre des segments, mais à la relation englobante. Cette relation englobante rassemble des propositions dans des unités désignées sous le terme de cadres de discours. Ces cadres sont initiés par une expression détachée dite *introduceur*, contraignant l'interprétation

de l'ensemble des propositions du cadre selon les critères donnés par l'expression qui l'introduit. L'exemple 24 présente un cadre spatial où la proposition de l'exemple vérifie le critère de contrainte spatial initié par l'expression *Dans le calvados*. Nous revenons plus précisément sur ces travaux au chapitre 7 dans la perspective de modéliser les phénomènes itératifs en termes de cadres.

Dans le Calvados, la moitié des emplois du secteur des TIC correspond à une activité industrielle.

EXEMPLE 24

2.3.2 Les cadres temporels

La notion de cadres temporels a fait l'objet de travaux spécifiques, notamment dans le cadre de la thèse de Terran (2002). L'auteur y caractérise les introducteurs de cadres temporels par la possibilité qu'ils ont d'étendre une référence temporelle à une portion de texte de taille variable. On y retrouve les principes des travaux de Charolles où les cadres peuvent entretenir des relations d'inclusion : on a un nouveau cadre temporel, compatible avec le cadre précédemment ouvert. Ces cadres s'organisent alors selon une structure arborescente. On y retrouve également le rôle prédominant de la position de l'introducteur dans l'installation et la portée d'un cadre ; sont considérés comme des introducteurs de cadres temporels les compléments antéposés et détachés.

Elle distingue également des indices de reconduction qui sont des éléments qui permettent d'intégrer une portion de texte dans un cadre. Elle montre que le temps verbal est une marque de stabilité qui permet d'indiquer la rémanence de la portée alors que le type des procès n'intervient pas de façon déterminante dans cette rémanence. Elle précise que :

bien qu'il soit possible d'identifier des marques de clôture, la fermeture d'un cadre n'est un processus ni tout à fait définitif, ni irréversible : la mémoire conserve une trace de sa lecture, ce qui permet à d'autres constituants de le réactiver au besoin dans la suite du texte.

(Terran, 2002, p.42)

Comme nous le verrons plus loin dans ce manuscrit, nous sommes arrivé au même constat en ce qui concerne le traitement des itérations.

2.3.3 La dynamique énonciative de surface

Dans (Battistelli et Chagnoux, 2008), les auteurs appréhendent la temporalité des textes au travers de mécanismes relevant de la notion de prise en charge énonciative. Le concept de prise en charge repose sur le fait de « dire ce qu'on croit (être vrai) ». Ainsi, les énonciateurs d'un énoncé assurent implicitement la véracité de leurs propos. Les exemples extraits de Chagnoux (2008) illustrent la diversité des mécanismes langagiers par laquelle un énonciateur peut exposer des faits ou des opinions sans en assumer complètement la prise en charge (*cf.* exemple 25).

- (a) **Selon** le ministre canadien, il y aurait huit ressortissants canadiens parmi les victimes.
- (b) La police libanaise **affirme que** l'aviation israélienne a entièrement détruit un lieu de prière et de rassemblement chiite à Budaiï, près de Baalbeck, dans l'est du Liban.
- (c) « Les deux prisonniers israéliens ont été transférés dans un lieu sûr », **a déclaré** le Hezbollah.

EXEMPLE 25 – Exemples extraits de (Chagnoux, 2008).

Les auteurs proposent une approche qui considère le texte comme un ensemble de segments, nommé « référentiel », qui sont des blocs discursifs homogènes à l'instar des cadres de Charolles (1997), avec pour but de modéliser la dynamique discursive du texte représenté par les relations hiérarchiques qu'entretiennent les référentiels. Cette dynamique discursive est visualisée sous forme de graphe orienté dont les noeuds sont les référentiels du texte et les arcs sont les transitions entre référentiels.

<seg1>À propos du lieu de l'enlèvement des deux soldats, les versions diffèrent. Les Israéliens **indiquent qu'**<seg2>ils ont été capturés près de la ferme collective de Zarit en territoire israélien tout près de la frontière libanaise.</seg2> De son côté, la police libanaise **soutient que**<seg3>la capture s'est produite dans la région de Aïta al-Chaab en territoire libanais donc proche de la frontière libano-israélienne où une unité israélienne avait pénétré **le matin même**.<seg3></seg1>

EXEMPLE 26 – Exemple extrait de (Battistelli et Chagnoux, 2008) où les auteurs relèvent trois référentiels énonciatifs distincts :

- Le référentiel de l'énonciateur sur lequel sont validés trois faits - la divergence des versions de la capture et les deux déclarations antagonistes ;
- le référentiel des Israéliens introduit par « Les Israéliens indiquent qu' » et sur lequel est validé le fait que la capture a eu lieu près de la frontière libanaise (et donc en territoire israélien) ;
- le référentiel de la police libanaise introduit par « la police libanaise soutient que » et sur lequel est validé le fait que la capture a eu lieu en territoire libanais.

L'analyse des segments qu'elles proposent consiste à rechercher l'ouverture et la fermeture des référentiels qui se traduisent linguistiquement par la présence de marqueurs linguistiques (des marqueurs strictement grammaticaux comme les temps verbaux ou les groupes adverbiaux, des marqueurs syntaxico-sémantiques et des marqueurs typographiques). Ces marqueurs linguistiques fonctionnent sur des principes de cohésion ou de rupture. Les marqueurs de cohésion assurent l'homogénéité discursive en liant ensemble les segments au sein d'un même bloc cohérent. À l'inverse des marqueurs de cohésion, les marqueurs de rupture indiquent les transitions entre référentiels distincts.

L'analyse repose ici sur les principes similaires aux travaux sur les cadres de discours, c'est-à-dire qu'il ne s'agit donc pas simplement de repérer des marqueurs mais plutôt de repérer des configurations de marqueurs permettant de conclure ou non à l'intégration de segments dans un bloc. Nous orienterons notre travail, sur la facette de l'analyse automatique des phénomènes itératifs, en nous fondant sur ces mêmes principes.

Dans le cadre de ce travail, nous n'avons pas approfondi cette question de la charge énonciative. Il serait intéressant néanmoins par la suite d'étudier cette question, notamment au niveau des modèles (*SdT*, ...). Prenons l'exemple 27, nous avons deux itérations qui diffèrent uniquement par le nombre d'explosions mais qui fait référence au même « événement ». Il nous

semble que la notion de charge énonciative porte une information qui doit aboutir à une seule interprétation de ces deux itérations.

Selon les autorités, il y a eu trois explosions dans le centre de Bagdad au lever du jour. De leurs côtés, les riverains **affirment qu'il y eu au minimum une dizaine d'explosions.**

EXEMPLE 27 – Exemple construit d'une énoncé illustrant les itérations dans un contexte de prise en charge énonciative.

EN RÉSUMÉ

Nous avons exposé dans ce chapitre trois grandes orientations de modèles traitant de la temporalité : des modèles essentiellement référentiels, aspectuels et discursifs. Les différents modèles que nous avons présenté ont des approches différentes du traitement des phénomènes itératifs : les modèles référentiels ont une approche quantificationnelle⁵ et les modèles aspectuo-temporels représentent l'itération à l'aide d'une duplication de la structure représentative d'un procès. Pour les modèles discursifs, nous avons constaté une absence de travaux sur la dimension textuelle de ces phénomènes.

De ce constat, nous avons mis en place une étude fine sur la manifestation textuelle des itérations en corpus (*cf.* chapitre 6) dans une double perspective, d'une part, décrire la configuration textuelle des phénomènes itératives ; d'autre part, de caractériser les marqueurs nécessaires à une analyse automatique.



5. Notons que les modèles référentiels présentés, et plus généralement, les travaux ayant une approche quantificationnelle des itérations ne prennent pas, ou peu, en compte la notion d'aspect, qui est un de nos axes de recherches dans le cadre de cette thèse.

Deuxième partie

DE LA SDT
AUX REPRÉSENTATIONS DISCURSIVES

PRÉAMBULE

L'objectif de cette partie est de présenter le contexte théorique dans lequel s'inscrit notre étude des phénomènes itératifs. Ce contexte est issu de plusieurs collaborations sur les ordres de grandeur et de répétition dans le cadre du projet TCAN OGRE (cf. Lévy *et al.* (2005)). À l'issue de ce projet, un groupe de travail autour de Laurent Gosselin, Patrice Enjalbert, Gérard Becher et Yann Mathet s'est formé sur la problématique des phénomènes itératifs, et auquel nous avons participé. Cette collaboration a donné lieu à la formalisation de trois modélisations des phénomènes itératifs⁶. Ces modélisations reposent sur une double appréhension commune d'un énoncé itératif :

- celle du *procès modèle* (ou occurrence type) d'une part : l'entité événementielle en quelque sorte générique qui est présentée comme répétée dans le temps ;
- celle du *procès global* (ou série itérative) de l'autre : l'itération proprement dite, conçue comme ensemble des occurrences du modèle, et donnant accès à ces occurrences.

Nous exposons dans la suite de cette partie les travaux issus de ce groupe qui traitent du phénomène de l'itération avec trois regards distincts.

ORIENTATION LINGUISTIQUE : Laurent Gosselin a étendu la *SdT* que nous avons présentée au chapitre 2, en intégrant une nouvelle définition de l'itération. Elle est conçue comme un macro-procès englobant les occurrences.

ORIENTATION LOGIQUE : Patrice Enjalbert et Gérard Becher proposent, dans la continuité de cette extension de la *SdT* et de manière complémentaire, une formalisation algébrique des itérations en se basant sur la notion de série.

ORIENTATION INFORMATIQUE : Yann Mathet élabore un modèle non plus fondé sur la *SdT* mais sur une représentation cognitive de l'analyse d'un énoncé itératif en proposant une modélisation objet qui prend en considération de façon prépondérante l'aspect discursif dudit énoncé.



6. Ces modélisations sont présentées de manières détaillées dans un ouvrage commun à paraître et intitulé « Aspects de l'itération : une approche pluridisciplinaire ».

L'ITÉRATION DANS LE MODÈLE *SdT*

Dans ce chapitre, nous exposons le traitement des phénomènes itératifs au sein de la *SdT*. Ce travail est une extension récente de la *SdT*, et est encore en cours d'élaboration. Comme nous l'avons vu au chapitre 2, ce modèle, issu des travaux de Gosselin (1996, 2005), pose la représentation de la temporalité autour de quatre intervalles et de leur mise en relation à l'aide de règles. Il se fonde notamment sur la visée aspectuelle des procès pour construire une représentation.

SOMMAIRE

3.1	Inscription temporelle de l'itération	51
3.2	Sources de l'itération	53
3.3	Calcul des visées aspectuelles	55

3.1 INSCRIPTION TEMPORELLE DE L'ITÉRATION

Laurent Gosselin part du constat que « l'itération résulte d'un marquage tantôt lexical, tantôt grammatical, ou peut même n'être pas explicitement marquée » expliquant l'incapacité pour les approches classiques de classer l'itération dans la catégorie de l'aspect lexical ou dans celle de l'aspect grammatical. Il propose donc de requalifier l'opposition classique entre aspect lexical et grammatical au profit d'une distinction purement sémantique, qui oppose l'aspect conceptuel à la visée aspectuelle.

L'aspect conceptuel correspond à l'appréhension du procès comme entité sémantique, éventuellement complexe, structurée. Le procès est le résultat d'un processus de subsomption d'une portion découpée dans le flux des changements et des situations intermédiaires. Ces considérations conduisent Laurent Gosselin à poser, au niveau de l'aspect conceptuel, l'existence d'une catégorie de « Procès au sens large », englobant à la fois les procès au sens strict, les séries itératives comme macro-procès, et les phases dès lors qu'elles sont catégorisées comme sous-procès au moyen de coverbes. La visée aspectuelle opère la monstration du procès par l'intermédiaire d'une « fenêtre de monstration » ou « intervalle de visibilité » qui en donne à voir tout ou partie est comme opération de monstration d'un procès préalablement construit.

Dans le modèle *SdT*, l'itération est donc conçue comme une série itérative. Cette série correspond à une agglomération de procès identiques, c'est-à-dire une série d'occurrences d'un élément itéré. Ce nouveau concept permet de rendre compte du fait que l'itération peut faire simultanément l'objet de deux visées aspectuelles. L'énoncé de l'exemple 28 permet d'illustrer ce dernier propos. Intuitivement, nous constatons que la série itérative est montrée comme inaccomplie (elle a encore lieu au moment où le locuteur parle) tandis que les procès itérés sont perçus dans leur globalité (valeur aoristique).

Depuis deux mois, il mangeait en dix minutes.

EXEMPLE 28 – Exemple extrait de (Gosselin, 2011 – à paraître).

En termes d'intervalles, il est associé à chaque procès, au sens large, un intervalle de procès ainsi qu'un intervalle de référence. La notation de ces intervalles est la suivante : IPs et IRs pour la série itérative, et IP_i et IR_i pour les procès itérés.

Analysons l'énoncé de l'exemple 28 dont la structure globale est illustrée par la figure 14. La visée est aoristique sur les procès itérés (IR CO IP), et inaccomplie sur la série (IPs RE IRs) ; d'où la compatibilité de circonstanciers apparemment contradictoires : *en dix minutes* évalue la distance entre IP_1 et IP_2 (la durée du procès modèle), tandis que *depuis deux mois* porte sur l'intervalle qui sépare IPs_1 (le début de la série) de IRs_1 (le moment considéré), dans la structure :

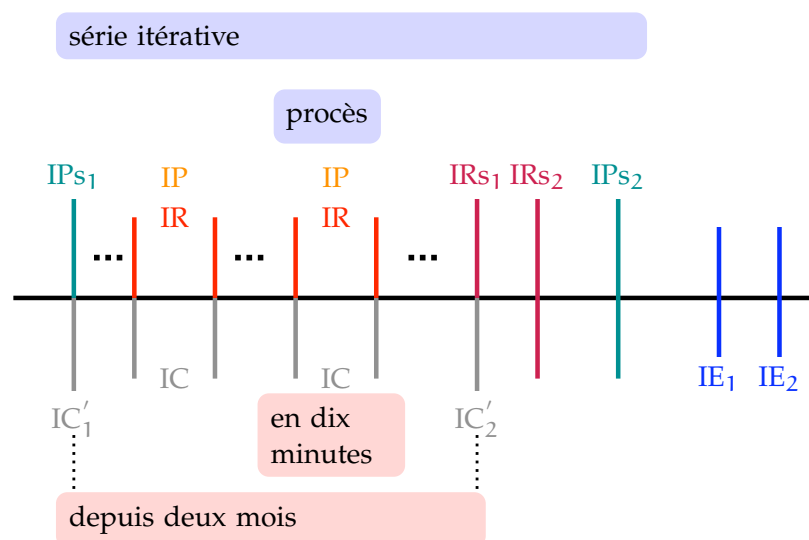


FIGURE 14 – Structure aspectuo-temporelle de l'itération simple de l'exemple 28.

Par ailleurs, le modèle distingue deux constructions possibles des itérations. D'une part, les *itérations circonstancielles*, qui résultent d'une expression temporelle comme *tous les jeudis* présentant une structure itérative intrinsèque (cf. exemple 29a) ; d'autre part, les *itérations de procès*, qui peuvent résulter d'un opérateur sur une occurrence type (cf. exemple 29b) ou bien de conflit entre des instructions linguistiques. Nous présentons les différentes sources de l'itération à la section suivante.

- | |
|---|
| <p>(a) Il va à la piscine <i>tous les jeudis</i>.
 (b) Il va <i>souvent</i> à la piscine.</p> |
|---|

EXEMPLE 29

Cette distinction ne suppose pas de modification dans le résultat de la structure globale. En effet, il y a construction d'une série d'intervalles disjoints dans les deux cas.

3.2 SOURCES DE L'ITÉRATION

L. Gosselin met en évidence deux types de sources de l'itération. D'une part, les marqueurs itératifs à proprement parler. Nous pouvons noter que ces différents marqueurs ont fait l'objet de nombreuses études, en particulier le travail de Mascherin (2007), qui a réalisé une étude du RE-, ou encore celui de Lim (2002), présenté dans le chapitre 1, qui s'est intéressée à un ensemble de marqueurs dénotant plus particulièrement la fréquence.

Les principaux types de marqueurs sont :

- les lexèmes verbaux intrinsèquement itératifs (par ex. : hachurer) ;
- les affixes itératifs (*-iller* dans sautiller, *-ailler* dans tournailler, *re-* dans recommencer) ;
- les déterminants du SN objet dans certains types de groupes verbaux (manger des/deux fraises) ;
- les périphrases itératives (avoir coutume de) ;
- les circonstanciels temporels (le mardi, chaque fois que) ;
- les adverbes itératifs fréquentatifs (parfois, rarement) ;
- les adverbiaux itératifs répétitifs (trois fois, à cinq reprises) ;
- les adverbes présuppositionnels (encore).

D'autre part, les itérations peuvent résulter de conflits entre les instructions linguistiques (cf. exemple 28) ou de contraintes pragmatico-référentielles (cf. exemple 30). Dans ce dernier exemple, le conflit provient de la durée exprimée

et de la durée continue de l'activité *nager*, qui sont incompatibles pour un être humain.

Paul a nagé pendant 10 ans.

EXEMPLE 30

Reprenons l'énoncé de l'exemple 28 pour illustrer un conflit :

Depuis deux mois, il mangeait en dix minutes.

Cet énoncé présente un conflit entre l'imparfait et *en dix minutes*. Ils codent les instructions suivantes :

- *imparfait* : $IP_1 < IR_1$, $IR_2 < IP_2$ (aspect inaccompli)
- *en + durée* : $IC_1 = IP_1 = IR_1$, $IC_2 = IP_2 = IR_2$
- *depuis + durée* : $IC_{1'} = IP_1$ ou IP_2 , $IC_{2'} = IR_2$

Il y a donc conflit entre les instructions : $IP_1 = IR_1$ et $IP_1 < IR_1$

La résolution de ce conflit passe par l'itération. Elle consiste à créer une série de procès, notée $[IP_{s1}, IP_{s2}]$, à laquelle est associé un intervalle de référence $[IR_{s1}, IR_{s2}]$, de sorte que chacune des occurrences de procès puisse être vue de façon aoristique, tandis que la série itérative dans son ensemble est présentée sous l'aspect inaccompli (marqué par l'imparfait). Le circonstant *depuis deux mois* porte donc sur la série globale, et marque le décalage entre IP_{s1} et IR_{s1} , tandis que *en dix minutes* porte sur chaque occurrence de procès. L'ensemble des contraintes linguistiques est représentée dans la figure 15. Par souci de lisibilité, le modèle prévoit de distinguer les axes temporels pour la série itérative et la série de procès. Nous simplifions encore la représentation en symbolisant toutes les occurrences par une seule correspondant à un procès modèle.

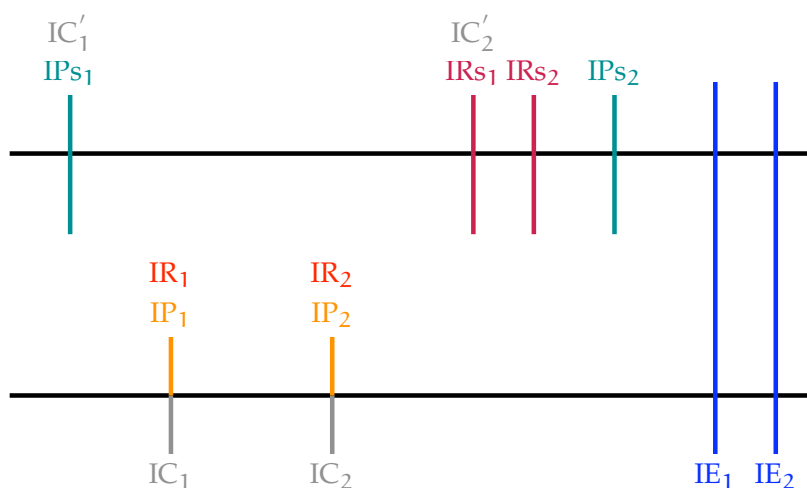


FIGURE 15 – Exemple d’une structure aspectuo-temporelle résultant d’un conflit.

3.3 CALCUL DES VISÉES ASPECTUELLES

Nous venons d’exposer qu’au sein de la *SdT*, l’itération relevait de l’aspect conceptuel et se traduisait par la construction d’un procès englobant les procès itérés. Se pose alors, au regard du calcul des visées aspectuelles, la question de la distinction entre l’aspect semelfactif et itératif.

Pour le calcul des visées aspectuelles, la *SdT* utilise les différents marqueurs aspectuo-temporels présents en énoncé pour calculer la structure associée à un procès semelfactif. Le modèle se repose sur les facteurs suivants :

- A. l’instruction associée à un temps verbal (par ex. l’imparfait qui marque l’inaccompli et le passé simple l’aoristique) ;
- B. l’instruction codée par un auxiliaire de visée aspectuelle (par ex. *être en train de* qui marque l’inaccompli, et *être sur le point de* le prospectif), ainsi que les co-verbes de phase et de modalité d’action ;
- c. les règles associées aux principes généraux présentées à la section 2.2.2.1 page 32 :
 - a) les propriétés d’anaphoricité de l’intervalle de référence ;
 - b) la corrélation globale entre visée aspectuelle et temps relatif ;
 - c) la présence de circonstanciels de durée totale impliquant l’accès aux bornes du procès et excluant donc l’inaccompli ;
 - d) la présence de [depuis + durée] qui exclut l’aspect aoristique.

Au sein du modèle, ces marqueurs sont également utilisés pour calculer la double visée aspectuelle d’un procès itératif. Du fait de cette double visée, il est nécessaire de déterminer l’interaction de ces facteurs pour savoir si la

valeur aspectuelle qu'ils portent s'appliquent à la série itérative ou aux procès itérés.

Le modèle prévoit deux nouveaux facteurs, en complément de ceux présentés ci-dessus, pour le calcul des visées aspectuelles des procès itératifs :

1. Les séries itératives intrinsèquement bornées (*il a gagné cette course 4 fois*), c'est-à-dire dont le nombre d'occurrences est déterminé, qui sont très difficilement compatibles avec la visée inaccomplie.
2. L'agglomération de procès différents qui fait l'objet d'une visée aoristique dès lors qu'elle est itérée (*cf. exemple 31*).

Chaque matin, en un quart d'heure, il se levait, s'habillait et déjeunait.

EXEMPLE 31 – Exemple (104) extrait de Gosselin (2011 – à paraître) illustrant la notion d'agglomération de procès différents au sein de la *SdT*.

L'ensemble des facteurs présentés interagissent de la façon suivante :

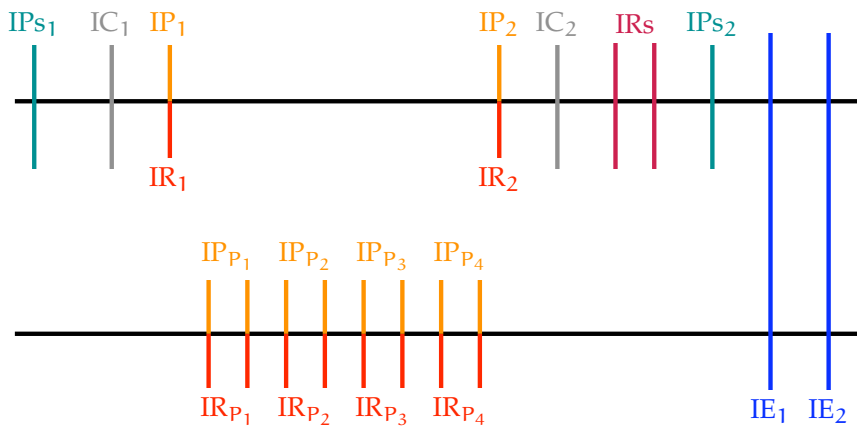
- I. Si le principe 1 ci-dessus ne s'y oppose pas, les instructions aspectuelles codées par les temps verbaux déterminent la visée aspectuelle qui affecte la série itérative englobante.
- II. Les auxiliaires de visée aspectuelle et les co-verbes portent soit sur la série itérative (*il est sur le point de venir le lundi*) soit sur le procès modèle (*chaque lundi, il est sur le point de venir*) en fonction de la structuration syntaxique et stratificationnelle de l'énoncé.
- III. Tout procès (au sens large) dont la visée aspectuelle n'est régie ni par les instructions marquées par le temps verbal ni par celle que code un éventuel auxiliaire ou co-verbe à valeur aspectuelle, voit sa visée aspectuelle déterminée par les principes généraux ((c.a) à (c.d)).

Application sur un exemple

Félicité tous les jours s'y rendait.
À quatre heures précises, elle passait au bord des maisons, montait la côte, ouvrait la barrière, et arrivait devant la tombe de Virginie.

EXEMPLE 32 – Exemple extrait d'Un cœur simple.

La série itérative est vue comme inaccomplie du fait de l'imparfait (et on pourrait lui adjoindre un circonstanciel de type *depuis deux mois*). L'agglomération de procès différents est subsumée par le procès *Félicité s'y rendre*, constituant une « élaboration ». Chacun des procès constitutifs de l'agglomération est vue de façon aoristique, du fait que nous ayons une succession de procès ; l'agglomération elle-même fait l'objet d'une visée aoristique, selon le principe 2 vu précédemment. Nous obtenons la représentation de la figure 16. Le premier axe inscrit les intervalles associés au procès et à la série itérative. Sur le second axe, nous avons développé le macro-procès correspondant à la relation d'élaboration du premier procès.



IR_{P_1} coïncide avec IP_{P_1} , *passer au bord des maisons* ;

IR_{P_2} coïncide avec IP_{P_2} , *monter la côte* ;

IR_{P_3} coïncide avec IP_{P_3} , *ouvrir la barrière* ;

IR_{P_4} coïncide avec IP_{P_4} , *arriver devant la tombe* ;

IR coïncide avec IP : agglomération de procès, correspondant au procès *s'y rendre* ;

IC recouvre IP ;

IRs est antérieur à IE (temps passé) ;

IPs recouvre IRs , inaccompli dû à l'imparfait.

FIGURE 16 – Exemple d'une représentation de la structure aspectuo-temporelle associée à l'exemple 32.

EN RÉSUMÉ

Nous avons rendu compte dans ce chapitre de l'intégration des phénomènes itératifs au sein de la *SdT*, en présentant les principes généraux du calcul aspectuel. Principes qui sont pré-existants dans la *SdT*, mais dont il a fallu étudier l'interaction dans le cadre de l'aspect itératif.

La contribution majeure de cette extension concerne donc ces deux aspects, d'une part l'inscription temporelle de l'itération, qui repose sur une double visée aspectuelle associée à une série itérative et aux procès itérés ; et d'autre part, les interactions des marqueurs aspectuo-temporels.

Nous consacrons la partie iv au calcul aspectuel des énoncés itératifs au travers de l'implémentation du système *CAsTe*.



UNE FORMALISATION ALGÈBRIQUE

Ce chapitre développe une approche des phénomènes itératifs basée sur les principes de la *SdT*. Cette approche est fondée sur un point de vue ensembliste fonctionnant en termes d'intervalles.

SOMMAIRE

4.1	Les expressions calendaires itératives	60
4.2	L'ontologie temporelle	61
4.3	La relation circonstancielle itérative	62
4.3.1	Déterminer les séries à mettre en relations	62
4.3.2	Calculer la relation circonstancielle	63
4.4	Les adverbess fréquents	66

INTRODUCTION

Le formalisme algébrique, que nous exposons ici, développé par Enjalbert et Becher (2011 – à paraître), se place dans la continuité de la *SdT*, en proposant de développer une approche des *procès itératifs* et des *relations circonstanciell*, basée sur des intervalles. Conformément au chapitre 3, qui développe l'idée qu'un processus itératif doit être appréhendé à deux niveaux (celui du processus « modèle » ou « type », répété au cours du temps, et celui de l'itération elle-même, processus « global » résultant de cette répétition) le formalisme transforme l'intervalle de processus global (celui de l'itération) en un intervalle généralisé, autrement dit une *série*, représentant l'ensemble des occurrences du processus type.

La notion de série est donc le noyau fondamental de cette formalisation. Elle repose sur un intervalle généralisé, qui consiste en la réunion d'un nombre fini d'intervalles convexes ou de points. L'enveloppe convexe est définie pour tout intervalle généralisé I composé de n intervalles J , par le plus petit intervalle englobant les constituants de l'intervalle généralisé.

Une série correspond à un cas particulier d'intervalle généralisé, dont les intervalles constitutifs ne se chevauchent pas, et sont donc structurés par une relation de succession. Les auteurs ont développé des relations et des opérations de base pour pouvoir manipuler les séries et les intervalles, telles que :

Relation :

- *inclusion* correspond à l'inclusion de tout intervalle d'une série S_1 dans un intervalle d'une série S_2 ;

Exemple : La série JOUR, constituée des jours, est incluse dans la série des semaines, noté SEMAINE.

- *extraction* exprime le fait qu'une série S_1 peut être constituée par tout ou partie des intervalles appartenant à une autre série S_2 , noté $S_1 < S_2$.

Exemple : S_1 est une série constituée des jours {*lundi, mardi, mercredi*}, extraite de la série S_2 des jours de la semaine.

Opération :

- *ratio* d'une série S_1 par rapport à une série S_2 ;

Exemple : Le ratio de la série JOUR par rapport à SEMAINE est de 7.

- *restriction* d'une série S_1 par rapport à une série S_2 (cf. figure 17) ;
- etc.

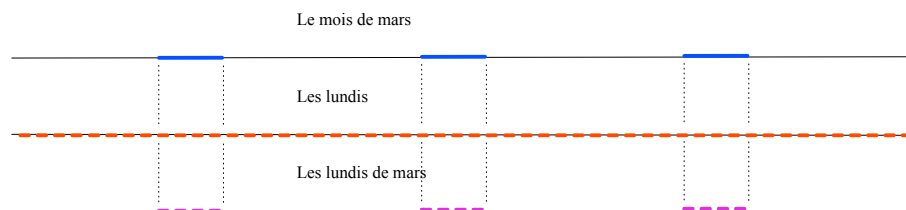


FIGURE 17 – Modélisation de l'opération de restriction de *Tous les lundis des mois de mars*. Le résultat de la restriction de la série des lundis (notée LUNDI) par rapport à la série des mois de mars (notée MARS) sera une sous-série des éléments extraits de LUNDI appartenant aussi à MARS, notée LUNDI/MARS. Pour des raisons de lisibilité, la figure illustre la restriction pour une seule année : le schéma est à répéter pour plusieurs années.

Cette formalisation de l'itération propose donc des outils mathématiques pour représenter ces structures et calculer les relations circonstanciennes impliquées dans la modélisation aspectuo-temporelle, autour de la notion de série. Nous en présentons ici les grandes lignes et renvoyons à l'ouvrage de référence pour plus de précisions ¹.

4.1 LES EXPRESSIONS CALENDAIRES ITÉRATIVES

Afin de pouvoir calculer les relations circonstanciennes, le modèle permet de décrire la sémantique des expressions calendaires itératives (ci-après ECI).

1. « Aspects de l'itération : une approche pluridisciplinaire – L'itération dans le modèle SdT : une formalisation algébrique », à paraître.)

Ce sont ces séries qui seront mises en relation avec la série portée par le processus itératif.

La sémantique d'une ECI décrit un intervalle circonstanciel non convexe qui marque la succession de périodes temporelles non contiguës. Elle prend la forme d'une fonction associant une ECI à un ensemble de séries temporelles. En effet, il est possible d'obtenir plusieurs séries pour une même ECI dans le cas de déterminants quantificateurs tels que *la plupart*, *certain*, *etc.*

Prenons l'exemple de l'ECI *la plupart des lundis de mars*. Le déterminant *la plupart des* permet de retenir certains intervalles de la série LUNDI/MARS. Le modèle prévoit la fonction PLUPART définie à l'équation 4.1 où *seuil* est une valeur à ne pas dépasser. En appliquant la fonction PLUPART à la série LUNDI/MARS, nous obtenons un ensemble de série illustré à la figure 18.

$$\text{PLUPART } S = \{S' \text{ tq } S' \prec S \text{ et } \|S'\|/\|S\| > \textit{seuil} \}$$

4.1 – Définition algébrique de la fonction PLUPART S.

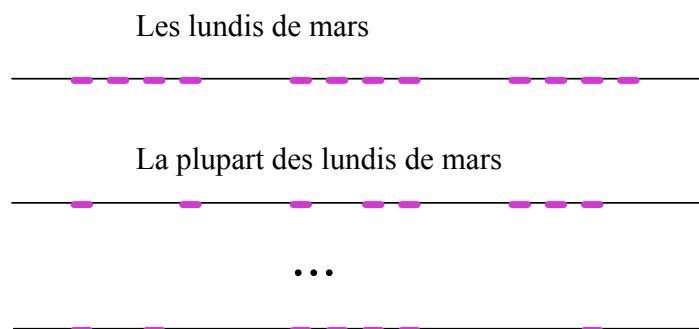


FIGURE 18 – Représentation de l'ECI *la plupart des lundis de mars*. Il y est représenté deux séries possibles de l'ECI, obtenues à partir de la fonction PLUPART sur la série LUNDI/MARS.

Pour construire cette sémantique, le modèle décrit un certain nombre des constructions intervenant dans les ECI en utilisant les opérations disponibles par le modèle algébrique.

4.2 L'ONTOLOGIE TEMPORELLE

L'approche présentée ici est basée sur les principes de la *SdT*. Afin de rendre compte de la notion aspectuelle prise en compte au niveau du processus

type et de l'itéré, les auteurs ont mis en place l'appareillage nécessaire en termes d'intervalles.

Au niveau « type », on retrouve les intervalles de la *SdT*, c'est-à-dire :

- l'intervalle de procès type, noté IP_t ;
- l'intervalle de référence type, noté IR_t .

Au niveau « itération », une distinction se fait avec la *SdT* : l'intervalle de référence, noté IR_i , demeure identique, c'est-à-dire convexe et classique contrairement à l'intervalle de procès qui est un intervalle généralisé du fait de la représentation de l'itération sous forme de série. L'aspect est caractérisé par les relations aspectuelles entre IP_t et IR_t pour le niveau type, et entre IP_i et IR_i pour le niveau itération.

Le formalisme reproduit un schéma similaire pour les compléments circonstanciels, à savoir deux composantes :

- une composante pour la série correspondant à la succession des plages temporelles évoquées par le complément circonstanciel, appelé intervalle circonstanciel itératif et noté IC_i ;
- une composante qui représente une période générique correspondant à un intervalle convexe, dit intervalle circonstanciel type, noté IC_t .

Les relations circonstanciennes sont représentées classiquement par une relation entre l'intervalle circonstanciel et l'intervalle de procès ou de référence².

4.3 LA RELATION CIRCONSTANCIELLE ITÉRATIVE

L'analyse pour le calcul des circonstanciennes itératives s'opère en trois temps :

1. déterminer les séries à mettre en relation ;
2. déterminer la relation entre le procès typique et l'intervalle circonstanciel type (IC_t) ;
3. calculer la relation circonstancielle, c'est-à-dire exprimer une constante sur les séries.

4.3.1 Déterminer les séries à mettre en relations

Le modèle repose sur le constat que le complément circonstanciel, noté CC , lui-même peut être itératif (*Tous les lundis, Jean allait à la piscine*) ou non (*La cigale chanta tout l'été*), et que diverses combinaisons sont par ailleurs possibles

2. Les relations utilisées dans ces travaux sont identiques à celles de la *SdT*, que nous avons présentée à la section 2.2.2. Pour rappel, il s'agit des relations d'antériorité, de postériorité, de simultanéité, de recouvrement, de coïncidence et d'accessibilité.

(*L'an dernier, tous les lundis, Jean allait à la piscine*), avec la possibilité d'avoir un adverbe fréquentiel (noté A_{freq}) qui vient modifier la relation circonstancielle (*Le lundi, Jean va souvent à la piscine* ou *L'an dernier, Jean allait souvent à la piscine*). Il en découle les trois configurations suivantes :

- i) association d'un procès (itératif) P et d'un CC lui même itératif, ce qu'une représentation symbolique donne : $CC_{iter} + P_{iter}$;
- ii) association d'un procès (itératif) P et d'un CC non itératif, dit encore *unitaire* : $CC_{unit} + P_{iter}$;
- iii) association d'un procès (itératif) P, d'un CC itératif et d'un CC unitaire : $CC_{unit} + CC_{iter} + P_{iter}$.

4.3.2 Calculer la relation circonstancielle

Le calcul de la relation circonstancielle, dans les configurations (i) et (iii) où le CC_{iter} intervient, se fonde sur le mécanisme de présupposition stratificationnelle de Nolke (1994). Il repose sur le fait que, dans un énoncé, certains éléments informationnels sont supposés connus et établis dans la situation de communication (*substrat* de l'énoncé) ; d'autres éléments constituent en revanche un apport d'information nouvelle (*foyer* de l'énoncé)³. L'exemple 33 présente deux énoncés impliquant cette stratification de l'information.

- (a) Le lundi, Jean allait à la piscine.
 - *Substrat* = le lundi ;
 - *Foyer* = Jean va à la piscine.
- (b) Jean allait à la piscine le lundi.
 - *Substrat* = Jean va à la piscine ;
 - *Foyer* = le lundi.

EXEMPLE 33

Ce mécanisme permet de rendre compte assez finement de la différence d'interprétation d'énoncés. Ainsi, l'exemple 33a spécifie que, immanquablement, le lundi, Jean allait à la piscine. Mais il est possible qu'il ait eu aussi cette pratique le dimanche, ou occasionnellement. L'exemple 33b spécifie quand à lui que seul le lundi était « jour de piscine ».

Dans ce formalisme, la représentation en substrat/foyer permet de déterminer le sens de la relation entre les séries associées au substrat et au foyer

3. La dualité de rôle peut être mise en évidence en considérant que l'énoncé constitue une réponse à une question, réelle ou fictive, le substrat figurant alors dans la question.

(notées respectivement I_{Substrat} et I_{Foyer}) telle que toute composante de la série du substrat est en relation avec une composante de la série du foyer, par une relation dite *type*.

Prenons l'énoncé de l'exemple 33a, appliquons lui le principe d'analyse du modèle :

- *stratification* : le substrat est le CC, donc le lundi, et le foyer est le procès ;
- *niveau « type »* : selon les règles de la *SdT*, la relation correspond au recouvrement de l'intervalle de procès par l'intervalle circonstanciel : $IC_t \text{ RE } IP_t$. Cette relation est issue de l'aspect aoristique (cf. figure 19). Nous détaillons les règles de calcul qui s'appliquent dans le chapitre 9.

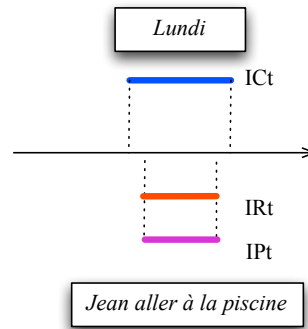


FIGURE 19 – Représentation d'un procès type.

- *niveau « itération »* (cf. figure 20) :
 - cas général : $\forall I \subset I_S, \exists J \in I_F$ tel que $I \text{ R}_t J$;
 - dans le cas présent : $\forall I \subset IC_i, \exists J \in IP_i$ tel que $I \text{ RE } J$.

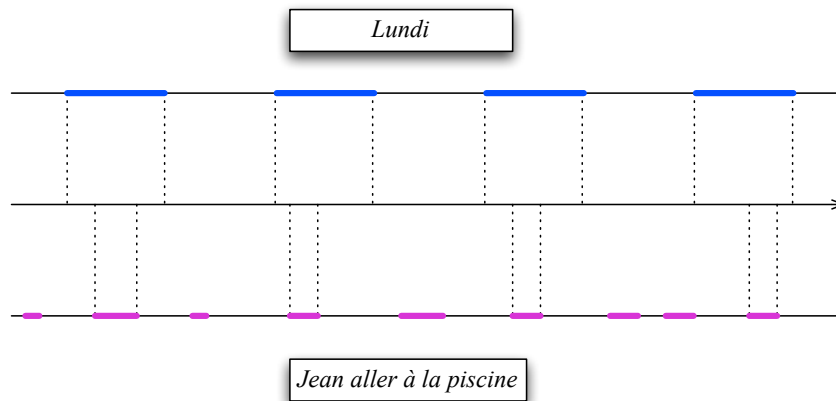


FIGURE 20 – Représentation de la série d'itérés de l'exemple 33a dont le substrat est le CC et le foyer est le procès.

Dans le cas de l'énoncé de l'exemple 33b, nous avons une inversion du substrat et du foyer. La relation au niveau type reste la même, c'est-à-dire une relation de recouvrement, tandis que la relation entre les séries se trouve inversée. Nous obtenons donc :

$\forall I \subset IP_i$ (*Jean aller à la piscine*), $\exists J \in IC_i$ (*lundi*) tel que $I \text{ RE } J$ (cf. figure 21).

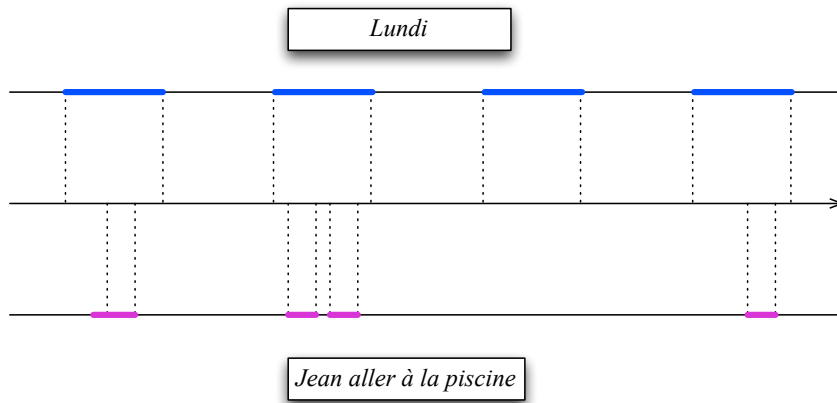
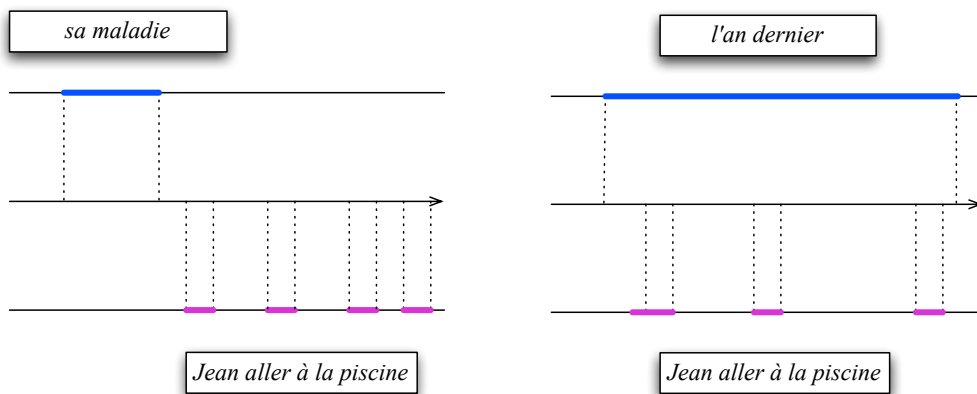


FIGURE 21 – Représentation de la série d'itérés de l'exemple 33b dont le substrat est le procès et le foyer est le CC.

En ce qui concerne le calcul de la relation circonstancielle incluant un CC_{unit} , le modèle repose sur l'hypothèse qu'il ne contraint que la période couverte par l'itération (cf. exemple 22b).



(a) Après sa maladie, Jean est allé régulièrement à la piscine

(b) L'an dernier, Jean allait à la piscine

FIGURE 22 – Représentation des configurations faisant intervenir un CC_{unit} .

4.4 LES ADVERBES FRÉQUENTATIFS

L'analyse des adverbess fréquentatifs est appréhendée comme précédemment en termes de stratification énonciative. Cette analyse repose sur la distinction de deux types de configurations :

- la première configuration dépend de la présence d'une composante itérative dans le substrat (notée $CC_{iter} + A_{freq} + P_{iter}$), l'information fréquentielle portant sur cette composante (cf. l'exemple 34a pour le CC ou l'exemple 34b pour le procès). Cette configuration est dite *non autonome* pour indiquer la présence d'une autre entité itérative dans l'énoncé ;
- la seconde configuration, dite *autonome*, dépend de la présence d'une composante itérative dans le foyer exclusivement (notée $CC_{unit} + A_{freq} + P_{iter}$). L'information fréquentielle exprimée par l'adverbe porte sur le procès lui-même (cf. exemple 34c).

- (a) Le lundi, Jean va souvent voir sa grand-mère.
- (b) Jean va souvent voir sa grand-mère le lundi.
- (c) L'été dernier, je suis allé régulièrement voir ma grand-mère.

EXEMPLE 34

L'hypothèse fondamentale de l'analyse des adverbess fréquentatifs repose sur le fait qu'ils ne créent pas d'intervalle circonstanciel mais qu'ils jouent le rôle de modifieur sur les séries mises en relation. Le modèle détaille plus précisément l'appareillage nécessaire pour rendre compte de ces configurations, et des usages sélectifs et restrictifs des adverbess fréquentatifs.

Configuration autonome

Dans cette configuration, la seule structure itérative disponible est le procès lui-même. L'information fréquentielle portée par l'adverbe opère donc une modification sur le procès. La nature de cette modification opérée par l'adverbe correspond à la spécification du degré d'occupation de l'intervalle du complément circonstanciel (par exemple la régularité des occurrences ou encore leur nombre). Formellement, le modèle spécifie la relation d'inclusion de l'intervalle de procès de l'itération (IPi) dans l'intervalle circonstanciel (IC), selon la sémantique associée à l'adverbe.

$$S \subseteq_A I$$

4.2 – « S est incluse dans I selon une distribution fréquentielle indiquée par A. »

Ainsi pour l'exemple 34c, la série des occurrences de « aller voir sa grand-mère » est incluse dans l'intervalle « l'été dernier ». Nous avons $S \subseteq_A I$, où S est la série des occurrences de « aller voir sa grand-mère » et I est « l'été dernier », avec A = régulièrement.

Configuration non autonome

Dans cette configuration, notée $CC_{iter} + A_{freq} + P_{iter}$, le modifieur porte sur la série se situant dans le substrat. La fonction de l'adverbe, dans ce cas, est d'opérer une sélection parmi les intervalles de la série. Dans l'exemple 34a, il opère une sélection parmi les lundis. Les auteurs parlent de « degré d'occupation » de l'intervalle non convexe (la série) donnée par le complément circonstanciel itératif. Formellement, le modèle introduit une famille de relations d'extraction de série, précisant ou quantifiant en quelque sorte la relation < :

$$S <_A S'$$

4.3 – S est une série extraite de S' avec une fréquence (une régularité, etc.) spécifiée par l'adverbe A.

Ainsi pour l'exemple 34a, nous avons $I <_A \text{LUNDI}$, où $I \in \text{LUNDI}$ et I contient une composante de la série du procès *Jean aller voir sa grand-mère*.

EN RÉSUMÉ

Cette formalisation des itérations est basée sur la notion d'intervalle au sein d'une série, cette série représentant l'itération. Elle intègre un appareillage permettant de décrire les expressions calendaires itératives, de calculer les relations entre la série et les compléments circonstanciels de l'énoncé, et enfin elle permet d'analyser et de représenter le rôle de *modifieur* que jouent les adverbes fréquentatifs. Le modèle est toutefois restreint à une analyse phrastique, et ne permet pas de prendre en compte la dimension textuelle d'un énoncé.



L'ITÉRATION SELON UNE APPROCHE OBJET

Le modèle auquel ce chapitre est consacré se démarque de la *SdT* en ce que la représentation sous formes d'intervalles n'est que secondaire. En effet, le *Modèle Objet des Phénomènes Itératifs* (dorénavant MOPI) propose une modélisation objet du *processus* de la construction du sens lié à l'itération, et s'inscrit par là-même dans le courant de la linguistique cognitive formelle. Nous présentons donc les fondements de ce modèle dans lequel s'inscrit en partie de notre travail.

SOMMAIRE

5.1	Hypothèse cognitive	69
5.2	Modélisation objet	72
5.2.1	L'itération	73
5.2.2	Les itérateurs	74
5.2.3	Le modèle itératif et ses constituants	75
5.2.4	Les sélections	76
5.3	Lien avec la <i>SdT</i>	77
5.4	Discussion	78

5.1 HYPOTHÈSE COGNITIVE

Le modèle MOPI (Mathet, 2007, 2011 – à paraître) repose sur l'hypothèse cognitive d'une double facette de l'itération. Selon cette hypothèse, les itérations temporelles peuvent être appréhendées selon un double point de vue : extensionnel et intensionnel. En d'autres termes, le modèle suppose que face à un procès (ou un agglomérat de procès) itéré, il est possible d'être amené à considérer ce tout :

- soit comme un ensemble de procès étalés dans le temps (point de vue extensionnel) ;
- soit comme une situation unique (point de vue intensionnel), issue d'une abstraction « canonique » de l'ensemble des itérés.

La figure 23 illustre cette double facette de l'itération à l'aide de l'énoncé de l'exemple 35. Par ailleurs, il permet de mettre en évidence le processus de construction d'une itération, que nous avons entrevu dans le chapitre précé-

Elle est allée sept fois à la montagne.

EXEMPLE 35

dent : par le biais d'un itérateur *sept fois* et d'un itérant *aller à la montagne*, nous obtenons la création de sept procès étalés dans le temps.

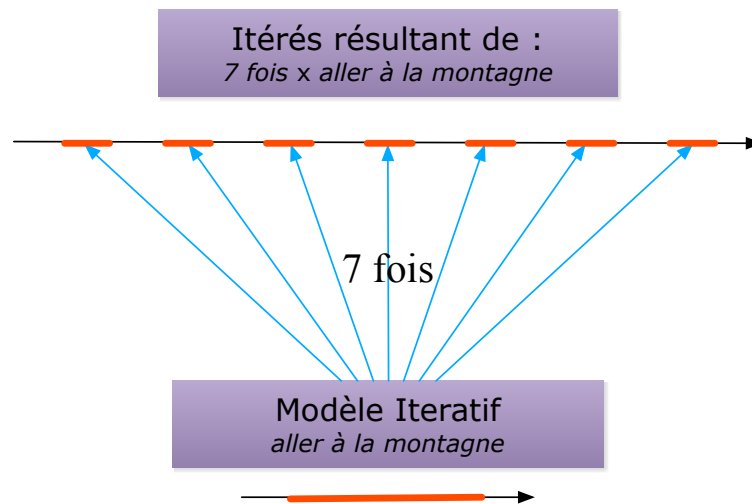


FIGURE 23 – Les *itérés* représentent le point de vue extensionnel de l'itération, tandis que l'*itérant* correspond à une abstraction des itérés (point de vue intensionnel).

De plus, cette double facette se voit confortée par l'observation de certaines configurations linguistiques rendant saillant le point de vue offert par l'une ou l'autre de ces deux facettes :

- A. Configuration privilégiant la facette extensionnelle : *Le trois septembre, elle est allée pour la première fois à la piscine, faisant une heure de brasse. Elle y est retournée trois semaines plus tard, pour une séance de deux heures environ, et une troisième fois hier matin, de 8 h à 10 h.*

L'attention est portée sur l'ensemble des itérés qui sont appréhendés de façon extensionnelle. Toutefois la facette intensionnelle transparait au travers de la lexie *fois* indiquant que l'on identifie bien des situations identiques.

- B. Configuration privilégiant la facette intensionnelle : *À cette époque, ma journée commençait vers 8 heures. Après une rapide douche, je descendais déjeuner, puis ...*

L'attention est ici portée sur une journée type en faisant abstraction de l'étalement répétitif de ces situations dans le temps.

- c. Configuration mettant en avant les deux facettes extensionnelle et intensionnelle : *Tous les jeudis, nous faisons une partie de cartes.*

Cette dernière configuration est à priori moins orientée que les deux précédentes. Il est possible en effet, selon le contexte, de glisser soit vers sa facette extensionnelle comme en A en la faisant suivre par « *En tout, il y a eu 12 jeudis. Le premier jeudi, on a joué de 20 h à 22 h, les trois jeudis suivants, à partir de 20 h 30, etc.* », soit vers sa facette intensionnelle comme en B en la continuant par *La partie commençait par le mélange puis la distribution des cartes. Puis, etc.*, où on parle d'une partie « canonique ».

Espaces mentaux et modèles itératifs

Le modèle MOPi se repose sur un certain nombre de principes et d'éléments des « Espaces Mentaux » de Fauconnier (1984). Le constat fait par l'auteur est que les deux visions possibles des itérations, extensionnelle et intensionnelle, semblent pouvoir correspondre à deux espaces mentaux différents. Si l'on place, de façon extensionnelle, des procès correspondant à une situation itérée dans l'espace mental de référence M, il est possible de construire un espace plus abstrait M', dans lequel cette situation itérée est appréhendée de façon unique.

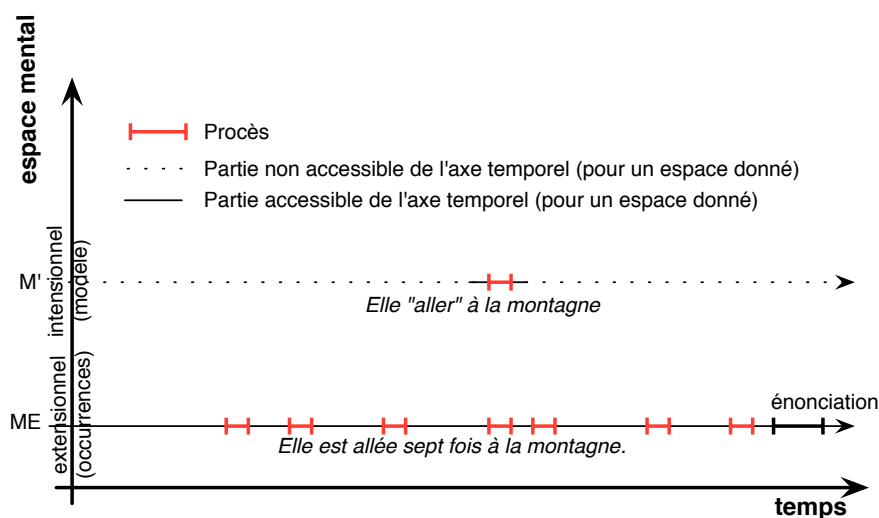


FIGURE 24 – La représentation de l'exemple 35 dans les « Espaces Mentaux » amène à la construction d'un espace mental ME contenant les 7 itérés de l'itération (vision extensionnelle). La version modèle de ces itérés se trouve positionnée au sein d'un espace M' (vision intensionnelle).

L'espace M' sert donc à construire et à manipuler la situation servant de modèle aux différents itérés présents dans M. Les propriétés de M' se dé-

duisent à la fois de celles de M et de l'introducteur itératif qui en est à l'origine. Elles sont d'ordre temporel d'une part, et référentiel d'autre part. Temporellement, M' ne permet qu'une appréhension réduite de l'axe temporel, d'une longueur inférieure ou égale à celle du plus grand itéré potentiellement constructible de M. Par ailleurs, M' n'a pas accès aux différents intervalles de ses espaces parents. Enfin, et de façon corollaire, M' n'a aucune visibilité sur les intervalles d'énonciation de ses espaces parents. Il en résulte un détachement de ces espaces mentaux vis-à-vis de la temporalité absolue (passé, présent et futur). Du point de vue de la référence, les éventuelles entités itérées dans M sont vues comme une entité unique dans M', entité qui ne garde que les propriétés communes à l'ensemble des entités itérées (nivellement par la non-spécificité).

5.2 MODÉLISATION OBJET

Contrairement au formalisme algébrique exposé au chapitre 4 qui décrit les itérations à un niveau local (circonscrit à une proposition, voire une phrase), le modèle MOPI s'est fondé sur une modélisation discursive des phénomènes itératifs. Prenons l'énoncé de l'exemple 36 afin d'en exposer les principes généraux.

Félicité tous les jours s'y rendait.
À quatre heures précises, elle passait au bord des maisons, montait la côte, ouvrait la barrière, et arrivait devant la tombe de Virginie.

EXEMPLE 36 – Extrait d'« Un cœur simple » de G. Flaubert

Dans la première phrase le marqueur « tous les jours » joue le rôle d'opérateur (au sens mathématique) de duplication du procès *Félicité s'y rendre*, qui est une abstraction non localisée temporellement de l'ensemble des itérés. Cette opération a pour conséquence de déclencher une itération, actuellement composée d'un itérateur et d'un procès « modèle » au sens du modèle MOPI.

La seconde phrase de notre énoncé vient élaborer le procès de la première avec l'agglomération des procès *passer*, *monter*, *ouvrir*, et *arriver*. La relation d'élaboration fait que ces différents procès font partie intégrante de l'itération. En conséquence, nous obtenons un procès « modèle » pour chacun des procès de l'agglomération, liés par des relations de succession.

En résumé, l'itération produite par l'analyse de cet énoncé est constituée d'un itérateur, de cinq procès « modèles », et de leurs relations respectives. Le

modèle MOPI inscrit les procès modèles dans un modèle itératif dont le rôle est de servir de modèle à ce qui est itéré, ce que nous avons précédemment désigné par *itérant*. La schématisation de l'itération est illustrée dans la figure 25.

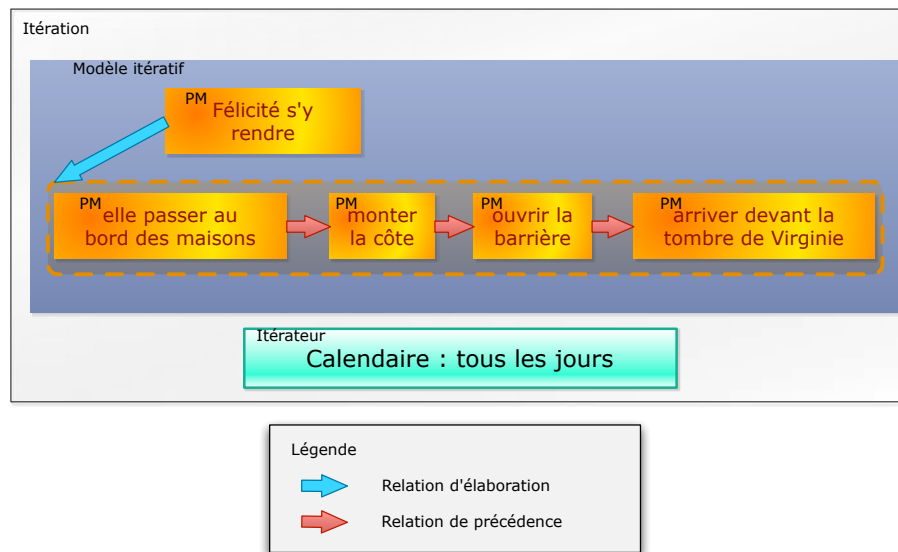


FIGURE 25 – Représentation d'une itération dans le modèle MOPI, constituée d'un itérateur et du modèle itératif, lui-même constitué de relations et de procès modèles (en jaune).

5.2.1 L'itération

Comme nous venons de le voir, au sein du modèle MOPI, l'itération est constituée d'un itérateur et d'un modèle itératif, le principe étant d'appliquer l'itérateur au modèle itératif afin de créer les itérés. L'itérateur est un objet permettant d'itérer des événements dans le temps via des marqueurs comme *tous les jeudis* ou encore *3 fois*. Le modèle itératif désigne ce qui est itéré ou plus exactement un modèle de ce qui itéré. Il est constitué d'un ou de plusieurs procès modèles, reliés les uns aux autres par des relations de différents ordres (relations temporelles, de causalité, etc.).

Une itération minimale dans le modèle MOPI est donc composée d'un itérateur et d'un modèle itératif, lui-même composé d'un ou plusieurs procès dits modèles.

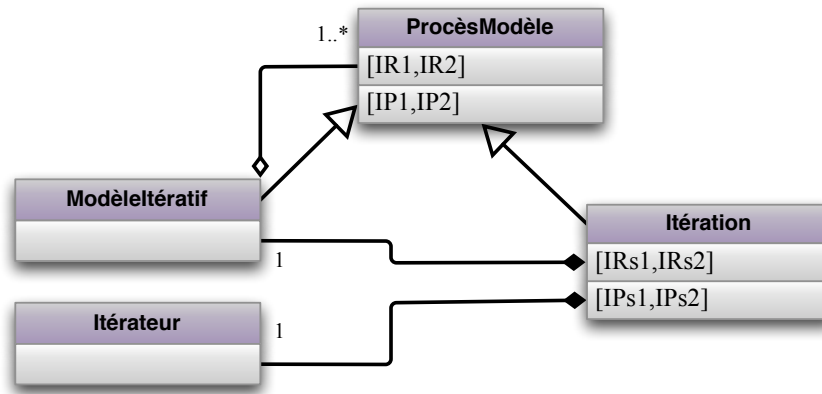


FIGURE 26 – Représentation UML de MOPI.

5.2.2 Les itérateurs

Le rôle d'un itérateur est de dupliquer des événements en proposant une suite, souvent ordonnée, de « positions » (connues ou non) de l'axe temporel à chacune desquelles on pourra faire correspondre temporellement un itéré. Le nombre de positions est porté par la sémantique associée à l'itérateur (cf. figure 23 page 70).

Les différentes classes d'itérateurs, recoupées aux données linguistiques proposées par L. Gosselin et que nous avons reportées à la section 3.2, sont les suivantes :

- A. Ensemble d'intervalles calendaires :
 - certains circonstants de localisation temporelle tels que *chaque mardi* (calendaire régulier), *un lundi sur trois* (calendaire fréquentiel), *lundi et mercredi derniers* (itérateur calendaire) ;
- B. Quantificatif :
 - déterminants du SN objet tels que *manger deux pommes* ;
 - adverbes itératifs répétitifs comme *trois fois*, *quelques fois* ;
- C. Événementiel :
 - subordonnée temporelle comme *quand ils viennent*, *dès qu'ils arrivent*, *lorsqu'il fait beau* (états sujets à changements vrai/faux) ;
 - Adverbes présuppositionnels comme *encore* ;
- D. Fréquentiel :
 - adverbes itératifs fréquentatifs comme *parfois*, *fréquemment* ;
- E. Composé : résultant de la juxtaposition de plusieurs itérateurs : *3 jeudis et une autre fois*.

Seuls deux types de données ne donnent pas lieu à une classe d'itérateur :

1. les verbes intrinsèquement itératifs dont le modèle ne rend pas compte ;
2. les conflits étant de natures très variées, ils peuvent donner lieu à différents types d'itérations, comme « Longtemps je me suis couché de bonne heure » peut donner lieu, pour des raisons pragmatico-référentielles, à la classe *CalendaireRégulier*, ou « à l'époque, il jouait du piano » à du *Fréquentiel*.

5.2.3 *Le modèle itératif et ses constituants*

Le rôle d'un modèle itératif est de servir de modèle à ce qui est itéré. Formellement, il est composé de 1 à N procès modèles, et de relations entre ces différents procès modèles.

Les relations Le modèle permet de reprendre les relations discursives telles que celle définie dans la SDRT, par exemple, la relation méronomique d'élaboration (*les parties de poker commençaient par la distribution des cartes*) ou la relation causale (*Marie est tombée car Paul l'a poussée*). Elles ne sont pas cependant formalisées en l'état actuel du modèle.

Les procès modèles Un procès modèle reprend à son compte les caractéristiques principales des procès classiques. Par ailleurs, une itération peut elle-même constituer un procès modèle, et donc être inscrite dans une autre itération, récursivement. En effet, certains énoncés imbriquent des itérations, comme peut en rendre compte l'énoncé de l'exemple 37. Ici le procès *aller sur le campus* a lieu deux fois, et se voit itéré à intervalles réguliers : le lundi.

Le lundi, je vais sur le campus deux fois.

EXEMPLE 37 – Propriété de récursion.

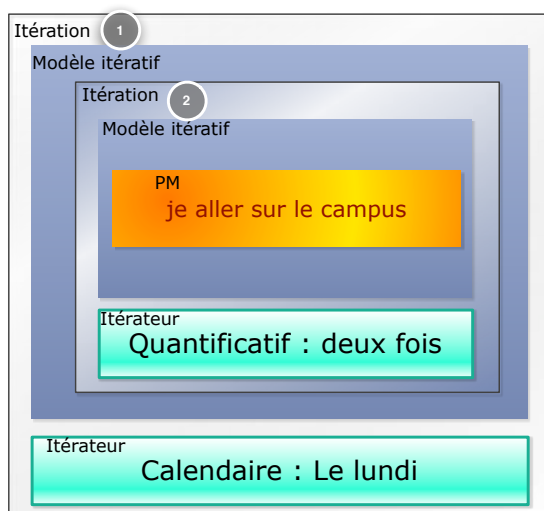


FIGURE 27 – Représentation du phénomène de récursivité d’une itération dans le modèle MOPI, où l’itération, notée 2, se voit être le procès modèle de l’itération, noté 1.

5.2.4 Les sélections

Les itérations permettent, comme nous venons de le modéliser, de créer un certain nombre d’itérés à partir d’un modèle itératif. Il est courant que, dans les énoncés, la suite apporte des modifications à certains voire tous les itérés de l’itération créée. Si l’on complète l’énoncé de l’exemple 36 par celui de l’exemple 38, l’adverbe fréquentatif *parfois* vient sélectionner certains itérés pour les enrichir d’une nouvelle information portée par le procès *Félicité être accompagnée de Madame Aubain*.

Félicité tous les jours s’y rendait.

À quatre heures précises, elle passait au bord des maisons, montait la côte, ouvrait la barrière, et arrivait devant la tombe de Virginie. [...]

Parfois, Félicité était accompagnée de Madame Aubain.

EXEMPLE 38

Le modèle MOPI prévoit un mécanisme supplémentaire pour gérer ce type de phénomènes, appelé *sélection*, permettant, une fois l’itération créée, de sélectionner un sous-ensemble de ses itérés. Elle opère une restriction sur une itération, et les procès attachés viennent modifier ou enrichir les itérés de l’itération ainsi sélectionnée. À la différence d’un itérateur, un sélecteur procède

à une restriction non plus sur un continuum temporel mais sur un ensemble discret d'itérés. La classe Sélection a deux attributs principaux :

- l'itération sur laquelle elle opère ;
- son sélecteur, qui est la fonction de restriction de l'ensemble des itérés.

Ce sélecteur peut être de deux types : fréquentiel ou quantitatif. Le sélecteur fréquentiel exprime la fréquence de sélection des itérés selon la valeur associée au sélecteur (cf. exemple 38). Comme pour les itérateurs fréquents, la densité exprimée par le sélecteur est sujette à l'interprétation du lecteur et du contexte dans lequel il s'inscrit. Le sélecteur quantitatif sélectionne un certain nombre d'itérés, sans préciser lesquels, comme dans l'exemple 39, suite possible de l'exemple 36 introduisant le contexte itératif.

Madame Aubain a accompagné Félicité trois fois.

EXEMPLE 39

La figure 28 illustre la sélection de l'énoncé de l'exemple 39.

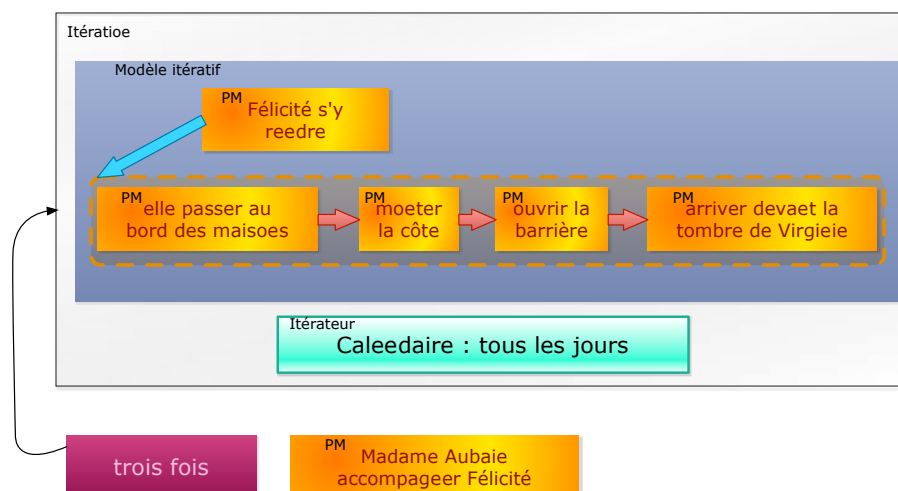


FIGURE 28 – Représentation d'une sélection d'itérés de l'itération de l'introduction au modèle. Elle est constituée d'un sélecteur et d'un processus modèle, est rattachée à une itération.

5.3 LIEN AVEC LA SDT

De manière sous-jacente, le modèle MOPI permet de rendre compte des structures aspectuo-temporelles de la *SdT*. Selon la *SdT*, tout énoncé se voit associer un intervalle d'énonciation $[IE_1, IE_2]$ correspondant au moment où

l'énoncé est produit. Dans le modèle *MOP1*, cet intervalle est placé dans l'espace de plus haut niveau, l'espace énonciatif *ME*. Le modèle associé à l'espace parent *MP* les intervalles liés à une « série de procès itérés », respectivement nommés $[IP_1, IP_2]$ pour ce qui est de l'ensemble des itérés, $[IR_1, IR_2]$ pour ce qui est de l'intervalle de référence de la série, et $[IC_1, IC_2]$ pour les intervalles circonstanciels liés à la « série de procès itérés ». Quant aux intervalles liés au modèle d'occurrence, respectivement nommés $[IP_1, IP_2]$ pour ce qui est du procès, $[IR_1, IR_2]$ pour ce qui est de l'intervalle de référence, et $[IC_1, IC_2]$ pour les intervalles circonstanciels appartenant au modèle d'occurrence (*souvent, il vient le dimanche*), ils se placent dans un espace mental modèle *MM* du présent modèle. En ce qui concerne le calcul à proprement parler des structures aspectuo-temporelles, les règles de la *SdT* s'appliquent à l'identique dans le cadre de ce modèle.

5.4 DISCUSSION

Dans la suite de nos travaux, nous nous inscrivons complètement dans cette modélisation. Nous reprenons à notre compte la terminologie et les concepts exposés ici. Nous proposons notamment une modélisation textuelle des phénomènes itératifs permettant le passage à la représentation cognitive, avec l'objectif de réaliser un traitement automatique amenant à la construction de ces représentations.

Nous souhaitons toutefois discuter le rapprochement effectué avec la *SdT* qui nous semble limiter la puissance de représentation de *MOP1*. En effet, la *SdT* telle que nous l'avons exposé dans ce manuscrit est fondée sur la notion de procès porté par un verbe. Or, comme nous l'avons esquissé dans l'introduction de ce mémoire, l'itération n'est pas portée uniquement par le verbe mais peut l'être également par un substantif¹.

À notre connaissance, actuellement, la *SdT* ne permet pas de représenter la temporalité associée aux itérations fondées sur un substantif événementiel, en particulier dans le cadre d'une phrase averbale². Là où il nous semble que *MOP1* permet d'analyser et de représenter de tel phénomènes, comme nous allons le démontrer ci-après. Précisons d'ores et déjà notre propos : nous ne

1. Nous démontrons également ce point dans le cadre de notre analyse des phénomènes itératifs sur corpus, dont nous présentons les résultats dans le chapitre suivant (*cf.* chapitre 6).

2. Rappelons que les substantifs recouvrent différentes sous-catégories :

- les substantifs objets qui regroupent des objets comptables tel que des chemises, des morceaux et des objets massifs comme le sucre et l'eau ;
- les substantifs événementiels classés en trois classes (état, action et événement) ;
- les substantifs qui font référence à des individus tels que la femme, les députés.

soutenons nullement l'incapacité de la *SdT* à représenter la temporalité des structures itératives liées à un substantif événementiel. Nous indiquons toutefois que l'intégration de ce type de structures itératives au sein de la *SdT* n'est pas triviale, et pose de nombreuses questions, comme nous l'exposerons dans les perspectives de ce manuscrit.

Les substantifs événementiels

Un substantif dénote un événement si il est compatible avec un verbe aspectuel (Gross et Kiefer, 1995). Il existe deux types de substantifs événementiels : des substantifs qui sont issus des verbes (déverbaux) et des substantifs événementiels simples (non dérivés de verbe)³. Nous avons des formes telles que *tremblement/trembler*, *déraillement/dérailler*, *occupation/occuper* pour le premier type, et des formes comme *orage*, *tempête*, *accident*, *cérémonie*, *match* pour le second. Un des verbes possédant un spectre large est *avoir lieu*.

- (a) Un tremblement de terre a eu lieu hier soir.
(b) Un accident aura lieu sur cette route demain.

EXEMPLE 40

Il est également possible de déterminer la valeur durative ou ponctuelle des substantifs à l'aide des verbes aspectuels. Nous avons notamment les *commencer*, *s'interrompre*, *continuer*, *durer*, *terminer* qui indiquent un événement duratif tandis que des verbes comme *se passer*, *survenir*, *se produire* indiquent un événement ponctuel (cf. exemples 41a et 41b). Nous noterons que les événements duratifs peuvent être compatibles avec la seconde série de verbes aspectuels.

- (a) La cérémonie a commencé à 10 heures du matin.
(b) Le coup de feu s'est produit au matin.

EXEMPLE 41 – Exemples extrait de Kiefer (1998, p.56).

3. Notons que les substantifs dérivés des verbes ne possèdent pas systématiquement une structure événementielle. Des substantifs comme *définition*, *invitation* sont les formes nominalisées des verbes *définir* et *inviter* respectivement mais ils ne sont pas des substantifs événementiels (Kiefer, 1998, p.57)

Comme nous venons de le voir au sein de ce chapitre, le modèle MOPI permet une modélisation objet de la construction du sens liée aux phénomènes itératifs. Au sein de MOPI, une itération est représentée par un itérateur qui itère un modèle itératif, lui-même composé de procès modèle. Dans l'hypothèse où l'itération est portée par un substantif événementiel (dans une phrase averbale ou non), elle sera représentée, à l'instar des verbes, par un procès modèle.

- (a) Trois attentats ont eu lieu à Bagdad.
 (b) Après une série de meurtres dans plusieurs états, le FBI a arrêté un suspect.

EXEMPLE 42

Dans les exemples 42a et 42b, nous retrouvons deux itérations fondées sur des substantifs événementiels. Nous avons respectivement un itérateur de quantification, marqué par le numéral *trois*, qui itère le procès : *un attentat avoir lieu à Bagdad.*, et un itérateur fréquentiel, marqué par le substantif *série*, qui itère le substantif *meurtre*. Les figures 29a et 29b représentent respectivement la modélisation au sein de MOPI des exemples 42a et 42b.

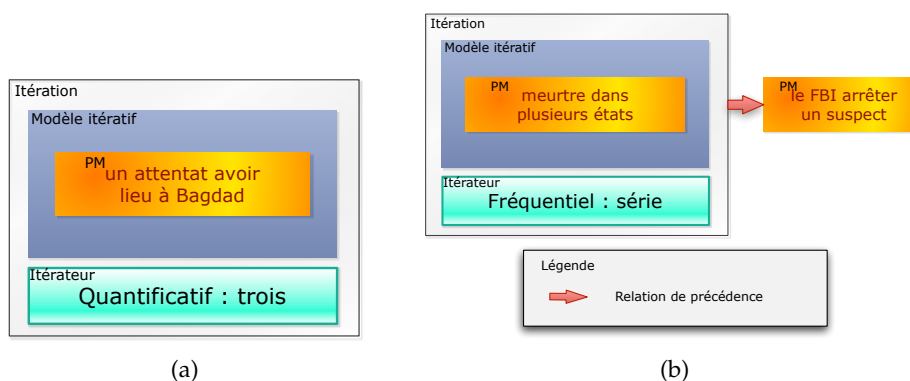


FIGURE 29 – Représentation des exemples 42a et 42b.

Notons que les itérateurs diffèrent, lexicalement, de ceux prévus initialement, bien que la classification reste valide. Les nouvelles formes vont s'intégrer aux différentes classes existantes. Sans avoir la volonté de l'exhaustivité, nous pouvons citer :

- les fréquentsiels : les déterminants indéfinis pluriels (*des*), les déterminants nominaux (*une série de, une suite de, une pluie de*) ;

- les quantificationnels : les adjectifs numéraux (*3, vingt-cinq*), les déterminants pluriels (*plusieurs voyages, en plusieurs occasions*) ;
- les calendaires : la pluralité des noms calendaires tels que *les fêtes de Noël* ;
- les événementiels : les locutions adverbiales de type (*à chaque compétition, à tous les coups*).

EN RÉSUMÉ

Le modèle MOPi est une approche de haut niveau schématisant le processus d'interprétation des phénomènes itératifs amenant à la représentation cognitive de ces phénomènes. Notons que le modèle sert à représenter l'interprétation d'un énoncé itératif qu'il faut au préalable analyser. Le modèle possède toutefois certaines limites, puisqu'il ne rend pas compte notamment de certains itérateurs comme les verbes intrinsèquement itératifs tel que *sautiller* ou *hachurer* dont l'action peut être vue comme atomique.

Dans la suite de notre étude, nous inscrivons dans le cadre de ce modèle. Nos travaux consistent pour une part à construire une représentation textuelle des phénomènes itératifs, étape préalable à une analyse de ces phénomènes dans les principes de ce modèle. Nous souhaitons d'ores et déjà apporter une précision sur la terminologie que nous allons employer par la suite. Nous dénommons par la suite les itérateurs et les sélecteurs par le terme de « déclencheurs » (de phénomènes itératifs) lorsque nous voulons y faire référence, indépendamment du type de phénomènes initiés. Nous appelons également les itérations et les sélections sous le terme de « structures itératives ».



Troisième partie

LES PHÉNOMÈNES ITÉRATIFS EN CORPUS

PRÉAMBULE

La partie que nous allons présenter maintenant correspond au cœur de ce travail de thèse. L'objectif que nous nous sommes fixés au début de ce mémoire est d'automatiser la découverte des phénomènes itératifs en corpus. Afin d'atteindre cet objectif, nous avons suivis une méthodologie en trois étapes. Dans un premier temps, nous nous sommes intéressés à la manifestation de l'itération au sein d'un corpus. Le but est de prendre la mesure de la complexité textuelle de ces phénomènes tant dans l'approche du repérage (la source des itérations) que dans l'organisation discursive de ces phénomènes. Dans un second temps, nous avons étudié des modèles de représentation du discours pouvant modéliser les structures itératives d'un énoncé dans toutes leurs diversités textuelles, afin d'établir un modèle d'annotation, notamment pour la phase automatique. Enfin, dans un troisième temps, nous exposons la mise en œuvre d'un analyseur des structures itératives prenant en compte les résultats des deux premiers aspects.



OBSERVATION DES PHÉNOMÈNES ITÉRATIFS

Le présent chapitre se focalise sur la manifestation textuelle des itérations au sein d'un corpus. L'objectif est de caractériser précisément les phénomènes itératifs en répondant aux questions suivantes : quelle est la portée d'une itération dans le discours ? Quel type d'organisation discursive prennent les phénomènes itératifs ? Nous avons donc réalisé une étude sur corpus annoté que nous présentons ainsi que les résultats obtenus.

SOMMAIRE

6.1	Constitution et description du corpus [ITER]	87
6.1.1	Les articles du journal [LE MONDE]	88
6.1.2	Le sous-corpus [ROMAN]	89
6.2	Annotation manuelle du corpus	90
6.2.1	Méthode d'annotation	90
6.2.2	Outil d'aide à la tâche d'annotation	91
6.3	Premiers constats sur les phénomènes itératifs dans le corpus [ITER] .	92
6.3.1	Les structures itératives	92
6.3.2	Les substantifs événementiels	99
6.3.3	Les déclencheurs	100
6.4	Organisation textuelle des structures itératives	102
6.4.1	La position initiale des déclencheurs	103
6.4.2	La reprise événementielle	104
6.4.3	La distance entre les structures itératives	105
6.4.4	Exemples de configurations textuelles	107

6.1 CONSTITUTION ET DESCRIPTION DU CORPUS [ITER]

Pour décrire finement les phénomènes itératifs, en particulier sur la notion discursive, la linguistique de corpus propose une méthode adaptée : c'est à partir de textes réels et d'annotation que nous avons mené une étude exploratoire et descriptive sur la notion de structure itérative. Nous avons donc constitué, indépendamment de l'analyse linguistique que nous voulions réaliser, un corpus dont le critère de regroupement est conditionné par la taille du développement textuel. Les textes sur lesquels nous travaillons ont été collectés en deux temps. Nous nous sommes intéressé tout d'abord aux romans, au tra-

vers du projet Gutenberg et de l'ABU¹. Ces deux projets mettent à disposition un large choix d'œuvres en version numérique, et dont les droits sont tombés dans le domaine public. Puis dans un second temps, nous nous sommes intéressés, suite à son acquisition par le laboratoire, d'un corpus d'articles du journal LE MONDE. La sélection de ces deux types de collections de documents a été faite pour les deux raisons suivantes :

- les articles ont l'avantage de présenter un développement relativement court permettant une étude des petits phénomènes, circonscrite à un paragraphe ;
- *a contrario*, les romans présentent eux un développement plus long, ce qui nous a permis de prendre en compte des structures de plus gros grains, et les indices relatifs.

Finalement, nous obtenons un corpus dont la taille est conséquente si l'on considère d'un côté que ces données seront annotées manuellement et de l'autre que nous mènerons des analyses discursives et sémantiques complexes sur ces données (*cf.* chapitre 8, page 135). Afin de distinguer par la suite les différents corpus, nous les nommons : le corpus dans son ensemble est nommé « corpus [ITER] » ; il est composé de deux sous-corpus, [ROMAN] et [LMD].

6.1.1 Les articles du journal [LE MONDE]

Le corpus [LE MONDE] couvre une période de 20 ans allant de 1987 à 2006. Le corpus étant dans des formats différents, il fallut réaliser des pré-traitements afin de normaliser le corpus dans une même format. Le choix s'est porté sur XML pour gagner en interopérabilité². L'intégralité du corpus est composée de 904 263 articles, et chaque article est classé par domaine de secteur (art, économie, étranger. . .). Dans le cadre de cette étude nous avons sélectionné un ensemble d'articles correspondant au mois de Janvier de l'année 2006, soit un total de 2 053 articles. Afin de limiter la tâche d'annotation, nous avons sélectionné un sous-ensemble de 244 articles, selon les trois secteurs suivants : *Sport*, *Analyses et débats*, et *Une*, qui est annoté manuellement. Nous présentons la méthode d'annotation que nous avons suivi à la section 6.2.1 page 90.

1. www.gutenberg.org – abu.cnam.fr

2. Le travail de normalisation fut effectué par Jean-Luc Manguin, ingénieur de recherche au laboratoire Greyc.

SECTEUR	[LMD]
Analyses et débats	51
Sport	117
Une	76

TABLEAU 1 – Nombres d’articles par secteur pour le sous-corpus [LMD].

6.1.2 *Le sous-corpus* [ROMAN]

Le corpus [ROMAN] est constitué de plusieurs œuvres littéraires du 19^e. Nous avons annoté les textes suivants³ :

- *La fille aux yeux d’or*, Honoré de Balzac, 1835 ;
- *Le Petit Chose*, Alphonse Daudet, 1868 ;
- *Un cœur simple*, extrait du recueil les « Trois Contes », Gustave Flaubert, 1877 ;
- *Une vie*, Guy de Maupassant, 1883 ;
- *J’accuse*, Émile Zola, 1898.

Taille du corpus [ITER]

Le corpus [ITER] constitue un fonds textuel de 201 397 mots, soit approximativement 14 000 phrases annotées selon leur caractère itératif ou non⁴. Ce qui constitue un corpus de travail d’une taille conséquente pour mettre en évidence les caractéristiques des structures itératives. Le tableau 2 décrit chacun des sous-corpus selon leur taille absolue et relative.

3. Nous avons annoté uniquement le chapitre 1 pour *La fille aux yeux d’or*, et les chapitres 1 et 2 pour *Le Petit Chose*.

4. Le nombre de propositions est approximatif, il a été calculé automatiquement en comptant uniquement les noyaux verbaux du corpus.

	[ROMAN]	[LMD]	[ITER]
Nombre de mots	108 422	92 975	201 397
Pourcentage des mots	53,83 %	46,17 %	100 %
Nombre de propositions	17 587	10 756	28 565
Pourcentage des propositions	67,51 %	38,43 %	100 %
Nombre de phrases	8 239	5 027	13 266
Pourcentage des phrases	62,11 %	37,89 %	100 %

TABLEAU 2 – Taille du corpus [ITER], en termes de phrases, de propositions et de mots.

6.2 ANNOTATION MANUELLE DU CORPUS

L’objectif de cette annotation manuelle est de rendre compte de ce qu’est une structure itérative, du point de vue textuel. En effectuant cette annotation, nous avons une double visée. Premièrement, nous cherchons à caractériser les manifestations textuelles des phénomènes itératifs. Deuxièmement, à partir de cet aperçu des structures itératives en corpus, nous souhaitons établir des indices lexicaux et sémantiques permettant de les repérer et de les délimiter automatiquement au sein d’un énoncé.

6.2.1 Méthode d’annotation

Suite à l’étude de la littérature du domaine (que nous avons présentée aux chapitres 1 page 15 et 3 page 51), et en particulier des travaux de Yann Mathet sur la modélisation des phénomènes itératifs (*cf.* chapitre 5), nous avons mis en place l’annotation des structures itératives. Nous cherchons à repérer les structures itératives et les marqueurs les déclenchant. Nous avons également annoté les expressions temporelles dont nous pensons qu’elles peuvent être des indices forts pour la délimitation des structures itératives. Le protocole d’annotation est donc le suivant :

- **Annoter les expressions temporelles**
- **Annoter les déclencheurs**, itérateurs et sélecteurs, selon la classification suivante issue de MOPI :
 - quantificationnel : à *cinq reprises*, *2 fois* ;
 - événementiel : à *chaque fois que*, *quand* ;
 - fréquentiel : *souvent*, *rarement* ;
 - calendaire : *tous les jours*, *le jeudi*.
- nous ajoutons quatre autres classes, issues de la *SdT* :

- présuppositionnel : *encore, déjà* ;
- verbe itératif : *relancer, criailler* ;
- temps : *présent d'habitude, imparfait itératif* ;
- conflit : provenant de l'incompatibilité d'expressions temporelles tel que *depuis deux mois* et *pendant cinq minutes*.
- **Annoter les structures itératives** :
 - les itérations ;
 - les sélections, et les relier aux structures itératives auxquelles elles se rattachent.

Nous avons effectué ce repérage manuel des zones textuelles sur notre corpus [ITER]. Nous attirons l'attention du lecteur sur le fait que nous sommes le seul annotateur dans cette phase d'annotation manuelle⁵. Rappelons que nous visons aussi bien à déterminer des structures de niveau discursif (supérieur à la phrase) qu'à caractériser des phénomènes très localisés (niveau de la phrase, voire de la proposition). Nous souhaitons au travers de cette annotation mettre en évidence un maximum de structures itératives différentes dans des configurations différentes.

Par ailleurs, en l'état, le corpus annoté obtenu suit l'évolution de nos réflexions sur quels types de structures est pertinent pour représenter la diversité des phénomènes itératifs dont nous présentons les tenants et aboutissants dans le chapitre suivant consacré à la modélisation des structures itératives. Il a donc pour seule finalité d'être le support à nos réflexions sur un modèle d'annotation et un traitement automatique.

6.2.2 Outil d'aide à la tâche d'annotation

Pour réaliser cette tâche, nous avons utilisé un outil développé au sein du Greyc dans le cadre du projet ANR Annodis : GLOZZ⁶. La plate-forme GLOZZ est un environnement d'annotation et d'exploration de corpus fortement configurable permettant d'adapter les annotations à des modèles théoriques, en particulier discursifs. L'objectif premier de la plate-forme est de prendre en charge des configurations structurellement variées, en s'appuyant sur un méta-modèle générique de l'organisation du discours, issu de Widlöcher (2008). Ce méta-modèle repose sur l'articulation entre les notions d'unité (une séquence d'éléments textuels adjacents), de relation (un rapport binaire entre deux unités) et de schéma (une configuration textuelle complexe récurrente impliquant

5. Les résultats que nous présentons dans la suite du chapitre doivent être lus en prenant en compte cet aspect de l'annotation. En effet, il est possible que nous ayons oublié d'annoter certains phénomènes itératifs.

6. www.glozz.org

unités et relations). Ce méta-modèle ayant fait l’objet de notre part d’une adaptation à la représentation textuelle des phénomènes itératifs, nous l’aborderons plus en détails dans le chapitre suivant.

À l’aide d’une interface graphique (cf. figure 30), la plate-forme permet donc de repérer des structures de différents niveaux textuels (syntagme, proposition, paragraphe. . .), de les relier entre eux, et d’y associer également des traits sémantiques. Nous renvoyons le lecteur à l’article (Widlöcher et Mathet, 2009) pour plus de détails sur la plate-forme.

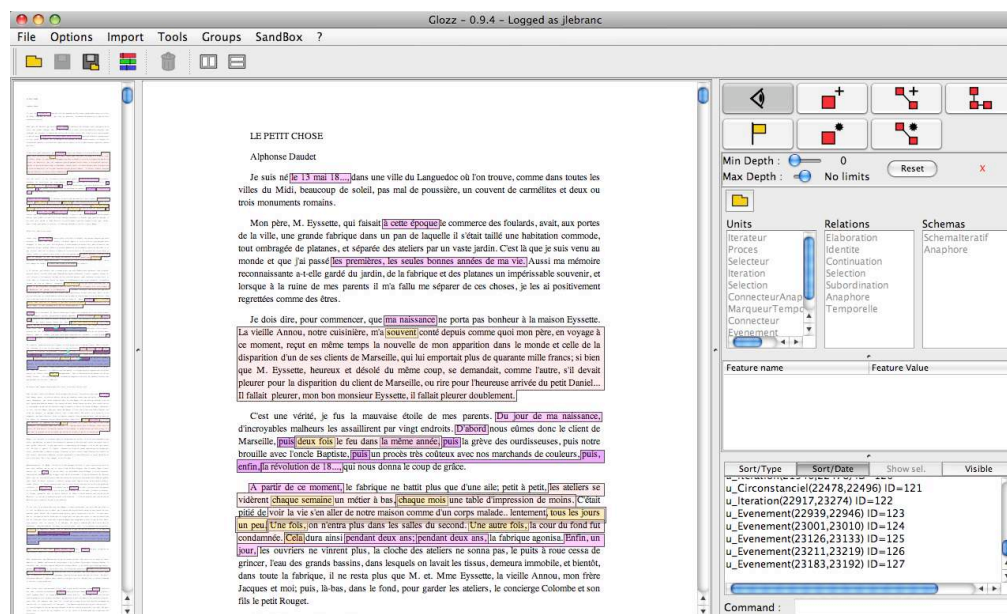


FIGURE 30 – Interface de GLOZZ permettant l’annotation et l’exploration de corpus.

6.3 PREMIERS CONSTATS SUR LES PHÉNOMÈNES ITÉRATIFS DANS LE CORPUS [ITER]

Un des objectifs de l’annotation manuelle du corpus est de permettre de mieux comprendre les phénomènes d’itérations, de les quantifier au sein des textes. Nous rappelons que les structures itératives peuvent être multi-procès.

6.3.1 Les structures itératives

Notre corpus [ITER] contient 650 structures itératives, soit approximativement 1 540 propositions (avec une moyenne de 2,3 propositions par structures itératives). Le tableau 3 récapitule le nombre des structures itératives en fonction des sous-corpus tandis que le tableau 4 présente la quantité de propo-

sitions « itératives » et « sélectives » (c'est-à-dire qui sont présentes dans une itération et une sélection respectivement).

	[ROMAN]	[LMD]	[ITER]
Itérations	328	165	493
Sélections	85	72	157
Total	413	237	650

TABLEAU 3 – Quantité de structures itératives dans le corpus [ITER].

- Pour le corpus [ROMAN], cette phase d'annotation manuelle a mis en évidence la présence de 413 structures itératives (itérations et sélections confondues), soit 1 227 propositions.
- Pour le corpus [LMD], cette phase d'annotation manuelle a mis en évidence la présence de 237 structures itératives, soit 306 propositions.

Ce nombre relativement important de structures nous a permis de mettre en évidence les principales configurations textuelles de ces structures, que nous présentons à la section suivante.

	[ROMAN]	[LMD]	[ITER]
Propositions itératives	862	213	1 075
Propositions sélectives	365	93	458
Total	1 227	306	1 533

TABLEAU 4 – Quantité de structures itératives dans le corpus [ITER].

L'exemple 43 présente une itération (en rouge) suivie de quatre sélections (en bleu) dont trois qui lui sont rattachées, la troisième étant rattachée à la seconde sélection. Nous indiquons le nombre de propositions (itératives et sélectives) au sein de chaque structure itérative. Nous pouvons constater que les structures itératives de notre exemple sont circonscrites aux paragraphes.

	L'après-midi, on s'en allait avec l'âne au-delà des Roches-Noires, du côté d'Hennequeville. Le sentier, d'abord, montait entre des terrains vallonnés comme la pelouse d'un parc, puis arrivait sur un plateau où alternaient des pâturages et des champs en labour. A la lisière du chemin, dans le fouillis des ronces, des houx se dressaient; çà et là, un grand arbre mort faisait sur l'air bleu des zigzags avec ses branches.	5 propositions
1	Presque toujours on se reposait dans un pré, ayant Deauville à gauche, Le Havre à droite et en face la pleine mer. Elle était brillante de soleil, lisse comme un miroir, tellement douce qu'on entendait à peine son murmure; des moineaux cachés pépiaient, et la voûte immense du ciel recouvrait tout cela. Mme Aubain, assise, travaillait à son ouvrage de couture; Virginie près d'elle tressait des joncs; Félicité sarclait des fleurs de lavande; Paul, qui s'ennuyait, voulait partir.	9 propositions
2	D'autres fois, ayant passé la Touques en bateau, ils cherchaient des coquilles. La marée basse laissait à découvert des oursins, des godefiches, des méduses; et les enfants couraient, pour saisir des flocons d'écume que le vent emportait. Les flots endormis, en tombant sur le sable, se déroulaient le long de la grève; elle s'étendait à perte de vue, mais du côté de la terre avait pour limite les dunes la séparant du Marais, large prairie en forme d'hippodrome.	6 propositions
3	Quand ils revenaient par là, Trouville, au fond sur la pente du coteau, à chaque pas grandissait, et avec toutes ses maisons inégales semblait s'épanouir dans un désordre gai.	3 propositions
4	Les jours qu'il faisait trop chaud, ils ne sortaient pas de leur chambre. L'éblouissante clarté du dehors plaquait des barres de lumière entre les lames des jalousies. Aucun bruit dans le village. En bas, sur le trottoir, personne. Ce silence épandu augmentait la tranquillité des choses. Au loin, les marteaux des calfats tamponnaient des carènes, et une brise lourde apportait la senteur du goudron.	6 propositions

EXEMPLE 43 – Exemple extrait d'*Un cœur simple* illustrant la répartition des propositions au sein des structures itératives.

À l'aide du tableau 5, nous pouvons remarquer que les phénomènes itératifs concernent moins de 6 % des propositions de notre corpus. Les phénomènes itératifs sont donc relativement rares au sein de notre corpus, en particulier dans le sous-corpus [LMD] qui couvre à peine 3 % des propositions du corpus.

PROPOSITION	[ROMAN]	[LMD]	[ITER]
Itérations	4,90 %	1,94 %	3,76 %
Sélections	2,08 %	0,85 %	1,60 %
Total	6,98 %	2,79 %	5,36 %

TABLEAU 5 – Proportions des propositions incluses dans des structures itératives.

Ce faible chiffre concernant la couverture des structures itératives dans le sous-corpus [LMD] peut s'expliquer par le nombre d'articles contenant des structures itératives, seulement 80 articles sur les 244 articles initialement sélectionnés. Si nous restreignons le calcul aux seuls 80 articles incluant des structures itératives, nous atteignons un taux proche de celui du sous-corpus [ROMAN], de l'ordre de 6,2 % des propositions du sous-corpus [LMD] (cf. tableau 6).

SECTEUR	ARTICLES	STRUCTURES ITÉRATIVES
Analyses et débats	27 (51)	82 (4,78 %)
Sport	45 (117)	144 (7,61 %)
Une	8 (76)	11 (7,32 %)
Total	80 (244)	237 (6,23 %)

TABLEAU 6 – Nombres d'articles contenant des structures itératives par secteur pour le sous-corpus [LMD], la répartition des articles sélectionnés initialement figurent entre parenthèses.

Les structures itératives : une diversité de taille

L'annotation manuelle a mis en évidence une variation importante dans la taille des structures itératives (en terme de propositions). Le figure 7 récapitule le pourcentage de structures itératives de type *itération* selon le nombre de proposition les constituants. Nous pouvons remarquer que plus de la moitié des structures sont constituées d'une seule proposition. Cette valeur atteint les 76 % dans le sous-corpus [LMD]. Notons que les itérations supérieur à cinq propositions sont présentes uniquement dans le sous-corpus [ROMAN].

TAILLE	[ROMAN]	[LMD]	[ITER]
1	46,95 %	76,36 %	56,80 %
2	23,78 %	19,39 %	22,31 %
3	7,93 %	3,03 %	6,29 %
4	6,71 %	0,00 %	4,46 %
5	5,49 %	1,21 %	4,06 %
6 et plus	8,84 %	0,00 %	5,88 %

TABLEAU 7 – Proportion des itérations dans les sous-corpus [ROMAN] et [LMD] selon le nombre de propositions.

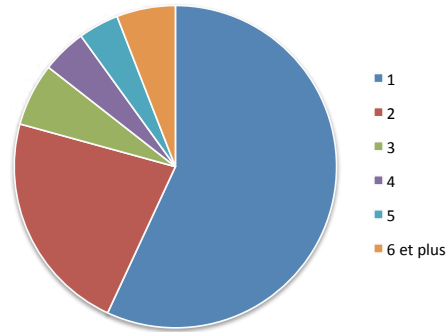


FIGURE 31 – Répartition des itérations dans le corpus [ITER] selon leur taille.

En ce qui concerne les structures itératives de type *sélection*, les résultats montrent une même proportion de structures composées d'une ou de deux propositions, proche de 25 %. Ils existent néanmoins une forte proportion de structures possédant cinq propositions et plus, en particulier au sein du sous-corpus [ROMAN] qui atteint les 25 %.

TAILLE	[ROMAN]	[LMD]	[ITER]
1	29,41 %	25,00 %	27,39 %
2	28,24 %	22,22 %	25,48 %
3	8,24 %	26,39 %	16,56 %
4	7,06 %	12,50 %	9,55 %
5	9,41 %	6,94 %	8,28 %
6 et plus	16,47 %	6,94 %	12,10 %

TABLEAU 8 – Proportion des sélections dans les sous-corpus [ROMAN] et [LMD] selon le nombre de propositions.

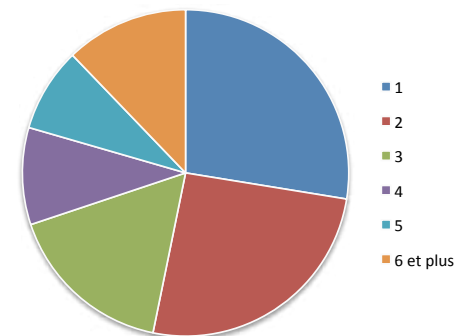


FIGURE 32 – Répartition des sélections dans le corpus [ITER] selon leur taille.

À l'issue de l'observation du corpus, et dans la perspective d'une analyse automatique, nous pouvons distinguer deux caractéristiques des structures itératives selon leur taille. Premièrement, les structures itératives composées d'une seule proposition sont généralement inscrites dans une phrase (*cf.* exemple 44).

Mme Aubain descendit le fossé, poussa Virginie, Paul ensuite, tomba plusieurs fois en tâchant de gravir le talus, et à force de courage y parvint.

EXEMPLE 44 – Extrait d'*Un cœur simple*.

A contrario, les structures itératives ayant une taille supérieure à un peuvent évidemment porter sur plusieurs phrases (cf. l'exemple 45), voire plusieurs paragraphes (cf. exemple 43 page 94).

Tous les jeudis, des habitués venaient faire une partie de boston. Félicité préparait d'avance les cartes et les chaufferettes. Ils arrivaient à huit heures bien juste, et se retiraient avant le coup de onze.

EXEMPLE 45 – Extrait d'*Un cœur simple*.

La seconde caractéristique concerne la manifestation textuelle des structures itératives ayant une taille supérieure à un. L'annotation manuelle a mis en évidence une propriété de discontinuité textuelle des propositions. De manière générale, les structures itératives de grande taille sont constituées d'une suite de propositions consécutives appartenant à la même structure itérative comme l'illustre l'exemple 45 de la présente page. Cependant dans certains exemples, comme l'exemple 46, les propositions d'une même structure itérative peuvent être textuellement discontinues, c'est-à-dire non-consécutives.

« Quand le temps était clair, on s'en allait de bonne heure à la ferme de Geffosses. La cour est en pente, la maison dans le milieu ; et la mer, au loin, apparaît comme une tache grise. Félicité retirait de son cabas des tranches de viande froide [...] »

EXEMPLE 46 – Extrait d'*Un cœur simple*.

Dans le court extrait de l'exemple 46, les deux propositions présentes dans la phrase : « *La cour est en pente, la maison dans le milieu ; et la mer, au loin, apparaît comme une tache grise.* » n'intègrent pas l'itération initié par le déclencheur événementiel, marqué par la subordonnée temporelle : « *Quand le temps était clair,* » . Les événements constituant la suite de l'énoncé composent pour leur part l'itération introduite par la subordonnée temporelle. Nous notons que c'est un phénomène relativement rare, puisque nous avons annoté seulement six

structures itératives présentant cette propriété dans le sous-corpus [ROMAN], et uniquement des itérations. Cela représente approximativement 2 % des itérations de ce sous-corpus.

Les structures itératives : le phénomène d'imbrication

Par ailleurs, nous avons constaté que certaines structures itératives pouvaient être imbriquées textuellement dans d'autres, au total cela concerne 21 structures itératives, comme l'illustre l'exemple 47.

« Chaque jour, après le déjeuner, Julien, tout en fumant sa pipe et se gargarisant avec du cognac [...] faisait plusieurs parties de bésigue avec sa femme. [...] »

EXEMPLE 47 – Exemple extrait d'*Un Cœur Simple* illustrant le phénomène de récursivité.

Toutefois, l'analyse des structures itératives présentant ce phénomène peut amener à des résultats différents. Dans certaines structures itératives, nous retrouvons le concept de récursivité présent dans MOPI et que nous avons exposé au chapitre précédent. Pour rappel, conceptuellement, la récursivité consiste en l'itération d'une itération. Notre exemple précédent (cf. exemple 47) présente ce premier cas. Nous avons l'itération d'une journée type, marquée par *Chaque jour*, contenant une seconde itération : *faire une partie de bésigue*.

Ce phénomène d'imbrication textuelle ne relève cependant pas toujours de l'itération d'une autre itération, mais de ce que nous pouvons appeler d'un ancrage temporel d'une itération sur une autre itération. Prenons l'exemple 48 où est présente une première itération, marquée par *chaque mardi*. Nous pouvons voir que les différentes propositions dont le verbe est souligné intègrent cette itération. Ce qui est également vrai pour la seconde itération, marqué par *quelquefois*. Cependant l'analyse de cette structure itérative complexe montre que les occurrences de la seconde itération peuvent ne pas avoir lieu systématiquement chaque mardi (car certains mardi il peut ne pas y avoir de soleil).

Nous noterons que cette distinction d'analyse dépasse quelque peu le cadre de chapitre consacré à la manifestation textuelle des structures itératives, et qui relève plus d'une nouvelle configuration à prendre en compte au sein de MOPI.

En conclusion, au sein de notre corpus, bien que les structures itératives soient principalement composées d'une ou deux propositions, approximati-

[Mme Aubain] fit un arrangement avec un loueur de voitures, qui la menait au couvent chaque mardi. Il y a dans le jardin une terrasse d'où l'on découvre la Seine. Virginie s'y promenait à son bras, sur les feuilles de pampre tombées. Quelquefois le soleil traversant les nuages la forçait à cligner ses paupières, pendant qu'elle regardait les voiles au loin et tout l'horizon, depuis le château de Tancarville jusqu'aux phares du Havre. Ensuite on se reposait sous la tonnelle.

EXEMPLE 48 – Extrait d'*Un cœur simple*.

vement 75 %, nous constatons une proportion importantes de structures itératives possédant plus de cinq propositions, qui concerne plus de 12 % des structures, avec une plus forte proportion au sein du sous-corpus [ROMAN]. Nous avons également constaté que dans notre corpus toutes les structures itératives sont contraintes à un paragraphe.

6.3.2 Les substantifs événementiels

Notre annotation a mis en évidence également, dans une proportion que nous avons sous-estimé, des structures itératives non plus fondées sur un verbe comme dans l'exemple 49a mais sur un substantif événementiel comme dans l'exemple 49b.

- (a) « Félicité tous les jours s'y rendait. »
 (b) « L'Instance équité et réconciliation (IER) a organisé des auditions publiques, [...] »

EXEMPLE 49

Pour rappel, un substantif dénote un événement si il est compatible avec un verbe aspectuel tel que *avoir lieu*⁷. Il est possible de distinguer ceux issus de verbes tels que *tremblement/trembler* ou *occupation/occuper*, et les déverbaux (non issus de verbes) tels que *match*, *débat*.

7. Nous rappelons que nous avons déjà abordé le cas des substantifs événementiels dans le chapitre précédent (cf. chapitre 5 page 79), en présentant leur intégration au modèle MORI.

	[ROMAN]	[LMD]	[ITER]
Itérations	27 (08,23 %)	68 (41,21 %)	95 (19,27 %)
Sélections	3 (03,53 %)	13 (18,06 %)	16 (10,19 %)
Total	30 (11,76 %)	81 (59,27 %)	111 (29,46 %)

TABLEAU 9 – Quantité et proportion des structures itératives fondées sur un substantif événementiel.

Le tableau 9 récapitule le nombre et la proportion des structures itératives fondées sur un substantif événementiel. Nous constatons qu'un tiers des structures itératives est concerné. Ce chiffre atteint même les 60 % dans le sous-corpus [LMD]. Il est donc nécessaire de prendre en compte ces structures itératives dans l'analyse automatique, en particulier en complétant la classification des itérateurs afin d'intégrer de nouvelles formes lexicales telles que *chaque*, *plusieurs*, *des match(s)*.

6.3.3 Les déclencheurs

L'annotation manuelle des déclencheurs a mis en évidence la répartition des déclencheurs par type, mais également de nouvelles formes lexicales. Le tableau 10 récapitule cette répartition des déclencheurs par type, et pour chacun des sous-corpus. Notons que nous répertorions également les déclencheurs amenant à la construction de structures itératives fondées sur les substantifs événementiels tels que nous les avons vu précédemment.

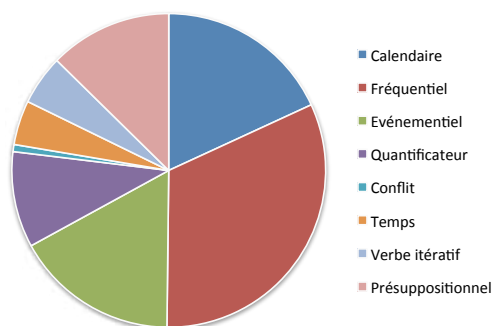


FIGURE 33 – Répartition des déclencheurs par type dans le corpus [ITER].

Nous pouvons remarquer que dans notre corpus les déclencheurs sont principalement de type *Fréquentiel*, à contrario, les déclencheurs de type *Verbe itératif*, *Temps* et *Conflit* sont peu utilisés. Nous remarquons également que plus de 95 % des structures itératives sont déclenchées par des formes pouvant être intégrées au sein d'un lexique. En effet, seules les structures déclenchées par la classe des *Temps*, c'est-à-dire l'imparfait itératif ou bien le présent d'habitude, nécessitent une interprétation en contexte par le lecteur pour déterminer l'itérativité de

l'énoncé.

TYPE DE DÉCLENCHEUR	[ROMAN]	[LMD]	[ITER]
Fréquentiel	38,03 %	19,88 %	32,47 %
Calendaire	22,87 %	9,04 %	18,63 %
Événementiel	18,62 %	13,86 %	17,16 %
Présuppositionnel	5,85 %	28,92 %	12,92 %
Quantificateur	5,59 %	20,48 %	10,15 %
Temps	5,32 %	3,01 %	4,61 %
Verbe itératif	3,46 %	3,01 %	3,32 %
Conflit	0,27 %	1,81 %	0,74 %

TABLEAU 10 – Proportion des déclencheurs par type.

Notre annotation a permis de mettre également en évidence de nouveaux types de sélecteurs, et dans des proportions importantes (cf. tableau 11), là où le modèle MOPI n'en proposait que deux : fréquentiel et quantificationnel. Nous retrouvons en réalité les mêmes types que pour les itérateurs, c'est-à-dire les calendaires, les événementiels, les présuppositionnels, et également les verbes itératifs, comme l'illustre l'exemple 50. De plus, nous avons remarqué également que certaines sélections ne possèdent aucun déclencheur mais sont déclenchées par une organisation discursive particulière. Nous abordons ces cas particuliers dans la section suivante sur l'organisation textuelle des structures itératives.

TYPE DE SÉLECTEUR	[ROMAN]	[LMD]	[ITER]
Fréquentiel	35,44 %	31,25 %	34,23 %
Événementiel	35,44 %	18,75 %	30,63 %
Calendaire	20,25 %	3,13 %	15,32 %
Présuppositionnel	5,06 %	40,63 %	15,32 %
Quantificateur	2,53 %	3,13 %	2,70 %
Verbe itératif	1,27 %	3,13 %	1,80 %

TABLEAU 11 – Proportion des déclencheurs de type « sélecteur ».

L'exemple présente une itération déclenchée par un itérateur calendaires (*toutes les cinq minutes*), puis nous avons une sélection initiée par sélecteur de type événementiel *chaque arrêt*.

[...] toutes les cinq minutes elle s'arrêtait, disant à la pauvre bonne patiente qui la soutenait : "Asseyons-nous, ma fille, je suis un peu lasse."
Et à chaque arrêt elle laissait sur un des bancs tantôt le tricot qui lui couvrait la tête, tantôt un châle, [...]

EXEMPLE 50 – Extrait d'*Un cœur simple*.

En complément des déclencheurs spécifiques aux structures itératives fondées sur les substantifs événementiels, l'annotation a permis également de mettre en évidence de nouvelles formes venant compléter notre classification actuelle. À titre d'exemple, nous avons répertorié les formes suivantes : *ni le premier ni le dernier* pour les fréquentiels, à *nouveau* pour les présuppositionnels, ou encore *notamment* et *comme* pour les événementiels.

En résumé

L'annotation manuelle du corpus [ITER] a permis de mettre en évidence les trois éléments majeurs suivants, qui sont à prendre en compte dans une analyse automatique.

- Une diversité de taille des structures itératives, avec une majorité de structure composée d'une seule proposition, ainsi que de la propriété de discontinuité textuelles des propositions.
- Une confirmation chiffrée de la part importante du rôle des substantifs événementiels dans la construction de structures itératives, qui jusqu'à maintenant n'était pas pris en compte dans les modèles que nous avons exposés dans la partie précédente.
- Une classification élargie des déclencheurs avec quatre classes supplémentaires par rapport à celle établie dans MOPI. En outre, nous avons constaté que plus de 95 % des déclencheurs peuvent être repérables à l'aide d'un lexique ou d'une grammaire adaptée.

6.4 ORGANISATION TEXTUELLE DES STRUCTURES ITÉRATIVES

Les éléments descriptifs présentés dans cette section font écho, à la fois aux observations du corpus annoté, et au modèle théorique sur lequel repose notre étude. L'observation approfondie des structures itératives nous a permis de distinguer les principales configurations textuelles que nous détaillons ci-après.

6.4.1 La position initiale des déclencheurs

Une des caractéristiques que l'on retrouve fréquemment au sein des structures itératives est la position du déclencheur en initiale de la proposition, voire du paragraphe (cf. exemple 51).

Du soir au matin, ce fut une colère formidable qui, ne sachant à qui s'en prendre, s'attaquait à tout, au soleil, au mistral, à Jacques, à la vieille Annou, à la Révolution, oh! surtout à la Révolution!...

EXEMPLE 51 – Extrait de *Le Petit Chose*.

Cette propriété est un des fondements sur lesquels reposent les travaux d'encadrement du discours de Charolles (1997). L'hypothèse de l'encadrement du discours définit un cadre de discours comme un regroupement de plusieurs propositions sous un critère sémantique véhiculé par une expression introductrice de cadre, situé à l'initiale d'une proposition, généralement en position détachée. Les unités regroupant ces propositions sont désignées par le terme de *cadre*. Les principaux travaux ont porté sur les cadres temporels tels que dans l'exemple 52.

Levé tous les jours à cinq heures, il a franchi comme un oiseau l'espace qui sépare son domicile de la rue Montmartre. [...] À neuf heures, il est au sein de son ménage, débite un calembour à sa femme, lui dérobe un gros baiser, déguste une tasse de café ou gronde ses enfants. À dix heures moins un quart, il apparaît à la mairie. Là, posé sur un fauteuil, comme un perroquet sur son bâton, chauffé par la ville de Paris, il inscrit jusqu'à quatre heures, sans leur donner une larme ou un sourire, les décès et les naissances de tout un arrondissement.[...]

EXEMPLE 52 – Extrait de *La fille aux yeux d'or* (1834-35).

Ici, l'exemple présente une itération décrivant la journée du personnage. Chaque circonstanciel temporel introduit un cadre où ces activités se déroulent. Par exemple, à neuf heures, le personnage est chez lui et réalise un certain nombre d'activités. Nous présentons plus largement ces travaux dans le chapitre suivant que nous consacrons à la représentation des structures textuelles afin d'élaborer un modèle d'annotation complet de ces structures en nous basant sur les résultats de l'annotation manuelle.

Nous avons comptabilisé dans le sous-corpus [LMD] une centaine de structures itératives ayant un déclencheur en position initiale, au sujet desquelles nous pouvons constater deux points. Premièrement, seules certaines classes de déclencheurs sont concernées. Le tableau 12 présente ces différentes classes et leur répartition. On constate que les déclencheurs de types « événementiel » et « fréquentiel » sont les plus fréquents.

TYPE	[ITÉRATEUR]	[SÉLECTEUR]	[TOTAL]
Événementiel	30	22	52
Fréquentiel	24	17	43
Calendaire	23	9	32
Quantificateur	3	1	4

TABEAU 12 – Répartition des déclencheurs en position initiale selon le type et le corpus.

Deuxièmement, plus de la moitié des structures itératives ayant une taille supérieure à deux sont introduites par un déclencheur en position initiale. De plus, au sein de notre corpus, les structures itératives de type « itération » constituées d’une seule proposition, et possédant un déclencheur en position initiale, sont systématiquement suivies d’une sélection, comme dans l’exemple 53. De ce dernier constat, nous pouvons conclure que les déclencheurs peuvent jouer un rôle d’introducteur de cadre.

Quelquefois je l’apercevais debout, sur la porte de la loge, du côté des ateliers ; il se tenait là tristement ; et lorsqu’il voyait que je le regardais, le malheureux poussait pour m’attendrir les plus effroyables rugissements, en agitant sa crinière flamboyante ; mais plus il rugissait, plus je me tenais loin.

EXEMPLE 53 – Extrait de *Le Petit Chose*.

6.4.2 La reprise événementielle

Par ailleurs, le rôle de la position du déclencheur intervient également dans d’autres configurations textuelles. Nous avons notamment répertorié au sein du sous-corpus [LMD], une dizaine de structures itératives dont les propositions se situent antérieurement à celle contenant le déclencheur, comme

l'illustre l'exemple 54 où la reprise événementielle est marquée par le pronom *Ce*.

Les Russes Tatiana Totmianina et Maxim Maririn ont remporté, mercredi 18 janvier, à Lyon, la médaille d'or de l'épreuve de couple du championnat d'Europe de patinage artistique. C'est leur cinquième titre d'affilée.

EXEMPLE 54 – Article extrait de la rubrique *Sport* du journal LE MONDE.

La reprise événementielle est généralement marquée par les pronoms tels que *ce*, *y* mais l'est également par certains syntagmes comme dans l'exemple 78 avec *cet événement*⁸.

[...] Le taureau avait acculé Félicité contre une claire-voie ; sa bave lui rejaillissait à la figure, une seconde de plus il l'éventrait. Elle eut le temps de se couler entre deux barreaux, et la grosse bête, toute surprise, s'arrêta.

Cet événement, pendant bien des années, fut un sujet de conversation à Pont-l'Évêque. Félicité n'en tira aucun orgueil, ne se doutant même pas qu'elle eût rien fait d'héroïque.

EXEMPLE 55 – Extrait d'*Un Cœur Simple*.

6.4.3 La distance entre les structures itératives

La dernière propriété dont nous souhaitons discuter aborde la notion de distance entre les structures itératives. Nous avons constaté au sein de notre corpus, que les structures itératives sont généralement circonscrites au paragraphe, comme l'illustre la exemple 56, où chaque encadré correspond à une structure itérative mais également à un paragraphe.

Nous avons également constaté que les structures itératives sont généralement textuellement contiguës à celle dont elles sont rattachées. Dans l'exemple 56 (reprise de l'exemple 43 page 94), nous avons les sélections (en bleu) 1, 2 et 4 qui sont contiguës et toutes les trois sélectionnent l'itération de l'exemple (en rouge). Nous avons annoté un contre-exemple où une structure de type « sélection » était distante de plusieurs paragraphes de la structure à

8. Notons que la reprise événementielle dans cet exemple n'est pas le moteur de l'itération.

laquelle elle se rattache. Cette propriété n'est pas a priori pas généralisable à l'ensemble des structures itératives, mais elle a potentiellement un intérêt dans la phase d'analyse automatique pour la circonscription des structures itératives.

L'après-midi, on s'en allait avec l'âne au-delà des Roches-Noires, du côté d'Hennequeville. Le sentier, d'abord, montait entre des terrains vallonnés comme la pelouse d'un parc, puis arrivait sur un plateau où alternaient des pâturages et des champs en labour. A la lisière du chemin, dans le fouillis des ronces, des houx se dressaient; çà et là, un grand arbre mort faisait sur l'air bleu des zigzags avec ses branches.

1 Presque toujours on se reposait dans un pré, ayant Deauville à gauche, Le Havre à droite et en face la pleine mer. Elle était brillante de soleil, lisse comme un miroir, tellement douce qu'on entendait à peine son murmure; des moineaux cachés pépiaient, et la voûte immense du ciel recouvrait tout cela. Mme Aubain, assise, travaillait à son ouvrage de couture; Virginie près d'elle tressait des joncs; Félicité sarclait des fleurs de lavande; Paul, qui s'ennuyait, voulait partir.

2 D'autres fois, ayant passé la Touques en bateau, ils cherchaient des coquilles. La marée basse laissait à découvert des oursins, des godéfiches, des méduses; et les enfants couraient, pour saisir des flocons d'écume que le vent emportait. Les flots endormis, en tombant sur le sable, se déroulaient le long de la grève; elle s'étendait à perte de vue, mais du côté de la terre avait pour limite les dunes la séparant du Marais, large prairie en forme d'hippodrome.

3 Quand ils revenaient par là, Trouville, au fond sur la pente du coteau, à chaque pas grandissait, et avec toutes ses maisons inégales semblait s'épanouir dans un désordre gai.

4 Les jours qu'il faisait trop chaud, ils ne sortaient pas de leur chambre. L'éblouissante clarté du dehors plaquait des barres de lumière entre les lames des jalousies. Aucun bruit dans le village. En bas, sur le trottoir, personne. Ce silence épandu augmentait la tranquillité des choses. Au loin, les marteaux des calfats tamponnaient des carènes, et une brise lourde apportait la senteur du goudron.

EXEMPLE 56 – Reprise de l'exemple 43 illustrant la contiguïté de structures itératives.

6.4.4 Exemples de configurations textuelles

Pour conclure cette section sur l'organisation textuelles des structures itératives, nous terminons en rendant compte de leur diversité textuelle rencontrées en corpus à l'aide d'exemples.

L'énumération

Nous avons constaté que certaines structures itératives étaient présentées sous formes d'énumérations. Certaines énumérations suivent un format type, et sont constituées de la ponctuation « : » puis suivie des propositions, comme l'exemple 58 où nous une structure itérative de type « itération » initiée par le substantif événementiel *ses premières surprises*, puis les deux propositions en italique qui représentent chacune une occurrence de l'itération.

Le tournoi a connu ses premières surprises lundi : *l'Américaine Venus Williams (tête de série no 10) a été battue par la Bulgare Tszvetana Pironkova (2-6, 6-0, 9-7), la Russe Elena Dementieva (no 9) a été éliminée par l'Allemande Julia Schruuff (7-5, 6-2).*

EXEMPLE 57 – Article extrait de la rubrique *Sport* du journal LE MONDE.

Il existe également des énumérations où chaque proposition est enchaînée à l'aide de connecteurs temporels tels que *et, puis, d'abord, ensuite* (cf. exemple 58) ou bien intégrant des circonstanciels temporels introduisant des cadres temporels (cf. exemple 59).

Du jour de ma naissance, d'incroyables malheurs les assaillirent par vingt endroits. D'abord nous eûmes donc le client de Marseille, puis deux fois le feu dans la même année, puis la grève des ourdisseuses, puis notre brouille avec l'oncle Baptiste, puis un procès très coûteux avec nos marchands de couleurs, puis, enfin, la révolution de 18..., qui nous donna le coup de grâce.

EXEMPLE 58 – Extrait de *Le Petit Chose*.

Sa carrière semble placée sous le sceau des maladies et des blessures. En 2004, après avoir dominé la saison 2003, où elle avait atteint la première place du classement mondial, elle avait été touchée par un cytomégalovirus. Sans force, exsangue, elle avait dû lutter pendant un an pour refaire surface. Elle avait alors réalisé, début 2005, une extraordinaire saison sur terre battue, avec, à la clé, un succès à Roland-Garros. Puis une blessure à la jambe droite avait de nouveau entravé sa progression.

EXEMPLE 59 – Article extrait de la rubrique *Sport* du journal LE MONDE.

Certaines énumérations sont présentées sous formes d'une série de phrases (cf. exemple 60. Dans le cas de cet exemple, nous avons également une extension à plusieurs paragraphes, combinée au premier type d'énumération.

Les écueils se sont succédé. *Le non au référendum sur la Constitution européenne a été un coup de boutoir pour une diplomatie plaçant traditionnellement la construction européenne au coeur du rayonnement français dans le monde. L'année s'est terminée sur l'image désastreuse de la crise des banlieues, marquée par une incapacité des responsables français à convaincre, à l'étranger, qu'il ne s'agissait pas d'une faillite du modèle national d'intégration.*

On peut énumérer, pêle-mêle, d'autres moments délicats survenus l'an passé : *l'isolement français en Europe sur le dossier de l'OMC, l'échec de la candidature de Paris aux Jeux olympiques, les hésitations sur l'ouverture de négociations avec la Turquie pour l'adhésion à l'Union européenne, les âpres disputes autour du budget européen, l'échec du sommet de Barcelone sur le « dialogue euroméditerranéen ».*

EXEMPLE 60 – Article extrait de la rubrique *Analyses et Débats* du journal LE MONDE.

Ces exemples ne recouvrent pas l'ensemble des configurations possibles que peuvent prendre les énumérations, ce qui dépasseraient le cadre de notre étude. Par ailleurs, il reste une question en suspens qui concerne le statut des propositions dans le cadre de ces exemples. En effet, selon nous, la structure énumérative leur confère un nouveau statut. Jusqu'à maintenant, les propositions que nous rencontrions été considérées comme des procès s'intégrant naturellement dans la structure itérative en cours, indépendamment des différentes propriétés existantes (discontinuité, récursivité, ...). Nous constatons que l'énumération de ces procès, en rapport avec une structure itérative, leur

donne un statut de sélection des occurrences de la structure itérative, venant enrichir les itérés un à un. De ce fait, nous envisageons chaque proposition (ou ensemble de proposition, selon les énoncés) d'une énumération comme une structure itérative de type « sélection » venant préciser une occurrence. Nous reviendrons sur ce point dans le chapitre suivant au moment de l'élaboration du modèle d'annotation des structures itératives (cf. section 7.3.4.2 page 130).

Notons que certains exemples, où l'itération n'est pas présente, amènent à la construction de la structure itérative de type « itération » comme l'exemple 61 dans lequel les deux procès « sélection » (en italique) sont rattachés. En effet, l'itération n'est construite qu'à la lecture du deuxième procès. Dans notre exemple, nous avons la construction d'une itération constituée d'un procès modèle *un jeune être mort*, dont les deux procès *être trouver mort* et *être tuer* viennent préciser les occurrences de l'itération.

Les membres du collectif anti-Jeux olympiques se sont rendus, le 13 janvier, au siège d'Amaury Sport Organisation (ASO), société organisatrice de l'épreuve, pour lui remettre le « Prix de l'indignité humaine », symbolisé par un pneu maculé de sang. *Quelques heures auparavant, un jeune Guinéen avait trouvé la mort, renversé par un concurrent du Dakar. Le lendemain, un garçonnet était tué par un camion d'assistance, au Sénégal cette fois.*

EXEMPLE 61 – Article extrait de la rubrique *Analyses et Débats* du journal LE MONDE.

Un phénomène d'« alternative »

Certains exemples de structures itératives présentent également un phénomène que nous pouvons qualifier d'« alternative » entre une série d'événements, qui intègre toutefois la même structure itérative. Prenons l'exemple 52, reporté ci-dessous, où nous retrouvons la structure itérative marquée par le calendrier *tous les jours* et qui intègre les propositions des phrases suivantes. Dans la seconde phrase, nous avons une série de verbes, mis en opposition, notamment par la conjonction de coordinations *ou*. Pour mieux saisir ce phénomène, nous pouvons ré-interpréter la phrase comme certaines fois *il débite un calembour*, d'autres fois *il gronde ses enfant*, ou bien encore certaines fois il réalise les quatre événements.

Levé tous les jours à cinq heures, il a franchi comme un oiseau l'espace qui sépare son domicile de la rue Montmartre. [...] À neuf heures, il est au sein de son ménage, débite un calembour à sa femme, lui dérobe un gros baiser, déguste une tasse de café ou gronde ses enfants. [...]

Extrait de *La fille aux yeux d'or*.

L'illustration

Nous avons également annoté des structures itératives présentant des structures itératives sous formes d'illustrations à l'aide de marqueurs spécifiques tels que *notamment* ou *comme*. L'exemple 62 présente une itération déclenchée par le marqueur *à de multiples reprises*, et une sélection qui vient illustrer une des occurrences à l'aide de l'adverbe *notamment*.

Dominique de Villepin le suit comme il l'a fait à de multiples reprises, notamment en matière fiscale en septembre 2005.

EXEMPLE 62 – Article extrait de la rubrique *Analyses et Débats* du journal LE MONDE.

Les itérations présupposées

Enfin, nous avons également rencontré des itérations présupposées comme dans l'exemple 63 où le déclencheur est de type « Présuppositionnel » (ici *de nouveau*), induisant donc l'existence d'autres occurrences. Occurrence qui est spécifiée dans la phrase suivante par un second marqueur présuppositionnel (ici *déjà*).

Dans un entretien diffusé, dimanche 9 janvier, sur la chaîne CBS, le vainqueur de la coupe du monde 2004-2005 et vedette attendue des prochains Jeux olympiques de Turin (10-26 février 2006) s'est de nouveau illustré en dehors des pistes en déclarant qu'il avait disputé des courses en état d'ébriété. Fin 2005, il s'était déjà prononcé en faveur de la légalisation des produits dopants.

EXEMPLE 63 – Article extrait de la rubrique *Sport* du journal LE MONDE.

À la lumière des résultats de notre annotation manuelle des structures itératives, et dans la perspective d'un traitement automatique, nous pouvons faire les deux constats suivants :

- Les structures itératives sont principalement de petite taille, à 75 % constituées d'une seule proposition.
- Les principaux déclencheurs sont de types « fréquentiel » et « événementiel », pour lesquels la constitution de grammaires ou de ressources reste triviale.

Néanmoins, nous avons également mis en évidence une diversité importante dans les manifestations textuelles : discontinuité textuelle, imbrication des structures, présentation des phénomènes itératifs sous formes de structures énumératives, . . . complexifiant la tâche pour les 25 % de structures itératives de taille supérieure à deux. Comme nous le verrons au chapitre 8 consacré à la mise en œuvre d'une analyse automatique, cette grande diversité nécessite une délimitation précise de l'analyse, ainsi qu'un ensemble d'indices permettant cette analyse.

Par ailleurs, la diversité dans la manifestation textuelle des structures itérative amène également la question de la représentation de ces structures. *Quelle modèle va permettre de rendre compte le plus précisément possible de la complexité de ces phénomènes ?* à la fois dans la perspective d'une analyse automatique mais également d'une analyse manuelle plus large. Nous apportons une réponse dans le chapitre suivant en discutant de deux modèles différents, et nous présentons un modèle d'annotation possible.



REPRÉSENTATION DES STRUCTURES ITÉRATIVES

Dans le chapitre 5, nous avons vu que le modèle MOPi propose une conceptualisation sous forme d'objet des phénomènes itératifs, en se fondant sur un modèle cognitif – et que le modèle a pour finalité de formaliser la sémantique d'une itération, et non sa représentation textuelle. Ce chapitre se place dans ce dernier aspect, et a pour but d'exposer la représentation textuelle des structures itératives que nous adoptons. À cette fin, nous introduirons les deux plans selon lesquels cette représentation fut établie, c'est-à-dire une approche basée sur l'organisation du discours d'une part; et sur un méta-modèle de l'autre.

SOMMAIRE

7.1	Organisation du discours	113
7.1.1	Cadre du discours	113
7.1.2	Cadre et structure itérative	116
7.2	Unité - Relation - Schéma	120
7.2.1	Les unités discursives	121
7.2.2	Les relations discursives	122
7.2.3	Les schémas discursifs	123
7.3	Représentation textuelle des structures itératives	125
7.3.1	Les procès	125
7.3.2	Les compléments circonstanciels de temps	126
7.3.3	Les déclencheurs des structures itératives	127
7.3.4	Les structures itératives	128
7.4	Exemples d'annotation	132

7.1 ORGANISATION DU DISCOURS

7.1.1 *Cadre du discours*

Dans (Charolles, 1997), Charolles décrit un mode particulier d'organisation du discours fondé sur les notions d'univers et de cadre de discours. Il montre comment une expression détachée dite *introduceur* peut initier un segment textuel appelé *cadre*. Un tel cadre sera constitué d'un ensemble de propositions dont l'interprétation est contrainte par les critères donnés par l'expression qui

l'introduit, s'inscrivant ainsi dans la famille de la sémantique vériconditionnelle¹.

Charolles se fonde sur les univers de discours dans son analyse. Un univers de discours correspond à un ensemble de circonstances attachées à un événement, fait ou procès. Ces circonstances sont généralement introduites par « des groupes syntaxiques périphériques adjoints à la phrase ». Ces syntagmes sont qualifiés d'introducteurs. Ils ont pour rôle d'installer des univers, et de régir la mise en relation des univers entre eux. Nous pouvons faire un parallèle avec les itérateurs déclenchant une itération qui introduisent ce que nous appellerions un cadre itératif et qui correspondrait à l'itération.

Une propriété essentielle de ces introducteurs dans le modèle de Charolles, présente dans les phénomènes itératifs, est qu'ils peuvent porter sur plusieurs propositions, et non pas seulement sur celle à laquelle ils sont rattachés. L'ensemble des propositions soumises à la portée d'un introducteur forme ainsi un cadre de discours.

Le rôle tenu par les introducteurs de discours sont liés à deux familles d'opérations que nous pouvons distinguer parmi celles décrites par Charolles. Une première famille intervient dans l'organisation des univers de discours :

- l'installation proprement dite des univers de discours ;
- l'intégration sous un univers ouvert d'une ou plusieurs propositions.

Une seconde famille d'opérations concerne la mise en relation des univers entre eux :

- la projection d'univers dits « parents » ;
- l'unification d'un nouvel univers avec un univers dit « virtuel » ;
- la subordination des univers les uns sous les autres.

Nous allons pouvoir observer dans l'exemple suivant que les différentes opérations sont liées à des phénomènes discursifs relatifs aux constructions formées de plusieurs cadres :

Le premier introducteur de cet extrait est « en général » qui va installer un univers dit générique, contrairement aux autres introducteurs comme « chez nous », « au Japon » faisant référence à des zones géographiques qui vont ouvrir des univers spatiaux. L'introducteur « en général » ouvre lui aussi un univers spatial mais *a posteriori*. En effet, nous ne le déterminons qu'à la lecture des introducteurs suivants. Charolles a donc adopté une approche incrément-

1. La sémantique vériconditionnelle est l'étude du sens en termes de condition de vérité. C'est essentiellement à Frege, puis à Tarski, qu'on doit d'avoir rapproché signification et vérité. Selon eux, comprendre une phrase, c'est connaître les conditions dans lesquelles elle est vraie. Autrement dit, déterminer le sens d'une phrase, c'est être à même de préciser ses conditions de vérité. La question n'est donc pas de savoir si une phrase donnée est vraie ou fausse, mais de préciser quelles conditions doivent être remplies pour que la phrase soit vraie. ([Sémanticlopédie](#))

En général, les gens se serrent la main droite quand ils se rencontrent ou se séparent, ou bien ils s'embrassent. Hello, bonjour, namaste ! **Chez nous**, un baiser est surtout une preuve d'amour et de tendresse à l'égard de quelqu'un de cher, mais **chez certains peuples**, c'est un salut courant. **En Inde**, les gens se saluent mains jointes sur la poitrine, comme s'ils priaient. **Au Japon**, les gens s'inclinent à plusieurs reprises, face à face, en joignant les mains. **En France**, les hommes faisaient le baisemain aux femmes mariées en signe de respect, et les jeunes filles la révérence, mais cette coutume se perd de plus en plus.

EXEMPLE 64 – Exemple extrait de (Charolles, 1997).

tale qui prévoit des possibilités de ré-analyse pouvant conduire à la mise à jour de l'interprétation préalablement construite.

Nous allons, à partir de cet introducteur, décrire les opérations du modèle. Nous allons commencer par l'opération de projection d'espaces associés, qui selon Charolles, accompagne l'installation des univers.

L'opération d'unification intervient seulement une fois qu'un univers est installé. Elle consiste à reconnaître l'univers parent auquel l'univers est rattaché. Dans le cas de notre exemple, cela revient à reconnaître l'univers introduit par « chez nous » comme une instance de l'univers parent spécifique projeté par « en général » et permet *a posteriori* de reconnaître que ce dernier est un univers spécifique.

L'opération de subordination intervient quand un univers de discours inclut d'autres univers plus spécifiques. C'est le cas avec l'univers introduit par « chez nous », qui au lieu de provoquer la fermeture de l'univers introduit par « en général » constitue un univers de niveau inférieur car élabore les propos de l'univers précédent.

Finalement, l'ensemble de ces cadres spatiaux spécifiques sont subordonnés au premier univers (générique), pour produire une structure reproduite sur la figure 34 page suivante.

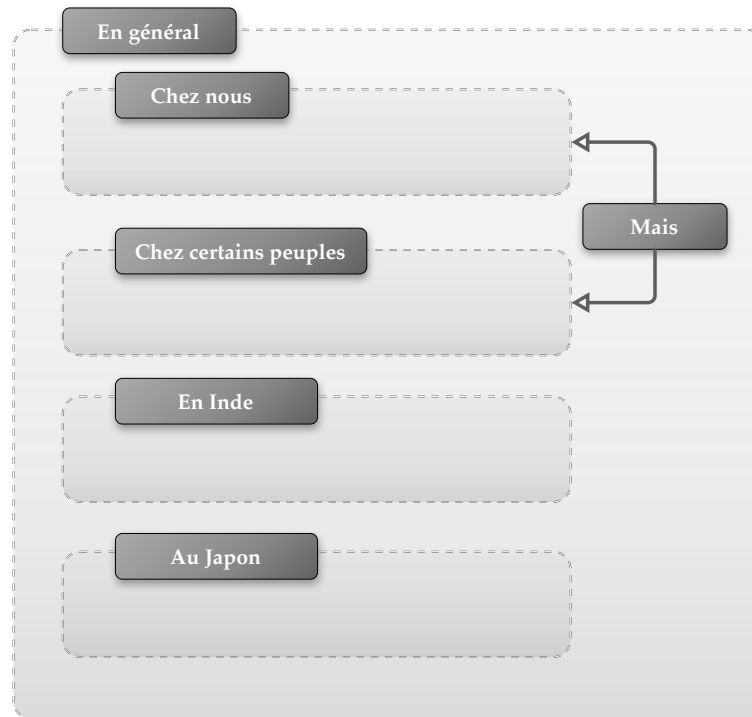


FIGURE 34 – Analyse partielle des cadres de discours de l'extrait précédent.

Pour conclure, il nous semble important de noter, bien que nous ne le détaillons pas, que le terme de « cadre de discours » n'est pas constitué uniquement de constructions sous formes d'univers de discours. Charolles définit les quatre types de cadres suivants : les univers de discours, les champs thématiques (initiés par des adverbiaux du type au « au sujet de X », « concernant X »), les domaines qualitatifs (« Pour faire plaisir à ma mère », « Pour embêter mon père, ») et les espaces de discours (initiés selon des aspects métalinguistiques de l'énonciation « en somme », « bref » ou de l'organisation du discours (« d'une part », « d'autre part »)).

7.1.2 Cadre et structure itérative

La partie qui suit traite de la validité des cadres de discours appliqués aux phénomènes itératifs, et surtout dans quelle mesure il nous faudra l'adapter. À cette fin, nous avons étudié différents cas d'applications du modèle que nous illustrons par l'analyse de l'exemple 65 page ci-contre.

Levé tous les jours *à cinq heures*, il a franchi comme un oiseau l'espace qui sépare son domicile de la rue Montmartre. [...] *À neuf heures*, il est au sein de son ménage, débite un calembour à sa femme, lui dérobe un gros baiser, déguste une tasse de café ou gronde ses enfants. *À dix heures moins un quart*, il apparaît à la mairie. Là, posé sur un fauteuil, comme un perroquet sur son bâton, chauffé par la ville de Paris, il inscrit *jusqu'à quatre heures*, sans leur donner une larme ou un sourire, les décès et les naissances de tout un arrondissement.[...] *Libéré à quatre heures* de son service officiel, il apparaît pour répandre la joie et la gaieté au sein de la boutique la plus célèbre qui soit en la Cité. Aussi, dès qu'il arrive, agace-t-il les demoiselles de comptoir,[...] ou, plus souvent encore *avant de dîner*, il sert une pratique, copie une page du journal ou porte chez l'huissier quelque effet en retard. *À six heures*, tous les deux jours, il est fidèle à son poste. Inamovible basse-taille des chœurs, il se trouve à l'Opéra, prêt à y devenir soldat, Arabe, prisonnier, sauvage, paysan, ombre, patte de chameau, lion, [...] *À minuit*, il redevient bon mari, homme, tendre père, il se glisse dans le lit conjugal, [...] Enfin, s'il dort, il dort vite, et dépêche son sommeil comme il a dépêché sa vie. N'est-ce pas le mouvement fait homme, l'espace incarné, le protégé de la civilisation? Cet homme résume tout [...]

EXEMPLE 65 – Extrait de *La fille aux yeux d'or* (1834-35).

L'itérateur « tous les jours » installe un univers U_C que nous qualifions d'univers itératif avec un déclencheur de type calendaire qui fait référence aux jours de la semaine. Il est immédiatement suivi du marqueur temporel « à cinq heures » qui va ouvrir un univers U_{H_1} que nous qualifions d'univers horaire (en comparaison avec l'univers spatial) dans lequel vont s'intégrer les propositions qui suivent. Cet univers U_{H_1} est subordonné à l'univers parent U_C car il l'élabore.

Le marqueur « à neuf heures » ouvre un nouvel univers U_{H_2} et ferme le précédent car les deux marques sont du même type et ne peuvent pas être valides en même temps.

Les marqueurs « à dix heures moins un quart » et « Libéré à quatre heures » ouvrent respectivement U_{H_3} et U_{H_4} et ferment respectivement U_{H_2} et U_{H_3} . Dans U_{H_4} , un nouveau marqueur est présent « avant de dîner » qui élabore le propos de U_{H_4} . Nous avons donc création d'un univers $U_{H_{S1}}$ subordonné à U_{H_4} . Cet univers contient un itérateur « plus souvent encore » qui va créer un « cadre itératif » comprenant les diverses propositions. U_{H_4} est fermé par

la marque « A six heures » qui ouvre un univers U_{H_5} . Cet univers contient un « cadre itératif » vérifié « tous les deux jours ».

Enfin, le dernier marqueur temporel « à minuit » ferme l'univers U_{H_5} et ouvre un nouvel univers U_{H_6} . Cet univers est fermé par un changement de sujet entre l'univers et le reste du passage (le sujet passe de « il » à « cet homme »). Ce changement de sujet a pour conséquence de fermer aussi l'univers parent U_C . La figure 35 de la présente page représente une modélisation des univers créés en appliquant et adaptant le modèle de Charolles.

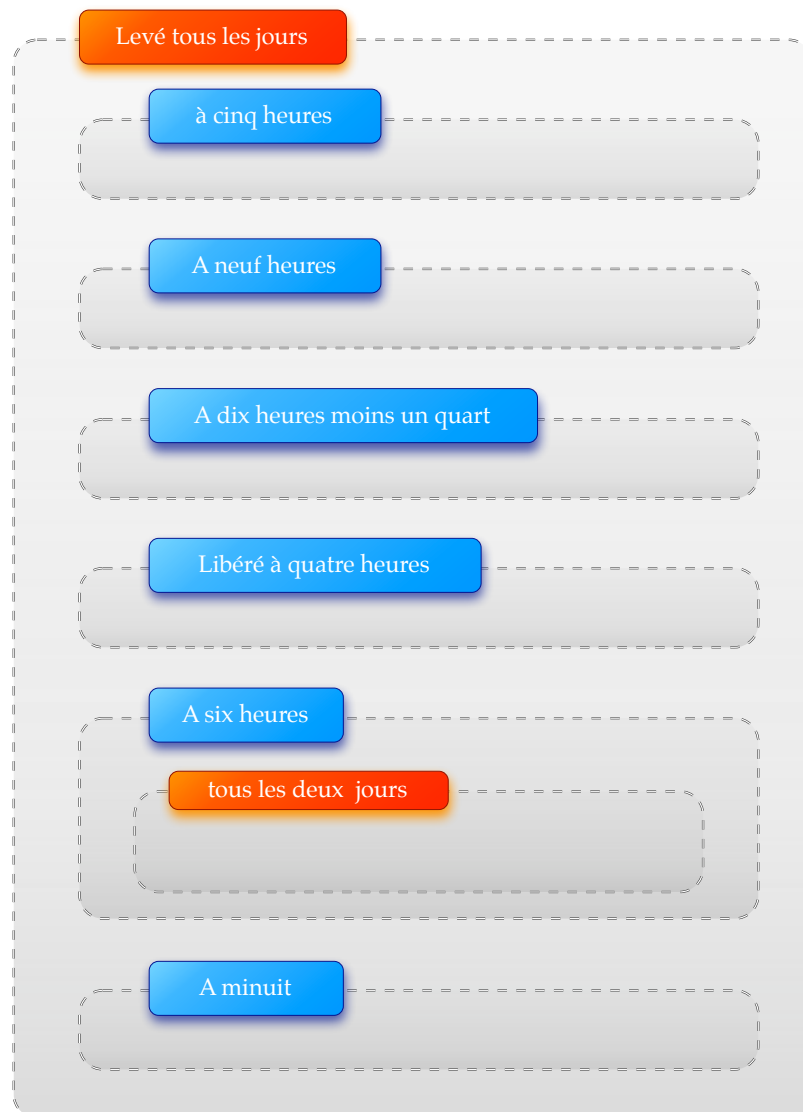


FIGURE 35 – Représentation des univers de discours sur un texte itératif.

Un cadre itératif serait donc composé d'un introducteur de cadre qui serait un itérateur selon la définition donnée au chapitre 5, et de n propositions

(événements) vérifiant la sémantique de l'introducteur. Ici, chaque événement devra donc vérifier/valider la périodicité du cadre, c'est-à-dire avoir lieu « tous les jours ».

Conclusion de l'analyse

Suite à l'approfondissement des différents cas d'applications étudiés, et au regard des différentes propriétés mises en évidence à la section 6.4, nous avons abouti à deux interrogations qui mettent en doute la faisabilité de l'application immédiate de ce modèle à notre objet d'étude.

1. Quelle représentation pour les configurations présentant des discontinuités ?
2. Quelle validité pour les cas de subordination comme présenté à la figure 36 ?

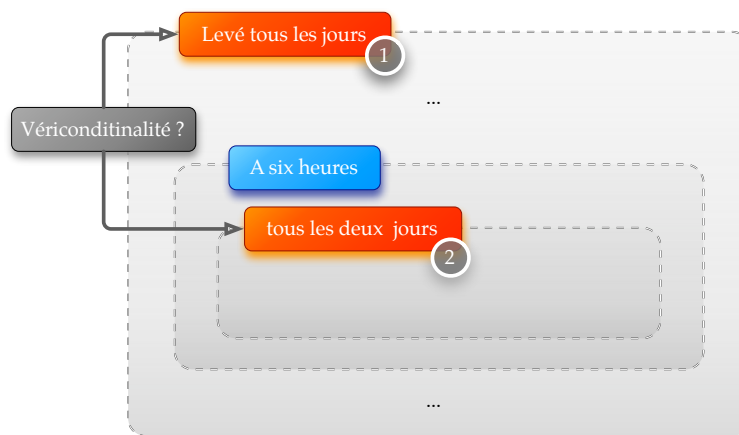


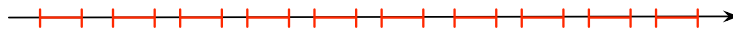
FIGURE 36 – Représentation du conflit de subordination entre deux cadres.

La première interrogation porte donc sur la représentation des configurations discontinues. En effet, il existe de nombreux travaux portant sur les univers de discours, nous pouvons citer sans avoir la volonté d'être exhaustif : les cadres spatiaux (Charolles *et al.*, 2005; Sarda, 2005) « En Normandie, ... », les cadres temporels (Charolles et Péry-Woodley, 2005; Le Draoulec et Péry-Woodley, 2003, 2005) « Depuis les années 2000, ... » ou encore les cadres praxéologiques (Vigier, 2005) « En informatique, ... ». Cependant, à notre connaissance, aucun travail ne traite de cas de discontinuités au sein du modèle des univers de discours. Il serait donc difficile de représenter un cas tel que l'exemple 65 sans apporter des amendements au modèle. Potentiellement, ces amendements pourraient s'inscrire dans deux axes :

- permettre d'ajouter des « trous » ;
- modifier la sémantique des univers, en particulier sur le principe de vériconditionnalité.

La seconde interrogation à laquelle nous nous sommes heurté porte sur l'imbrication des structures itératives². La figure 36 modélise ce propos sur l'exemple 65. Nous remarquons que nous avons un cadre introduit par « Levé les jours » auquel est subordonnée un second cadre itératif introduit par « tous les deux jours ». La problématique ainsi posée est que les propositions constituant le cadre 2 ne valident plus le cadre 1. En effet, l'introducteur de cadre 1 réalise une sélection de jours sur la série associée à l'introducteur du cadre 1 (voir figure 37).

Tous les jours



Tous les deux jours



FIGURE 37 – Représentation de la sélection de *tous les deux jours* sur *tous les jours* posant la problématique du respect du principe de vériconditionnalité.

Finalement, nous en arrivons, en l'état de nos connaissances, à la conclusion que l'hypothèse de l'encadrement de discours ne peut pas s'appliquer aux structures itératives, sans apporter des amendements au modèle. Toutefois, comme nous avons pu le voir au chapitre précédent, certains travaux sur l'organisation du discours nous semblent intéressants dans le cadre de la « structuration » des itérations, en particulier la notion de cadres temporels. Nous reviendrons sur ces derniers lors de la présentation des indices permettant une circonscription automatique (*cf.* chapitre 8).

Afin de représenter textuellement les structures itératives, nous nous orientons donc vers des travaux associant une sémantique moins contraignante vis-à-vis du contenu et de sa représentation formelle, en l'occurrence le meta-modèle *U-R-S*.

7.2 UNITÉ - RELATION - SCHÉMA

Les travaux de (Widlöcher, 2008), sur l'analyse des structures rhétoriques du discours, visent à déterminer les éléments structurels fondamentaux sur lesquels s'élabore l'organisation du discours. En se fondant sur différentes struc-

2. Rappelons l'imbrication consiste à subordonner un cadre itératif à un autre cadre itératif.

tures d'études telles que les structures énumératives, thèse/illustration, ou thèse/antithèse/synthèse, il définit un ensemble d'éléments primitifs nécessaires à la description formelle des motifs discursifs de ces différentes configurations. Il distingue trois éléments fondamentaux : les *unités discursives*, les *relations discursives* et les *schémas discursifs* que représente la figure 38 de la présente page.

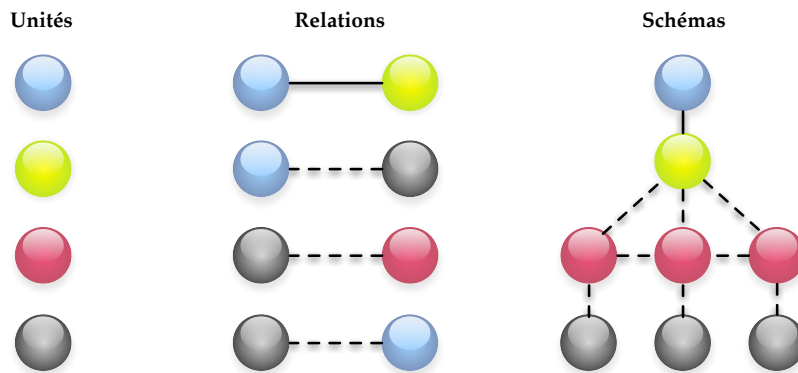


FIGURE 38 – Unités, relations et schémas.

L'exemple 66 va nous permettre de détailler les différents éléments du modèle *U-R-S*. Au travers de cet exemple itératif, nous voulons définir les éléments du modèle en esquissant une modélisation de structures itératives sans pour autant entrer dans les détails de la modélisation.

Chaque lundi matin, le brocanteur [...] étalait par terre ses ferrailles. Puis la ville se remplissait d'un bourdonnement de voix, [...] Vers midi, au plus fort du marché, on voyait paraître sur le seuil un vieux paysan de haute taille, [...]

EXEMPLE 66 – Exemple dans une version diminuée extrait d'*Un Cœur Simple*.

7.2.1 Les unités discursives

Les unités discursives correspondent à des zones textuelles « délimitées ». Elles procèdent au regroupement d'un ensemble d'unités d'ordre inférieur. Ces ensembles sont caractérisés par une certaine homogénéité de sens, d'objet ou de fonction d'un point de vue interprétatif.

Chaque lundi matin, le brocanteur [...] étalait par terre ses
 ferrailles. Puis la ville se remplissait d'un bourdonnement
 de voix, [...]. Vers midi, au plus fort du marché, on voyait
 paraître sur le seuil un vieux paysan de haute taille, [...].

FIGURE 39 – Les unités.

La figure 39 présente quatre segments consécutifs dont le premier représente un itérateur et les trois suivants des procès. Rappelons que la délimitation textuelle des trois procès correspond aux trois propositions de l'énoncé. Intuitivement, nous pouvons remarquer que l'unité « procès » procède au regroupement des unités, non marqué sur la figure, que sont sujet, verbe et syntagme, tandis que l'unité « itérateur » serait une unité vue comme « atomique » du complément circonstanciel *chaque lundi matin*.

7.2.2 Les relations discursives

Les relations discursives correspondent à l'existence d'un rapport entre des unités discursives, relations pouvant être orientées ou au contraire symétriques selon le statut que l'on attribue à ces dernières. La figure 40 de la présente page présente ce rapport existant entre les trois procès, marqué ici par une relation temporelle de *succession*.

Chaque lundi matin, le brocanteur [...] étalait par terre ses
 ferrailles. Puis la ville se remplissait d'un bourdonnement
 de voix, [...]. Vers midi, au plus fort du marché, on voyait
 paraître sur le seuil un vieux paysan de haute taille, [...].

FIGURE 40 – Les relations.

7.2.3 Les schémas discursifs

Les schémas discursifs constituent le troisième type d'éléments de discours. Ils correspondent à des *patterns* discursifs de plus haut niveau, caractérisés par une certaine unité fonctionnelle ou sémantique dépendante d'unités et d'un faisceau de *relations* entre ces unités. La figure 41 illustre cette structure de haut niveau représentant, en rouge, une itération. Ce schéma contient donc les quatre unités, l'« itérateur » et les trois « procès » ainsi que les relations temporelles les liant.

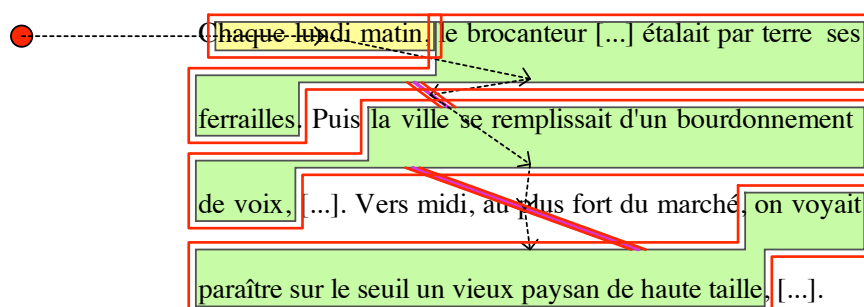


FIGURE 41 – Les schémas.

À l'issue de la description de ce modèle, nous pouvons remarquer qu'il s'applique à des exemples simples de structures itératives. Il nous reste toutefois à apporter deux autres points importants à préciser sur ce modèle triptyque, fondamentaux dans le cas des phénomènes itératifs, en particulier pour les cas de structures complexes, la récursivité et la discontinuité, textuelle, entre les segments discursifs.

La récursivité

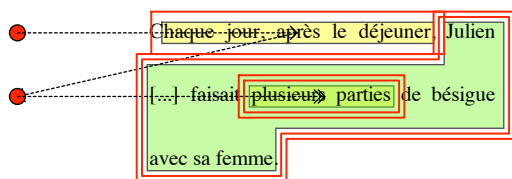


FIGURE 42 – La récursivité.

Chaque élément peut entrer dans la composition d'un élément de même type. Cela est fortement lié à un phénomène de *crystallisation* selon lequel un élément quelconque peut être vu comme une unité. Il permet donc de représenter des cas d'imbrications tel que sur la figure 42 qui illustre un schéma imbriquée dans un autre schéma, exemplifiant l'itération d'une itération présente au sein de l'exemple 47, et reporté ci-dessous.

« Chaque jour, après le déjeuner, Julien [...] faisait plusieurs parties de bésigue avec sa femme. [...] »

EXEMPLE 67 – Report de l'exemple 47 page 98 extrait d'*Un Cœur Simple* illustrant le phénomène de récursivité.

La discontinuité entre les segments discursifs

Une unité et un schéma au sens du modèle *U-R-S* sont des unités logiques vues conceptuellement comme atomiques. Cette atomicité est une vision purement conceptuelle. Du point de vue du texte, un schéma est composé d'unités inférieures pouvant être séquentielles ou non. Ce propos est illustré par la figure 43, où le schéma itératif est composé d'un itérateur et d'un procès sans intégrer le procès représentant la proposition subordonnée de l'énoncé.

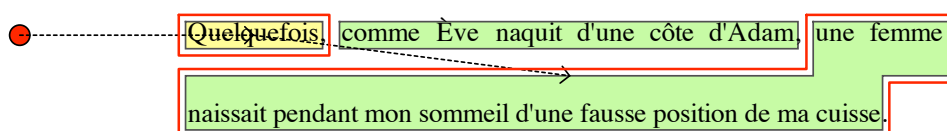


FIGURE 43 – La discontinuité.

Conclusion

Finalement, le modèle *U-R-S* remplit les critères nécessaires pour formaliser et modéliser les structures itératives d'un énoncé. Il permet notamment de représenter les différentes configurations de la topologie grâce à la propriété de récursivité du modèle et de la gestion des configurations discontinues.

7.3 REPRÉSENTATION TEXTUELLE DES STRUCTURES ITÉRATIVES

Nous abordons dans cette section la formalisation des structures itératives autour des trois catégories élémentaires du modèle *U-R-S* que sont unités, relations et schémas³. Nous nous fondons sur les propriétés énoncées au chapitre précédent (c'est-à-dire *la propriété de discontinuité* et *le phénomène d'imbrication*), et les différentes organisations textuelles présentées. L'objectif est d'établir un modèle d'annotation des structures itératives permettant de prendre en compte de leurs différentes organisations textuelles possibles. Nous avons conçu ce modèle d'annotation à la fois pour une annotation manuelle dans la perspective par exemple d'une campagne d'annotation plus large que celle déjà effectuée, mais également pour notre phase d'annotation automatique. Nous illustrons les choix de représentations des structures itératives et des ces constituants (les procès, les compléments circonstanciels et les déclencheurs) à l'aide de l'exemple 68.

Tous les jeudis, des habitués venaient faire une partie de boston. Félicité préparait d'avance les cartes et les chaufferettes. Ils arrivaient à huit heures bien juste, et se retiraient avant le coup de onze.
Certaines fois, ils faisaient une partie de bésigue.

EXEMPLE 68 – Exemple extrait d'*Un Cœur Simple*, dont le second paragraphe est une suite possible.

7.3.1 Les procès

Le « procès » est l'unité discursive principale des structures itératives. On retrouve cette acception dans différents travaux sous la notion d'éventualité Vendler (1967). Textuellement, cette unité correspond à une proposition d'un énoncé englobant le sujet, le verbe et les constituants de la proposition (*cf.* figure 44). Les procès peuvent être également annotés à l'aide d'un schéma dans l'hypothèse où les éléments constituants le procès sont non contigus comme avec la première proposition de notre exemple.

Sémantiquement, le procès est l'unité comportant l'événement de la proposition qui est itérée dans le cas où il est un des constituants d'une itération. Cet événement peut être porté aussi bien par un substantif que par un verbe. Il porte également les instructions fondamentales pour le calcul aspectuel : le

3. Nous ne détaillons pas ici les différents indices amenant à la construction de ces structures qui sont détaillés au sein du chapitre 8 consacrés à leur mise en œuvre.

Tous les jeudis, des habitués venaient faire une partie de boston,

Félicité préparait d'avance les cartes et les chaufferettes, Ils arrivaient

à huit heures bien juste, et se retiraient avant le coup de onze,

Certaines fois, ils faisaient une partie de bésigue.

FIGURE 44 – Représentation du modèle d'annotation des différents procès de l'énoncé.

temps grammatical porté par le verbe et le type de procès (« état », « activité », « accomplissement », « achèvement »).

7.3.2 Les compléments circonstanciels de temps

Nous avons vu, tout au long de ce manuscrit, que les compléments circonstanciels ont un rôle important dans l'analyse des phénomènes itératifs : au sein de la *SdT*, dans le calcul aspectuel et plus précisément son interaction entre les intervalles circonstanciels et les autres intervalles de l'énoncé, et au sein du formalisme algébrique : dans son rôle du calcul des séries. Nous détaillons leur rôle dans le calcul des représentations aspectuo-temporelles dans la partie iv.

La représentation des marqueurs temporels au sein d'un modèle d'annotation des phénomènes itératifs est donc essentielle. Nous distinguons les différents circonstanciels, comme suit, pour prendre en compte les contraintes du calcul aspectuel :

- circonstanciel de durée : *pendant 2 mois, en deux heures, depuis deux jours* ;
- localisation anaphorique : *à ce moment, en même temps* ;
- localisation déictique : *ce matin, maintenant* ;
- localisateur date complète : *le 14 juillet 1789* ;
- localisateur date indéterminé : *[par] un/une + repère calendaire*.

Il est également nécessaire de déterminer si ces marqueurs sont ponctuels ou non.

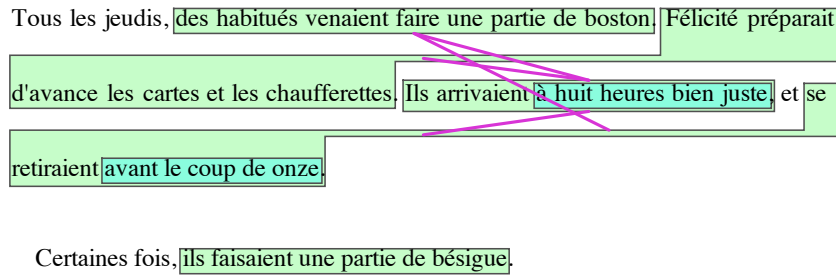


FIGURE 45 – Représentation du modèle d’annotation des compléments circonstanciels de temps, ainsi que les relations entre les procès.

7.3.3 Les déclencheurs des structures itératives

Une des conclusions de notre étude de corpus est que, d’un point de vue textuel, il existe peu de différence entre un itérateur et un sélecteur. Cette distinction apparaît uniquement lorsque l’énoncé est analysé. De plus, le rôle du déclencheur peut être sous-spécifié dans la mesure où il intègre par la suite une structure itérative typée « itération » ou « sélection », et donc le déclencheur se trouve déterminé.

Textuellement, un déclencheur recouvre deux réalités différentes. La première est représentable par une unité discursive de type « déclencheur » de la forme « *tous les jeudis* » (en jaune sur la figure 46), « *souvent* », « *trois fois* » . . .

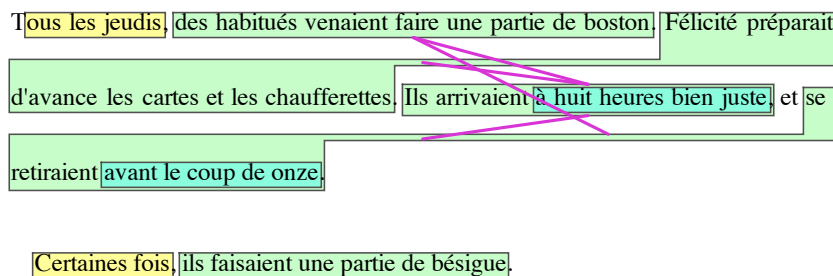


FIGURE 46 – Représentation du modèle d’annotation des différents procès de l’énoncé (en vert) et des déclencheurs (en jaune).

La seconde relève d’un conflit au sein de la proposition. Les exemples 108a et 108b illustrent ces conflits d’ordre pragmatique entre des unités composant le procès et déclenchant ainsi une itération du procès. Nous les représentons à l’aide d’un schéma constitué des unités en conflits (cf. figure 47 page suivante).

- (a) Depuis deux mois, Adrien mange en 10 minutes.
 (b) Longtemps, je me suis couché de bonne heure.

EXEMPLE 69

L'exemple 108a est composé de deux unités de types « marqueur temporel » tandis que l'exemple 108b qui met en conflit l'adverbe « *Longtemps* » et le verbe « *se coucher* » ayant la qualité d'être ponctuel, ou tout du moins d'être temporellement atomique en regard de la durée exprimée par « *Longtemps* ».

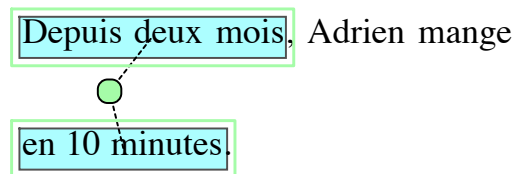


FIGURE 47 – Représentation du modèle d'annotation d'un déclencheur issu d'un conflit.

Comme nous l'avons vu précédemment, nous distinguons les déclencheurs selon la classification suivante :

- Issue de MOPI :
 - quantificationnel : à *cinq reprises*, *2 fois* ;
 - événementiel : à *chaque fois que*, *quand* ;
 - fréquentiel : *souvent*, *rarement* ;
 - calendaire : *tous les jours*, *le jeudi*.
- Issue de la *SdT* :
 - présuppositionnel : *encore*, *déjà* ;
 - verbe itératif : *relancer*, *criailler* ;
 - temps : *présent d'habitude*, *imparfait itératif* ;
 - conflit, que nous venons de présenter ci-dessus.

7.3.4 Les structures itératives

Nous l'avons esquissé précédemment, notamment au travers de l'étude menée sur le corpus annoté présenté à la section 6.4, les itérations et les sélections sont des structures de haut-niveau englobant un certain nombre d'unités inférieures que nous formalisons à l'aide d'un schéma. Nous représentons donc une itération et une sélection respectivement par un schéma dit *itératif* (ou *ScI*) et un schéma dit *sélectif* (ou *ScS*).

7.3.4.1 Les itérations

Comme nous l'avons présenté au chapitre 5, une itération, conceptuellement, est constituée d'un modèle itératif intégrant un ou plusieurs procès modèles, et potentiellement une autre itération, liés les uns aux autres par un ensemble de relations temporelles⁴.

Afin de décrire les différentes configurations possibles, un schéma itératif est nécessairement constitué d'une part d'un déclencheur selon la description que nous avons effectuée précédemment (cf. section 7.3.3); d'autre part, il se compose de procès. Le rôle associé à cette structure est équivalent à celui du modèle itératif dans le modèle MORI.

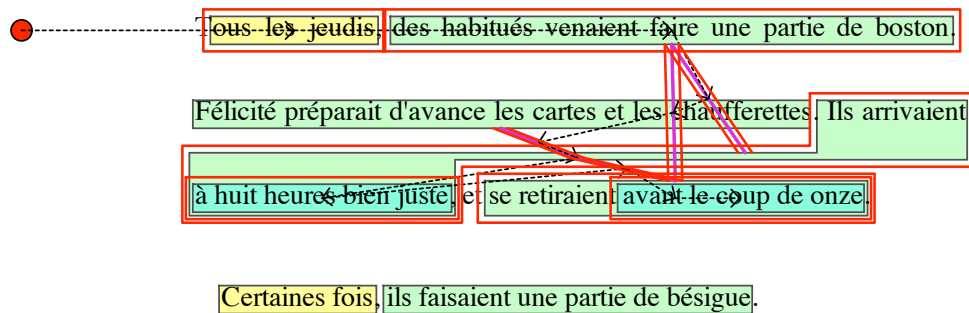


FIGURE 48 – Représentation du modèle d'annotation des structures itératives.

Le schéma itératif peut être constitué soit d'un unique procès, décrivant ainsi une itération circonscrite à une proposition (cf. figure 49a), soit de plusieurs procès décrivant ainsi une itération pouvant s'étendre sur plusieurs paragraphes (cf. figure 49b), tout en tenant compte du fait que certains procès peuvent ne pas appartenir à l'itération (cf. figure 49c). De plus, il est possible que cette structure se compose d'un schéma itératif représentant donc une itération comme dans la figure 49d où le ScI (1) composé notamment de l'itérateur « Le jeudi » est constitué aussi de deux autres schémas itératifs (2 et 3).

4. Les relations temporelles font références au relation d'Allen (1983) et mettent en relations deux unités quelconques.

Mme Aubain descendit le fossé, poussa Virginie, Paul
 ensuite, tomba plusieurs fois en tâchant de gravir le talus,
 et à force de courage y parvint.

(a) Itération minimale

Tous les jeudis, des habitués venaient faire une partie de boston.
 Félicité préparait d'avance les cartes et les chaufferettes. Ils arrivaient
 à huit heures bien juste, et se retiraient avant le coup de onze.

(b) Itération sur plusieurs procès

Quelquefois, comme Ève naquit d'une côte d'Adam,
 une femme naissait pendant mon sommeil d'une fausse
 position de ma cuisse.

(c) Itération discontinue

Le jeudi, je vais à la piscine. Je fais dix séries de 500
 mètres en crawl et puis deux séries de 500 mètres de dos.

(d) Itération imbriquée

FIGURE 49 – Représentation des différentes configurations que peut prendre le modèle d'annotation des schémas itératifs.

7.3.4.2 Les sélections

La sélection est le second type de structure itérative. Nous l'avons vu au chapitre 5, conceptuellement, une sélection est nécessairement rattachée à une itération ou à une autre sélection. Elle n'a pas d'existence en dehors puisqu'elle vient sélectionner un sous-ensemble, voire l'ensemble des itérés d'une itération pour réaliser des amendements à ces différents procès itérés.

Textuellement, nous avons constaté au chapitre précédent que les sélections sont nécessairement constitués de procès, et possiblement d'un déclencheur. Comme l'illustre la figure 50, nous représentons une sélection par un « schéma sélectif » (en bleu), constitué d'une unités discursives du type « procès », d'un déclencheur, et d'un schéma itératif.

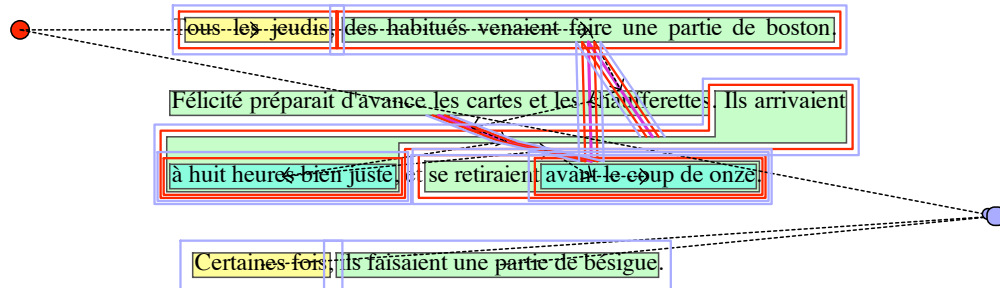


FIGURE 50 – Représentation du modèle d'annotation du schéma sélectif (en bleu) constitué du sélecteur (en rouge) du procès, et d'un schéma itératif.

Cette représentation, à l'instar des itérations, évolue en fonction des exemples tels que dans la figure 51, où les sélections sont initiées sans déclencheurs mais dûes à la structure énumérative des propositions.

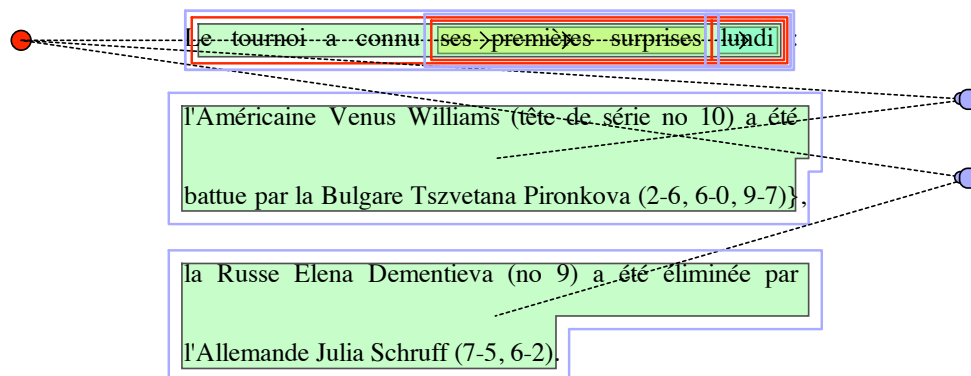


FIGURE 51 – Représentation du modèle d'annotation du schéma sélectif dans le cadre d'une présentation des propositions en structure énumérative.

7.4 EXEMPLES D'ANNOTATION

Nous proposons d'illustrer ce modèle d'annotation sur l'extrait de *La fille aux yeux d'or* traité en début de chapitre (cf. exemple 65 page 117).



FIGURE 52 – Représentation du modèle d'annotation sur l'extrait de *La fille aux yeux d'or* traité en début de chapitre. La figure ne présente pas les procès et le rattachement des sélections à l'itération principale (marqué par *tous les jours*) afin de rendre le schéma plus lisible. Une version complète de l'annotation se trouve en annexe C page 245.

EN RÉSUMÉ

Nous avons présenté le modèle d'annotation que nous avons établi pour la représentation des structures itératives en nous fondant sur le modèle *Unité-Relation-Schéma*. Il nous a permis de positionner textuellement les propriétés principales des structures itératives, et ainsi de rendre compte de leur complexité du point de vue textuel. Nous utilisons également notre modèle d'annotation dans la tâche d'analyse automatique que nous avons mise en œuvre, et que nous présentons au chapitre 8.

Par ailleurs, dans les choix de modélisation que nous avons effectués, nous avons essayé de rester fidèle aux structures conceptuelles de MOPI. En effet, chaque objet du modèle d'annotation possède son équivalent au sein de MOPI, nous permettant ainsi d'analyser directement les structures itératives annotées.



ANALYSEUR DES STRUCTURES ITÉRATIVES (ADSI)

Ce chapitre détaille la mise en œuvre de notre analyseur automatique des structures itératives. L'objectif consiste à produire un prototype permettant d'annoter les structures itératives, en les repérant et les circonscrivant. Les résultats permettront de constituer un corpus pour *in fine* appliquer le calcul des relations aspectuo-temporelles que nous détaillons dans la partie suivante.

Les différents modules de l'analyseur ont été construits de manière à respecter le plus possible la modélisation triptyque présentée à la section 7.3 (page 125). Ils servent également à valider certaines hypothèses élaborées suite à l'étude de corpus sur les indices linguistiques et discursifs potentiellement pertinents pour la délimitation automatique des structures itératives que nous présentons ici.

SOMMAIRE

8.1	Introduction	135
8.2	Traitements de base pour une analyse en TAL	137
8.2.1	Segmentation : du texte aux mots	137
8.2.2	Segmentation en phrases	138
8.3	Repérage des constituants	139
8.3.1	Traitement des procès	139
8.3.2	Traitement du temps	142
8.3.3	Traitement de l'aspect	145
8.3.4	Traitement des déclencheurs	147
8.4	Analyse des structures itératives	150
8.4.1	Méthode d'analyse	150
8.4.2	Du point de vue de l'implémentation	153

8.1 INTRODUCTION

L'étude de corpus a permis de mettre en évidence une grande diversité des configurations textuelles des structures itératives : discontinuité textuelle, imbrication des structures, présentation des phénomènes itératifs sous formes de structures énumératives. Traiter l'ensemble des configurations que nous avons mis à jour est une tâche qui n'est pas triviale, en particulier lorsque les itérations sont présentées sous formes d'énumérations. Nous avons donc cir-

conscrit notre analyseur à l'étude des structures itératives en rendant compte prioritairement des discontinuités textuelles et de l'imbrication des structures itératives.

Nous avons également constaté, dans cette étude, que les structures itératives ayant une taille supérieure à deux sont introduites par un déclencheur en position initiale. Nous partons de ce constat pour la phase de circonscription des structures itératives. Nous avons suivi une méthode d'analyse similaire à celle mise en œuvre dans les travaux de Frédérik Bilhaut (Bilhaut, 2006) sur l'analyse des expressions spatiales et temporelles, dans le cadre des univers de discours de Charolles (1997)¹. Il pose le problème de la circonscription des cadres en terme de portée, c'est-à-dire de la délimitation de la borne finale du cadre initiée par un introducteur à l'aide d'un certain nombre d'indices exploitables de façon automatique qui forment dans leur ensemble un **faisceau** raisonnablement fiable.

Notre analyseur se décompose en deux phases :

- le repérage des différents constituants des structures itératives, à savoir les procès, les circonstanciels de temps et les déclencheurs ;
- et la circonscription des structures itératives a proprement dite, qui se fonde sur les constituants de la première phase.

La plate-forme LINGUASTREAM

Nous avons développé cet analyseur au sein de LINGUASTREAM, plate-forme dédiée au traitement du langage naturel développée au GREYC (CAEN) par Frédérik Bilhaut et Antoine Widlöcher (2006)². LINGUASTREAM est fondée sur la notion de chaîne de traitement. Le principe d'un traitement sous forme de chaîne consiste à décomposer un problème, ici linguistique, en plusieurs sous-tâches bien délimitées. Dans LINGUASTREAM, ces tâches prennent l'apparence de modules de traitement *a priori* distincts (c'est-à-dire qui nécessitent souvent des formalismes différents), connectés entre eux et qui collaborent de manière transparente pour l'utilisateur. LINGUASTREAM permet de construire des représentations sémantiques à chacune des expressions repérées, en prenant en compte la variabilité du grain des indices linguistiques (mot, syntagmes, etc.). Ce processus permet de s'abstraire de la linéarité des

1. Nous avons présenté dans la section 7.1 l'hypothèse de l'encadrement du discours due à Michel Charolles (1997), qui décrit un mode particulier d'organisation du discours en identifiant des segments (dits « cadres de discours ») homogènes par rapport à un critère sémantique spécifié par une expression détachée en initiale de phrase (dite introducteur de cadre ou IC).

2. www.linguastream.org

textes, et de traiter uniquement des représentations sémantiques et abstraites particulières.

La chaîne de traitement construite avec LINGUASTREAM a pour objectif le repérage des indices amenant à la construction d'une structure itérative, et la circonscription textuelle de cette dernière. Un des intérêts majeurs d'une chaîne de traitement est la modularité et la décomposition des objectifs en sous-tâches spécialisées relevant très souvent de niveaux d'analyse variés. Notre analyseur est décomposé de la manière suivante :

- A. le traitement des procès ;
- B. le traitement de l'aspect ;
- C. le traitement du temps ;
- D. le traitement des déclencheurs ;
- E. le traitement des structure itératives.

8.2 TRAITEMENTS DE BASE POUR UNE ANALYSE EN TAL

Préalablement à tout traitement automatiquement des textes, il est indispensable de réaliser un certain nombre de traitements de base. Ainsi, notre analyse nécessite une étape de segmentation des textes en mots, et nous projetons une analyse morpho-syntaxique sur l'ensemble des mots délimités, à l'aide de l'étiqueteur TREETAGGER. Ces traitements sont traditionnels en TAL, nous les présentons ici, principalement pour la facette d'évaluation qualitative que nous avons effectués. Cette évaluation est nécessaire pour évaluer plus précisément nos propres analyseurs dans la perspective de leurs améliorations.

8.2.1 *Segmentation : du texte aux mots*

C'est une tâche relativement bien définie et de nombreux outils existent. Nous utilisons l'outil TREETAGGER pour réaliser cette tâche, couplé à un étiquetage morpho-syntaxique. Il est développé à l'« *Institute for Computational Linguistics of the University of Stuttgart* »³. Il effectue par défaut une segmentation, une lemmatisation et un marquage morpho-syntaxique des unités lexicales. Il associe à chaque mot les traits suivants : un lemme (lemma), une étiquette morphologique (tag) et une sous-étiquette morphologique (stag). Le TREETAGGER est intégré au sein de la plateforme LINGUASTREAM.

Les résultats sont relativement corrects en ce qui concerne l'étiquetage. Nous avons relevé quelques bruits qui sont des cas classiques d'ambiguïtés

3. www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

catégorielles. Dans le premier exemple, le mot « matches » est étiqueté comme un verbe au présent.

*... les **matches** ont été d'un niveau appréciable.*

Ici, nous avons une erreur classique, et quasi-systématique, de la forme de surface « est » qui est considéré comme le verbe « être ».

*Ainsi, d'avril à novembre, a-t-on plus de chances de les observer sur le côté **est** des atolls.*

Une erreur également rencontrée est liée à un mauvais étiquetage, dû parfois à une impossibilité de distinguer deux formes. Dans l'exemple qui suit, la valeur du lemme obtenue pour la forme « suis » est : « suivre | être ».

*Je **suis** restée pétrifiée.*

Certaines erreurs proviennent de l'absence d'étiquetage d'une valeur comme l'exemple qui suit le temps grammatical du verbe « avoir » n'est pas renseigné.

*Si c'était facile, on l'**aurait** déjà fait.*

Dans l'exemple suivant, le verbe « estimer » est considéré comme un nom et n'est pas annoté comme un verbe. Certaines erreurs sont liées aux fautes d'orthographe des auteurs comme pour la préposition « à » qui est annotée comme un verbe dû à l'oubli de l'accent.

*Pierre-Dominique Schmidt, ambassadeur de Belgique **a** Paris, **estime** que le chanteur « **a** la fibre de chez nous ».*

L'impact de ces erreurs va porter principalement sur le traitement de l'aspect et des événements. En effet, si un verbe n'est pas détecté, le calcul de l'aspect ou de type de procès ne pourra pas s'appliquer.

8.2.2 Segmentation en phrases

Nous avons constaté au chapitre 6 que les structures itératives étaient composées de propositions ainsi que certains déclencheurs, en particulier la classe des « Événementiels ». Pour des raisons techniques, nous avons réalisé qu'une segmentation partielle en propositions où nous repérons uniquement les subordonnées (relatives, causales et temporelles). Ce choix est motivé par le rôle de ces propositions dans notre analyse : les subordonnées temporelles peuvent potentiellement déclencher une itération.

Nous avons donc recentré notre segmentation sur les phrases, une délimitation plus précise pouvant se faire dans un second temps via une plate-forme dédiée telle que GLOZZ. La phrase est un niveau de segmentation qui potentiellement limite fortement la portée des structures itératives. Le repérage des

phrases est donc pour nous une étape importante et relativement triviale pour le français. Il nécessite toutefois de prendre en compte les éléments entre parenthèses ou les discours rapportés. Une fois ce pré-traitement appliqué, nous considérons une phrase comme une suite de mots située entre une marque de début de paragraphe ou une balise de fin de phrase et un signe de ponctuation fort (« ? », « ! », « ... ») ou une balise de fin de paragraphe.

8.3 ADSI : LE REPÉRAGE DES CONSTITUANTS

Les différents analyseurs permettant le repérage des constituants d'une structure itérative reposent sur plusieurs lexiques que nous avons constitué spécifiquement pour cette tâche, d'une part ; et font l'objet d'une grammaire locale d'unification en ProLog (composant EDCG de LINGUASTREAM), d'autre part. Ces lexiques n'ont toutefois pas pour objectif d'être complet mais de couvrir, dans un premier temps, les différentes configurations rencontrées en corpus afin d'établir un prototype fonctionnel. Nous présentons les lexiques et les informations sémantiques qu'ils comportent lors de leurs utilisations au sein d'un traitement. Notons également que les différents lexiques constitués reprennent certains des différents marqueurs de fréquence qu'a listé Lim (2002, pages 229-251).

8.3.1 *Traitement des procès*

Notre premier traitement, mis à part la segmentation phrastique, a consisté à repérer les procès. Nous rappelons que conceptuellement les structures itératives sont constitués de procès qui portent les événements itérés. En raison d'une certaine simplification technique liée aux difficultés de la segmentation en propositions, nous avons fait le choix de représenter textuellement les procès par les événements de l'énoncé. Nous avons donc focalisé notre traitement sur la détection des noyaux verbaux et des substantifs événementiels.

8.3.1.1 *Les noyaux verbaux*

Nous nous sommes basés sur l'étiquetage du TREETAGGER pour détecter les formes simples et les formes composées, à l'aide d'une grammaire Prolog. Au final, nous analysons les noyaux verbaux et les principaux temps de l'indicatif : l'imparfait, le présent, le participe-présent, le participe-passé, le passé simple, le futur, le plus-que-parfait, le passe-composé et l'infinitif. Nous récupérons également les formes négatives et les voix passives. Lorsque les noyaux ver-

baux sont introduits par les pronoms personnels, nous récupérons également le genre et le nombre.

Les différentes erreurs issues de cette annotation sont étroitement liées à celle du `TREETAGGER`, que nous avons évoqué à la ??, notamment les erreurs de catégorisation des verbes. Nous avons également certaines formes qui sont détectées comme deux noyaux verbaux à la place d'un seul, par exemple les verbes de modalités. Cela provient de certaines règles implémentées qui ne prennent pas en compte ces cas, par un oubli de notre part.

Passé le mois d'avril, [nous ne pouvons pas] [disputer] la CAN pour la simple raison qu'il pleut en Afrique

8.3.1.2 Les substantifs événementiels

Pour rappel, il existe deux types de substantifs événementiels : des substantifs qui sont issus des verbes dits « déverbaux » (*tremblement/trembler, déraillement/dérailer, occupation/occuper*) et des substantifs événementiels simples, non dérivés de verbe (*orage, tempête, accident, cérémonie, match*). Nous avons rendu compte, dans l'étude de corpus, de leur importance dans le cadre des structures itératives. En effet, nous avons approximativement 30 % des structures itératives qui sont constituées de substantifs événementiels dont 60 % pour le sous-corpus [LMD].

Afin de les annoter, nous avons constitué un lexique dérivé de *VerbAction* créé par (Hathout *et al.*, 2002). Le lexique *VerbAction* contient 9393 couples de nom-verbe, où le nom est dérivé du verbe. *VerbAction* énumère également le genre et le nombre de chaque entrée. Comme ces informations morphologiques ne sont pas pertinentes dans notre processus de détection, nous avons simplement extrait les paires de lemme pour constituer notre lexique. Le verbe correspondant à chaque nom est également mentionné dans le lexique.

```

<couple>
  <verb>
    <lemma>persifler</lemma>
    <tag>Vmn---</tag>
  </verb>
  <noun gender="masculine" number="singular">
    <lemma>persiflage</lemma>
    <tag>Ncms</tag>
  </noun>
</couple>
<couple>
  <verb>
    <lemma>philosophhailler</lemma>
    <tag>Vmn---</tag>
  </verb>
  <noun gender="feminine" number="singular">
    <lemma>philosophhaillerie</lemma>
    <tag>Ncfs</tag>
  </noun>
</couple>

```

EXEMPLE 70 – Deux entrées extraites du lexique *VerbAction*. Nous conservons uniquement les couples de lemmes verbe-nom, c'est-à-dire *persifler* – *persiflage* et *philosophhailler* – *philosophhaillerie*.

Nous avons complété ce lexique des lemmes des substantifs non déverbaux tels que *match*, *épreuve*, *course* ou *campagne* qui sont présent en nombre dans le corpus [LMD], en particulier dans les articles de sport. Il a été créé en semi-automatique dans un processus en deux étapes. Tout d'abord, les substantifs événementiels potentiels ont été trouvés par l'extraction des résultats de requêtes sur l'intégralité du corpus LE MONDE, puis nous les avons validé manuellement.

Nous avons cherché les substantifs en cooccurrence avec un verbe aspectuel, en particulier avec *avoir lieu* qui possède un large spectre, mais également avec les verbes *commencer*, *s'interrompre*, *continuer*, *durer*, *terminer* qui indique un événement duratif ainsi que les verbes comme *se passer*, *survenir*, *se produire* qui indique un événement ponctuel (cf. exemples 71a et 71b)⁴.

4. Pour compléter le lexique, nous avons utilisé de simple expressions régulières telles que *un * a eu lieu*, *lors de la ** et *le * s'est produit* où * est susceptible d'être un nominal de l'événement.

- (a) La cérémonie a commencé à 10 heures du matin.
 (b) Le coup de feu s'est produit au matin.

EXEMPLE 71 – Exemples extrait de Kiefer (1998, p.56).

Notons que du fait de la polysémie des mots, le lexique n'est pas sans contraintes sur les résultats de l'annotation. En effet, la polysémie des mots implique que certaines entrées du lexique peuvent relever d'une structure événementielle ou non, par exemple le substantif *campagne* (cf. exemple 72).

- (a) Le suicide des agriculteurs gangrène **les campagnes**.
 (b) De **sa campagne européenne** 2005, l'Olympique lyonnais a gardé un goût amer...

EXEMPLE 72 – Le substantif *campagne* possède a minima les deux acceptions suivantes :
 – Étendue de pays plat et découvert (cf. 72a) ;
 – Expédition militaire ; ensemble des opérations militaires menées sur un théâtre déterminé (cf. 72b, substantif événementiel).

8.3.2 *Traitement du temps*

L'analyse du temps repose sur les temps morphologiques présentés précédemment, et sur les expressions temporelles : les adverbiaux temporels et les subordonnées temporelles.

8.3.2.1 *Les adverbiaux temporels*

Le traitement des adverbiaux temporels regroupe des unités lexicales très variées : des adverbes (« aujourd'hui », « hier »), des syntagmes nominaux (« les années trente », « tous les ans »), des syntagmes prépositionnels (« depuis plusieurs jours », « depuis longtemps », « en 1960 », « des années trente aux années cinquante »). L'analyseur temporel assigne aux expressions temporelles repérées une annotation sémantique spécifique⁵. Celle-ci est composée d'une structure de traits qui distingue notamment les types d'informations suivantes :

- le type du circonstant : en+durée, pendant+durée, localisateur date incomplète (*en mai*), localisateur date complète (*en 1789*) et depuis+durée ;

⁵ L'analyseur repose sur une grammaire Prolog de Gérard Becher que nous avons repris et adapté.

– la valeur temporelle pour les circonstanciels déictiques (*passé, présent, futur*). Lorsque l’expression temporelle joue le rôle de déclencheur tels que « tous les jours » ou « le jeudi », nous ajoutons un trait sémantique indiquant ce rôle de déclencheur calendaire.

prep :	pendant
typeArg :	intervalle_standard
stypeArg :	pendant+duree
grain :	an
valeur :	5
precision :	exacte

(a) Structure de trait du circonstant *pendant cinq ans*.

typeArg :	iteration
declencheur :	calendaire
det :	tous les
grain :	an
valeur :	4

(b) Structure de trait du déclencheur calendaire *tous les quatre ans*.

FIGURE 53 – Exemple de structures de trait des circonstanciels semelfactif et itératif.

L’analyseur couvre de nombreuses et diverses formes d’expressions temporelles telles que celles référencées dans le tableau 13.

le 15 janvier, le 13/12, janvier, dimanche
les années 80, le 20ème siècle, fin janvier
dimanche soir, la nuit de mardi à mercredi
l’été, le printemps 2002, en 2007, Entre janvier et avril
15h30, treize heures trente
demain, la semaine prochaine, lundi prochain, Noël de l’année prochaine
à la 68e minute, Il y a treize mois, Onze année après, dès les premiers instants

TABLEAU 13

Les expressions temporelles très complexes ne sont pas totalement intégrées dans l’analyseur.

Les clubs mettent les joueurs à la disposition de l’équipe de France [pendant douze semaines] sur quarante-deux lors de [la saison sportive].

Certaines expressions nécessitent également des ressources complémentaires que nous n’avons pas mises en place, notamment les entités nommées événementielles (*le Festival de Deauville, la Coupe du monde, les Jeux Olympiques*). L’expression temporelle n’est alors que partiellement annotée :

On peut arriver à 20 matches [les années] de Coupe du monde.

8.3.2.2 Les subordonnées temporelles

Nous distinguons les subordonnées temporelles directes et indirectes, respectivement introduites par une conjonction ou une locution conjonctive telle que *quand, comme, tandis que, pendant que* ou au moyen d'un substantif désignant une période de temps telle que *le jour où, un jour où, l'année où, au moment où, après le jour où*. Cette distinction a son importance dans le cadre du calcul aspectuel (cf. section B.4 page 238). Chaque marqueur de subordination porte une relation temporelle spécifique que nous présentons à la section B.4, par exemple *pendant que* marque une relation d'accessibilité. Nous avons répertorié ces différentes conjonctions au sein d'un lexique ainsi que les relations temporelles associées (les traits sémantiques sont référencés à la figure 54).

[type :	directe
	itérateur :	événementiel
	relation :	coïncidence

FIGURE 54 – Exemple de structure de trait de la locution conjonctive à *chaque fois que*.

Il est nécessaire de distinguer les différentes conjonctions introduisant les subordonnées. Nous pouvons distinguer trois types différents :

- les conjonctions telles que *alors que* ou *tandis que* qui ne peuvent pas déclencher d'itérations⁶ ;
- les conjonctions qui déclenchent de manière systématique des structures itératives telles que *à chaque fois* ou *les jours où* ;
- les conjonctions introduites par des conjonctions telles que *quand, lorsque* ou *dès que* qui potentiellement déclenchent une structure itérative.

Nous détaillons ces dernières à la section 8.3.4.1 où nous présentons les critères permettant de caractériser quand ils ont un rôle de déclencheur ou non.

L'annotation des subordonnées temporelles reste qualitativement bonne, même dans des phrases possédant plusieurs subordonnées, pas seulement temporelles.

Ils sont toujours mordus par les offres alléchantes [que les agents leur proposent] [alors que très souvent leurs promesses sont de la démagogie].

Toutefois, nous ne gérons pas les incises présentes dans des subordonnées, ce qui peut amener à une clôture prématurée de la subordonnée.

6. Notons que nous parlons ici uniquement du rôle de déclencheur de la conjonction. Il est possible qu'un autre déclencheur au sein de la proposition permette de créer une structure itérative comme dans l'énoncé suivant : *Ils sont toujours mordus par les offres alléchantes que les agents leur proposent alors que très souvent leurs promesses sont de la démagogie* .

On le vit prostré, la tête dans les deux mains, [lorsque l'arbitre de la rencontre], le Tunisien Mourad Daami, invalida un but de son équipe.

8.3.3 Traitement de l'aspect

Le traitement de l'aspect consiste à analyser les différents marqueurs aspectuels qui permettront notamment de réaliser le calcul des représentations aspectuo-temporelles des structures itératives que nous présentons dans la partie suivante. Ces marqueurs sont les suivants :

- les verbes et leurs temps morphologiques (analysé dans le traitement précédent) ;
- les marqueurs aspectuels ;
- les types de procès.

8.3.3.1 Les marqueurs aspectuels

Les adverbes aspectuels tels que *déjà*, *de/à nouveau* sont des adverbes qui amènent à la construction d'une itération, toujours présupposée. Certains adverbes tels que *encore* ou *toujours* sont particulièrement intéressants car la construction de l'itération dépend du contexte. Il est possible de classer ces adverbes selon le type de procès du verbe qu'ils accompagnent. Si ils sont combinés à un prédicat télique, ils déclenchent alors une itération⁷.

- (a) Marie lit encore.
Marie est encore fâchée.
- (b) Marie a lu encore un livre.
Marie a encore déclamé le poème.

EXEMPLE 73 – L'exemple 73a présente l'adverbe *encore* dans un rôle de continuation de procès tandis que l'exemple 73b présente les procès comme la répétition d'un événement présupposé.

Nous avons donc mis en place, à l'aide d'un lexique, une détection des adverbes aspectuels, que nous combinons au type de procès, pour les adverbes *toujours* et *encore*. Nous repérons également les auxiliaires de visée aspectuelle tels que *être sur le point de* et *être en train/cours de* qui marquent le prospectif et l'inaccompli respectivement.

7. Pour rappel, un prédicat est télique si il est compatible avec *en+durée* et incompatible avec *pendant+durée*. Nous détaillons l'ensemble des tests linguistiques liés aux types de procès en annexe A.

8.3.3.2 Les types de procès

Pour rappel, Vendler (1967) définit les quatre types de procès : *état*, *activité*, *accomplissement* et *achèvement*. Les types de procès sont définis selon leurs traits *ponctuel*, *dynamique* et *borné* (cf. tableau 16).

TYPE DE PROCÈS	DYNAMIQUE	BORNÉ	PONCTUEL
état	-	-	-
activité	+	-	-
accomplissement	+	+	-
achèvement	+	+	+

TABLEAU 14 – Définition des types de procès selon leurs traits *dynamique*, *borné* et *ponctuel*.

La valeur du type de procès dépend du sens du verbe et de son environnement actanciel. Laurent Gosselin propose au sein de la *SdT* une classification des verbes et de leurs types de procès. Nous avons constitué un lexique dont nous présentons la démarche en annexe A. En résumé, un verbe est associé à une classe qui est constituée de quatre types d'associations différentes auxquelles sont associées un type de procès. Les quatre types d'associations sont identifiés grâce à :

- A. la présence d'un syntagme nominal objet introduit par un article partitif ou indéfini pluriel :

Il attelait des chevaux.

- B. l'absence de syntagme nominal objet, de subordonnée complétive objet, et de subordonnée interrogative indirecte :

Je roule (sur l'autoroute).

- C. la présence d'un syntagme nominal objet introduit par un article démonstratif, possessif ou défini, ou d'une subordonnée complétive objet, ou d'une subordonnée interrogative indirecte :

Il accapare les blés de la province.

- D. la présence d'un syntagme nominal objet introduit par un article indéfini :

Le bûcheron casse un rondin de bois.

Il est donc nécessaire de repérer ces quatre types d'associations et de déterminer le type de procès à l'aide du lexique constitué. Nous avons donc mis en place une analyse des syntagmes nominaux, spécifiant les points suivants :

- le genre : masculin ou féminin ;
- le nombre : singulier ou pluriel ;
- le type de l'article : démonstratif, possessif, partitif, indéfini ou défini.

8.3.4 Traitement des déclencheurs

Le traitement de déclencheurs repose en partie sur les différentes analyses que nous venons de présenter. En effet, nous déterminons le rôle de déclencheurs des marqueurs au moment de leurs analyses respectives. Nous avons présenté précédemment les déclencheurs issue des différentes classes, utilisés lors de l'étude de corpus : *calendaire, présuppositionnel*.

Les déclencheurs des classes *fréquentielles* (souvent, parfois, rarement,...), *quantificatinnelles* (deux fois, à plusieurs reprises,..) reposent sur des grammaires Prolog, et ne posent pas de problèmes de repérage.

[valeur :	souvent
	itérateur :	fréquentiel
	sporadicité :	75

(a) Structure de trait du fréquentiel *souvent*.

[valeur :	à plusieurs reprises
	itérateur :	quantificatinnel
	quantification :	flou

(b) Structure de trait du quantificatinnel à *plusieurs reprises*.

FIGURE 55 – Exemple de structures de trait des déclencheurs.

Les *verbes « intrinsèquement » itératifs* tels que (virevolter, hachurer, récrire, réacquérir, ...) sont listés au sein d'un lexique qui comporte actuellement 564 entrées. Les verbes, sous leur formes lemmatisées, sont facilement distinguables de par leurs préfixes (*re-*) ou leurs suffixes (*-ailler, -oleter, -iller, -onner, -asser, -eter, -oter, -ouiller* et *-oyer*). Nous avons créée cette ressource en nous basant sur *Lexique 3*, en filtrant sur les affixes. Ce lexique comporte également certains verbes tels que *fumer* impliquant une habitude. Notons que ce lexique ne possède pas de traits sémantiques spécifiques, car à terme il devra être intégré au lexique sur les classes de verbes (*cf.* annexe A).

En ce qui concerne les déclencheurs de type « temps morphologiques », c'est-à-dire le présent ou l'imparfait d'habitude, nous ne les traitons pas fautes d'avoir trouvé des critères satisfaisant permettant de les distinguer des cas non itératifs.

8.3.4.1 *Les déclencheurs événementiels*

Nous nous intéressons dans cette section aux subordonnées temporelles dont la conjonction ou la locution conjonctive ne déclenche pas systématiquement une structure itérative. Pour rappel, nous avons présenté les autres cas dans la section précédente intitulée « Les subordonnées temporelles » (cf. page 142).

Comme les exemples 74a et 74b le démontrent la conjonction *quand* (ou *lorsque*) peut introduire une subordonnée temporelle qui sera le déclencheur d'une itération (cf. exemple 74a) ou non (cf. exemple 74b). Afin de déterminer les cas où cela est vrai, nous avons caractérisé deux critères que nous présentons par la suite :

- le type de procès que la subordonnée introduit (état irréversible ou non). Nous qualifions d'état irréversible tout état qui ne peut avoir lieu qu'une fois au cours du temps comme « être mort », « être adulte » ... ;
- le rapport entre le temps morphologique de la principale et de la subordonnée.

(a) Quand j'étais stressé, j'avais mal à la tête.

(b) Quand j'étais petit, j'étais timide.

EXEMPLE 74 – L'exemple 74a présente une subordonnée temporelle qui peut se répéter au cours du temps tandis que l'exemple 74b introduit une période temporelle contiguë et révolue.

Le type de proposition introduit par la conjonction

Ce premier critère consiste à déterminer si la subordonnée temporelle introduit un état irréversible. En effet, il a été remarqué durant le projet, que la présence d'un état irréversible dans une subordonnée temporelle interdit une lecture itérative du procès de la principale associée à cette subordonnée. Actuellement, nous n'avons pas implémenté ce critère pour deux raisons : d'une part, il serait nécessaire de constituer une ressource importante ; et d'autre part, nous n'en avons pas détecté dans notre étude de corpus.

Le temps entre la proposition principale et la subordonnée

Le second critère est issu des travaux d'Annes Condamines sur les subordonnées temporelles (cf. section 1.3.2 page 21), et confirmé par une étude sur corpus. Il consiste à regarder le rapport du temps existant entre celui de la prin-

principale et celui de la subordonnée. Les temps favorisant une lecture itérative de la subordonnée temporelle sont référencés au tableau 15.

PROPOSITION PRINCIPALE	PROPOSITION SUBORDONNÉE
présent	présent
présent	passé composé
passé composé	présent
imparfait	imparfait
imparfait	plus que parfait
plus que parfait	imparfait

TABLEAU 15 – Couple des temps morphologiques favorisant permettant une lecture itérative d’une subordonnée temporelle.

Les exemples de 75a à 75d illustrent les différents cas possibles du tableau 15.

- (a) Lorsque je cours un 100 mètres, j’utilise des chaussures à clous.
- (b) Je regarde la télévision quand j’ai terminé mes leçons.
- (c) J’ai terminé mon travail lorsque je vais voir ma copine.
- (d) Marie m’appelait quand elle terminait son travail.

EXEMPLE 75

Les exemples 76a et 76b montrent que d’autres couples de temps de l’indicatif ne permettent pas le déclenchement d’une lecture itérative du procès de la principale. L’exemple 76a fait intervenir deux passés composés tandis que l’exemple 76b présente un verbe au plus-que-parfait dans la principale et au passé composé dans la subordonnée.

- (a) Luc m’a appelé quand il a eu terminé son travail.
- (b) J’avais terminé mon travail lorsque je suis rentré chez moi.

EXEMPLE 76

8.4 ANALYSE DES STRUCTURES ITÉRATIVES

L'analyse des structures itératives nécessite de distinguer les structures composées d'une seule proposition de celles s'étendant à plusieurs. Le premier cas ne pose pas de problème implémentatoire spécifique les déclencheurs se trouvant dans la majorité des cas en position intégrée dans la même proposition. Nous nous attardons plus particulièrement au second cas.

8.4.1 Méthode d'analyse

Nous avons présenté en début de chapitre que l'étude des cadres de discours revêt des problématiques similaires aux nôtres, en particulier sur la portée des introducteurs, et de la clôture de cadres, en ce qui nous concerne les déclencheurs et les structures itératives. Nous devons cependant prendre en compte deux critères supplémentaires, que nous avons déjà évoqués dans l'étude de corpus et que l'on peut observer sur les exemples :

- La possible discontinuité textuelle des propositions intégrant une structure itérative.

« Quand le temps était clair, on s'en allait de bonne heure à la ferme de Geffosses. **La cour est en pente, la maison dans le milieu ; et la mer, au loin, apparaît comme une tache grise.** Félicité retirait de son cabas des tranches de viande froide [...] »

EXEMPLE 77 – Extrait d'*Un cœur simple*. La phrase en gras n'appartient pas à la structure itérative.

- La possible position postérieure de l'introducteur par rapport à la structure itérative.

Les Russes Tatiana Totmianina et Maxim Maririn ont remporté, mercredi 18 janvier, à Lyon, la médaille d'or de l'épreuve de couple du championnat d'Europe de patinage artistique. C'est leur cinquième titre d'affilée.

EXEMPLE 78 – Article extrait de la rubrique *Sport* du journal LE MONDE.

La première phase de notre méthode, la détection des déclencheurs en position initiale, est la moins problématique. L'analyse des déclencheurs, présen-

tée dans la section précédente, fournit l'ensemble des marqueurs candidats, auxquels on attribue ou non le statut d'introducteur en fonction de critères positionnels relativement simples. Les introducteurs sont en effet caractérisés par leur position détachée en tête de phrase, séparée des propositions suivantes par une ponctuation faible (généralement la virgule). Notons qu'il est possible que les déclencheurs soient en position initiale mais que toutefois la ponctuation fasse défaut, par exemple :

parfois on convertit toutes les forces de cette âme en habileté, en splendeur pour agir sur des êtres [...]

Le problème de la portée, c'est-à-dire de la délimitation de la borne finale des structures, est plus complexe. En effet, il n'existe généralement pas de marque explicite permettant de fermer cette structure. Comme nous l'avons montré dans l'étude de corpus, l'apparition d'un nouveau déclencheur ne constitue pas systématiquement un indice de terminaison, en raison de l'éventualité de l'imbrication des structures itératives et de la discontinuité textuelle au sens de la page précédente.

Parfois nous allions jusqu'au viaduc, dont les enjambées de pierre commençaient à la gare et me représentaient l'exil et la détresse hors du monde civilisé, *parce que chaque année, en venant de Paris, on nous recommandait de [...]*. Nous revenions par le boulevard de la gare,

EXEMPLE 79 – Exemple illustrant la récursivité des structures itératives. Ici *Parfois* introduit une première structure itérative dont l'énoncé en italique, qui contient une structure itérative déclenchée par *chaque année*, ne permet pas de terminer la première structure itérative. En effet, la suite de l'énoncé initiée par *Nous revenions* est une partie intégrante de l'itération introduite par *Parfois*.

L'étude linguistique fait apparaître cependant un faisceau d'indices exploitables de façon automatique. Le faisceau d'indices ainsi élaboré est particulièrement hétérogène, puisqu'il regroupe à la fois des critères de surface, morphosyntaxiques et sémantiques. On exploite tout d'abord des critères liés à la structure logique du document, et plus précisément le découpage en paragraphes. Nous avons observé en corpus qu'une structure itérative s'étend très rarement au-delà du paragraphe auquel il appartient (*cf.* section 6.4.3). Cette limite peut donc être utilisée de façon fiable en dernier recours pour clore la structure.

Nous utilisons un critère lié à la cohésion des temps des verbes, en nous appuyant sur une analyse linguistique de Le Draoulec et Péry-Woodley (2001) sur l'impact de ce paramètre sur la cohésion textuelle. Le changement de temps

verbal peut être considéré comme indice de terminaison d'une structure. La mise en œuvre automatique de ce test dépend évidemment de l'analyse morphologique effectuée en amont au travers l'annotation des noyaux verbaux. Afin d'éviter la fermeture prématurée des structures, nous relâchons partiellement ce critère en ne considérant que le temps de l'auxiliaire. On assimile ainsi le passé composé au présent, le plus-que-parfait à l'imparfait, etc.

[Dès que le soleil descendait,] les gens du hameau sortaient un à un de la lisière des bois et revenaient par la route avec leurs brouettes et leurs charges de fagots : élaguer les taillis et élever des vaches pie-noires semblait être leur unique occupation. Ils saluaient Grange en passant sous le châtaignier, avec des remarques météorologiques sagaces - de la guerre il n'était jamais question - [quelquefois] il invitait le fils Biheureau à boire un verre, et faisait des frais de conversation. La mélancolie passait vite, et il lui poussait une pointe d'importance : il se faisait l'effet d'un vidame débonnaire descendu de son donjon pour boire au frais avec les manants de sa châtellerie.

EXEMPLE 80 – Exemple illustrant la cohésion des temps des verbes.

Par ailleurs, nous prenons en compte la discontinuité des structures itératives en permettant des « trous » dans cette cohésion textuelle, c'est-à-dire que nous autorisons certains verbes à déroger à cette règle. Une étude plus complète serait nécessaire pour déterminer la taille de ce « trou », en l'état actuel, nous avons fixé cette taille à celle d'une phrase, comme dans l'exemple 77.

Nous avons également envisager d'utiliser la cohésion des marqueurs aspectuo-temporels pour affiner ce critère, en nous appuyant sur l'analyse de Condamines (1992) indiquant que seuls les prédicats de types *activité* et *état transitionnel* ont la capacité d'être itérés, les accomplissements et les achèvements nécessitant une configuration bien précise (cf. section 1.3.2 page 21). Toutefois, tel que nous avons implémenté l'analyse aspectuelle, elle reste moins fiable qu'une analyse morphologique. Afin d'éviter les clôtures prématurées du fait d'un manque de précision de l'analyseur, nous ne nous reposons pas dessus dans la suite de ce traitement.

La méthode exploite aussi un critère de cohésion sémantique, ce que nous permet l'analyseur sémantique des expressions temporelles. Nous utilisons d'une part les valeurs symboliques produites par ces derniers pour évaluer la compatibilité référentielle d'un déclencheur calendaire avec les expressions temporelles qui lui succèdent en discours ; et d'autre part, les connecteurs tem-

poriels tels que *d'abord, et puis, ensuite*. Dans l'exemple 81 que nous avons simplifié pour l'illustration, nous calculons une chaîne temporelle des expressions temporelles par rapport au calendrier *tous les jours*. Nous pouvons également calculer les conflits entre les différents déclencheurs calendaires, cela nous permet de rendre compte de la relation de sélection entre *tous les deux jours* et *tous les jours*. On considère l'apparition d'une expression incompatible comme un indice de rupture, puisqu'elle n'appartient pas à l'univers défini par la chaîne temporelle.

Levé tous les jours à cinq heures, il a franchi comme un oiseau l'espace qui sépare son domicile de la rue Montmartre. [...] À neuf heures, il est [...] À dix heures moins un quart, il apparaît [...] À six heures, tous les deux jours, il est fidèle à son poste. À minuit, il redevient bon mari, [...]

EXEMPLE 81 – Extrait de *La fille aux yeux d'or* (1834-35).

Enfin, la méthode se repose sur un dernier critère qui est la reprise anaphorique événementielle afin de prendre en compte les structures itératives dont le déclencheur est en position finale et non plus initiale (cf. exemple 78). Nous appliquons donc les différents critères que nous venons d'exposer mais en cherchant ici la borne initiale de la structure itérative.

8.4.2 Du point de vue de l'implémentation

Comme nous l'avons vu dans la section précédente sur le repérage des constituants des structures itératives, nous détectons les procès (les noyaux verbaux et les substantifs événementiels), les circonstanciels de temps et les déclencheurs. L'analyse de ces constituants a permis d'associer aux expressions reconnues une représentation de leur « sens » sous la forme de structures de traits. L'analyse des structures itératives s'appuie sur ces annotations comme critère de circonscription.

Sur cette base, la détection des déclencheurs dans un rôle d'introducteur peut être mise en place à l'aide de critères essentiellement positionnels. Les contraintes exprimées sont fondamentalement séquentielles, puisque nous recherchons des zones de texte vérifiant des motifs imposant la présence, dans un ordre fixé, d'éléments immédiatement successifs. Ces règles sont donc simplement exprimables à l'aide du formalisme d'expressions régulières (MRE) de la plate-forme LINGUASTREAM.

Pour la détermination de la portée, la méthode présentée dans la section précédente s'appuie sur des critères tels que la cohésion des temps verbaux, sur la structuration en paragraphes, et sur des calculs sémantiques de cohérence. Pour construire les structures itératives, nous utilisons le formalisme CDML (Constraint-based Discourse Modeling Language), pendant implémentatoire du modèle *unité – relation – schéma*, développé par Antoine Widlöcher (2008) (cf. section 7.2 page 120). CDML fournit un moyen formel et déclaratif de décrire, et d'analyser automatiquement, par contraintes, les structures du discours. Son principe fondamental consiste à permettre de préciser un ensemble de contraintes devant être satisfaites par des objets textuels pour devenir indices de la présence d'objets textuels de plus haut niveau. Par exemple, un déclencheur en position initiale et une suite de procès construisent une structure itérative si les verbes respectent la règle de cohésion verbale (cf. figure 56 pour un exemple de règle)⁸. Notons que CDML est encore en cours d'élabo-

```

Rule {type : "StructureIterative"} :
  start({type : "declencheurPositionInitiale"})
  end({type : "phrase"})
  homogeneity(comparator : portee)
  not presence(pattern : {type : "declencheurPositionInitiale"}, amount : 2)
  size(mode : #LONGEST)

Comparator portee ({type : "verbe"} as $v1, {type : "verbe"} as $v2) :
  $v1/temps = $v2/temps

```

FIGURE 56 – Exemple de grammaire CMDL permettant d'implémenter le modèle *Unités, relations et schémas* de Widlöcher (2008).

ration, et qu'il n'intègre pas encore la notion de *schéma*. L'application de ce formalisme nous oblige à déroger au modèle d'annotation que nous avons établi dans le chapitre précédent, sans pour autant remettre en doute la validité des choix que nous avons effectués. Actuellement, les structures itératives sont représentées par des unités textuelles, similaires aux cadres employés dans Charolles.

8. Nous renvoyons le lecteur à Widlöcher (2008) pour plus de détails sur la formalisation et les règles de grammaires de CDML.

{Sci:1A} {sentence:1A} {declencheur:1A} Dès que le soleil [descendait] , {declencheur:1} [+] les gens du hameau [sortaient] [+] un à un de la lisière des bois et [revenaient] [+] par la route avec leurs brouettes et [+] leurs charges de fagots : [élaguer] [+] les taillis et [élever] [+] des vaches pie - noires [semblait] [être] [+] leur unique occupation . {sentence:1} {sentence:2A} Ils [saluaient] [+] Grange en [passant] [+] sous le châtaignier , [+] avec des remarques météorologiques sagaces {sentence:2} {sentence:3A} - [+] de la guerre il [n' était] [+] jamais question - {sentence:3} {Sci:1} {Sci:1A} {sentence:4A} {declencheur:1A} quelquefois {declencheur:1} il [invitait] [+] le fils Bihoreau à [boire] [+] un verre , et [faisait] [+] des frais de conversation . {sentence:4} {sentence:5A} [+] La mélancolie [passait] vite , et il [lui poussait] [+] une pointe d' importance : il [se faisait] [+] l' effet d' un vidame débonnaire descendu [+] de son donjon pour [boire] au frais [+] avec les manants de sa châtellerie . {sentence:5} {Sci:1}

{Sci:2A} {sentence:6A} {declencheur:2A} Quand il [était] de retour avant la tombée de la nuit , {declencheur:2} il [manquait] {declencheur:3A} rarement {declencheur:3} de [descendre] [+] dans le fortin pour une courte inspection ; c' [était] ce qu' il [appelait] [jeter] [+] un coup d' oeil au blockhaus . {sentence:6} {sentence:7A} A [dire] vrai [+] le coup d' oeil [était] sans nécessité aucune , [+] le bloc restant fermé à clef toute la journée , mais il [lui était] venu [+] une manie bizarre : il [aimait] [se tenir] là quelques instants [+] à la chute du jour . {sentence:7} {Sci:3A} {sentence:8A} {declencheur:4A} Quand il [était] de bonne humeur , {declencheur:4} il s' en [plaisantait] lui - même : il [se disait] qu' il [ressemblait] [+] à ces officiers mécaniciens vieilliss [+] dans le métier qui [préfèrent] [descendre] [fumer] [+] leur cigarette dans les fonds du navire . {sentence:8} {Sci:3} {Sci:4A} {sentence:9A} {declencheur:5A} Lorsqu' il [avait rabattu] sur lui [+] la lourde porte de coffre - fort , {declencheur:5} il [s' arrêtait] [+] un instant sur le seuil , et [jetait] [+] sur les murs et [+] sur le plafond écrasé qui [faisait] [rentrer] d' instinct [+] la tête dedans les [épaules] [+] un coup d' oeil qui [n' allait] [+] jamais sans malaise : il [était envahi] [+] par une sensation intense de dépaysement . {sentence:9} {Sci:4} {Sci:2}

FIGURE 57 – Exemple d’annotation des structures itératives issues de notre analyseur.

EN RÉSUMÉ

Nous venons de décrire la phase d’analyse automatique de notre travail. Premièrement, l’analyseur couvre l’annotation des différents constituants des structures itératives, c’est-à-dire les procès, les circonstanciels de temps et les déclencheurs, tels que nous les avons défini au chapitre 7. Deuxièmement, l’analyseur se fonde sur ces constituants pour circonscrire les structures itératives. La méthode que nous employons se concentre sur les structures dont les déclencheurs sont en position initiale. Cette méthode exploite des critères de nature différente puisqu’elle regroupe des critères de surfaces, morphosyntaxiques et sémantiques.

Nous devons terminer la phase d’évaluation de notre l’analyseur tant d’un point de vue quantitatif que qualitatif pour rendre compte des limites de cette analyse, et pour envisager des améliorations possibles. Notons que la complexité des traitements nécessaires à l’analyse automatique de telles structure complique considérablement la tâche d’évaluation. Comme nous venons de le voir dans ce chapitre, notre méthode s’appuie sur une chaîne de traitements complexe, où l’analyse des structures itératives constitue l’ultime phase, s’appuyant sur les différents résultats antérieurement obtenus. De ce fait, toute erreur dans une phase intermédiaire de l’analyse peut avoir des conséquences sur toutes les étapes qui suivent. Conscient de cette problématique, l’évaluation mise en place prend le partie d’évaluer chaque traitement de manière indépendante.



Quatrième partie

LES PHÉNOMÈNES ITÉRATIFS
ET LE CALCUL ASPECTUEL

PRÉAMBULE

La présente partie expose notre approche du calcul aspectuel fondé sur la *Sémantique de la Temporalité* de Laurent Gosselin dont nous avons présenté les fondements au chapitre 2 page 32, et l'extension réalisée pour prendre en compte les phénomènes itératifs au chapitre 3 page 51.

Nous présentons au chapitre 9 les principes du calcul aspectuel des procès itératifs en faisant le parallèle avec le calcul des procès semelfactifs. De ces principes, nous avons mis en œuvre l'implémentation du système de Calcul Aspectuo-Temporel (*CAsTe*), qui permet d'obtenir la représentation aspectuo-temporelle du texte, que nous présentons au chapitre 10. Afin de réaliser cette implémentation, nous avons réécrit, en nous fondant notamment sur les travaux d'implémentation de Person (2004), les règles issues de la *SdT* pour les procès itératifs que nous présentons en annexe B.

Par ailleurs, nous aborderons également nos derniers travaux qui sont encore en cours d'élaboration sur l'extension d'un formalisme XML existant sur le temps, *TIMEML* (*cf.* chapitre 11), qui permet de structurer les connaissances temporelles contenues dans un texte. Notre objectif est de compléter un formalisme existant en intégrant notre conception des phénomènes itératifs, c'est-à-dire la distinction entre la série itérative et le procès modèle notamment marquée par la notion de double visée aspectuelle.



LE CALCUL ASPECTUO-TEMPOREL

Nous avons fondé notre calcul aspectuo-temporel des procès itératifs sur les travaux de la sémantique du temps et de l'aspect de Laurent Gosselin dans lesquels il conçoit les procès itératifs dans une double appréhension : un procès modèle, d'une part ; et une série itérative, d'autre part. Nous abordons dans ce chapitre les principes de ce calcul, que nous avons mis en œuvre et que nous présentons dans le chapitre suivant, en faisant notamment le parallèle avec le calcul des procès semelfactifs.

SOMMAIRE

9.1	Principes du calcul de la représentation	161
9.2	Des procès semelfactifs aux procès itératifs	163
9.3	La temporalité absolue des procès modèles	166
9.4	Les circonstanciels de temps	167
9.5	La dimension textuelle	170
9.5.1	La portée des connecteurs temporels	171
9.5.2	La récursivité	172
9.5.3	La sélection	174

9.1 PRINCIPES DU CALCUL DE LA REPRÉSENTATION

Dans la *SdT*, le processus de construction de la représentation aspectuo-temporelle du texte se fonde sur des marques linguistiques du texte. Des règles de construction d'éléments de représentation sont définies pour chacune des marques pertinentes. Pour rendre compte de la polysémie des marques au sein du modèle, le mode de calcul est compositionnel holiste, et donc combine les principes de compositionnalité (la signification du tout est déterminée par celle de ses parties) et de contextualité (la signification d'une expression est au moins partiellement déterminée par le contexte dans lequel elle apparaît).

Ces marques de type temporelles et aspectuelles se répartissent sur divers éléments de chaque proposition du texte. Il s'agit donc de préciser quels sont ces éléments linguistiques qui fournissent les instructions nécessaires à la construction de la représentation aspectuo-temporelle du texte. Il existe trois catégories de marques :

- les marqueurs aspectuo-temporels : les verbes et leurs compléments, les temps morphologiques, les compléments circonstanciels (*pendant deux jours, le 14 juillet*), des locutions adverbiales (*hier, la veille*), les propositions subordonnées temporelles (*tant que P, au moment où P*), les connecteurs temporels (*ensuite, puis*) et les adverbes aspectuels ;
- la construction syntaxique des propositions qui entretiennent entre elles des relations constituant des contraintes pour la construction de la représentation temporelle, par exemple la subordonnée temporelle ;
- les contraintes pragmatico-référentielles comme l'identité de procès, les relations d'élaboration ou de causalité, etc.

Les instructions, en se fondant sur les marques ci-dessus, produisent les éléments de la représentation globale du texte. Ces éléments correspondent aux relations entre les intervalles du modèle et les informations sur les procès. En cas de conflit de ces instructions, la *SdT* prévoit des règles de résolution qui amènent à des effets de sens en discours qui sont la dilatation (*Il est arrivé en cinq minute.*), le glissement vers la phase préparatoire ou finale (*il rentra à 8h35, ici la phase finale*), l'interruption (*Il courut le marathon pendant deux heures, puis il s'effondra.*) et l'itération (*Depuis deux mois, il mange en cinq minutes.*). Ce principe de résolution de conflits permet de prendre en compte la polysémie des marqueurs. La figure 58 expose le principe du calcul temporel au sein de la *SdT*.

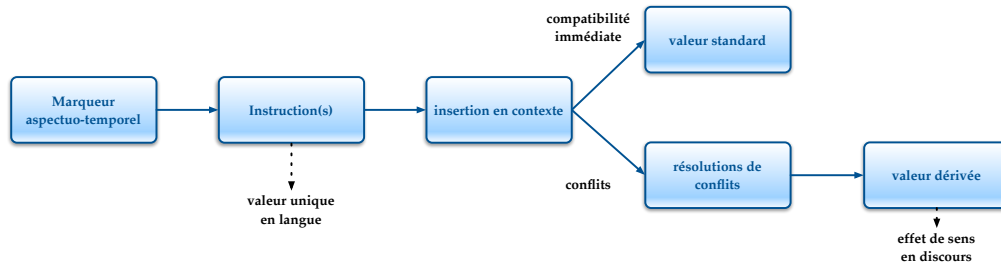


FIGURE 58 – Schéma récapitulatif du principe du calcul temporel au sein de la *SdT* : les marqueurs aspectuo-temporels codent des instructions, c'est-à-dire des relations entre les intervalles du modèle *SdT*. Le modèle prévoit alors deux fonctionnements. Le premier correspond à un fonctionnement standard où les instructions construisent directement les représentations aspectuo-temporelles. Le second correspond à l'existence de conflits entre les instructions, amenant à des effets de sens en discours, dont l'itération est un des modes de résolutions.

Dans son extension de la *SdT* aux phénomènes itératifs, Laurent Gosselin conçoit l'itération comme une série itérative qui correspond à une agglomération de procès identiques, c'est-à-dire une série d'occurrences d'un élément itéré. Il introduit ainsi, à la différence des procès classiques, une double visée aspectuelle : la première est relative à la série itérative, et la seconde au procès modèle.

Notre calcul se place dans cette lignée, en conservant la notion de série itérative et de procès modèle, ce dernier étant donc une abstraction des itérés de la série. Notons que selon nous les itérés de l'itération ne sont pas accessibles sans y faire référence explicitement. De ce fait, nous ne calculons aucune relation aspectuo-temporelle sur les itérés, nous limitant donc au calcul de relations aspectuo-temporelles au niveau de la série itérative et du procès modèle.

La problématique du calcul aspectuel des procès itératifs consiste donc à prendre en compte la double visée aspectuelle. En effet, là où les marqueurs aspectuo-temporels codent sur une seule visée aspectuelle dans le cas des procès semelfactifs, ils portent sur deux visées pour les procès itératifs. Dans une perspective d'analyse automatique, il est donc nécessaire de déterminer si la valeur aspectuelle associée à un marqueur porte sur la série itérative ou sur le procès modèle.

Nous reprenons les travaux effectués par Laurent Gosselin que nous avons présentés au chapitre 3 où il met en évidence les interactions des différents marqueurs aspectuels par rapport aux procès itératifs. Nous nous basons en particulier sur le constat suivant qu'il fait : les valeurs aspectuelles associées aux temps morphologiques sont affectées à la série tandis que le procès modèle recouvre dans la majorité des cas une visée d'aoristique (*cf.* chapitre 3 page 55). Par exemple, pour une itération au passé simple telle que l'exemple 82, nous avons les instructions suivantes :

- la valeur temporelle est le passé : IRs ANT IE ;
- la valeur aspectuelle de la série est aoristique : IPs CO IRs ;
- la valeur aspectuelle du procès modèle est aoristique : IP CO IR.

La figure 59 page suivante représente un chronogramme de l'exemple 82. Cette figure nous permet de poser une première question sur l'adaptation du calcul aux procès itératifs : *Le procès modèle, du fait qu'il est une abstraction des itérés, est-il positionné par rapport au moment de l'énonciation ?* Question à laquelle nous apportons une réponse négative et détaillée dans la section suivante.

Madame Aubain tomba plusieurs fois en tachant de gravir le talus.

EXEMPLE 82

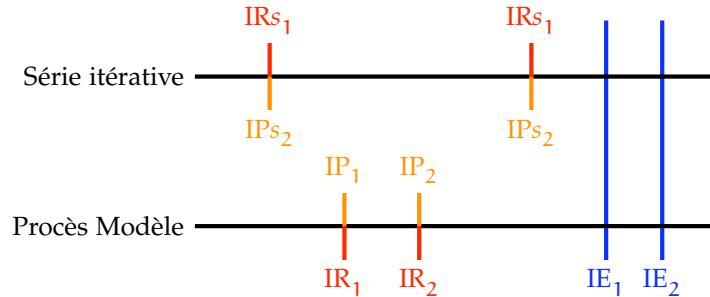


FIGURE 59 – Représentation du chronogramme de l'exemple 82 au passé simple.

Cette étape du calcul permet donc de déterminer les relations entre les intervalles de procès, de référence et d'énonciation. Les différentes règles associées aux temps morphologiques sont détaillées en annexe B ainsi que l'ensemble des règles que nous avons mises au point. Il reste toutefois à étudier des points précis préalablement à une implémentation. Nous consacrons la suite de ce chapitre à l'étude de trois cas.

Premièrement, il s'agit d'étudier la portée des circonstants temporels semelfactifs et itératifs par rapport à la double visée aspectuelle. Dans le cas des procès semelfactifs tels que *Hier, je suis allé à la piscine*, la *SdT* prévoit la création d'un intervalle circonstanciel qui est en relation avec l'intervalle de référence (c'est le cas dans notre exemple) ou de procès. Pour les procès itératifs, il est nécessaire de distinguer d'une part les circonstants temporels semelfactifs qui créent un seul intervalle ; et d'autre part, les circonstants temporels itératifs créant deux intervalles tels que *tous les jeudis* : un sur la série itérative (*l'ensemble des jeudis*) et un sur le procès modèle (*un jeudi modèle*). Par ailleurs, nous devons également déterminer si les circonstanciers portent sur la série itérative, sur le procès modèle ou encore sur un autre circonstanciel. Prenons les exemples 83a et 83b. Nous avons dans le premier exemple un intervalle circonstanciel semelfactif associé à *depuis 2008* et qui est en relation avec l'intervalle de référence de la série tandis que dans le second exemple, nous avons un circonstanciel semelfactif en relation avec le procès modèle.

- (a) Depuis 2008, les crises se succèdent au niveau mondial.
 (b) Il terminait régulièrement son travail à 20h.

EXEMPLE 83

Deuxièmement, il s'agit de déterminer si les procès semelfactifs et itératifs d'un énoncé sont en relation avec la série itérative (*cf.* exemple 84a) ou avec les procès modèles (*cf.* exemple 84b). L'exemple 84a présente les procès semelfactifs en relation avec la série itérative (Proces_{entrer} ANT Iteration_{boire} ANT Proces_{sortir}) tandis que dans l'exemple 84b les relations sont au niveau des procès modèles (ProcesModele_{aller faire} ANT ProcesModele_{rentrer}).

- (a) Il est entré dans un bar, il a bu trois bières, puis il est sorti.
 (b) Chaque lundi, il allait faire ses courses, et il rentrait chez lui.

EXEMPLE 84

Enfin, troisièmement, il s'agit de proposer une représentation des phénomènes de récursivité et des sélections en termes d'intervalles pour le calcul aspectuel.

Notation

Nous employons les notations suivantes pour présenter les différentes règles de calcul :

Les séries itératives :

- les intervalles de procès sont notés IPs ;
- les intervalles de références sont notés IRs ;

Les procès modèles :

- les intervalles de procès sont notés IP ;
- les intervalles de références sont notés IR ;

Les circonstanciels :

- les intervalles circonstanciels, liés aux circonstanciels semelfactifs, sont notés IC ;

- les intervalles circonstanciels, associés aux déclencheurs, notamment de type calendaire (*tous les jours*), sont notés ICs pour la série et IC_m pour le modèle.

9.3 LA TEMPORALITÉ ABSOLUE DES PROCÈS MODÈLES

Dans la *SdT*, le positionnement temporel des séries itératives est indéniable (cf. chapitre 3 page 51). Elles ont, à l’instar des procès classiques, leurs intervalles de procès mis en relation avec l’intervalle d’énonciation. *Quid des procès modèles ?* La question qui se pose est donc de déterminer si les procès modèles possèdent une valeur temporelle absolue ou non, c’est-à-dire si ils sont en relation avec l’intervalle d’énonciation.

Nous pensons pour notre part que les procès modèles ne sont pas liés à l’intervalle d’énonciation. Nous étayons ce propos avec l’exemple suivant : « *En 2010, je suis souvent allé au cinéma. »*. Nous avons une itération dont l’ensemble des itérés est placé temporellement avant le moment d’énonciation, contraint par l’intervalle de procès de la série. Le procès modèle est *a priori* également antérieur au moment d’énonciation. Si notre énoncé se poursuit par la phrase « *D’ailleurs, j’y retournerai en 2012.* », nous avons les itérés de cette nouvelle itération qui sont postérieurs à l’intervalle d’énonciation. Le procès modèle reste toutefois le même que pour la première itération, il devrait donc se situer postérieurement à l’intervalle d’énonciation. La figure 60 présente une schématisation simplifiée de la configuration de notre exemple où le procès modèle est lié aux itérés des deux itérations. Nous pouvons constater que le procès modèle ne peut pas être à la fois antérieur et postérieur à l’intervalle d’énonciation.

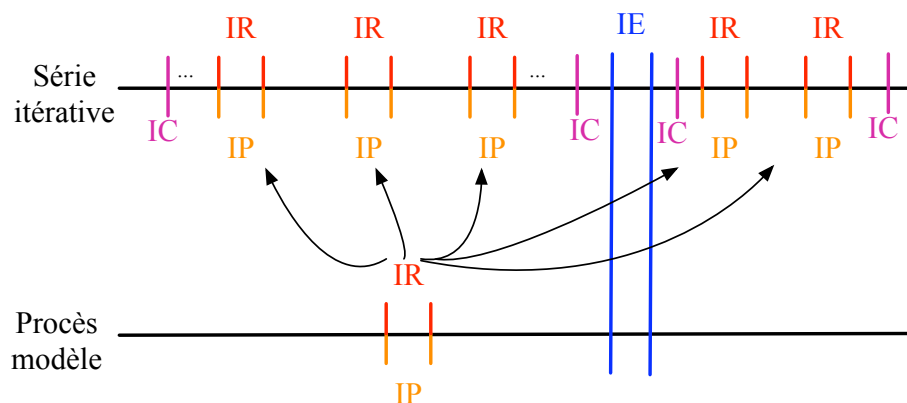


FIGURE 60 – Représentation simplifiée de l’exemple « *En 2010, je suis souvent allé au cinéma. D’ailleurs, j’y retournerai en 2012.* ».

Nous proposons donc une solution où les procès modèles sont atemporels, seuls les itérés, et les séries, étant positionnés par rapport à l'intervalle d'énonciation. Nous schématisons l'a-temporalité des procès modèle par une ligne pointillée entre l'axe de la série et des procès modèles. Pour notre exemple précédent, nous avons la structure aspectuo-temporelle représentée sur la figure 61, où le procès modèle est non contraint par rapport à l'intervalle d'énonciation.

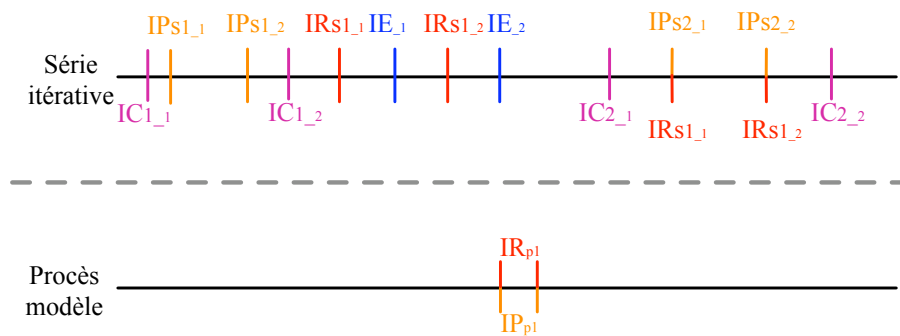


FIGURE 61 – Représentation de l'atemporalité des procès modèles de l'exemple « En 2010, je suis souvent allé au cinéma. D'ailleurs, j'y retournerai en 2012. ». Les intervalles de la série itérative associés au procès *je aller au cinéma* sont IPs_1 , IRs_1 et IC_1 tandis que ceux associés à *je y retourner* sont IPs_2 , IRs_2 et IC_2 .

De plus, cette solution suit les principes de MOP1 où les procès modèles sont utilisés pour créer les itérés. Notons que certaines configurations de phénomènes itératifs ont leurs procès potentiellement temporalisés comme dans « *Jean allait souvent au cinéma avec son père.* ». Dans l'hypothèse où Jean a perdu son père, il n'est plus possible d'avoir d'itérés dans le futur. Selon Yann Mathet dans Mathet (2011 – à paraître), seule la création de nouvel itéré est interdite, ne remettant pas ainsi en cause la modélisation choisie.

9.4 LES CIRCONSTANCIELS DE TEMPS

Nous abordons maintenant la problématique de la portée des circonstancielles temporelles dans le cadre des procès itératifs que nous avons introduits en section 9.2. Elle nécessite de déterminer, en plus du fait qu'elle porte soit sur l'intervalle de référence, soit sur l'intervalle de procès, si elle porte sur la série itérative ou sur le procès modèle comme l'illustre l'exemple 85.

- (a) En 10 minutes, il a lu trois fois ce livre.
 (b) Jean mange régulièrement à 19 heures 30.

EXEMPLE 85 – Dans l'exemple 85a, le circonstanciel de durée « en 10 minutes » porte sur la série itérative tandis que dans l'exemple 85b, le circonstanciel semelfactif « à 19 heures 30 » porte sur le procès modèle.

La position syntaxique du complément circonstanciel permet de déterminer le type d'intervalle en relation avec l'intervalle circonstanciel. Si celui-ci est en position détachée, ce qui correspond généralement à la séparation par une virgule avec le reste de la proposition (*cf.* exemple 85a), l'intervalle circonstanciel est en relation avec un intervalle de référence, sinon il est en relation avec l'intervalle de procès. Notons qu'une subordonnée temporelle est considérée comme un circonstant détaché. Nous obtenons donc la règle suivante :

Si
 – il existe un circonstant temporel
Alors
Si le circonstant temporel est en position détachée
Alors
 – IC est en relation avec IR ou IRs
Sinon
 – IC est en relation avec IP ou IPs

Règle 1 – Règle des circonstants temporels.

La détermination du choix entre la série itérative et le procès modèle dépend de la sémantique du calendrier et du procès. Cette détermination nécessite donc des connaissances sur le monde pour la lever. Nous pouvons toutefois lever l'ambiguïté dans trois cas.

Le premier cas correspond à un circonstanciel ayant une lecture uniquement semelfactive (par exemple, *cette semaine* contrairement à *à huit heures*). Le circonstanciel sera rattaché systématiquement à la série itérative (*cf.* l'exemple 86 où la subordonnée temporelle est semelfactive).

Quand Marie était jeune, elle fumait tous les jours.

EXEMPLE 86

Si

- le circonstanciel est semelfactif

Alors

- IC RC Ix, avec Ix ∈ {IPs, IRs }

Règle 2 – Cas de la portée d'un circonstanciel semelfactif.

Le second cas correspond à la présence d'un circonstanciel et un déclencheur calendaire. En fonction de la durée exprimée par le circonstanciel, nous aurons un rattachement à la série, dans le cas où elle est plus grande que celle exprimée par le calendaire (*cf.* exemple 87b), ou le rattachement se fera au niveau du procès modèle (*cf.* exemple 87a) dans le cas contraire.

- (a) J'ai lu ce livre tous les jeudis pendant 10 minutes.
- (b) J'ai lu ce livre tous les jeudis pendant 10 ans.

EXEMPLE 87

Si

- la durée du circonstanciel est inférieure à celle du calendaire

Alors

- IC RC Ix, avec Ix ∈ {IP, IR }
- IC_m ACCESS IC

Sinon

- IC RC Ix, avec Ix ∈ {IPs, IRs }
- IC_s ACCESS IC

Règle 3 – Cas de la portée entre un circonstanciel et un déclencheur calendaire.

Le troisième cas correspond à la présence de deux circonstanciels entraînant un conflit, et se résolvant par l'itération du procès comme dans l'exemple 88.

Il a mangé en cinq minutes pendant deux mois.

EXEMPLE 88

Si

- il y a conflit entre les circonstanciels

Alors

- création d'une itération
- IC_{\min} RC I_X , avec $I_X \in \{IP, IR\}$
- IC_{\max} RC I_X , avec $I_X \in \{IPs, IRs\}$

Règle 4 – Cas de la portée entre deux circonstanciels.

9.5 LA DIMENSION TEXTUELLE

La dimension textuelle consiste, dans la perspective du calcul aspectuel, à mettre en relation les intervalles des procès de l'énoncé. Les travaux de (Person, 2004) définissent déjà des règles mettant en relation les procès d'un énoncé. Un calcul des itérations était réalisé de manière partielle en reliant les intervalles associés à la série avec ceux des autres procès. La mise en relation des intervalles dépend principalement :

- de la structure syntaxique permettant de déterminer la précédence ou la succession des procès selon l'ordre d'apparition pour les propositions simples, selon le temps pour les subordonnées ;
- des connecteurs temporels qui relient les procès selon le type du connecteur :
 - une relation de succession (*et, ensuite, ...*) ;
 - une relation de précédence (*d'avance, ...*) ;
 - une relation de simultanéité (*quand, lorsque, ...*) ;
- de la cohérence temporelle qui correspond à la dépendance contextuelle de l'intervalle de référence, ce qui implique que chaque intervalle de référence de la représentation temporelle du texte doit posséder un intervalle antécédent avec lequel il coïncide.

Ces trois points sont traités dans le cadre de la *SdT* classique et intégrés dans l'implémentation de Person (2004). Nous avons donc concentré notre étude sur les trois points suivants :

1. la portée des connecteurs temporels ;
2. la récursivité des itérations ;
3. le phénomène de sélection des itérés.

Avant de détailler ces trois points, nous souhaitons apporter une précision sur les relations entre procès modèles. Ils respectent les mêmes règles que pour les énoncés non itératifs. Par défaut, il existe une relation de succession entre les procès composant l'itération.

9.5.1 La portée des connecteurs temporels

La problématique est de déterminer le rattachement des connecteurs temporels sur la série itérative ou sur le procès modèle. Dans la majorité des cas, la position du déclencheur, détachée ou intégrée, lève l'indétermination. La relation s'établit entre le procès non itéré et la série dans les cas où le déclencheur est en position intégrée, comme l'illustre l'exemple 89. Lorsque le déclencheur

Madame Aubain descendit le fossé, [. . .], tomba plusieurs fois en tâchant de gravir le talus, et à force de courage y parvint, puis courut vers la barrière.

EXEMPLE 89 – Exemple extrait d'*Un cœur simple*.

est en position détachée, la relation s'établit entre les procès modèle de la série (cf. l'exemple 90).

Quand je passais à côté de lui, il frisait sa moustache d'un air féroce et roulait de gros yeux, comme s'il eût voulu sabrer un cent d'Arabes.

EXEMPLE 90 – Exemple extrait de *Le Petit Chose*.

- Si*
- le déclencheur est en position intégrée
 - il existe un connecteur temporel entre P_i et P_j
- Alors*
- IP_{S_i} RC IP_j
- Sinon*
- IP_i RC IP_j

Règle 5 – Règle sur la portée des connecteurs temporels.

Sur l'exemple 89, nous obtenons une relation de succession entre les intervalles de procès (*descendre*, *parvenir* et *courir*) et de la série itérative (*tomber*), comme illustré sur la figure 62.

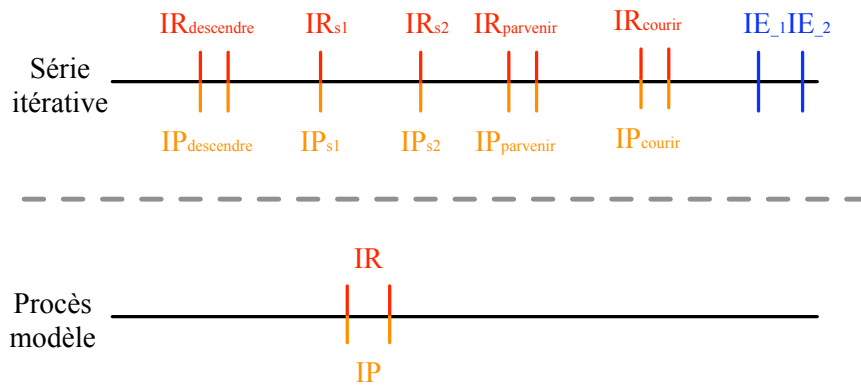


FIGURE 62 – Exemple de chronogramme de la portée des connecteurs.

9.5.2 La récursivité

La récursivité correspond à l'imbrication d'une itération au sein d'une autre. Deux cas sont à distinguer : les deux déclencheurs sont dans des propositions séparées ou au sein de la même. Dans le premier cas, nous avons une mise en relation traditionnelle entre les procès de l'itération principale et la série de la seconde itération, qui est considérée comme un procès itéré dans la première. Le second cas est plus intéressant dans le sens où il faut déterminer les priorités de déclenchement des itérations.

Nous proposons une première solution qui est centrée autour de la notion de position détachée ou intégrée des déclencheurs. Dans l'hypothèse où nous avons un déclencheur en position détachée et le second en position intégrée, le premier effectuera une itération de la série initiée par le second déclencheur comme dans l'exemple 91a. Dans l'hypothèse où nous avons cette fois les deux déclencheurs en position intégrée, l'ordre d'apparition dans la proposition détermine la construction des itérations. Le premier déclencheur itère la série initiée par la seconde comme dans l'exemple 91b. Dans l'hypothèse où nous avons cette fois les deux déclencheurs en position détachée, il semble que le second déclencheur apporte uniquement une précision sur le premier (*cf.* l'exemple 91c), ne déclenchant pas d'itération. Il est peut être vu comme un sélecteur.

- (a) Je vais souvent au médecin quand je suis malade.
- (b) Souvent, je vais au médecin quand je suis malade.
- (c) Régulièrement, quand il fait beau, nous allons à la plage.

EXEMPLE 91

Une seconde solution, qui reste à étudier plus largement, repose sur les interactions possibles entre les différents types de déclencheurs. En effet, il semble que certaines compositions de déclencheurs amènent toujours au même type de construction récursive. Par exemple, si la phrase est composée de deux déclencheurs de type fréquentiel et quantificationnel, nous obtenons systématiquement l'itération par le fréquentiel de l'itération déclenchée par le quantificationnel (*cf.* l'exemple 92), quelque soit sa position détachée ou intégrée. Ceci est également le cas pour une phrase composée d'un calendaire et d'un quantificationnel.

- (a) Souvent/Tous les jeudis, je fais dix tours de piste
- (b) Je fais souvent/tous les jeudis dix tours de piste.

EXEMPLE 92

En ce qui concerne l'aspect, les règles établies présentées en annexe B s'appliquent également dans le cas des structures récursives. Notons toutefois que les exemples, que nous avons rencontrés en corpus, présentaient des visées aoristiques à la fois pour la seconde série itérative et le procès modèle associé.

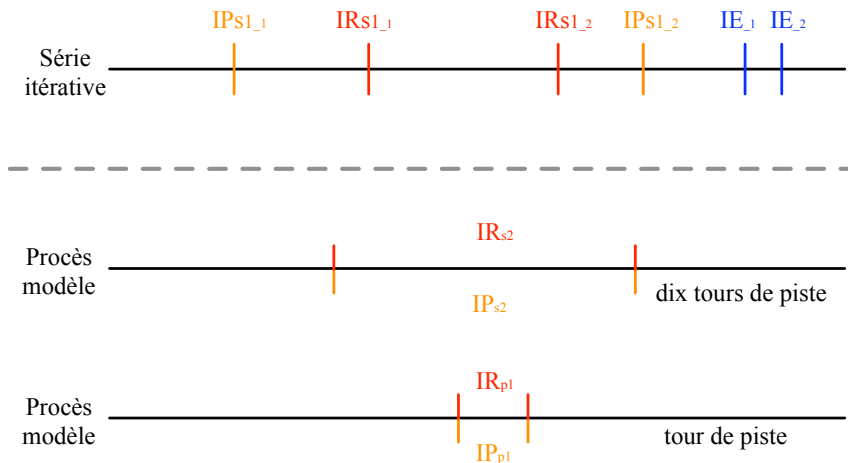


FIGURE 63 – Exemple de chronogramme de deux itérations imbriquées, par exemple « Je fais souvent dix tours de piste. ».

9.5.3 La sélection

Quid de la sélection dans la *SdT*? C'est un phénomène peu abordé nous semble-t-il par Laurent Gosselin. La difficulté de son traitement peut se formuler par la question suivante : quels itérés sont sélectionnés? En fonction du sélecteur, il nous est possible de répondre à la question levant par la même occasion la difficulté.

Trois types de sélections sont à différencier :

- les sélections appliquées à l'ensemble des itérés comme dans l'exemple 93a, nous avons une bijection entre la série de l'itération et celle de la sélection ;
- les sélections qui précisent les itérés choisis avec des sélecteurs tels que *la première fois, les trois dernières fois* ayant un repérage directe ou relatif (cf. l'exemple 93b) ;
- les sélections dont le sélecteur à une sémantique floue comme *parfois, souvent*, et qui ne permettent pas de déterminer les itérés avec précision (cf. l'exemple 93c). Il existe un ensemble de séries d'itérés solutions.

Dans la perspective du calcul aspectuel, nous proposons une solution dans la continuité de celle mise en œuvre pour les itérations. Nous calculons donc les relations aspectuelles sur une version modèle des itérés sélectionnés. La problématique liée au phénomène de sélection ne se pose donc plus en termes de calcul aspectuel mais en une opération de correspondance entre séries. Nous n'avons pas approfondi cette opération de correspondance entre séries qui dépasse le cadre de ce chapitre consacré au calcul aspectuel. Elle fait toutefois l'objet d'une étude dans les travaux de P. Enjalbert que nous avons pré-

Tous les jeudis, des habitués venaient faire une partie de boston. . . .

- (a) À chaque fois, Félicité préparait les cartes d'avance.
- (b) Les trois dernières fois, nous avons joué au moins quatre heures.
- (c) Souvent, nous finissions vers une heure du matin.

EXEMPLE 93

sentés au chapitre 4. Pour rappel, le modèle introduit une famille de relations d'extraction de série, dont le sélecteur précise ou quantifie la relation, pour formaliser la série issue de la sélection. Nous reprenons cette relation d'extraction entre la série itérative et la série associée à la sélection pour définir une sélection (cf. règle 6)¹.

Si

– il existe une sélection S appliquée à une itération I

Alors

– il y a une relation d'extraction : $S <_{\text{Adverbe}} I$

Règle 6 – Règle d'extraction des itérés de l'itération.

Selon les principes de la *SdT*, nous affectons à la sélection un couple d'intervalles de procès et de référence à la série sélective (IPs et IRs) puis un second couple pour chaque procès modèle (IP et IR), à l'instar des itérations. Dans le traitement aspectuel des sélections, les règles établies pour les itérations s'appliquent au cas des sélections. Il faut toutefois prendre en compte trois points :

1. Les intervalles de procès de la série itérative et de la série sélective sont en relation d'ACCESSIBILITÉ.
2. Les procès modèles sélectionnés par le sélecteur sont reportés dans la sélection, et les intervalles de procès sont en relation d'identité avec leur équivalent de l'itération.
3. Les procès modèles de la sélection sont en relation entre eux selon les mêmes règles que pour les itérations.

Nous en déduisons la règle 7 sur les relations d'identité entre les intervalles associés à l'itération et à la sélection. Les procès modèles sélectionnés sont identiques entre les deux séries, nous avons donc une relation d'identités entre chaque intervalle de procès.

¹. Cet emprunt amène plus largement à la question du rapprochement des modèles présentés à la partie ii.

Si

– il existe une sélection S de l'itération I

Alors

– il y a une relation d'accessibilité : $IP_{S_S} \text{ ACCESS } IP_{S_I}$

– il y a une relation d'identité entre les procès modèles

sélectionnés : $IP_S \text{ CO } IP_I$

Règle 7 – Règle d'identité des intervalles de procès associées à l'itération et à la sélection.

Afin de distinguer les sélections des autres structures aspectuo-temporelles, nous projetons les intervalles sur un second axe. L'exemple 94 présente une itération du procès *Je descendre chercher de l'eau* déclenchée par le calendrier *tous les jours*. Puis dans la seconde phrase est présent une sélection de l'itération, via le déclencheur fréquentiel *Quelquefois*. Il enrichit ainsi les procès sélectionnés du procès *un palefrenier nettoyer ses harnais*. La représentation aspectuo-temporelle des structures itératives est illustrée par la figure 64.

Tous les matins, je descendais chercher de l'eau dans la cour. Quelquefois, un palefrenier en casaque rouge nettoyait ses harnais près de la pompe.

EXEMPLE 94

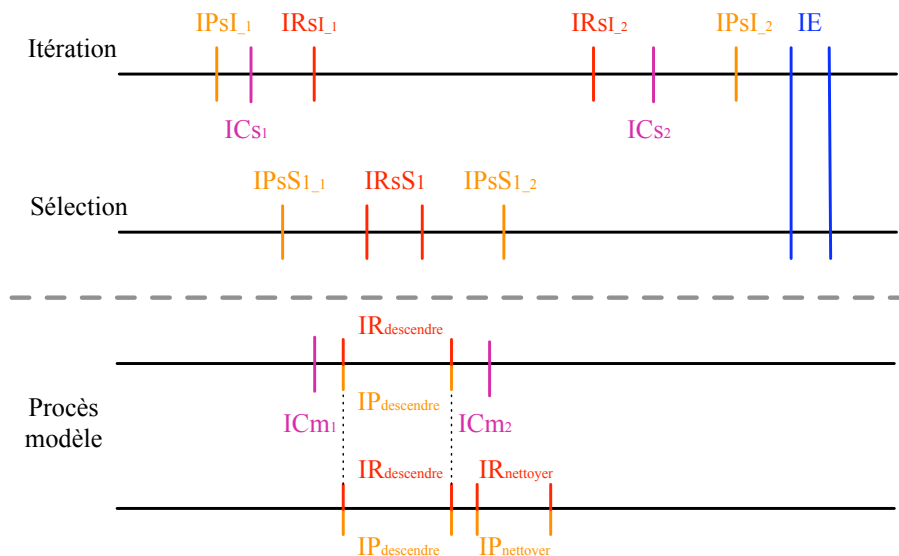


FIGURE 64 – Exemple de chronogramme faisant apparaître les procès modèles de l'itération et de la sélection. Notons que les lignes pointillées illustrent la relation d'identité des intervalles.

EN RÉSUMÉ

Nous avons présenté les principes du calcul aspectuel qui reposent sur les mêmes marqueurs et des règles similaires aux procès semelfactifs. Nous nous sommes intéressés d'une part aux règles permettant de déterminer dans quelles conditions la valeur aspectuelle d'un marqueur est affectée à la série itérative ou au procès modèle ; et d'autre part, nous avons proposé une représentation aspectuo-temporelle des sélections dans les termes de la *SdT*.

Les différentes règles présentées ci-dessus, ainsi que celles exposées en annexe B, sont intégrées dans le système *CAsTe* qui permet d'obtenir la représentation aspectuo-temporelle du texte que nous présentons au chapitre suivant.



LE SYSTÈME CASTE

Le présent chapitre présente l'implémentation du système *CAste*, pour Calcul Aspectuo-Temporel, qui permet d'obtenir la représentation aspectuo-temporelle du texte. *CAste* se fonde sur les règles de calcul exposées en annexe B et est semi-automatique afin de demander, si besoin, les informations manquantes ou de désambiguïser certains conflits. Nous exposons les principes sur lesquels reposent notre application, puis nous commentons deux extraits analysés.

SOMMAIRE

10.1	Description des informations aspectuo-temporelles	180
10.2	Principes de fonctionnement de <i>CAste</i>	183
10.3	Résultats commentés	186

INTRODUCTION

Le calcul aspectuo-temporel repose sur les marqueurs aspectuo-temporels qui déclenchent l'activation des règles. Il est donc nécessaire de pouvoir les identifier et y faire référence de façon exhaustive. Nous devons prendre en compte les marqueurs constituant la proposition :

- les verbes et leurs compléments ;
- les temps morphologiques ;
- les circonstanciels temporels ;
- les adverbes aspectuels.

Ainsi que :

- certaines constructions syntaxiques :
 - subordination ;
 - propositions coordonnées par des connecteurs temporels ou causaux ;
- la proximité des propositions ;
- les relations pragmatico-référentielles.

Notre système est circonscrit au calcul des structures aspectuo-temporelles. Il est nécessaire au préalable de spécifier l'ensemble des marqueurs et des relations nécessaires au calcul.

Le calcul temporel nécessite un traitement permettant de trouver les marqueurs présents dans le texte et de déterminer leurs valeurs. Notre système peut prendre indifféremment un texte annoter manuellement ou automatiquement, spécifiant les informations aspectuo-temporelles selon la description ci-après.

10.1 DESCRIPTION DES INFORMATIONS ASPECTUO-TEMPORELLES

Le système nécessite donc en entrée : d'une part, la structure de l'énoncé en paragraphes, en phrases et en propositions ; d'autre part, les informations temporelles utiles au calcul, notamment les structures itératives, les déclencheurs ou encore les circonstanciels temporels. Nous associons aux différentes structures composant un énoncé, des couples d'attributs et de valeurs contenant les traits sémantiques nécessaire au calcul, que nous décrivons ci-dessous. Le système calcule en sortie les relations existantes entre les différents intervalles associées aux procès et aux structures itératives, qui sont représentées par un triplet : (intervalle, relation, intervalle), par exemple (IPs, COÏNCIDE, IRs) ¹.

Les propositions

Identifiant

- id : identifiant des propositions, permettant le rattachement des différentes structures de l'énoncé comme le circonstant temporel ;

Structure syntaxique

- type : la proposition est soit indépendante, principale ou subordonnée ;
- si la proposition est une subordonnée temporelle, on renseigne les trois champs suivants :
 - directe : la subordonnée est-elle directe ou non ;
 - relation : la relation existante entre la subordonnée temporelle et la principale ;

Traits du verbe

- temps : le temps morphologique du verbe ("présent", "imparfait", "passé-composé", "plus-que-parfait", "passé-simple", "futur", "participe-passé", "participe-présent") ;

1. Les entrées et sorties du système *CAsTe* sont au format XML. Afin de valider les fichiers d'entrées, nous avons établi une DTD présentée en section D.1 page 247.

- typeProces : le type de procès est *état nécessaire, état contingent, activité, accomplissement ou achèvement*² ;
- performatif : le verbe est-il performatif ;

Structure itérative

- idSi : l'identifiant de la structure itérative à laquelle est rattachée la proposition.

Les circonstanciels temporels

Ils sont liés à une proposition :

Identifiants

- id : l'identifiant du circonstant dans l'énoncé ;
- proposition : l'identifiant correspondant à celui de la proposition à laquelle il est rattaché ;

Information syntaxique

- détaché : détermine si la position syntaxique est détachée ou intégrée ;

Caractéristiques

- type : le type du circonstant correspondant à la catégorisation explicitées dans le chapitre sur les règles de calcul (en + durée, pendant + durée, localisateur date incomplète, localisateur date complète) ;
- relation : la relation circonstancielle avec l'intervalle sur lequel porte le circonstant ;
- autonome : les circonstanciels autonomes correspondent aux dates complètes (en 1789), aux dates indéterminées (un jour, une fois) et aux subordinées temporelles ;
- valeur_temporelle : la valeur temporelle pour les circonstants déictiques ("passé", "présent", "futur").

Les déclencheurs

Les déclencheurs sont liés à une proposition mais également à une structure itérative :

Identifiants

- id : l'identifiant du déclencheur dans l'énoncé ;
- proposition : l'identifiant de la proposition à laquelle il est rattaché ;

2. Pour rappel, le type de procès est calculé à partir du lemme du verbe et de son environnement actanciel. Afin d'automatiser le calcul, nous avons constitué une ressource lexicale que nous présentons en annexe A.

- *structure_iterative* : identifiant de la structure itérative dont dépend la proposition ;

Caractéristiques

- *détaché* : détermine si la position syntaxique est détachée ou intégrée ;
- *type* : "calendaire", "fréquentiel", "quantificationnel", "événementiel", "conflit" ou "autre".

Les structures itératives

Les structures itératives sont indépendantes des paragraphes, des phrases ou encore des propositions. Le rattachement des propositions aux structures itératives (soit une itération, soit une sélection) se fait au niveau de la proposition, via l'identifiant de la structure itérative, noté *idSi*. Dans le cas de la sélection, elle est nécessairement rattachée à une autre structure itérative, dont elle vient sélectionner certains des itérés.

Identifiants

- *id* : identifiant de la structure itérative ;
- *déclencheur* : identifiant du déclencheur ;

Sélection

- *structure_iterative* : identifiant de la structure itérative à laquelle est rattachée la sélection.

Les relations

Les relations entre propositions sont principalement celles issues des connecteurs temporels. Ces marqueurs sont explicitement présents dans le texte. Nous retrouvons également les relations de discours telles que la causalité ou encore l'élaboration.

Identifiants

- *cible* : identifiant de la proposition cible ;
- *source* : identifiant de la proposition source ;

Caractéristiques

- *type* : "succession", "précédence", "simultanéité", "causalité", "élaboration"...

Le calcul des représentations aspectuo-temporelles repose sur le système expert JESS, qui signifie Java Expert System Shell³. JESS est un moteur de règles et un environnement de script fondé sur le système expert CLIPS, qui a été initialement développé par la NASA. Le paradigme utilisé par JESS repose sur l'application de manière continue d'un ensemble de règles sur un ensemble de faits.

En pratique, le système charge dans la base de fait les résultats issus du traitement précédent qui annote les marqueurs aspectuo-temporels (les verbes, les syntagmes nominaux objets, les compléments locatifs téliques, les temps morphologiques, les circonstants temporels et les connecteurs temporels) et la structure syntaxique. Puis, il calcule les différentes représentations, notamment les type de procès, les valeurs des temps morphologiques, les valeurs des circonstants temporels et les relations temporelles. La figure 65 illustre le fonctionnement global du système.

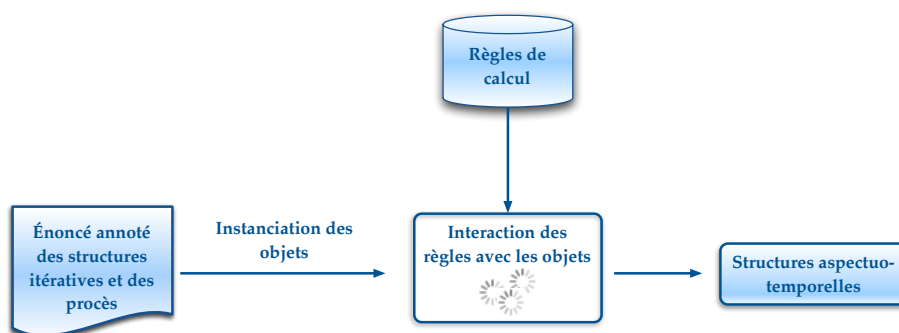


FIGURE 65 – Schéma du fonctionnement de notre calculateur de contraintes temporelles

Le cœur du système est constitué de JESS et des règles de calcul qui se déclenchent dès que leurs prémisses sont satisfaites. Ils simulent l'interaction entre les éléments considérés comme pertinents. Nous résumons les interactions du système sur la figure 66.

Le système *CAsTe* prévoit également une intervention de l'utilisateur sous forme d'une demande de la valeur d'un trait particulier, en cas d'ambiguïté. Actuellement, nous avons un seul cas pour le participe présent qui nécessite une intervention de l'utilisateur pour renseigner la valeur temporelle associée. La demande d'intervention d'un utilisateur se faisant au niveau des règles, il est envisageable de l'étendre à d'autres règles.

3. JESS fut mis au point par Ernest Friedman-Hill, et est accessible à l'adresse : www.jessrules.com.

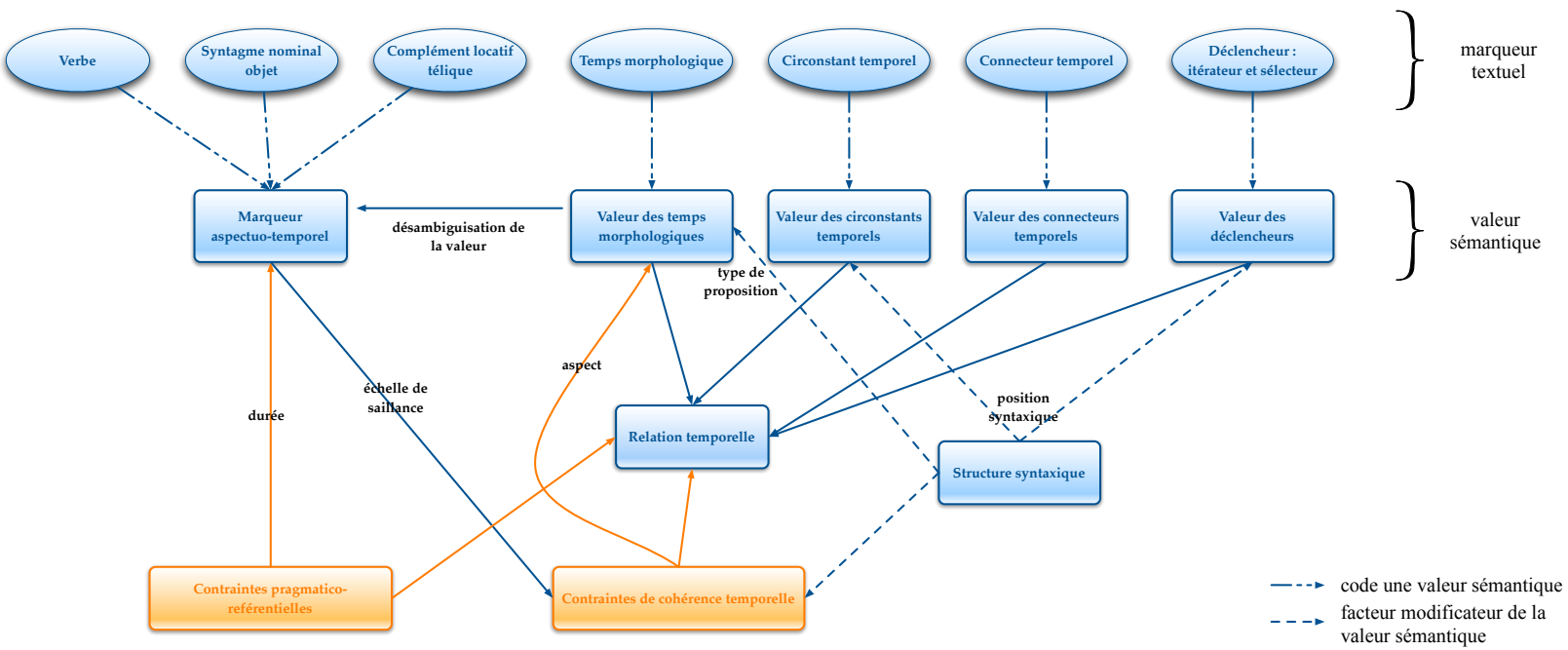


FIGURE 66 – Schéma des interactions des différentes valeurs sémantiques associées aux marqueurs textuels.

Notre système repose sur deux étapes en rendant prioritaires certaines règles. La première étape (en bleu sur le schéma des interactions ci-dessus) détermine les séries itératives et les procès associés à chaque proposition, et permet la création des intervalles nécessaires. Les relations temporelles produites concernent les relations entre intervalles issus de la même proposition, c'est-à-dire les contraintes entre les intervalles de procès, de références et circonstanciels, s'ils existent. Nous pouvons déjà déduire à cette étape certaines relations sur le texte par l'intermédiaire des règles sur les connecteurs et sur les subordinées. Par exemple, la présence du connecteur *et* entre deux propositions, nous permet de mettre en relation de *précédence* deux intervalles procès.

La seconde étape (en orange) calcule la cohérence temporelle par l'intermédiaire des contraintes liées à la dépendance contextuelle de l'intervalle de référence, en recherchant un intervalle antécédent selon l'échelle de saillance relative et la proximité relative des propositions (*cf.* figure 67), d'une part ; et aux propriétés de discontinuité et d'imbrication portées par les structures itératives, d'autre part. Puis nous déterminons la succession des procès après que les intervalles de référence ont été saturés.

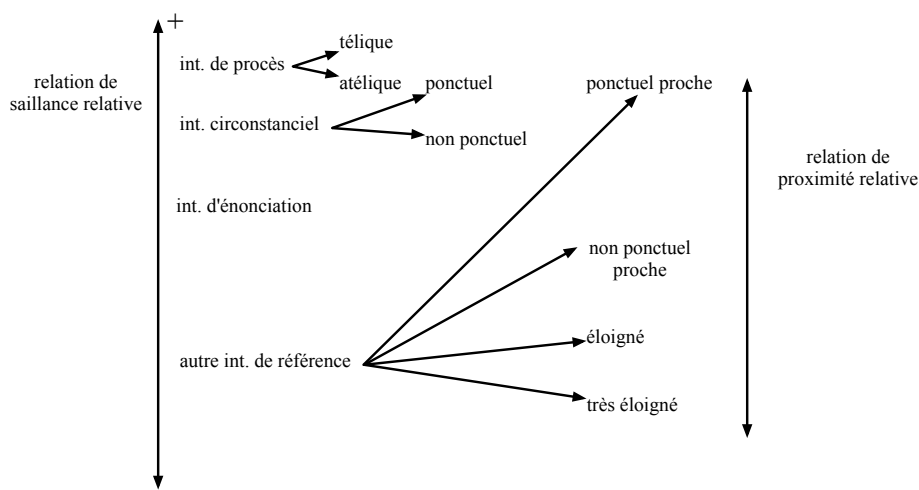


FIGURE 67 – Échelle de saillance relative pour le calcul du degré de saillance des intervalles d'après Gosselin (1996).

La lecture des relations temporelles étant difficile, les relations temporelles peuvent être visualisées sous forme d'un chronogramme. Nous avons repris une application développée par Y. Mathet en 2002 pour les relations entre procès non itérés que nous avons adaptée pour intégrer les phénomènes itératifs. Cette interface dynamique permet de faire glisser les bornes des intervalles en

fonction des contraintes calculées (cf. la figure 69 pour un aperçu des chronogrammes issus de l'application).

Dans une démarche de validation de nos résultats, cette application possède l'avantage de nous permettre de vérifier la véracité des relations calculées. En conséquence, nous pouvons réaliser un retour sur les règles implémentées, soit en les modifiant, soit en complétant le jeu de règles de nouvelles règles.

10.3 RÉSULTATS COMMENTÉS

Nous proposons de détailler deux exemples analysés extraits d'*Un cœur simple* de G. Flaubert.

Le premier extrait

*Tous les jeudis, des habitués venaient faire une partie de boston [P₁].
Félicité préparait d'avance les cartes et les chaufferettes [P₂]. Ils arrivaient
à huit heures bien juste [P₃] et se retiraient avant le coup de onze [P₄].*

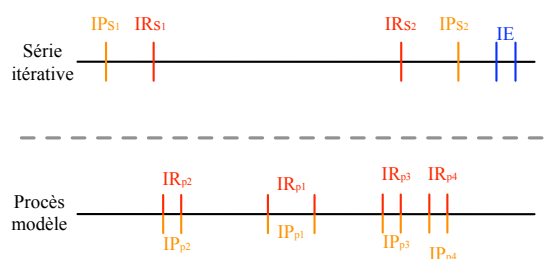
Un cœur simple, Gustave Flaubert

Le premier extrait d'*Un cœur simple* de Flaubert, que nous avons utilisé à plusieurs reprises dans ce manuscrit, est introduit par un déclencheur calendaire, *Tous les jeudis*, en position détachée. Nous avons donc la création d'une série itérative composée de quatre procès modèles (P₁, P₂, P₃, P₄).

L'extrait étant à l'imparfait, la série itérative est affectée d'une visée d'inaccomplie tandis que chaque procès modèle possède une visée aoristique.

Nous avons donc les relations suivantes :

- IP_S ANT IE
- IP_S RE IR_S
- IP_{P₁} CO IR_{P₁}
- IP_{P₂} CO IR_{P₂}
- IP_{P₃} CO IR_{P₃}
- IP_{P₄} CO IR_{P₄}



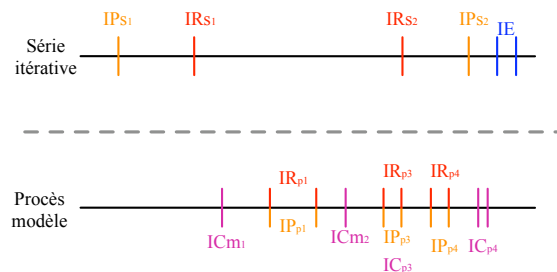
Il existe également trois circonstants temporels qui déclenchent chacun la production d'un intervalle circonstanciel :

- le déclencheur calendaire *tous les jeudis* qui produit à la fois un intervalle circonstanciel sur la série itérative et sur les procès modèles ;

- le circonstant à *huit heures bien juste* présent dans P_3 est un circonstant localisateur ponctuel;
- le circonstant *avant le coup de onze* présent dans P_4 est un circonstant localisateur à valeur d'antériorité.

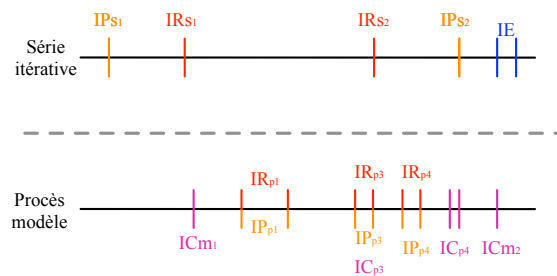
Nous obtenons donc les relations suivantes :

- IC_S ACCESS IR_S
- IC_m ACCESS IR_{P_1}
- IC_{P_3} CO IP_{P_3}
- IP_{P_4} ANT IC_{P_4}



De plus, les circonstants à *huit heures bien juste* et *avant le coup de onze* sont des moments de la journée du jeudi. Il existe donc une relation de recouvrement entre l'intervalle circonstanciel du jeudi modèle (IC_m) et de chaque intervalle circonstanciel.

- IC_m RE IC_{P_3}
- IC_m RE IC_{P_4}
- IC_{P_3} ANT IC_{P_4}



Il existe un connecteur de précédence (*d'avance*) entre les propositions P_1 et P_2 . La relation temporelle produite est la suivante :

- IP_{P_2} PREC IP_{P_1}

La critère de succession permet d'inférer les relations suivantes :

- IP_{P_2} PREC IP_{P_3}
- IP_{P_3} PREC IP_{P_4}

Il existe une identité de procès entre, d'une part, P_3 et le commencement de P_1 et d'autre part, entre P_4 et l'achèvement de P_1 , que nous ne calculons pas. Ces relations nécessiteraient une description plus approfondie des différentes phases d'un procès duratif.

L'exemple 95 de la page suivante illustre le résultat de l'analyse aspectuo-temporelle au format XML.

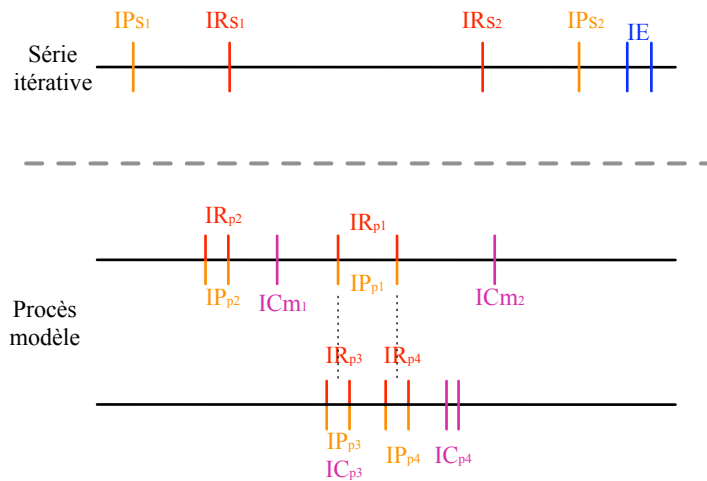


FIGURE 68 – Chronogramme du texte de Flaubert. Pour rappel, les procès sont : P₁ venir faire, P₂ préparer, P₃ arriver et P₄ se retirer.

```

<?xml version="1.0" encoding="utf-8"?>
<enonce valeur="IE">
  <contrainte cible="IPp4" source="IPp3" relation="PREC"/>
  <contrainte cible="IPp1" source="IPp2" relation="PREC"/>
  <contrainte cible="ICn2p3" source="ICm1" relation="RE"/>
  <contrainte cible="ICn2p3" source="ICm1" relation="RE"/>
  <iteration id="1">
    <contrainte cible="IRsI1" source="IPsI1" relation="RE"/>
    <contrainte cible="IE" source="IRsI1" relation="ANT"/>
    <contrainte cible="IRsI1" source="ICsI1" relation="ACCESS"/>
    <proces id="1" type="activite" label="Tous les jeudis,des habitués
venaient faire une partie de boston.">
      <contrainte cible="IRp1" source="IPp1" relation="CO"/>
      <contrainte cible="IRp1" source="ICm1" relation="ACCESS"/>
    </proces>
    <proces id="2" type="accomplissement" label="Félicité préparait d'avance
les cartes et les chaufferettes.">
      <contrainte cible="IRp2" source="IPp2" relation="CO"/>
    </proces>
    <proces id="3" type="activite" label="Ils arrivaient à huit heures bien
juste,">
      <contrainte cible="IRp3" source="IPp3" relation="CO"/>
      <contrainte cible="IPp3" source="ICn2p3" relation="CO"/>
    </proces>
    <proces id="4" type="activite" label="et se tiraient avant le coup de
onze.">
      <contrainte cible="IRp4" source="IPp4" relation="CO"/>
      <contrainte source="IPp4" cible="ICn3p4" relation="ANT"/>
    </proces>
  </iteration>
</enonce>

```

EXEMPLE 95 – Représentation XML de l'analyse aspectuo-temporelle de l'extrait de Flaubert.

Le second extrait

Notre second extrait d'*Un cœur simple* de Flaubert illustre le phénomène de discontinuité textuelle que nous avons présenté précédemment.

Quand le temps était clair [P₁], on s'en allait de bonne heure à la ferme de Geffosses [P₂].

La cour est en pente, la maison dans le milieu [P₃]; et la mer, au loin, apparaît comme une tache grise [P₄].

Félicité retirait de son cabas des tranches de viande froide [P₅], et on déjeunait dans un appartement [P₆] faisant suite à la laiterie [P₇]. Il était le seul reste d'une habitation de plaisance, maintenant disparue [P₈]. Le papier de la muraille en lambeaux tremblait aux courants d'air [P₉]. Mme Aubain penchait son front, accablée de souvenirs [P₁₀], les enfants n'osaient plus parler [P₁₁]. "Mais jouez donc !" disait-elle [P₁₂]; ils dé-campaient [P₁₃]. »

Un cœur simple, Gustave Flaubert

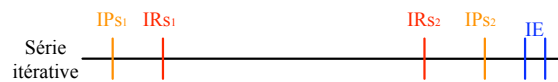
L'extrait est introduit par un déclencheur événementiel au travers de la subordonnée temporelle, *Quand le temps était clair*, en position détachée. Remarquons que les procès [P₃] et [P₄] n'appartiennent pas à la série itérative du fait qu'ils soient au présent, et illustrent par la même le phénomène de discontinuité textuelle. Nous avons donc la création d'une série itérative et de onze procès modèle, tandis que les procès [P₃] et [P₄] sont des procès « simples ».

Les procès de la série sont à l'imparfait, la série est donc affectée d'une visée d'inaccompli tandis que chaque procès modèle possède une visée aoristique.

Nous obtenons donc les relations suivantes sur la série itérative :

– IR_S ANT IE

– IP_S RE IR_S

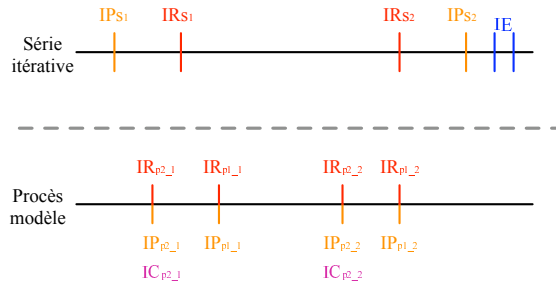


et les relations suivantes sur les procès modèles :

– IP_{P_i} CO IR_{P_i}, $i \in [1..13]$, sauf pour $i = 3,4$

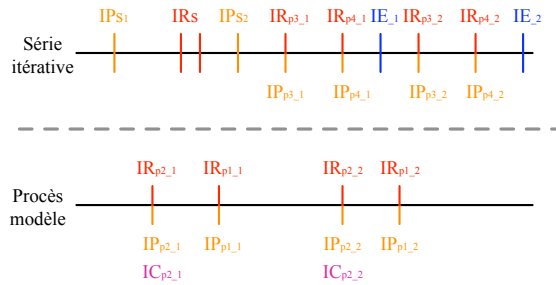
De plus, le déclencheur est une subordonnée temporelle directe, nous avons donc également la création d'un intervalle circonstanciel qui est en relation avec l'intervalle de référence de la subordonnée, [P₁], et avec l'intervalle de référence de la principale, [P₂] :

- IC_{P_2} SIMUL IR_{P_1}
- IC_{P_2} CO IR_{P_2}



En ce qui concerne les deux procès n'appartenant pas à la série itérative ([P₃] et [P₄]), nous avons deux procès au présent qui sont donc en relation de simultanéité avec l'intervalle d'énonciation. De plus, ils ne posent pas de contraintes vis-à-vis de la série. Nous obtenons donc les relations suivantes :

- IR_{P_3} SIMUL IE
- IP_{P_3} RE IR_{P_3}
- IR_{P_4} SIMUL IE
- IP_{P_4} RE IR_{P_4}



La critère de succession permet d'inférer les relations suivantes :

- IP_{P_2} PREC IP_{P_5}
- IP_{P_5} PREC IP_{P_6}
- IP_{P_6} PREC IP_{P_7}
- IP_{P_7} PREC IP_{P_8}
- IP_{P_8} PREC IP_{P_9}
- IP_{P_9} PREC $IP_{P_{10}}$
- $IP_{P_{10}}$ PREC $IP_{P_{11}}$
- $IP_{P_{11}}$ PREC $IP_{P_{12}}$
- $IP_{P_{12}}$ PREC $IP_{P_{13}}$

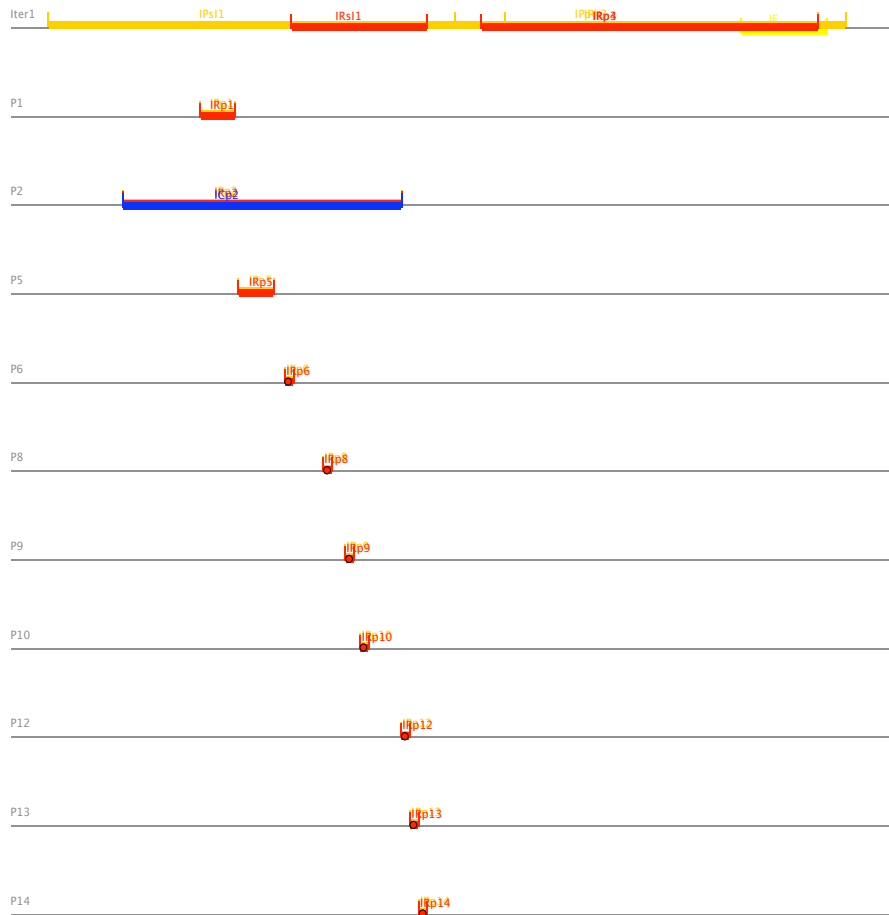


FIGURE 69 – Illustration du chronogramme du second extrait de Flaubert résultant de notre application de visualisation des relations entre intervalles.

EN RÉSUMÉ

Nous venons de présenter les principes de fonctionnement de notre calcul aspectuo-temporel qui repose sur le système expert JESS. Notre système est circonscrit au calcul des structures aspectuo-temporelles. En effet, il est nécessaire au préalable de spécifier l'ensemble des marqueurs et des relations nécessaires au calcul. Nous avons fait ce choix pour deux raisons. D'une part, nous souhaitons nous affranchir de la précision des résultats de notre analyseur automatique en nous offrant la possibilité de réaliser des annotations manuelles. D'autre part, certaines informations aspectuo-temporelles sont dépendantes du contexte, et nécessitent des ressources encyclopédiques conséquentes pour réaliser une désambiguïsation automatique, nous confortant ainsi dans l'idée d'intégrer une annotation manuelle intermédiaire entre l'analyse automatique des structures itératives et du calcul aspectuo-temporel.

Cette première implémentation nous a permis de tester notre approche théorique fondé sur une analyse de la *SdT*, en particulier sur le passage du calcul des procès sémelefactifs aux procès itératifs, présenté au chapitre précédent, et qui nous a amené à formaliser des règles de calcul (*cf.* annexe B). Notre système est en cours d'implémentation : il n'intègre pas encore les calculs aspectuo-temporel des structures itératives de type « sélection ». Par ailleurs, les phases de test se sont limités à une quinzaine d'énoncés annotés manuellement.

Notre prochaine étape consistera à intégrer le calcul des sélections, et plus largement le calcul des procès non itératifs à notre système. Par ailleurs, nos premières expérimentations donnent les résultats attendus, mais du fait de la taille du corpus ils ne sont pas probant. Il nous reste donc à faire une évaluation sur un corpus plus large.



UN FORMALISME XML POUR LES STRUCTURES ITÉRATIVES

Dans les chapitres précédents, nous avons développé les annotations manuelles et automatiques des phénomènes itératifs que nous avons réalisé. Dans la perspective d'unifier ces annotations, nous nous sommes intéressés à `TIMEML` qui est un formalisme XML permettant de structurer les connaissances temporelles contenu dans un texte. À cette fin, nous en exposons les limites par rapport à notre définition des phénomènes itératifs, et les différents amendements réalisés pour pouvoir les y intégrer.

SOMMAIRE

11.1	Qu'est-ce que <code>TIMEML</code>	193
11.1.1	Les entités de <code>TIMEML</code>	194
11.1.2	Les relations de <code>TIMEML</code>	196
11.2	Les phénomènes itératifs dans <code>TIMEML</code>	197
11.2.1	Les structures itératives	197
11.2.2	Les déclencheurs	198
11.3	<code>TIMEML ++</code> : traitement des phénomènes itératifs	200
11.3.1	Les structures itératives	200
11.3.2	Les déclencheurs	201

11.1 QU'EST-CE QUE `TIMEML`

`TIMEML` est le fruit d'un groupe de travail dirigé par James Pustejovsky, dans (Pustejovsky *et al.*, 2003), qui vise à standardiser les annotations sémantiques reliées à la temporalité dans un texte en langage naturel. Originellement, il fut conçu pour répondre à quatre problèmes liés à l'annotation des événements et des expressions temporelles :

1. Identifier un événement et l'ancrer dans le temps ;
2. Ordonner les événements les uns par-rapport aux autres (Propriété d'ordonnancement lexical ou discursive) ;
3. Reasonner contextuellement avec les expressions temporelles (telles que « la semaine dernière » et « deux semaines avant ») ;

4. Raisonner sur la persistance des événements (combien dure un événement ou le résultat du dernier événement).

Le guide d'annotation issu des premiers travaux est tourné vers l'anglais, mais les éléments du standard sont *a priori* indépendants de la langue considérée. Il existe notamment des travaux sur l'italien de Bertinetto (2003) ou encore sur le coréen de Shin *et al.* (1997). De plus, un guide d'annotation pour le français a été élaboré (Bittar, 2010). Depuis 2007, dans le cadre de l'ISO TC 37/SC4 *Semantic Annotation Framework (SemAF)*, TimeML a été adopté pour établir une spécification formelle appelée ISO-TimeML (Lee *et al.*, 2007), pour le balisage des informations temporelles en langues naturelles.

11.1.1.1 Les entités de TIMEML

Actuellement, afin de couvrir l'essentiel des informations temporelles que l'on peut associer à un texte, le standard traite de trois sortes d'entités annotables : les adverbiaux (objet TIMEX₃), les éventualités (EVENT), et les signaux (SIGNAL).

Les TIMEX ont plusieurs attributs, comme le type (date, heure, durée), la valeur absolue correspondant à la localisation temporelle (*cf.* exemple 96a), le temps d'ancrage, pour les localisations qui sont calculées relativement à une autre (*cf.* exemple 96b) et la quantité ou la fréquence, pour des localisations qui correspondent à plus d'un moment dans le temps (*cf.* exemple 96c).

```
(a) <TIMEX tid="1" type="DATE" value="2003-06-21"> 21 Juin  
    2003</TIMEX>  
(b) John est parti  
    <TIMEX tid="2" type="DURATION" value="P2D">  
    deux jours</TIMEX>après l'attaque  
(c) <TIMEX tid="3" type="SET" value="XXXX-04-4"  
    quant="EVERY" >tous les jeudis d'avril</TIMEX>
```

EXEMPLE 96 – Exemples de trois configurations possibles de TIMEX

Les EVENT correspondent aux événements d'un énoncé. Ici, le terme *événement* fait référence au procès dans la terminologie employé dans la *SdT*. On retrouve cette acception dans différents travaux sous la notion d'éventualité comme Vendler (1967), ou encore procès que nous reprenons à notre compte. Ces événements sont marqués textuellement soit par un verbe soit par un nom. Ils ont pour attribut une classe parmi : occurrence (la plupart des événements), état, action ou état intensionnel (ceux-ci supposent un second événement qui

ne se réalise pas nécessairement, comme dans *je voudrais qu'il vienne*), action d'énonciation ou de perception (dire, révéler, entendre, voir, ...) et finalement les événements aspectuels (commencer, finir de, continuer, ...). Le formalisme permet de distinguer l'événement lui-même de sa réalisation en attribuant un identifiant à l'événement et à l'instance. L'exemple 97 présente deux événements de la classe OCCURENCE l'un représenté textuellement par un nom, l'autre par un verbe.

```
John
<EVENT eid="e1" eiid="ei1" class="OCCURRENCE"...>
est parti</EVENT>
deux jours après
<EVENT eid="e2" eiid="ei2" class="OCCURRENCE"...>
l'attaque</EVENT>
```

EXEMPLE 97 – EVENT : « John est parti deux jours après l'attaque »

L'entité EVENT possède également les attributs suivants : le temps grammatical, la forme (nom, verbe, adjectif), l'aspect (Progressive, Imperfective, Perfective, ...). L'exemple 98 complète les différents attributs de l'exemple 97. On remarquera que pour l'événement *l'attaque*, marqué par un nom, les attributs *temps* et *aspect* ont des valeurs nulles.

```
John
<EVENT eid="e1" class="OCCURRENCE" eiid="ei1" tense="PAST"
aspect="NONE" pos="VERB">est parti</EVENT>
deux jours après
<EVENT eid="e2" eiid="ei2" tense="NONE" aspect="NONE"
pos="NOUN">l'attaque</EVENT>
```

EXEMPLE 98 – EVENT : « John est parti deux jours après l'attaque »

Enfin, la balise SIGNAL sert à identifier une forme (généralement une préposition) marquant une relation temporelle entre événements et localisations temporelles.

11.1.2 Les relations de TIMEML

Le standard prévoit, en plus des entités, la formalisation des relations existantes entre ces entités. Ces relations sont de trois ordres : temporelles (TLINK), aspectuelles (ALINK), et subordonnées (SLINK).

Les TLINKS ou TEMPORAL LINK représentent les relations temporelles existants entre des événements, des marqueurs temporels ou entre l'un et l'autre. Ces relations correspondent à celles des travaux d'Allen (1983) sur les intervalles, entre autres la *succession*, l'*antériorité* et le *recouvrement*.

```
Bill
<EVENT eid="e1" eiid="ei1" class="I_STATE" pos="VERB"
tense="PRESENT" aspect="NONE" polarity="POS">veut</EVENT>
<EVENT eid="e2" eiid="ei2" class="OCCURRENCE" pos="VERB"
aspect="NONE" tense="NONE" vform="INFINITIVE" polarity="POS">
enseigner</EVENT>
<TIMEX3 tid="t1" type="SET" value="XXXX-WXX-1"> le lundi</TIMEX3>
<TLINK eventInstanceID="ei2" relatedToTime="t1"
relType="IS_INCLUDED"/>
<SLINK eventInstanceID="ei1" relType="INTENSIONAL"
subordinatedEventInstance="ei2"/>
```

EXEMPLE 99 – TLINK et SLINK : « Bill veut enseigner le lundi »

Finalement, les relations SLINK sont utilisées pour introduire des relations entre deux instances d'événements, et les ALINKS représentent les relations entre les verbes aspectuels et les événements auxquels ils réfèrent.

```
Le bateau
<EVENT eid="e1" eiid="ei1" class="ASPECTUAL" pos="VERB"
tense="PAST" aspect="NONE" polarity="POS"> commença </EVENT> à
<EVENT eid="e2" eiid="ei2" class="OCCURRENCE" pos="VERB"
tense="NONE" aspect="NONE" vform="INFINITIVE" polarity="POS">
couler </EVENT>
<ALINK eventInstanceID="ei1" relType="INITIATES"
relatedToEventInstance="ei2"/>
```

EXEMPLE 100 – ALINK : « Le bateau commença à couler »

11.2 LES PHÉNOMÈNES ITÉRATIFS DANS TIMEML

Afin de pouvoir représenter les phénomènes itératifs tels que nous les avons définis, nous avons besoin de pouvoir :

- représenter les structures itératives, c'est-à-dire les itérations et les sélections ;
- représenter l'ensemble des itérateurs et sélecteurs ;
- représenter le procès ;
- et distinguer le procès global du procès modèle, en particulier pour des raisons de visées aspectuelles différentes.

11.2.1 Les structures itératives

Le standard permet de représenter des itérations en se manifestant au niveau d'un événement, c'est-à-dire via l'entité EVENT en renseignant l'attribut CARDINALITY (cf. exemple 101a). Il est également possible, à notre connaissance, d'exprimer la cardinalité des événements en créant toutes les entités. Par exemple, pour l'événement *enseigner deux fois*, le formalisme permet de créer deux entités de type EVENT (cf. exemple 101b).

```
(a) John a enseigné 150 fois l'an passé
<EVENT eiid="ei1" eventID="e1" tense="PAST" aspect="NONE"
pos="VERB" signalID="s1" cardinality="150" />

(b) John a enseigné deux fois lundi
<EVENT eiid="ei1" eventID="e1" tense="PAST" aspect="NONE"
pos="VERB" signalID="s1" />
<EVENT eiid="ei2" eventID="e1" tense="PAST" aspect="NONE"
pos="VERB" signalID="s1" />
```

EXEMPLE 101 – La représentation des itérations dans TIMEML.

Le standard ne distingue cependant pas l'instance d'un événement (c'est-à-dire le procès modèle) de l'instance d'une itération (c'est-à-dire la série itérative), hormis par l'attribut *cardinality*. Or cette distinction est primordiale comme nous avons pu le voir au chapitre 3. En effet, nous avons montré que dans certains cas, en particulier celui des phénomènes itératifs, il existait une double visée aspectuelle (cf. exemple 102). De plus, dans le cas d'une anaphore événementielle, il faut pouvoir déterminer si le rattachement se fait au niveau la série ou du procès modèle.

John enseignait le lundi et le mardi

VISÉE SÉRIE : inaccompli

VISÉE MODÈLE : aoristique

EXEMPLE 102

En ce qui concerne les sélections, le standard (*cf. Saurí et al. (2006, page 46)*) permet de réaliser une sous-série d'événement en utilisant les cardinalités respectives des entités `EVENT` et une relation `TLINK` de type `INCLUDES` ou `IS_INCLUDED` (*cf. exemple 103*).

Les policiers ont cherché dans les meurtres de 14 femmes. Dans six affaires, les suspects ont déjà été arrêté.

EXEMPLE 103 – Les deux événements sont créés pour chaque ensemble : le premier annote les meurtres, avec une cardinalité de 14, et le second annote les affaires, qui dans ce contexte fait référence aux meurtres, avec une cardinalité de 6. Les deux événements seront reliés par une relation temporelle de type `IS_INCLUDED` (ou `INCLUDES` selon la direction).

11.2.2 Les déclencheurs

Le second point à prendre en compte est la représentation des déclencheurs. Actuellement, en se basant sur la classification du modèle MOPI que nous avons présenté au chapitre 5, le standard permet uniquement l'annotation des « quantificateurs » et des « calendaires ». Elle se fait respectivement par l'entité `EVENT` (*cf. exemple 101*), et par l'entité `TIMEX`, avec l'attribut `TYPE` positionné à `SET`, permettant de représenter l'ensemble des unités temporelles décrivant l'expression temporelle (*cf. exemple 96c*). Il faut remarquer cependant que le formalisme de représentation des expressions temporelles n'est pas complet. Il ne permet pas de couvrir des expressions telles que *Un mardi sur deux, tous les 3èmes mardis du mois, trois jours ouvrables, ou encore le 4ème des 6 jours de campagne*. Rappelons qu'actuellement, la représentation des déclencheurs se fait au sein de l'entité `TIMEX`, et qu'il s'agit de décrire la valeur de l'expression par un patron, dans l'exemple 105, « pendant trois heures » est représentée par « *PT3H* ».

Par ailleurs, la représentation des expressions temporelles de type `SET` ne distingue pas l'ensemble (E) de ses éléments ($x \in E$). A priori, cette indistin-

tion provient de la sous-spécification des relations. En fonction des entités reliées, nous avons une lecture sur un élément (x) (cf. l'exemple 104a) ou sur l'ensemble (E) (cf. l'exemple 104b).

- (a) John enseigna *pendant 3 heures* tous les jeudis.
 (b) John enseigna tous les jeudis *pendant 10 ans*.

EXEMPLE 104

Enfin, le standard, par choix des concepteurs, ne permet pas de prendre en compte les adverbes de fréquences (*rarement, souvent...*) ou bien les conjonctions de subordination (*quand, lorsque, à chaque fois que...*), c'est-à-dire les itérateurs de types « fréquentiels » et « événementiels » (cf. Ferro *et al.* (2005)).

Bilan

Pour conclure, dans la version actuelle de TIMEML, un énoncé itératif est représenté comme dans l'exemple 105. L'itération est portée à la fois par l'entité TIMEX et l'instance de l'événement avec l'attribut CARDINALITY. Les relations temporelles annotées lient l'expression « pendant 3 heures » à l'instance de l'événement, par une relation de simultanéité, d'une part; et à l'expression « tous les jeudis », déclencheur de l'itération, par une relation d'inclusion, d'autre part. La figure 70 schématise les entités et les relations mises en œuvre.

```
John
<EVENT eid="e1" eiid="ei1" cardinality="EVERY">
enseigne</EVENT>
<TIMEX tid="t1" type="DURATION" value="PT3H"
anchorTimeID="t2">pendant 3 heures</TIMEX>
<TIMEX tid="t2" type="SET" quant="EVERY"
value="WXX-XXXX-4">tous les jeudis</TIMEX>,
<TLINK eventInstanceID="ei1" relatedToTime="t1"
relType="SIMULTANEOUS"/>
<TLINK timeID="t1" relatedToTime="t2" relType="IS_INCLUDED"/>
puis il
<EVENT eid="e2" eiid="ei2" cardinality="EVERY">va
manger</EVENT> au RU.
<TLINK eventInstanceID="ei1" relatedToEvent="ei2"
relType="BEFORE"/>
```

EXEMPLE 105 – « John enseigne pendant 3 heures tous les jeudis puis il va manger au RU. »

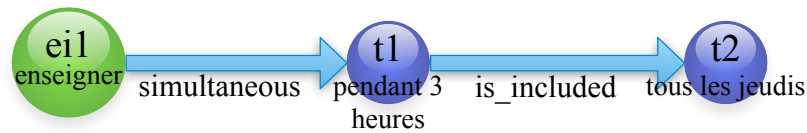


FIGURE 70 – Représentation des relations entre les entités formalisant une itération dans le standard TIMEML.

11.3 TIMEML ++ : TRAITEMENT DES PHÉNOMÈNES ITÉRATIFS

Nous allons aborder dans cette section l’adaptation du standard TIMEML aux phénomènes itératifs. Nous exposons les modifications apportées au modèle pour solutionner les limites exposées précédemment. Nous dérogeons aux recommandations proposées dans Lee *et al.* (2007, page 46), reportées ci-dessous, qui nous semblent toutefois difficile à suivre de par la nature des phénomènes étudiés. En effet, les phénomènes itératifs ne reposent pas uniquement sur des marqueurs de surfaces mais également sur leur interprétation en contexte comme dans le cas des conflits pragmatico-référentiels (*cf.* la section intitulée *Les déclencheurs « conflits »*).

ISO-TimeML is conceived as a highly surface-based language, in the sense that it does not aim at annotating meaning but at providing a way to normalize temporally relevant expressions. The annotators for each language are encouraged to develop language-specific ISO-TimeML specs having that in mind.

Nous décrivons en annexe D (*cf.* la section D.2 page 249) la syntaxe complète associée à l’adaptation de TIMEML que nous présentons ci-après.

11.3.1 Les structures itératives

Afin de représenter les structures itératives telles que nous les avons définies dans ce manuscrit, nous proposons une entité STRUCTURE à laquelle nous intégrons un attribut obligatoire TYPE, permettant de réaliser la distinction entre l’itération et la sélection. La valeur de cet attribut est soit « iteration », soit « selection ». Pour rendre compte de l’ensemble des informations nécessaires, nous avons les attributs suivants :

- l’attribut SCID attribue un identifiant à l’entité STRUCTURE ;
- l’attribut TYPE permet distinguer l’itération de la sélection ;
- l’attribut TRIGGERID permet d’identifier le déclencheur associé à la structure ;

- l'attribut `ASPECT` renseigne la valeur aspectuelle de la structure (aoristique, accompli, inaccompli et prospectif) ;
- l'attribut `STRUCTUREINSTANCEID` permet de représenter les liens entre deux structures itératives, notamment dans les cas d'imbrications d'itérations et de sélection.

Nous avons montré au sein de ce manuscrit qu'une structure itérative est composée nécessairement d'événements. Nous conservons donc l'entité `EVENT`, et ses couples d'attributs et de valeurs, et nous ajoutons l'attribut `STRUCTUREINSTANCEID` qui permet d'identifier la structure itérative dont dépend l'événement.

L'exemple 106 illustre la représentation d'une itération dans le cadre de cette extension, où nous retrouvons le déclencheur de l'itération *tous les jeudis*, et l'événement itéré *venir faire une partie de boston*, relié par une relation de simultanéité. La seconde entité de type `STRUCTURE` représente une sélection déclenchée par *Souvent* venant sélectionner certains des itérés de l'itération.

```
<TRIGGER tid="t1" value="TousLesN(Jeudi)" type="SET">
Tous les jeudis</TRIGGER>, des habitués
<EVENT eid="e1" eiid="ei1" aspect="aoristique"
structureInstanceID="sc1">venaient faire</EVENT>
une partie de boston.
<SCHEMA type="iteration" scid="sc1" triggerID="t1"
aspect="inaccompli"/>
<TLINK eventInstanceID="ei1" relatedToTime="t1"
relType="SIMULTANEAOUS"/> <TRIGGER trid="tr2"
value="often">Souvent</TRIGGER>, l'un des habitués<EVENT
eid="e5" eiid="ei5" structureInstanceID="sc1"
aspect="aoristique">apportait</EVENT>le repas.
<STRUCTURE type="selection" scid="sc2" eventID="e5"
triggerID="tr2" aspect="inaccompli"/>
```

EXEMPLE 106 – « Tous les jeudis, des habitués venaient faire une partie de boston. [...] Souvent l'un des habitués apportait le repas. »

11.3.2 Les déclencheurs

Comme nous l'avons vu précédemment, le standard ne couvre que partiellement les déclencheurs tels que nous avons pu les définir au sein de ce manuscrit. En effet, le standard permet de représenter uniquement les itérateurs issus de la classe des calendaires, les concepteurs ayant fait le choix de

ne pas annoter les adverbes de fréquences ou encore les conjonctions de subordinations. Afin de représenter l'ensemble des déclencheurs, nous proposons une nouvelle entité pour en rendre compte.

Nous représentons les différents déclencheurs par l'entité TRIGGER auquel est adjoit un ensemble de couples d'attributs et valeurs pour notamment distinguer les différents types de déclencheurs. L'entité est donc composée d'un identifiant TRID, d'un attribut TYPE et d'un attribut VALUE contenant la valeur portée par le déclencheur analysé.

Les déclencheurs « aspectuels »

Les déclencheurs de types aspectuels, tels que *encore*, présupposent l'existence d'autres instances de l'événement. Nous proposons de spécifier le nombre d'itérés au moyen d'un patron simple, correspondant aux nombres minimums de répétitions présupposées suivis du symbole "+". Dans l'exemple 107, il y a au minimum deux instances de *casser une vitre*.

```
Il a <TRIGGER trid="t1" type="ASP"
value="2+">encore</TRIGGER> cassé une vitre
```

EXEMPLE 107

Les déclencheurs « conflits »

Nous rappelons que les déclencheurs de type « conflits » mettent en jeu deux marqueurs, soit des compléments circonstanciels (*cf.* exemple 108a) soit un complément circonstanciel et un événement ponctuel (*cf.* exemple 108b).

- (a) Depuis deux mois, Adrien mange en 10 minutes.
- (b) Longtemps, je me suis couché de bonne heure.

EXEMPLE 108

Nous distinguons les deux cas au sein de TRIGGER de type CONFLICT. Dans le premiers cas, il est nécessaire de renseigner les deux attributs (TIMEXID1 et TIMEXID2) identifiants les circonstanciels représentés par l'entité TIMEX. Dans le second cas, nous identifions les marqueurs entrant en jeu, c'est-à-dire un circonstanciel (toujours marqué par une entité TIMEX) et l'événement

ponctuel (marqué par l'entité EVENT), à l'aide des attributs TIMEXID et EVENTID, respectivement. L'exemple 109 illustre ces représentations XML.

- ```
(a) <TIMEX tid="t1" type="DURATION" ...>Depuis deux mois</TIMEX>,
 Adrien mange <TIMEX tid="t2" type="DURATION" ...>en 10
 minutes</TIMEX>.
 <TRIGGER trid="c1" value="2+" type="CONFLICT" timexID1="t1"
 timexID2="t2" />

(b) <TIMEX tid="t1" type="DURATION" ...>Longtemps</TIMEX>, je
 <EVENT eid="e1" eiid="ei1" class="OCCURRENCE" ...> me suis
 couché de bonne heure
 <TRIGGER trid="c2" value="2+" type="CONFLICT" timexId="t1"
 eventId="e1" />
```

EXEMPLE 109

#### *Les déclencheurs « quantificationnels »*

Les déclencheurs de types quantificationnels, pour rappel, correspondent à des syntagmes tels que « à trois reprises », « 5 fois », « quelques fois », ... L'attribut TYPE se voit affecté de la valeur QUANT, et l'attribut VALUE contient la valeur de la répétition portée par le quantifieur (cf. exemple 110). Dans les cas où la sémantique associée au quantifieur n'est pas précise comme pour « plusieurs fois », la valeur VALUE doit être : 2+, comme dans le cas des déclencheurs aspectuels <sup>1</sup>.

- ```
(a) Madame Aubain tomba<TRIGGER trid="n1" value="3"
    type="QUANT">à trois reprises</TRIGGER>
(b) Nos <TRIGGER trid="n2" value="2" type="QUANT">2
    voyages</TRIGGER> à Hawaii furent merveilleux.
```

EXEMPLE 110

Les déclencheurs « événementiels »

Les déclencheurs événementiels correspondent aux conjonctions introduisant des subordonnées temporelles telles que « lorsque », « à chaque fois que »,

1. Nous préconisons de manière générale de positionner l'attribut VALUE à 2+ lorsque la quantité d'itérée n'est pas renseignée.

« dès que », ... Nous attribuons la valeur EV à l'entité TYPE tandis que l'attribut VALUE spécifie la relation temporelle existante entre la subordonnée introduite par le marqueur événementiel et la proposition principale. L'exemple 111 illustre une représentation du marqueur *lorsque* qui spécifie une relation de simultanéité entre les deux propositions. Remarquons qu'il peut exister une contradiction apparente entre la valeur de l'introducteur et la principale due au fait le procès principal peut dans certains cas, comme ici avec « John enseigne », être en fait une partie seulement d'un scénario itéré plus large inféré à partir de ce procès. Par ailleurs, le syntagme *d'abord* apporte une information supplémentaire sur la relation existante entre les événements sans réfuter la relation portée par le marqueur.

```

<TRIGGER trid="ev1" value="SIMULTANEOUS" type="EV">
Lorsque</TRIGGER>
John <EVENT eid="e1" structureInstanceID="sc1"
aspect="aoristique">enseigne</EVENT>au campus 1,
<STRUCTURE type="iteration" structureID="sc1" triggerID="ev1"
aspect="inaccompli" /> il
<EVENT eid="e2" eiid="ei2" structureInstanceID="sc1"
aspect="accompli">dépose</EVENT> <SIGNAL
sid="s2">d'abord</SIGNAL> les enfants à l'école.
<TLINK eventInstanceID="ei1" relatedToEvent="ei2"
relType="BEFORE" signalID="s2"/>

```

EXEMPLE 111 – « Lorsque John enseigne au campus 1, il dépose d'abord les enfants à l'école. »

Les déclencheurs « calendaires »

Les déclencheurs calendaires au sein de TIMEML sont représentés par l'entité TIMEX. Cependant certaines expressions ne sont pas exprimables dans ce formalisme, notamment pour des exemples tels que *un mardi sur deux, tous les troisièmes mardis du mois*. Nous choisissons de les représenter au sein de l'entité TRIGGER, en spécifiant le type par la valeur CAL. Afin de couvrir tous les expressions calendaires, nous utilisons la formalisation algébrique développée par P. Enjalbert et G. Becher et présentée au chapitre 4. Pour rappel, le formalisme possède un ensemble de mécanismes permettant de calculer les séries associées à une expression calendaire itérative. La fonction permettant d'obtenir la série correspond à la valeur de l'attribut VALUE, et nous obtenons les représentations suivantes pour les expressions associées :

- (a) un mardi sur deux
SUR(1, 2, MARDI)
- (b) tous les troisièmes mardis du mois
RESTRICT₃(MARDI, MOIS)

EXEMPLE 112 – Exemple de représentations d’expressions temporelles à l’aide du formalisme algébrique de Enjalbert et Becher (2011 – à paraître).

Par ailleurs, comme nous l’avons exposé au cours de ce manuscrit, il est possible de se référer temporellement au déclencheur calendaire soit à un niveau global (cf. l’exemple 113a), soit au niveau modèle (cf. l’exemple 113b). Afin de distinguer les deux niveaux, nous ajoutons un identifiant au niveau modèle, TMID, auquel il pourra être fait référence.

- (a) J’allais à la piscine tous les mardis. Maintenant, j’y vais le jeudi.
- (b) Tous les jeudis, John enseigne pendant 3 heures puis il va manger au RU.

EXEMPLE 113

De manière canonique, la représentation d’une itération issue d’un déclencheur calendaire est illustrée par l’exemple 114.

```
<TRIGGER trid="t1" tmid="tmi1" type="CAL"
value="TousLESN(JEUDI)"> Tous les jeudis</TRIGGER>, des
habitues venaient faire une partie de boston.
```

EXEMPLE 114

Les déclencheurs « fréquents »

Nous finissons par la représentation des déclencheurs fréquents (« rarement », « parfois », ...), qui s’avère être plus complexe que pour les déclencheurs précédents. En effet, le calcul de la fréquence dépend de nombreux paramètres tels que l’agent mis en œuvre dans l’énoncé, le contexte ou bien encore l’action réalisée. Par exemple, dans un contexte où nous avons deux individus, Pierre et Jean, la phrase *Je vais souvent au cinéma* aura une interprétation différente selon qu’il s’agit de Pierre (fréquence d’une fois par mois, par exemple) ou Jean (une activité bimensuelle, par exemple). Dans un se-

cond exemple comme *Jean va plus souvent au cinéma que Pierre* où l'on compare les habitudes des deux protagonistes, nous possédons une information qu'il existe un ratio entre leur habitudes, mais nous ne sommes pas à même de le calculer. Dans le cadre de la représentation des fréquentiels, nous proposons donc de représenter la valeur sémantique associée au fréquentiel par une fonction calculant cette fréquence. Le calcul se fait alors dans un outil externe qui intégrerait un certain nombre de connaissances sur le monde ou sur le texte permettant ainsi de définir la valeur de *souvent* et d'effectuer notamment des calculs de comparaisons de fréquences.

Du point de vue XML, l'attribut `TYPE` de l'entité `TRIGGER` est renseigné par la valeur `FREQ` tandis que l'attribut `VALUE` contient la valeur de la répétition portée par le fréquentiel (cf. exemple 115) que nous représentons par une fonction.

```
Pierre va <TRIGGER trid="tr1" type="FREQ"
value="souvent">souvent</TRIGGER> au cinéma.
```

EXEMPLE 115

EN RÉSUMÉ

Nous avons donc présenté une extension, en conservant les fondamentaux, de `TIMEML`, et en proposant deux entités `STRUCTURE` et `TRIGGER` représentant respectivement les structures itératives et les déclencheurs. Nous avons élaboré cette extension sur la base de notre étude de corpus, et des modèles théoriques sur lesquels reposent nos travaux, que nous avons présentés en partie ii.

Par ailleurs, en plus de la représentation des structures itératives, nous proposons une nouvelle prise en compte des calendaires en nous fondant sur les travaux sur les expressions calendaires de Enjalbert et Becher (2011 – à paraître) présenté au chapitre 4. Cette représentation sous formes de fonction telles que `SUR(1, 2, MARDI)` (un mardi sur deux) permet d'exprimer plus largement les calendaires qu'un système de formatage de date ou de période contraint à l'expression de période contiguë.

Cette extension est encore en cours d'élaboration. Notre prochaine étape consistera à annoter un corpus plus large que nos exemples d'études afin d'affiner la formalisation proposée, par exemple le corpus *FR-TimeBank* constitué par André Bittar (2010).

Pour conclure ce chapitre, nous exposons un exemple complet, de la présente page, annoté sur l'énoncé itératif suivant, issu d'*Un cœur simple* de Flaubert légèrement amendé pour l'illustration :

Tous les jeudis, des habitués venaient faire une partie de boston. Félicité préparait d'avance les cartes et les chaufferettes. Ils arrivaient à huit heures bien juste et se retiraient avant le coup de onze. Souvent l'un des habitués apportait le repas.

```

<TRIGGER trid="tr1" tmid="tr1m" value="TousLesN(JEUDI)">
Tous les jeudis</TRIGGER>, des habitués
<EVENT eid="e1" eiid="ei1" structureInstanceID="sc1"
aspect="aoristique">venaient faire</EVENT>une partie de boston.
<STRUCTURE scid="sc1" type="iteration" triggerID="tr1"
aspect="inaccompli"/>
<TLINK eventInstanceID="sc1" relatedToTime="tle"
relType="SIMULTANEAOUS"/>
<TLINK eventInstanceID="ei1" relatedToTime="t1" relType="IS_INCLUDED"/>
Félicité<EVENT eiid="ei2" eid="e2" structureInstanceID="sc1"
aspect="aoristique">préparait</EVENT>
<SIGNAL sid="s1">d'avance</SIGNAL>
les cartes et les chaufferettes.
<TLINK eventInstanceID="ei2" relatedToEvent="ei1" relType="BEFORE"
signalID="s1"/>
Ils<EVENT eid="e3" eiid="ei3" structureInstanceID="sc1"
aspect="aoristique">arrivaient</EVENT>
<TIMEX tid="t2" value="PT20H" type="TIME">à huit heures bien
juste</TIMEX>,
<TLINK eventInstanceID="ei3" relatedToTime="t2"
relType="SIMULTANEAOUS"/>
<TLINK eventInstanceID="ei3" relatedToEvent="ei1" relType="BEGINS"/>
et<EVENT eid="e4" eiid="ei4" structureInstanceID="sc1"
aspect="aoristique">se retiraient</EVENT>
<TIMEX tid="t3" value="PT23H" type="TIME">avant le coup de
onze</TIMEX>.
<TLINK eventInstanceID="ei4" relatedToTime="t3"
relType="SIMULTANEAOUS"/>
<TLINK eventInstanceID="ei4" relatedToEvent="ei1" relType="ENDED"/>
<TRIGGER trid="tr2" value="often">Souvent</TRIGGER>, l'un des
habitués<EVENT eid="e5" eiid="ei5" structureInstanceID="sc1"
aspect="aoristique">apportait</EVENT>le repas.
<STRUCTURE type="selection" scid="sc2" eventID="e5" triggerID="tr2"
aspect="inaccompli"/>

```

EXEMPLE 116



CONCLUSION

CONCLUSION & PERSPECTIVES

Nous avons présenté dans ce mémoire l'étude que nous avons menée sur l'expression de l'itération en langue. Travaux visant d'une part à une observation sur corpus de ces phénomènes, et d'autre part, à terme, à une analyse automatique sémantique et discursive pour aboutir *in fine* à un calcul des relations aspectuo-temporelles. Nous avons en outre fait apparaître différentes perspectives de recherche que nous allons maintenant évoquer.



Un premier volet a concerné l'analyse de l'inscription discursive des phénomènes itératifs. Nous avons mené en premier lieu une étude de corpus, sur des articles du journal « Le Monde » et des œuvres littéraires, qui nous a permis de mettre en évidence certaines propriétés des structures itératives, notamment le phénomène de récursivité, la discontinuité des propositions d'une structure ou encore l'importance de la position détachée des déclencheurs sur la taille des structures itératives.

Sur la base de cette étude, nous avons établi un modèle d'annotation afin de rendre compte de ces propriétés discursives, en nous fondant sur le triptyque *unité – relation – schéma* d'Antoine Widlöcher qui permet de modéliser les représentations textuelles, et d'y apporter une sémantique propre. Ce modèle d'annotation, développé au sein de la plate-forme d'annotation GLOZZ, nous a permis de stabiliser et spécifier les annotations nécessaires pour la phase automatique, d'une part; et d'établir une bijection entre les segments textuels et leurs instanciations au sein du modèle MOPI, d'autre part.

En l'état actuel de notre implémentation, la circonscription des structures itératives repose essentiellement sur les temps morphologiques et les types de procès, dont ce dernier pose encore des problèmes liés à la couverture du lexique que nous avons créé. La mise en œuvre de l'analyse automatique a toutefois donné des premiers résultats intéressants qualitativement, que nous devons encore confirmer par une évaluation précise de notre analyseur.

Nous pouvons d'ores et déjà envisager un vecteur d'amélioration de notre analyseur, et potentiellement du calcul aspectuel, qui reposerait sur l'intégration d'une ressource plus élaborée qu'un simple lexique telle qu'un réseau sémantique ou une ontologie. Nous envisageons d'utiliser une telle ressource, d'une part, pour la circonscription, en reliant des concepts liés présents dans

des propositions différentes comme dans l'exemple 117 où il existe une relation entre *une partie de boston* et *des cartes* ; et d'autre part, pour réaliser une désambiguïsation automatique des relations aspectuo-temporelles, dans l'exemple 117 nous avons l'action *venir* qui peut se décomposer en *arriver*, contraignant la borne initiale de l'intervalle de procès de *venir*, et en *retirer*, contraignant la borne finale.

Tous les jeudis, des habitués venaient faire une partie de boston. Félicité préparait d'avance les cartes et les chaufferettes. Ils arrivaient à huit heures bien juste, et se retiraient avant le coup de onze.

EXEMPLE 117 – Extrait d'*Un cœur simple*.



Nous avons développé dans un second volet le calcul aspectuo-temporel des énoncés itératifs. Nous nous sommes fondé sur les travaux de la sémantique du temps et de l'aspect de Laurent Gosselin dans lesquels il conçoit les procès itératifs dans une double appréhension : un procès modèle, d'une part ; et une série itérative, d'autre part.

Nous avons présenté les principes du calcul aspectuel des procès itératifs en faisant le parallèle avec le calcul des procès semelfactifs. Nous avons exposé notre calcul qui repose essentiellement sur l'affectation de la valeur des temps morphologiques à la série itérative, le procès modèle étant majoritairement aoristique, d'une part ; et sur le calcul de la portée sur la série ou sur le procès modèle des autres marqueurs aspectuels, d'autre part. Nous avons abordé également une représentation aspectuo-temporelle des structures itératives de type « sélection ». Ces analyses ont donné lieu à l'implémentation du système *CAsTe* qui calcule les représentations aspectuo-temporelles des procès itératifs, et se repose notamment sur un lexique des types de procès associés aux verbes, que nous avons constitué en nous basant sur la classification établie par Laurent Gosselin.

L'implémentation du calcul aspectuel que nous avons proposé ne couvre pas en l'état l'ensemble des cas existants. Nous envisageons donc de poursuivre l'intégration des propriétés décrites dans la *SdT* et non traitées dans le cadre de cette thèse telles que les phases ou encore la coupure modale. Cette dernière consiste à isoler le possible de l'irrévocable. Laurent Gosselin considère qu'elle n'est plus opérée seulement par le moment présent mais également par le moment de référence. Plus précisément, dans le cadre du modèle *SdT*, c'est la borne finale de l'intervalle de référence (IR) qui va opérer la cou-

pure modale entre l'irrévocable et le possible. Dans l'exemple 118, nous avons l'itération *faire trois tours de terrain* dont la fin du procès est située dans le domaine du possible, ne sachant pas si le procès est allé jusqu'à son terme.

Il était en train de faire trois tours de terrain, quand son téléphone se mit à sonner.

EXEMPLE 118

Nous souhaitons également nous pencher sur les substantifs événementiels dans le cadre du calcul aspectuel². Préalablement à une opérationnalisation, le traitement de cette problématique amène à se poser des questions étroitement liées à la *SdT*, qui à notre connaissance n'a pas été abordés au sein de cette théorie. Par exemple, *Les substantifs événementiels possèdent-ils un aspect ?*

La littérature le démontre, notamment les travaux de (Kiefer, 1998; Gross, 1996; Lefeuvre et Nicolas, 2004). Par exemple, les travaux de Kiefer s'intéressent à l'aspectualité des substantifs déverbaux, en appliquant des locutions temporelles (pendant X temps, dans/en X temps, à X, pour (une durée de) X, de X) pour déterminer la structure événementielle et en déduire ensuite l'aspect.

- (a) L'écriture de cette lettre en deux heures est surprenante.
- (b) L'arrivée de Pierre à 3 heures.
- (c) La survie de dix ans de Pierre à sa mère.

EXEMPLE 119 – Exemples extrait de Kiefer (1998).

À partir de ces travaux, il est possible de déterminer les traits de type *ponctuel*³, *dynamique* et *borné*⁴ pour les substantifs que l'on retrouve pour les verbes. Ces traits permettent d'en déduire les types de procès au sein de l'ontologie de Vendler (état, activité accomplissement et achèvement), et *a fortiori* de créer des ressources équivalentes à celles établies pour les verbes. Une première piste pourrait être de déterminer si ils existent une réciprocity entre les types de procès associés aux verbes et aux substantifs qui en dérivent.

2. Pour rappel, il existe deux types de substantifs événementiels : des substantifs déverbaux (*tremblement/trembler, déraillement/dérailer, occupation/occuper*) et des substantifs événementiels simples (*orage, tempête, accident, cérémonie, match*).

3. Le substantif est ponctuel si il est compatible avec *en+durée*, et si il caractérise le temps pris par la réalisation de l'événement, p.e. : *Victoire en deux heures*.

4. Le substantif est borné si il est compatible avec *en+durée*, et non borné si il est compatible avec *pendant+durée*.

Ceci n'est qu'une question parmi d'autres dont les deux plus importantes, nous semblent-ils, sont : *Quelles types d'intervalles pour la représentation des substantifs ?* et *Quelle est la visée aspectuelle sur ces substantifs ?* Ces questions restent ouvertes dans le cadre de la *SdT* et nécessitent donc une étude complète, tant du point de vue des procès itératifs que sémelfactifs.



ANNEXES

CONSTITUTION D'UN LEXIQUE DES TYPOLOGIES DE PROCÈS

L'analyse aspectuo-temporelle d'un énoncé itératif repose en partie sur le type du procès. Dans le cadre de notre analyse automatique, il est donc nécessaire de posséder une ressource contenant les verbes et les types de procès qui leurs sont associés. Le calcul des valeurs reposant sur des règles linguistiques faisant intervenir la sémantique, la constitution de ce lexique est difficilement automatisable. Nous avons donc constitué un lexique manuellement, et ce chapitre expose la méthode de calcul de cette typologie et la démarche que nous avons suivie.

SOMMAIRE

A.1	Classification	219
A.2	Exemple avec <i>Rouler</i>	220
A.3	Démarche & résultat	221

INTRODUCTION

Notre analyse aspectuo-temporelle repose principalement sur le type de procès porté par le verbe. Pour rappel, Vendler (1967) définit les quatre types de procès : *état*, *activité*, *accomplissement* et *achèvement*. Les types de procès sont définis selon leurs traits *ponctuel*, *dynamique* et *borné* (cf. tableau 16).

TYPE DE PROCÈS	DYNAMIQUE	BORNÉ	PONCTUEL
état	–	–	–
activité	+	–	–
accomplissement	+	+	–
achèvement	+	+	+

TABLEAU 16 – Définition des types de procès selon leurs traits *dynamique*, *borné* et *ponctuel*.

Le type de procès permet de mettre en évidence certains conflits de la langue qui amènent à la construction d'itération. Citons les trois conflits suivants :

- Si le procès est [ponctuel] et la valeur aspectuelle du procès [inaccompli], alors il y a glissement vers l'itération ou la dilatation.

Il mourrait.

- Si le type de procès est [achèvement] et qu'il existe un circonstant temporel associé de type [pendant + durée], alors la valeur du glissement de sens est [itération].

Il élimina la concurrence pendant deux ans.

- Si le type de procès est [accomplissement] et il existe un circonstant temporel associé de type [pendant + durée], alors la valeur du glissement de sens est [interruption ou itération].

Il a prévenu sa femme de son arrivée pendant deux heures (en l'appelant toutes les 30 minutes).

Dans le cadre de la *SdT*, Laurent Gosselin redéfinit les types de procès selon les traits sémantiques suivants :

- série de changements, c'est-à-dire une série d'événements qui correspond à un trait dynamique positif dans l'ontologie de Vendler (1967);
- absence de bornes ou bornes extrinsèques qui correspond à un trait borné négatif;
- borne intrinsèque qui correspond à un trait borné positif;
- changement atomique qui correspond à un trait ponctuel positif.

Le tableau 17 présente la définition des types de procès proposés par Laurent Gosselin.

TYPE DE PROCÈS	STRUCTURE EXTERNE	STRUCTURE INTERNE
état nécessaire	absence de bornes	absence de changements
état contingent	bornes extrinsèques	absence de changements
activité	bornes extrinsèques	série de changements
accomplissement	bornes intrinsèques	série de changements
achèvement	bornes intrinsèques	changement atomique

TABLEAU 17 – Définition des types de procès proposée par Laurent Gosselin.

Règles linguistiques

Afin de déterminer les différents traits sémantiques associés au verbe, il est nécessaire d'appliquer les cinq règles suivantes :

- la compatibilité du prédicat avec un circonstancier de localisation temporelle implique la présence des bornes du procès, seuls les états nécessaires en étant dépourvus;

Mon chat est mort lundi dernier.

**La terre était ronde la semaine dernière.*

- la compatibilité du prédicat à l'imparfait ou au présent avec la locution [être en train de V_{inf}] indique que la situation est constituée d'une série de changements ;

Il est en train d'éliminer les virus de mon ordinateur.

**La voiture était en train d'être rouge.*

- le caractère extrinsèque des bornes est révélé par la compatibilité du prédicat au passé composé avec [pendant + durée] ;

Il a tourné en rond pendant deux heures.

**Il a parcouru la distance Paris–Caen en deux heures.*

- le caractère intrinsèque des bornes est indiquée par la compatibilité du prédicat au passé composé avec [en + durée] ;

Il a parcouru la distance Paris–Caen en deux heures.

**Il a tourné en rond en deux heures.*

- le fait que l'expression [mettre n temps à V_{inf}], conjuguée au passé composé, soit équivalente à [mettre n temps avant de V_{inf}] révèle que le procès est un changement atomique.

Il a mis cinq minutes à atteindre le sommet de la montagne. \Leftrightarrow Il a mis cinq minutes avant d'atteindre le sommet de la montagne.

Jean a mis cinq minutes à ouvrir la porte. \Leftrightarrow Jean a mis cinq minutes avant d'ouvrir la porte.

A.1 CLASSIFICATION

La valeur du type de procès est donc définie par les cinq règles linguistiques présentées ci-dessus. Néanmoins, cette valeur ne dépend pas uniquement du sens du verbe mais également de son environnement actanciel. La classification définie dans la *SdT* est composée de neuf classes, notées de *a* à *i*. Chaque classe est constituée de quatre types d'associations différentes auxquelles sont associées un type de procès. Les quatre types d'associations sont identifiés grâce à :

1. la présence d'un syntagme nominal objet introduit par un article partitif ou indéfini pluriel :

Il attelait des chevaux.

2. l'absence de syntagme nominal objet, de subordonnée complétive objet, et de subordonnée interrogative indirecte :

Je roule (sur l'autoroute).

3. la présence d'un syntagme nominal objet introduit par un article démonstratif, possessif ou défini, ou d'une subordonnée complétive objet, ou d'une subordonnée interrogative indirecte :

Il accapare les blés de la province.

4. la présence d'un syntagme nominal objet introduit par un article indéfini :

Le bûcheron casse un rondin de bois.

Le tableau 18 résume les types de procès obtenus à partir de la classe du verbe et des quatre types d'associations qui viennent d'être définis.

CLASSE	TYPE 1	TYPE 2	TYPE 3	TYPE 4
a	état nécessaire	état nécessaire	état nécessaire	état nécessaire
b	état contingent	état contingent	état contingent	état contingent
c	activité	activité	activité	activité
d	activité	activité	activité ou accomplissement	activité ou accomplissement
e	activité	activité	activité ou accomplissement	accomplissement
f	activité	activité	accomplissement	accomplissement
g	activité	activité ou accomplissement	accomplissement	accomplissement
h	activité	accomplissement	accomplissement	accomplissement
i	achèvement	achèvement	achèvement	achèvement

TABLEAU 18 – Tableau de correspondance entre classe et type de procès.

A.2 EXEMPLE AVEC *rouler*

Pour déterminer la classe du verbe, il faut identifier le type de procès pour les quatre types d'association, et donc appliquer les tests linguistiques. Prenons l'exemple du type 2, avec la phrase *Je roule (sur l'autoroute)* :

- *Le procès possède-t-il des bornes ?*

J'ai roulé (sur l'autoroute) lundi dernier est acceptable. Le prédicat est donc bien compatible avec un circonstanciel de localisation. Nous concluons que le procès possède des bornes.

- *Le procès est-il constitué d'une série de changements ?*

Je suis en train de rouler (sur l'autoroute) est acceptable. Le procès est donc constitué d'une série de changements.

– *Quel est le type des bornes de l'intervalle de procès ?*

Je roule (sur l'autoroute) pendant 10 minutes est acceptable et **Je roule (sur l'autoroute) en dix minutes** n'est pas acceptable. Le procès possède donc des bornes extrinsèques.

– *Le procès est-il un changement atomique ?*

J'ai mis dix minutes à rouler (sur l'autoroute) n'est pas acceptable et n'a donc pas le même sens que **J'ai mis dix minutes avant de rouler (sur l'autoroute)**. Le procès n'est pas un procès atomique.

Le tableau 19 présente les types de procès pour le verbe *rouler* en fonction du type d'association. Grâce au tableau 18 de correspondance entre les classes et les types de procès, nous déduisons que le verbe *rouler* est de classe *f*.

TYPE D'ASSOCIATION	PHRASE TEST	TYPE DE PROCÈS
Type 1	Je roule des tapis	Activité
Type 2	Je roule (sur l'autoroute)	Activité
Type 3	Je roule ce tapis.	Accomplissement
Type 4	Je roule une cigarette.	Accomplissement

TABLEAU 19 – Types de procès pour chaque type d'association du verbe *rouler*.

A.3 DÉMARCHE & RÉSULTAT

Méthode

Afin de constituer cette ressource, nous avons sélectionné les 1250 verbes les plus fréquents de notre corpus [LMD], et les avons répartis entre quatre annotateurs. L'annotateur devait pour chaque verbe établir une phrase d'exemple pour les quatre types d'associations. Puis il devait, pour chaque type d'associations, appliquer les cinq règles linguistiques présentées ci-dessus. Enfin, il devait déterminer la classe grâce au tableau 18, en mettant en correspondance les quatre types de procès spécifiés.

Afin de faciliter la constitution du lexique, nous avons développé une application permettant de renseigner pour chaque verbe les phrases d'exemples et de cocher les tests linguistiques valides. Le programme détermine les types de procès pour chaque association, puis en déduit la classe du verbe. La figure 71 donne une capture d'écran de notre application, nommée *Classeverbe*.

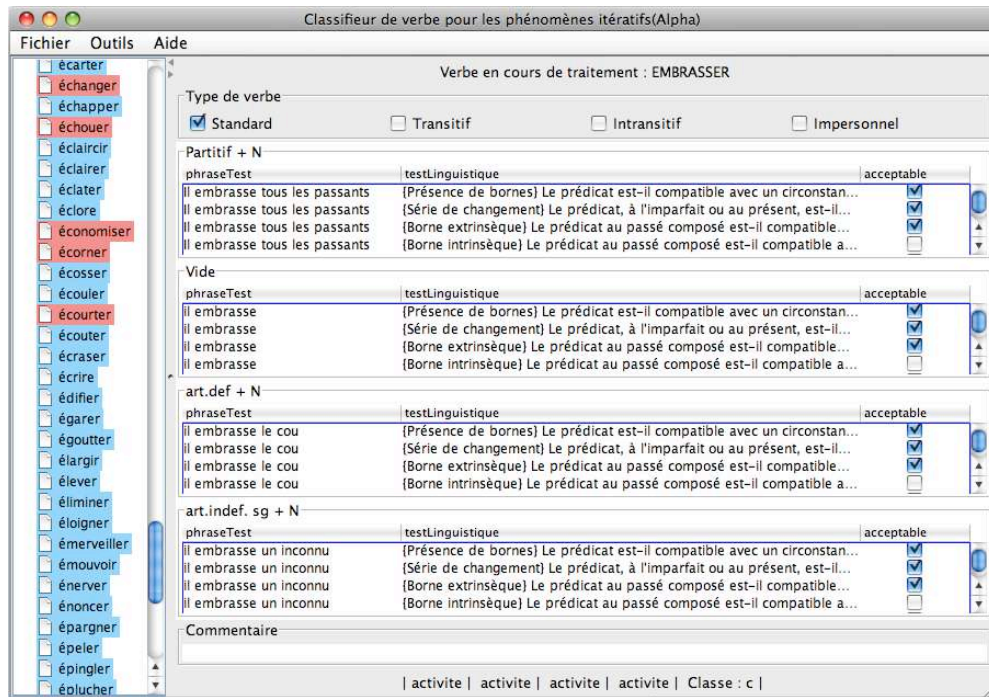


FIGURE 71 – Capture d’écran de l’application de constitution du lexique des typologies de procès.

Résultats

Une première phase d’annotation a mis en évidence deux problèmes. Premièrement, nous avons constaté que les quatre types d’association ne peuvent s’appliquer à l’ensemble des verbes. Nous distinguons donc les verbes de types :

- *transitif direct*, nécessitant un COD tel que *j’enfile la veste* ;
- *intransitif* et *transitif indirect*, interdisant un COD tel que *les pommes tombent* ou autorisant un COI tel que *Je parle à Paul* ;
- *impersonnel*, dont le sujet est impersonnel tel qu’*il pleut des cordes* ;
- *standard*, c’est-à-dire tous les autres verbes.

Deuxièmement, nous nous sommes confronté à la question de la valeur pronominale des verbes¹. Il existe différents types de constructions pronominales : certains verbes n’ont que la forme pronominale, par exemple *s’enrhumer*. D’autres ont les deux formes avec des sens au moins partiellement distincts, il peut donc s’agir de procès de natures différentes :

- ennuyer quelqu’un/s’ennuyer ;
- lever quelque chose/se lever ;
- distinguer quelque chose/se distinguer.

1. Nous reprenons ici en partie une discussion par email avec Laurent Gosselin.

D'autres ont les deux formes avec des sens équivalents, par exemple *regarder quelqu'un/se regarder* (= regarder soi-même) : il s'agit donc simplement du verbe *regarder*. Enfin, il existe des constructions pronominales qui constituent des diathèses (ou voix), comme le pronominal de sens passif : *ce chapeau se remarque de loin, ce linge se lave à la machine*. Dans ce cas on a affaire aux verbes *remarquer* et *laver*.

Afin de prendre en compte ces deux problèmes, nous avons, d'une part, intégré la possibilité de renseigner la catégorie *transitive directe, intransitive/transitivité indirecte* ou *impersonnelle* du verbe. D'autre part, nous avons fait le choix de dupliquer l'entrée du verbe pour introduire sa forme pronominale dans les cas nécessaires.

Actuellement, nous avons annoté et vérifié 825 verbes dont le détail par catégorie est reporté au tableau 20. L'application et le lexique, au format XML, sont librement disponibles à l'adresse suivante : <http://jlebranc.perso.info.unicaen.fr/classeverbe.html>

CATÉGORIE	NOMBRE
Transitif direct	199
Intransitif & transitif indirect	27
Impersonnel	0
Standard	599

TABLEAU 20 – Répartition des verbes annotés selon leur catégorie.



RÈGLES DE CALCUL ASPECTUO-TEMPOREL

Ce chapitre expose le jeu de règles que nous avons intégré au sein du système *CAsTe* que nous avons présenté au chapitre 10. Nous nous sommes fondé, pour constituer ces règles, sur les travaux de la *SdT* ainsi que sur le travail d'implémentation de Cédric Person (2004) sur les procès semelfactifs. Nous regroupons ici les instructions liées à chacun des marqueurs, les principes sur les conflits, les principes généraux de bonne formation, permettant de construire automatiquement les représentations aspectuo-temporelles associées aux énoncés itératifs. Nous rappelons que les principes du calcul sont présentés au chapitre 9 page 161.

SOMMAIRE

B.1	Les phénomènes itératifs	226
B.2	Les temps morphologiques	226
B.3	Les déclencheurs	233
B.4	La subordination temporelle	238
B.5	Les relations circonstancielles	239
B.6	Les circonstants de durée de type [depuis+durée]	243

Notation

Nous rappelons les notations que nous utilisons pour présenter les différentes règles de calcul :

Les séries itératives :

- les intervalles de procès sont notés *IPs* ;
- les intervalles de références sont notés *IRs* ;

Les procès modèles :

- les intervalles de procès sont notés *IP* ;
- les intervalles de références sont notés *IR* ;

Les circonstanciels :

- les intervalles circonstanciels, liés aux circonstanciels sémelfactifs, sont notés IC ;
- les intervalles circonstanciels, associés aux déclencheurs, notamment de type calendaire (*tous les jours*), sont notés ICs pour la série et IC m pour le modèle.

B.1 LES PHÉNOMÈNES ITÉRATIFS

L'analyse d'un énoncé itératif amène la construction d'intervalles associés à la série itérative. Chaque procès de cet énoncé se voit également associer un couple d'intervalle IP et IR.

Si
– l'énoncé est de type itératif
Alors
– création de IPs et IRs

Règle 8 – Règle de construction des intervalles de la série.

Si
– le procès appartient à la série itérative
Alors
– création de IP et IR du procès modèle associé

Règle 9 – Règle de construction des intervalles du procès.

Les valeurs aspectuelles portées par le temps morphologique affectent les intervalles correspondant à l'itération¹ (IPs et IRs), tandis que les procès modèles se voient affecter la valeur aspectuelle aoristique. Nous verrons par la suite que certaines configurations affectent cette valeur d'aoristique.

B.2 LES TEMPS MORPHOLOGIQUES

Nous avons donc, pour les différents temps morphologiques de l'indicatif, les règles suivantes :

1. Au chapitre 6, nous avons constaté que les itérations rencontrées en corpus sont composées d'un seul temps morphologique. De ce fait, les interactions possibles entre les règles de temps morphologique et la valeur aspectuelle sont limitées.

Le présent

Si

- le temps morphologique est présent

Alors

- la valeur temporelle est le présent : IRs SIMUL IE
- la valeur aspectuelle de la série est inaccomplie : IPs RE IRs
- la valeur aspectuelle du procès modèle est aoristique : IR co IP

Règle 10 – Cas des itérations au présent.

Dans le cas du présent, l'intervalle de référence de la série itérative peut recouvrir, coïncider ou être à cheval sur la borne droite ou gauche de l'intervalle d'énonciation, ce qui n'est pas le cas pour les autres temps morphologiques.

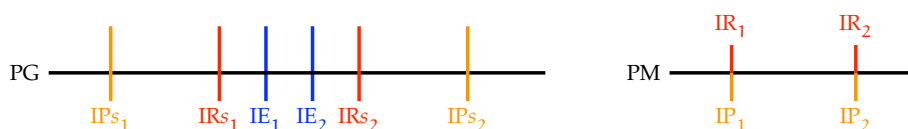


FIGURE 72 – Chronogramme d'un procès itératif au présent.

Tous les jours, je vais à la plage.

EXEMPLE 120

L'imparfait

Si

- le temps morphologique est l'imparfait

Alors

- la valeur temporelle est le passé : IRs ANT IE
- la valeur aspectuelle de la série est inaccomplie : IPs RE IRs
- la valeur aspectuelle du procès modèle est aoristique : IR co IP

Règle 11 – Cas des itérations à l'imparfait.

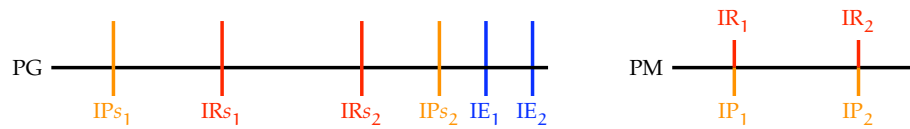


FIGURE 73 – Chronogramme d'un processus itératif à l'imparfait.

J'allais régulièrement à la plage.

EXEMPLE 121

Le passé simple

Si

– le temps morphologique est le passé simple

Alors

- la valeur temporelle est le passé : IRs ANT IE
- la valeur aspectuelle de la série est aoristique : IPs CO IRs
- la valeur aspectuelle du processus modèle est aoristique : IP CO IR

Règle 12 – Cas des itérations au passé simple.

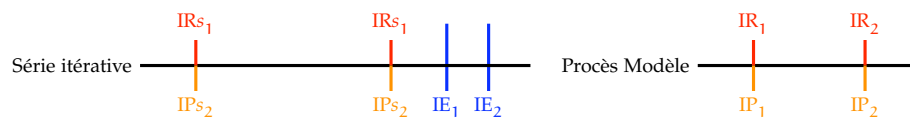


FIGURE 74 – Chronogramme d'un processus itératif au passé simple.

Madame Aubain tomba plusieurs fois en tachant de gravir le talus.

EXEMPLE 122

Le futur

Le futur, à l'instar des processus non itératifs, ne code pas de valeur aspectuelle particulière. Il existe cependant une valeur prédominante en contexte, l'aoristique.

Si

– le temps morphologique est le futur

Alors

– la valeur temporelle est futur : IRs post IE

– la valeur aspectuelle de la série est aoristique : IPs co IRs

– la valeur aspectuelle du procès modèle est aoristique : IP co IR

Règle 13 – Cas des itérations au futur.

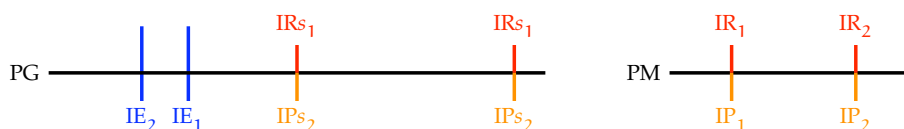


FIGURE 75 – Chronogramme d'un procès itératif au futur.

Nous irons quelquefois répandre des fleurs sur sa tombe sacrée.

EXEMPLE 123

Les temps composés

Dans le cas non itératif, les procès composés (le passé composé et le plus-que-parfait) peuvent prendre des valeurs d'accompli ou d'aoristique. La valeur prise dépend du fait que la saillance est liée à l'auxiliaire (accompli) ou au verbe (aoristique). Cette saillance peut être établie par le type de procès et les circonstants temporels présents dans la proposition.

Dans le cas des procès itératifs, nous représentons les temps composés par un seul couple d'intervalle de référence et de procès, que ce soit sur la série itérative ou sur le procès modèle. La série itérative peut prendre deux valeurs aspectuelles soit accompli, soit aoristique². Cependant la valeur accomplie nécessite la présence d'un procès télique (intrinsèquement bornée). Quant au procès modèle, il prend une valeur d'aoristique.

Nous avons donc les règles 14 et 15 suivantes :

2. La valeur d'accompli est testée par [depuis + durée] ou [il y a + durée que] qui exprime la durée écoulée entre la borne finale du procès et la borne initiale de l'intervalle de référence.

Si

– le temps morphologique est le passé composé

Alors

– la valeur temporelle est le présent : IRs SIMUL IE

Si le type de procès est *achèvement ou accomplissement* *Alors*

– la valeur aspectuelle de la série est accomplie : IPs ANT IRs

Sinon

– la valeur aspectuelle de la série est aoristique : IPs CO IRs

– la valeur aspectuelle du procès modèle est aoristique : IR CO IP

Règle 14 – Cas des itérations au passé composé.

L'exemple 124a présente la série itérative sous une visée aoristique comme l'illustre la figure 76a, tandis que l'exemple 124b présente une visée accomplie de la série itérative comme l'illustre la figure 76b.

- (a) Il est régulièrement allé voir sa mère à l'hôpital.
(b) Il a couru plusieurs fois le marathon de Paris.

EXEMPLE 124

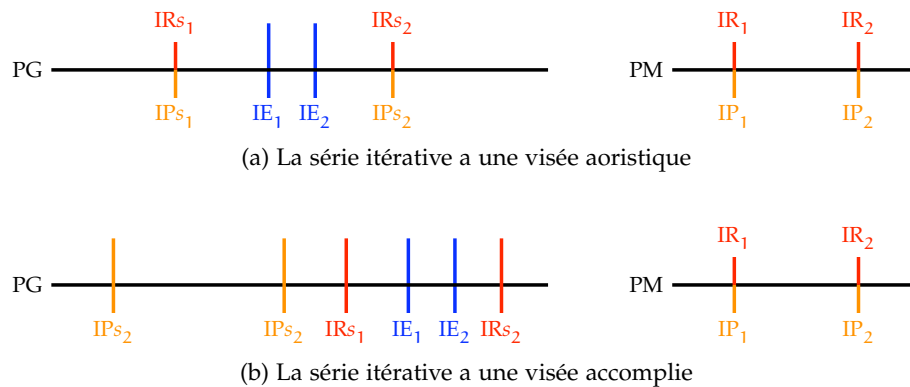


FIGURE 76 – Chronogramme d'un procès itératif au passé composé.

Si

– le temps morphologique est le plus-que-parfait

Alors

– la valeur temporelle est le passé : IRs ANT IE

Si le type de procès est *achèvement ou accomplissement* Alors

– la valeur aspectuelle de la série est accomplie : IPs ANT IRs

Sinon

– la valeur aspectuelle de la série est aoristique : IPs CO IRs

– la valeur aspectuelle du procès modèle est aoristique : IR CO IP

Règle 15 – Cas des itérations au plus-que-parfait.

L'exemple 125a présente la série itérative sous une visée d'aoristique comme l'illustre la figure 77a, tandis que l'exemple 125b présente une visée accomplie de la série itérative comme l'illustre la figure 77b.

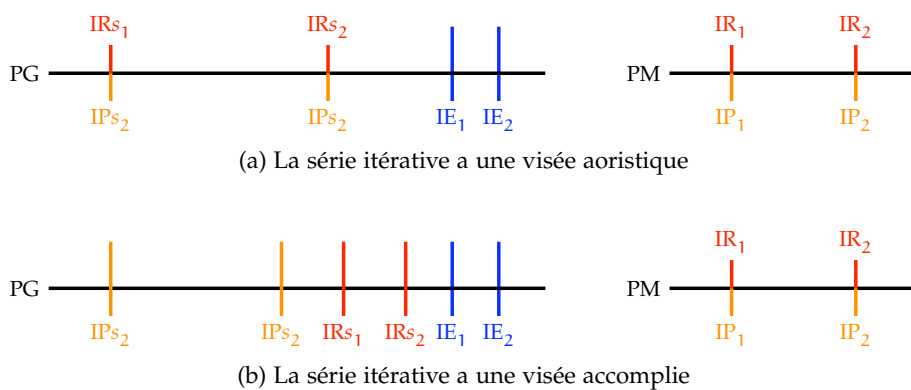


FIGURE 77 – Chronogramme d'un procès itératif au plus-que-parfait.

(a) Il avait régulièrement accompagné sa mère à l'hôpital (pour voir son père).

(b) Il avait couru plusieurs fois le marathon de Paris.

EXEMPLE 125

L'infinitif passé

Il convient d'ajouter la valeur associée à l'infinitif passé, concernant les propositions préfixées de la conjonction *après* ou *avant*, qui code une valeur aspectuelle d'aoristique sur la série itérative et sur le procès modèle. La position

temporelle dépend de la relation entretenue par la subordonnée et sa principale (cf. exemple 126 de la présente page). Cette valeur est la même dans les cas sémelfactifs, c'est-à-dire une valeur d'aoristique.

Si

– le temps morphologique est l'infinitif passé

Alors

– la valeur aspectuelle de la série est aoristique : IPs co IRs

– la valeur aspectuelle du procès modèle est aoristique : IR co IP

Règle 16 – Cas des itérations à l'infinitif passé.

La construction des procès à l'infinitif passé permet de déterminer que les intervalles liés au procès de la principale seront en relation avec la série.

(a) Avant de tirer sur son mari deux fois, elle a pris une longue respiration.

(b) Après avoir éternué tous les jours de la semaine, ...

EXEMPLE 126

Les auxiliaires de visée aspectuelle

Les auxiliaires de visée aspectuelle font références à des syntagmes tels que *être sur le point de* (marque le prospectif), *être en train de* (marque l'inaccompli). Nous déterminons une distinction d'affectation en fonction que le déclencheur, soit en position détachée (cf. exemple 127a) ou intégrée (cf. exemple 127b).

(a) Tous les jours, il s'apprêtait à venir à la maison quand le téléphone sonnait.

(b) Il s'apprêtait à venir tous les jours à la maison pour des travaux, quand il décéda.

EXEMPLE 127

Si

– l’itération possède un auxiliaire de visée aspectuelle

Alors

Si le déclencheur est intégré *Alors*

– la valeur affectant la série dépend de l’auxiliaire : IP_s RC IR_s

– une valeur d’aoristique affecte le procès modèle : IP CO IR

Sinon

– une valeur d’aoristique affecte la série : IP_s CO IR_s

– la valeur affectant le procès modèle dépend de l’auxiliaire :

IP RC IR

Règle 17 – Cas des itérations avec un auxiliaire de visée aspectuelle. La relation RC dépend de l’auxiliaire de visée aspectuelle.

Nous obtenons donc une valeur de prospectif pour le procès modèle de l’exemple 127a (c’est-à-dire IP POST IR), tandis que la valeur prospective est affectée à la série itérative dans l’exemple 127b (c’est-à-dire IP_s POST IR_s).

B.3 LES DÉCLENCHEURS

Les déclencheurs précisent les valeurs aspectuelles associées à la série itérative. Seuls les déclencheurs événementiels affectent également les valeurs des procès itérés. Ils précisent également, dans le cas des calendaires et des quantificationnels, les occurrences soit par le positionnement calendaire soit par le nombre, respectivement.

Les fréquentiels

Les déclencheurs de types fréquentiels ne posent aucune contrainte sur les visées aspectuelles aux séries itératives et aux procès modèles. Les exemples 128a et 128b se veulent illustrer la constance des valeurs issus des temps morphologiques. Notons qu’il existe toutefois une incompatibilité entre la visée accomplie et ce type de déclencheur. En effet, la visée accomplie nécessite que la série soit intrinsèquement bornée, et donc incompatible avec l’aspect fréquentatif.

(a) Félicité se rendait souvent sur la tombe de Virginie.

SÉRIE ITÉRATIVE : inaccomplie

PROCÈS MODÈLE : aoristique

(b) Félicité a fréquemment accompagné Mme Aubain sur la tombe de Virginie.

SÉRIE ITÉRATIVE : aoristique

PROCÈS MODÈLE : aoristique

EXEMPLE 128

Les calendaires

Les déclencheurs calendaires nécessitent la création d'intervalles de types circonstanciels sur la série itérative et sur le procès modèle. L'exemple 129 met en évidence la relation existante entre la série des jeudis venant contraindre temporellement la série, ainsi que celle existante entre le (ou les) procès modèle contraint par un jeudi dit modèle. Notons qu'il existe un cas dérogeant à cette règle, qui est la présence d'un ou plusieurs circonstanciels venant modifier les intervalles mis en relation. Nous traitons ce cas dans la section 9.4, intitulée *Les circonstanciels de temps*.

Le jeudi, j'accompagne ma grand-mère faire ses courses.

EXEMPLE 129

Si

– le déclencheur est de type calendaire

Alors

– création d'un *ICm*

Si le déclencheur est intégré *Alors*

– *ICm* ACCESS IP

Sinon

– *ICm* ACCESS IR

Règle 18 – Cas des déclencheurs de type calendaire. La relation RC est déterminée par les règles présentées section B.5.

La caractérisation des relations entre les intervalles est régie par les mêmes règles que pour les cas non itératifs (*cf.* section B.5) que nous avons adaptés aux cas itératifs.

Les quantificationnels

Les déclencheurs de types quantificationnels donnent une visée aoristique sur les séries itératives du fait que le quantificateur dénombre exactement les occurrences et montrent donc la série dans sa globalité. Notons qu'il existe une incompatibilité entre la visée inaccomplie et ce type de déclencheur.

Si

– le déclencheur est de type quantificationnel

Alors

– la valeur aspectuelle de la série est aoristique : IPs co IRs

Règle 19 – Cas des déclencheurs de type quantificationnel.

Toutefois, cette règle ne s'active pas lorsque la série itérative possède une visée accomplie. Cette exception provient de la valeur intrinsèquement bornée prise par la série à cette valeur aspectuelle, de sorte qu'elle n'est compatible qu'avec les itérations répétitives.

Les événementiels

Comme nous l'avons vu dans les chapitres précédents, les déclencheurs événementiels correspondent aux subordonnées temporelles³. Elles impliquent la création d'un intervalle circonstanciel. Elles sont introduites par des conjonctions telles que *lorsque*, *quand*, *à chaque fois que*, *dès que*⁴.

Au niveau des procès modèles, la conjonction établit une relation entre l'intervalle circonstanciel issue de la subordonnée et le procès de la principale. Les valeurs aspectuelles des procès modèles sont traitées comme pour les subordonnées temporelles non itératives que nous abordons à la section B.4. Au niveau de la série itérative, le déclencheur ne modifie pas la valeur aspectuelle,

3. Les différents critères amenant à la lecture itérative des subordonnées temporelles sont présentés au chapitre 8. Pour rappel, les subordonnées au futur ou à un temps composés et introduites par *quand* et *lorsque* ont une lecture semelfactive.

4. Les conjonctions portent la valeur de la relation entre la subordonnée et la principale. Ils sont définies dans les règles portant sur les relations : 24 à 31. Les conjonctions *quand* et *lorsque* portent la relation de coïncidence.

elle dépend de la valeur associée au procès de la principale établit par les règles sur les temps morphologiques.

Nous avons dans l'exemple 130a une valeur d'inaccompli qui est affectée à la série itérative dû à l'imparfait de la proposition principale. Nous distinguons toutefois le cas où la subordonnée porte un événement ponctuel tel que dans l'exemple 130b avec le verbe *arriver*. C'est alors la valeur aspectuelle portée par le procès de la subordonnée qui est affectée à la série itérative. La valeur aspectuelle associée à la principale est affectée au procès modèle.

- (a) Quand le temps était clair, on s'en allait de bonne heure à la ferme de Geffosses.
- (b) Il avait fait ses devoirs à chaque fois qu'elle rentrait du travail.

EXEMPLE 130

Si

- le déclencheur est de type événementiel
- le type de procès de la subordonnée est *achèvement*

Alors

- la valeur aspectuelle de la série est affectée par la valeur associée à la subordonnée

Règle 20 – Cas des déclencheurs de type événementiel.

Les autres déclencheurs

Dans le cadre de la *SdT*, Laurent Gosselin distingue encore deux types de déclencheurs : les présuppositionnels et les verbes intrinsèquement itératifs.

Les premiers font références aux adverbes (*encore, déjà*), les expressions (*à/de nouveau, une fois de plus,*) ou encore les adjectifs (*habituel, rare*) qui amènent à la construction d'une itération dont les itérés, à l'exception du dernier, sont présupposés, et donc ne sont pas aspectuellement accessibles. Les itérés sont accessibles uniquement si ils sont explicités via une quantification (*cf.* l'exemple 131b) ou une sélection d'un ou plusieurs itérés (*cf.* l'exemple 131c).

Si

– le déclencheur est de type présuppositionnel

Alors

– création uniquement de IPs et de IRs

– il n'existe qu'un procès construit dans la série itérative, le dernier

Règle 21 – Cas des déclencheurs de type présuppositionnel.

(a) Il a encore toussé. ...

(b) ... Cela fait dix fois ce matin.

(c) ... La première fois, il a toussé à deux heures du matin.

EXEMPLE 131

Le second type fait référence aux verbes marquant intrinsèquement l'itération tels que *hachurer*, *sautiller*. Les dits verbes présentent l'ensemble des itérés comme atomiques, et ne donnent pas accès aux itérés n'étant pas créés.

L'itérativité de ces verbes provient de leur sémantique, par exemple « sautiller : Faire de petits sauts », Le Littré. La création uniquement des intervalles de la série est dû à l'indivisibilité de la série. Comme pour les déclencheurs aspectuels, les itérés sont accessibles uniquement si ils sont explicités via une quantification ou une sélection d'un ou plusieurs itérés (*cf.* l'exemple 132).

Elle a sautillé pendant cinq minutes. Le premier saut était de trois centimètres.

EXEMPLE 132

Si

– le déclencheur est un verbe intrinsèquement itératif

Alors

– création uniquement des intervalles associés à la série (IPs et IRs)

Règle 22 – Cas des déclencheurs de type itératifs.

B.4 LA SUBORDINATION TEMPORELLE

Dans le cas des énoncés itératifs, la subordination temporelle n'implique pas de traitement spécifiques en dehors de son rôle de déclencheurs que nous

avons présenté précédemment. Le traitement est le même que pour les cas non itératifs :

1. Si les subordonnées sont introduites par une conjonction ou une locution conjonctive telle que *quand, comme, tandis que, pendant que*. La relation est alors IC co IR.
2. Si les subordonnées sont introduites au moyen d'un substantif désignant une période de temps telle que *le jour où, un jour où, l'année où, au moment où, après le jour où*. La relation est alors IC ACCESS IR.

Si

- il existe une subordonnée temporelle avec une relation circonstancielle RC associée à l'introducteur
- IR (reg) est l'intervalle de référence de la proposition régissante

Alors

- IC RC IR (reg)

Si la subordination est directe Alors

- IC co IR

Sinon

Si le circonstant est ponctuel Alors

- IC co IP

Sinon

- IC ACCESS IP

Règle 23 – Règle des subordonnées temporelles.

Remarquons deux points :

1. La subordination temporelle peut se faire également entre deux séries itératives sans jouer le rôle de déclencheur de l'itération (*cf.* exemple 133a). La subordonnée a pour rôle de période de référence de l'action de la principale.
2. La subordination temporelle, lorsque elle est semelfactive, se rattache à la série (*cf.* exemple 133b).

(a) Quand Marie venait souvent, Jean était rarement malade.

(b) Quand Marie était jeune, elle fumait tous les jours.

EXEMPLE 133

B.5 LES RELATIONS CIRCONSTANCIELLES

Les règles liées aux relations circonstancielle pour les cas itératifs doivent prendre en compte, comme pour les circonstanciels, leur portée sur la série ou sur le procès. Ce sont les circonstants temporels qui définissent l'intervalle circonstanciel, IC, et les relations qu'il entretient avec les autres intervalles de la représentation temporelle sont les suivantes :

1. Dans tous les cas, il existe une relation circonstancielle RC avec un intervalle lié à la proposition dans laquelle se trouve le circonstant. L'intervalle Ix en relation circonstancielle avec IC dépend de la portée du circonstant. La première relation est alors : IC RC Ix.
2. Si le circonstant est un complément circonstanciel non autonome, c'est-à-dire qu'il se repère soit à l'intervalle d'énonciation comme *demain* soit à un autre circonstanciel comme *la veille*⁵. Le circonstant nécessite une relation RA avec un intervalle antécédent Ia. Il faut alors ajouter la relation : IC RA Ia. Cette relation est définie par la valeur du circonstanciel, par exemple dans le cas de *la veille*, la relation par rapport à l'intervalle de procès ou de référence est antériorité⁶.

Les règles suivantes, issues des cas non itératifs, mettent en évidence les relations en fonction du type des circonstanciels.

Si

- le circonstanciel temporel est de la forme : lundi, cette année-là, pendant les vacances, pendant que P, ...

Alors

- il existe une relation d'ACCESSIBILITÉ

Règle 24 – Relation d'accessibilité avec IC.

(a) Je suis allé en mer pendant les vacances. (IC ACCESS IPs)

(b) Je mange régulièrement en cinq minutes. (IC ACCESS IP)

EXEMPLE 134 – Exemples associés à la règle 24

5. Les circonstanciels autonomes correspondent aux dates complètes (en 1789), aux dates indéterminées (un jour, une fois) et aux subordonnées temporelles.

6. Le lecteur pourra se reporter à (Person, 2004, p.79-81) pour plus de détails.

Si

- le circonstanciel temporel est de la forme : au cours de la matinée, dans le courant de la semaine, courant décembre, dans la soirée, tandis que P, alors que P, ...

Alors

- il existe une relation de RECOUVREMENT

Règle 25 – Relation de recouvrement avec IC.

- (a) Durant la révolution, les tunisiens organisèrent de nombreuses manifestations. (IC RE IRs)
- (b) Il court régulièrement durant 20 minutes. (IC RE IP)

EXEMPLE 135 – Exemples associés à la règle 25

Si

- le circonstanciel temporel est de la forme : pendant deux heures, en dix minutes, de 8 heures à midi, tant que P, ...

Alors

- il existe une relation de COÏNCIDENCE

Règle 26 – Relation de coïncidence avec IC.

- (a) Pendant 10 ans, elle a chanté dans toutes les grandes villes du monde. (IC RE IRs)
- (b) Il court le 3 000 mètres en dix minutes. (IC RE IP)

EXEMPLE 136 – Exemples associés à la règle 26

Si

- le circonstanciel temporel est de la forme : dès le début de la matinée, dès que P, sitôt que P, aussitôt que P, à partir du moment où P, ...

Alors

- il existe une relation de COÏNCIDENCE sur la borne gauche

Règle 27 – Relation de coïncidence sur la borne gauche avec IC.

Aussitôt qu'il arriva sur le seuil de la porte, elle tira trois fois. (IC ACCESS IPs)

EXEMPLE 137 – Exemple associé à la règle 27

Si

– le circonstanciel temporel est de la forme : jusqu'à midi, jusqu'à ce que P, jusqu'au moment où P, ...

Alors

– il existe une relation de COÏNCIDENCE sur la borne droite

Règle 28 – Relation de coïncidence sur la borne droite avec IC.

(a) Jusqu'à ce qu'il ait 18 ans, il jouait du panio. (IC ACCESS IRs)

(b) Il court habituellement jusqu'à midi. (IC ACCESS IP)

EXEMPLE 138 – Exemples associés à la règle 28

Si

– le circonstanciel temporel est de la forme : toute la matinée, tout le temps que P, ...

Alors

– il existe une relation de ACCESSIBILITÉ

Règle 29 – Relation d'accessibilité inversée avec IC.

J'ai fumé toute la matinée. (IPs ACCESS IC)

EXEMPLE 139 – Exemple associé à la règle 29

Si

– le circonstanciel temporel est de la forme : après le repas, après que P, ...

Alors

– il existe une relation d'ANTÉRIORITÉ

Règle 30 – Relation d'antériorité avec IC.

Aussitôt après l'attaque du Hamas, les ripostes israéliennes débutèrent.
(IC ANT IPs)

EXEMPLE 140 – Exemple associé à la règle 30

Si

– le circonstanciel temporel est de la forme : avant le repas, avant que P, ...

Alors

– il existe une relation de POSTÉRIORITÉ

Règle 31 – Relation de postériorité avec IC.

Avant que la révolution débute, le président organisa plusieurs
réunions de crise. (IC POST IPs)

EXEMPLE 141 – Exemple associé à la règle 31

B.6 LES CIRCONSTANTS DE DURÉE DE TYPE [DEPUIS+DURÉE]

Ces circonstants précisent la durée entre l'intervalle de procès et de référence, plus précisément les bornes initiales. Ils sont incompatibles avec les valeurs aspectuelles d'aoristique et de prospectif.

Si

– il existe un circonstant temporel de type [depuis + durée]

Alors

Si le circonstant temporel est rattaché à la série itérative *Alors*

– IC co [IP_{s1}, IR_{s1}]

Sinon

– IC co [IP₁, IR₁]

Règle 32 – Règle des circonstants temporels de type [depuis + durée].

Dans le cas où le circonstant temporel de type [depuis + durée] est dans un procès atélique et qu'il existe un circonstanciel ponctuel détaché, le circonstant temporel porte sur le procès.

Quand elle rentre du travail, il a fini ses devoirs depuis 10 minutes.

EXEMPLE 142

EN RÉSUMÉ

Nous venons de présenter les différentes règles de calcul des phénomènes itératifs, en reprenant les principes de la SdT et les formalisant. Nous avons traduit et formalisé, sous formes de règle, les valeurs aspectuelles des différents marqueurs entrant en jeu dans le traitement des structures itératives. Ces règles sont mises en œuvre dans un moteur de calcul, appelé *CAsTe* pour Calcul Aspectuo-Temporel, dont nous présentons les principes de fonctionnement au chapitre 10, et détaillons également des résultats.



Levé tous les jours à cinq heures, il a franchi comme un oiseau l'espace qui sépare son domicile de la rue Montmartre. Qu'il vente ou tonne, pleuve ou neige, il est au Constitutionnel et y attend la charge de journaux dont il a soumissionné la distribution. Il reçoit ce pain politique avec avidité, le prend et le porte. À neuf heures, il est au sein de son ménage, débite un calembour à sa femme, lui dérobe un gros baiser, déguste une tasse de café ou gronde ses enfants. À dix heures moins un quart, il apparaît à la mairie. Là, posé sur un fauteuil, comme un perroquet sur son bâton, chauffé par la ville de Paris, il inscrit jusqu'à quatre heures, sans leur donner une larme ou un sourire, les décès et les naissances de tout un arrondissement. Le bonheur, le malheur du quartier passe par le bec de sa plume, comme l'esprit du Constitutionnel voyageait naguère sur ses épaules. Rien ne lui pèse ! Il va toujours droit devant lui, prend son patriotisme tout fait dans le journal, ne contredit personne, crie ou applaudit avec tout le monde, et vit en hirondelle. À deux pas de sa paroisse, il peut, en cas d'une cérémonie importante, laisser sa place à un surnuméraire, et aller chanter un requiem au lutrin de l'église, dont il est, le dimanche et les jours de fête, le plus bel ornement, la voix la plus imposante, où il tord avec énergie sa large bouche en faisant tonner un joyeux Amen. Il est chanteur. Libéré à quatre heures de son service officiel, il apparaît pour répandre la joie et la gaieté au sein de la boutique la plus célèbre qui soit en la Cité. Heureuse est sa femme, il n'a pas le temps d'être jaloux ; il est plutôt homme d'action que de sentiment. Aussi, dès qu'il arrive, agace-t-il les demoiselles de comptoir, dont les yeux vifs attirent force chalands ; se gaudit au sein des parures, des fichus, de la mousseline façonnée par ces habiles ouvrières ; ou, plus souvent encore avant de dîner, il sert une pratique, copie une page du journal ou porte chez l'huissier quelque effet en retard. À six heures, tous les deux jours, il est fidèle à son poste. Inamovible basse-taille des chœurs, il se trouve à l'Opéra, prêt à y devenir soldat, Arabe, prisonnier, sauvage, paysan, ombre, patte de chameau, lion, diable, génie, esclave, eunuque noir ou blanc, toujours expert à produire de la joie, de la douleur, de la pitié, de l'étonnement, à pousser d'invariables cris, à se taire, à chasser, à se battre, à représenter Rome ou l'Égypte ; mais toujours – in petto –, mercier. À minuit, il redevient bon mari, homme, tendre père, il se glisse dans le lit conjugal, [...] Enfin, s'il dort, il dort vite, et dépêche son sommeil comme il a dépêché sa vie.

Extrait de *La fille aux yeux d'or* (1834-35).

Le vé-tout-les-jours à cinq heures, il franchi comme un oiseau l'espace qui sépare son domicile de la rue Montmartre. [Qu'il vende,] [fonne].
 pleuve, me son, il est au Constitutionnel et y a une charge de journaux, il a soumissionné la distribution, il reçoit ce pain politique avec avidité.
 le pied et le porte, à neuf heures, il est au sein de son ménage, il joue un calembour à sa femme, il dérobe un gros baiser, déguste une tasse de café.
 ou gronde ses enfants, à dix heures, nous un qui, il apparaît à la mairie, il pose sur un fauteuil comme un perroquet sur son bâton, chauffé par la
 ville de Paris, il inscrit jusqu'à quatre heures, sans leur donner une larme ou un sourire, les décès et les naissances de tout un arrondissement. Le
 bonheur, le malheur du quartier passe par le bec de sa plume, comme l'esprit du Constitutionnel voyageait naguère sur ses épaules. Rien ne lui pèse. Il
 va toujours droit devant lui, prend son patriotisme tout fait dans le journal, ne contredit personne, crie ou applaudit avec tout le monde, et vit en
 hirondelle. A deux pas de sa paroisse, il peut, en cas d'une cérémonie importante, laisser sa place à un surnuméraire, et aller chanter un requiem au
 lutrin de l'église, dont il est, le dimanche et les jours de fête, le plus bu-ormement, la voix la plus imposante, où il tord avec énergie sa large bouche en
 faisant tonner un joyeux Amen. Il est oba-té. Libéré à quatre heures de son service officiel, il apparaît pour répandre la joie et la gaieté au sein de la
 boutique la plus célèbre qui soit en la Cité. Heureuse est sa femme, il n'a pas le temps d'être jaloux ; il est plutôt homme d'action que de sentiment.
 Aussi, dès qu'il s'agit de sece-t-il les demoiselles de comptoir, dont les yeux vifs attirent force chalans, se gaudit au sein des parures, des riches, de la
 mousseline façonnée par ces habiles ouvrières, ou, plus souvent encore avant de dîner, il sert une pratique, copie une page du journal, se porte chez
 l'huissier quelque effet en retard, à six heures, tous les deux jours, il est fidèle à son poste, inamovible basse-taille des choeurs, il se trouve à l'Opéra,
 prêt à y devenir soldat, Arabe, prisonnier, sauvage, paysan, ombre, patte de chameau, lion, diable, génie, esclave, cun-que noir ou blanc, toujours
 expert à produire de la joie, de la douleur, de la pitié, de l'étonnement, à pousser d'invariables cri, à se tair, à chasser, à se haïr, à représenter Rome
 ou l'Égypte, mais toujours - in petto, mercier. A midi, il s'élève, bon-mari, homme, tendre-pois, il se glisse dans le lit conjugal, l'imagination
 encore tendue par les formes décevantes des nymphes de l'Opéra, et fait ainsi tourner, au profit de l'amour conjugal, les dépravations du monde et les
 voluptueux ronds de jambe de la Taglioni. Enfin, s'il dort, il dort vite, et dépêche son sommeil comme il a dépêché sa vie.



SYNTAXE ET SÉMANTIQUE XML

D.1 DTD POUR LE CALCUL-ASPECTUEL

Au chapitre 10, nous avons exposé l'implémentation du système *CAsTe*, permettant le calcul aspectuo-temporel d'énoncé itératif. Ce système prend en entrée un document XML devant valider le vocabulaire et sémantique associée à la DTD que nous exposons ci-dessous.

```
<!-- ELEMENT enonce -->
<!ELEMENT enonce ( ( paragraphe+ | structureIterative | relation )* ) >
```

```
<!-- ELEMENT paragraphe -->
<!ELEMENT paragraphe ( phrase+ )* >
<!ATTLIST paragraphe
  id #REQUIRED >
```

```
<!-- ELEMENT phrase -->
<!ELEMENT phrase ( ( proposition+ | relation )+ ) >
<!ATTLIST phrase
  id #REQUIRED >
```

```
<!-- ELEMENT structureIterative -->
<!ELEMENT structureIterative ( #PCDATA ) >
<!ATTLIST structureIterative
  id #REQUIRED
  type ( iteration | selection ) #REQUIRED
  declencheur CDATA #REQUIRED
  structureIterative CDATA #IMPLIED >
```

```
<!-- ELEMENT declencheur -->
<!ELEMENT declencheur ( #PCDATA ) >
<!ATTLIST declencheur
  id #REQUIRED
```

```

structureIterative CDATA #REQUIRED
detache ( true | false ) "false"
valeur CDATA #IMPLIED
proposition CDATA #REQUIRED
type ("calendaire"|"frequentiel"|"quantificationnel"|
      "evenementiel" | "conflit" | "autre") #REQUIRED>

<!-- ELEMENT relation -->
<!ELEMENT relation ( #PCDATA ) >
<!ATTLIST relation
  id #REQUIRED
  cible CDATA #REQUIRED
  source CDATA #REQUIRED
  type #REQUIRED >

<!-- ELEMENT circonstanciel -->
<!ELEMENT circonstanciel ( #PCDATA ) >
<!ATTLIST circonstanciel
  id #REQUIRED
  type #REQUIRED
  valeur CDATA #IMPLIED
  proposition CDATA #REQUIRED
  detache ( true | false ) "false"
  autonome ( true | false ) "false"
  valeur_temporelle ("passe" | "present" | "futur") #IMPLIED
  relation ("Coincidence" | "Anteriorite"|"Posteriorite"|"COpag"|
           "Recouvrement"|"Simultaneite"|"COpad"|"Accessibilite"|
           "AccessibiliteInverse") #REQUIRED >

<!-- ELEMENT proposition -->
<!ELEMENT proposition ( #PCDATA | proposition | declencheur | circonstanciel )* >
<!ATTLIST proposition
  propType CDATA #IMPLIED
  ponctuel ( true | false ) #IMPLIED
  directe ( true | false ) #IMPLIED
  relation CDATA #IMPLIED
  type CDATA #IMPLIED
  id CDATA #REQUIRED
  performatif ( true | false ) #IMPLIED

```

```

vClasse ('a' | 'b' | 'c' | 'd' | 'e' | 'f' | 'g' | 'h' | 'i') #IMPLIED
idIteration CDATA #REQUIRED
vLemme CDATA #REQUIRED
vCategorie ('type1' | 'type2' | 'type3' | 'type4') #IMPLIED
temps ("present"|"imparfait"|"passe-compose"|"plus-que-parfait"
|"passe-simple"|"futur"|"participe-passe"|"participe-present")
#REQUIRED>

```

D.2 DTD DE TIMEML POUR LES ITÉRATIONS

Nous exposons la DTD de l'extension de TIMEML que nous avons présenté au chapitre 11, et qui prend en compte les structures itératives telles que nous les avons définies. Nous avons adapté la DTD de la version française de TIMEML en incluant les modifications apportées dans le cadre de la normalisation de TIMEML, et nos propositions de syntaxes¹.

```

<!ELEMENT TimeML (#PCDATA | ALINK | CONFIDENCE | EVENT | SIGNAL |
SLINK | TIMEX3 | TLINK | TRIGGER | SCHEMA)* >
<!ATTLIST TimeML xsi:noNamespaceSchemaLocation CDATA #IMPLIED >
<!ATTLIST TimeML xmlns:xsi CDATA #IMPLIED >
<!ATTLIST TimeML comment CDATA #IMPLIED >

<!ELEMENT EVENT (#PCDATA) >
<!ATTLIST EVENT eid ID #REQUIRED >
<!ATTLIST EVENT eiid ID #REQUIRED >
<!ATTLIST EVENT class (ASPECTUAL | I_ACTION | I_STATE | OCCURRENCE |
PERCEPTION | REPORTING | STATE) #REQUIRED >
<!ATTLIST EVENT stem CDATA #IMPLIED >
<!ATTLIST EVENT comment CDATA #IMPLIED >
<!ATTLIST EVENT signalID IDREF #IMPLIED >
<!ATTLIST EVENT pos (ADJECTIVE | NOUN | VERB | PREPOSITION | NONE) #RE-
QUIRED >
<!ATTLIST EVENT tense (FUTUR | INFINITIF | NONE | PS | PC | PRESENT | PQP |
IMPARFAIT | PARTPRESENT | PARTPASSE) #REQUIRED >
<!ATTLIST EVENT aspect (NONE | INACCOMPLI | ACCOMPLI | AORISTIQUE |
PROSPECTIF) #REQUIRED >
<!ATTLIST EVENT cardinality CDATA #IMPLIED >
<!ATTLIST EVENT schemaInstanceID IDREF #IMPLIED >
<!ATTLIST EVENT polarity (POS | NEG) #REQUIRED >

```

1. La DTD de notre extension peut être téléchargée à l'adresse suivante :
<http://jlebranc.perso.info.unicaen.fr/timeml/timeml++.dtd>

<!ATTLIST EVENT modality CDATA #IMPLIED >
 <!ATTLIST EVENT comment CDATA #IMPLIED >

<!ELEMENT TIMEX3 (#PCDATA) >
 <!ATTLIST TIMEX3 tid ID #REQUIRED >
 <!ATTLIST TIMEX3 type (DATE | DURATION | SET | TIME) #REQUIRED >
 <!ATTLIST TIMEX3 value NMTOKEN #REQUIRED >
 <!ATTLIST TIMEX3 anchorTimeID IDREF #IMPLIED >
 <!ATTLIST TIMEX3 beginPoint IDREF #IMPLIED >
 <!ATTLIST TIMEX3 endPoint IDREF #IMPLIED >
 <!ATTLIST TIMEX3 freq NMTOKEN #IMPLIED >
 <!ATTLIST TIMEX3 functionInDocument (CREATION_TIME | EXPIRATION_TIME | NONE | MODIFICATION_TIME | PUBLICATION_TIME | RELEASE_TIME | RECEPTION_TIME) #IMPLIED >
 <!ATTLIST TIMEX3 mod (BEFORE | AFTER | ON_OR_BEFORE | ON_OR_AFTER | LESS_THAN | MORE_THAN | EQUAL_OR_LESS | EQUAL_OR_MORE | START | MID | END | APPROX) #IMPLIED >
 <!ATTLIST TIMEX3 quant CDATA #IMPLIED >
 <!ATTLIST TIMEX3 temporalFunction (false | true) #IMPLIED >
 <!ATTLIST TIMEX3 valueFromFunction IDREF #IMPLIED >
 <!ATTLIST TIMEX3 comment CDATA #IMPLIED >

<!ELEMENT SIGNAL (#PCDATA) >
 <!ATTLIST SIGNAL sid ID #REQUIRED >
 <!ATTLIST SIGNAL comment CDATA #IMPLIED >

<!ELEMENT ALINK EMPTY >
 <!ATTLIST ALINK lid ID #REQUIRED >
 <!ATTLIST ALINK relType (CONTINUES | CULMINATES | INITIATES | REINITIATES | TERMINATES) #REQUIRED >
 <!ATTLIST ALINK eventInstanceID IDREF #REQUIRED >
 <!ATTLIST ALINK relatedToEventInstance IDREF #REQUIRED >
 <!ATTLIST ALINK signalID IDREF #IMPLIED >
 <!ATTLIST ALINK syntax CDATA #IMPLIED >
 <!ATTLIST ALINK comment CDATA #IMPLIED >

<!ELEMENT SLINK EMPTY >
 <!ATTLIST SLINK lid ID #REQUIRED >
 <!ATTLIST SLINK relType (CONDITIONAL | COUNTER_FACTIVE | EVIDENTIAL | FACTIVE | MODAL | NEG_EVIDENTIAL) #REQUIRED >
 <!ATTLIST SLINK eventInstanceID NMTOKEN #REQUIRED >
 <!ATTLIST SLINK subordinatedEventInstance NMTOKEN #REQUIRED >
 <!ATTLIST SLINK signalID NMTOKEN #IMPLIED >
 <!ATTLIST SLINK syntax CDATA #IMPLIED >
 <!ATTLIST SLINK comment CDATA #IMPLIED >

```

<!ELEMENT TLINK EMPTY >
<!ATTLIST TLINK lid ID #REQUIRED >
<!ATTLIST TLINK relType (BEFORE | AFTER | INCLUDES | IS_INCLUDED | DUR-
ING | DURING_INV | SIMULTANEOUS | IAFTER | IBEFORE | IDENTITY | BEGINS |
ENDS | BEGUN_BY | ENDED_BY) #REQUIRED >
<!ATTLIST TLINK eventInstanceID IDREF #IMPLIED >
<!ATTLIST TLINK timeID IDREF #IMPLIED >
<!ATTLIST TLINK relatedToEventInstance IDREF #IMPLIED >
<!ATTLIST TLINK relatedToTime IDREF #IMPLIED >
<!ATTLIST TLINK signalID IDREF #IMPLIED >
<!ATTLIST TLINK origin CDATA #IMPLIED >
<!ATTLIST TLINK syntax CDATA #IMPLIED >
<!ATTLIST TLINK comment CDATA #IMPLIED >

```

```

<!ELEMENT CONFIDENCE EMPTY >
<!ATTLIST CONFIDENCE tagType (EVENT | TIMEX3 | SIGNAL | ALINK | SLINK |
TLINK | SCHEMA | TRIGGER) #REQUIRED >
<!ATTLIST CONFIDENCE tagID IDREF #REQUIRED >
<!ATTLIST CONFIDENCE attributeName CDATA #IMPLIED >
<!ATTLIST CONFIDENCE confidenceValue CDATA #REQUIRED >
<!ATTLIST CONFIDENCE comment CDATA #IMPLIED >

```

```

<!ELEMENT SCHEMA EMPTY >
<!ATTLIST SCHEMA scid ID #REQUIRED >
<!ATTLIST SCHEMA type (ITERATION | SELECTION) #REQUIRED >
<!ATTLIST SCHEMA value NMTOKEN #REQUIRED >
<!ATTLIST SCHEMA triggerID IDREF #REQUIRED >
<!ATTLIST SCHEMA aspect (NONE | INACCOMPLI | ACCOMPLI | AORISTIQUE |
PROSPECTIF) #REQUIRED >

```

```

<!ELEMENT TRIGGER (#PCDATA) >
<!ATTLIST TRIGGER trid ID #REQUIRED >
<!ATTLIST TRIGGER tmid ID #IMPLIED >
<!ATTLIST TRIGGER type (QUANT | FREQ | EV | CAL | ASP | CONFLIT) #RE-
QUIRED >
<!ATTLIST TRIGGER value NMTOKEN #REQUIRED >
<!ATTLIST TRIGGER instanceID IDREF #IMPLIED >

```



RÉFÉRENCES BIBLIOGRAPHIQUES

BIBLIOGRAPHIE

- James F ALLEN : Maintaining knowledge about temporal intervals. *Commun. ACM*, 26 (11):832–843, 1983. ISSN 0001-0782. URL <http://portal.acm.org/citation.cfm?id=358434>. (Cité aux pages 32, 129 et 196.)
- Nicholas ASHER : *Reference to abstract objects in discourse*. Kluwer, 1993. (Cité à la page 29.)
- Nicholas ASHER et Alex LASCARIDES : *Logics of Conversation (Studies in Natural Language Processing)*. Cambridge University Press, 2005. ISBN 0521659515. (Cité à la page 29.)
- Jon BARWISE et Robin COOPER : Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, 4(2):159–219, 1981. ISSN 01650157. URL <http://www.springerlink.com/index/n7231j788435026u.pdf>. (Cité à la page 19.)
- Delphine BATTISTELLI : *La temporalité linguistique : circonscrire un objet d'analyse ainsi que des finalités à cette analyse*. Hdr, Université de Nanterre - Paris X, novembre 2009. URL <http://tel.archives-ouvertes.fr/tel-00452464/en/>.
- Delphine BATTISTELLI et Marie CHAGNOUX : Vers un outil de visualisation de la dynamique textuelle : l'exemple des phénomènes citationnels et modaux, avril 2008. URL <http://hal.archives-ouvertes.fr/hal-00336859/en/>. article mis en forme. (Cité aux pages 42, 44 et 45.)
- Pier Marco BERTINETTO : Sulle proprietà tempo-aspettuali dell'infinito italiano. In M GIACOMO- MARCELLESI et A ROCCHETTI, éditeurs : *Il verbo italiano*, pages 113–165, Bulzoni, Roma, 2003. (Cité à la page 194.)
- Frédéric BILHAUT : *Analyse automatique de structures thématiques discursives : Application à la recherche d'information*. Thèse de doctorat, Université de Caen Basse-Normandie, 304 pages, 2006. URL <http://tel.archives-ouvertes.fr/docs/00/25/87/66/PDF/these-bilhaut.pdf>. (Cité à la page 136.)
- Frédéric BILHAUT et Antoine WIDLÖCHER : La plate-forme LinguaStream : un environnement intégré pour l'expérimentation en TAL. In *TALN 2006*, Leuven, Belgique, 2006. (Cité à la page 136.)
- André BITTAR : *Building a TimeBank for French : A Reference Corpus Annotated According to the ISO-TimeML Standard*. Thèse de doctorat, Université Paris Diderot (Paris 7), 301 pages, 2010. URL <http://www.linguist.univ-paris-diderot.fr/~abittar/docs/Bittar-PhD.pdf>. (Cité aux pages 194 et 206.)
- Joan BUSQUETS, Laure VIEU et Nicholas ASHER : La SDRT : une approche de la cohérence du discours dans la tradition de la sémantique dynamique. *Verbum*, XXIII (1):73–101, 2001. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.9814&rep=rep1&type=pdf>.

- Lauri CARLSON : Aspect and Quantification. In *Syntax and Semantics, Volume 14 : Tense and Aspect*, pages 31–64. Academic Press, 1981. (Cit      la page 15.)
- Marie CHAGNOUX : *Temporalit   et aspectualit   dans les textes fran  ais : mod  lisation s  mantico-cognitive et traitement informatique*. Th  se de doctorat, Universit   Paris IV Sorbonne, 256 pages, 2006. URL <http://marie.chagnoux.free.fr/Communications/TheseChagnoux.pdf>. (Cit   aux pages 41 et 42.)
- Marie CHAGNOUX : Informer sans s’engager : variations de prise en charge   nonciative dans les sujets d’actualit  , ao  t 2008. URL <http://hal.archives-ouvertes.fr/hal-00336855/en/>. article achev  . (Cit      la page 44.)
- Michel CHAROLLES : L’encadrement du discours : univers, champs, domaines et espaces. *Cahier de recherche linguistique*, 6(4):9–30, 1997. URL <http://www.lattice.cnrs.fr/IMG/pdf/cadresdisc.pdf>. (Cit   aux pages 42, 44, 103, 113, 115 et 136.)
- Michel CHAROLLES, Anne LE DRAOULEC, Marie-Paule P  RY-WOODLEY et Laure SARDA : Temporal and spatial dimensions of discourse organisation. *Journal of French Language Studies*, 15:203–218, 2005. (Cit      la page 119.)
- Michel CHAROLLES et Marie-Paule P  RY-WOODLEY : Les adverbiaux cadratifs. *Linguistique*, pages 1–6, 2005. URL http://w3.erss.univ-tlse2.fr:8080/index.jsp?perso=pery&subURL=articles/IntroductionLF_05.pdf. (Cit      la page 119.)
- Michel CHAROLLES et Denis VIGIER : Les adverbiaux en position pr  verbiale : port  e cadrative et organisation des discours. *Langue Fran  aise*, (148): 9–30, 2005. URL http://halshs.archives-ouvertes.fr/halshs-00373342/PDF/Charolles_Vigier_Langue_Francaise_2005_148.pdf.
- Anne CONDAMINES : Les conjonctions de subordination temporelle en fran  ais. *Cahiers de Grammaires*, 15:25–54, 1990. (Cit      la page 21.)
- Anne CONDAMINES : Expression de la r  p  tition dans une phrase subordonn  e temporelle. *La quantit   et ses modulations qualitatives*, 1992. (Cit   aux pages 21 et 152.)
- Diana CUKIERMAN et James DELGRANDE : Characterizing Temporal Repetition. *IEEE*, pages 80–87, 1996. URL <http://www.aaai.org/Papers/AAAI/1996/AAAI96-233.pdf>. (Cit      la page 19.)
- Diana CUKIERMAN et James DELGRANDE : A formalization of structured temporal objects and repetition. In *Proceedings of the Seventh International Workshop on Temporal Representation and Reasoning (TIME’00)*, pages 13–, Washington, DC, USA, 2000. IEEE Computer Society. ISBN 0-7695-0756-5. URL <http://portal.acm.org/citation.cfm?id=832248.832543>. (Cit      la page 19.)
- James CURRAN, Stephen CLARK et Johan Bos : Linguistically Motivated Large-Scale NLP with C&C and Boxer. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 33–36, Prague, Czech Republic, 2007. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/P07-2009>. (Cit      la page 29.)

- Laurence DANLOS, Bertrand GAIFFE et Laurent ROUSSARIE : Document Structuring à la SDRT. In *Proceedings of the European Workshop on Generation, ACL, Toulouse Proceedings of the European Workshop on Generation, ACL, Toulouse*, Toulouse, 2001. ACL. URL <http://halshs.archives-ouvertes.fr/halshs-00082836/PDF/80.pdf>.
- Jean-Pierre DESCLÉS : *Langages applicatifs, langues naturelles et cognition*. Paris : Hermès, 1990. (Cité à la page 40.)
- Jean-Pierre DESCLÉS et Zlatka GUENTCHÉVA : *Notion d'aspect*, chapitre Construction formelle de la catégorie grammaticale de l'aspect (essai), pages 195–237. Klincksieck, 1980. (Cité à la page 40.)
- David DOWTY : The effects of aspectual class on the temporal structure of discourse : semantics or pragmatics. *Linguistics and Philosophy*, 9:37–61, 1986. (Cité à la page 32.)
- Patrice ENJALBERT et Gérard BECHER : *Aspects de l'itération : une approche pluridisciplinaire*, chapitre L'itération dans le modèle SdT : une formalisation algébrique. 2011 – à paraître. (Cité aux pages 4, 18, 59, 60, 205 et 206.)
- Gilles FAUCONNIER : *Espaces mentaux. Aspects de la construction du sens dans les langues naturelles*. Minuit, 1984. (Cité à la page 71.)
- Lisa FERRO, Laurie GERBER, Inderjeet MANI, Beth SUNDHEIM et George WILSON : *TIDES - Standard for the Annotation of Temporal Expressions*. MITRE Washington C3 Center, McLean, Virginia, avril 2005. URL http://fofoca.mitre.org/annotation-guidelines/2005_timex2_standard_v1.1.pdf. (Cité à la page 199.)
- Laurent GOSSELIN : *Sémantique de la temporalité en français : un modèle calculatoire et cognitif du temps et de l'aspect*. Duculot, 1996. (Cité aux pages 3, 6, 32, 35, 39, 41, 51, 185 et 269.)
- Laurent GOSSELIN : *Temporalité et modalité*. de Boeck, 2005. (Cité aux pages 3, 32, 39 et 51.)
- Laurent GOSSELIN : *Aspects de l'itération : une approche pluridisciplinaire*, chapitre L'itération dans le modèle SdT. 2011 – à paraître. (Cité aux pages 7, 15, 52 et 56.)
- Laurent GOSSELIN : Aspect itératif et pluralité. In Dany AMIOT, Walter DE MULDER, Estelle MOLINE et Dejan STOSIC, éditeurs : *Hommages à Nelly Flaux*, pages 371–388. Ars Grammatica, Peter Lang, Berne, 2011. (Cité aux pages 4 et 274.)
- Laurent GOSSELIN, Patrice ENJALBERT, Gérard BECHER et Yann MATHET : *Aspects de l'itération : une approche pluridisciplinaire*. 2011 – à paraître. (Cité à la page 49.)
- Gaston GROSS : Prédicats nominaux et compatibilité aspectuelle. *Langages*, 30 (121):54–72, 1996. ISSN 0458-726X. URL http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726x_1996_num_30_121_1740. (Cité à la page 213.)
- Gaston GROSS et Ferenc KIEFER : La structure événementielle des substantifs. *Folia Linguistica*, 29(1-2):43–66, 1995. URL <http://www.reference-global.com/doi/abs/10.1515/flin.1995.29.1-2.43>. (Cité à la page 79.)

- Nabil HATHOUT, Fiammetta NAMER et Georgette DAL : An Experimental Constructional Database : The MorTAL Project. In Paul BOUCHER, éditeur : *Many Morphologies*, pages 178–209. Cascadilla, Somerville, Mass., 2002. (Cité à la page 140.)
- Hans KAMP : A Theory of Truth and Semantic Representation. In *Formal Methods in the Study of Language, Vol. I, Mathematische Centrum, Amsterdam. Reprinted in Groenendijk, Janssen and Stokhof (eds.) (1984) Truth, Interpretation and Information, Foris, 1981.*
- Hans KAMP et Uwe REYLE : *From Discourse to Logic : Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Studies in Linguistics and Philosophy. Kluwer Academic Publishers, Dordrecht, 1993. (Cité aux pages 19, 25, 27 et 28.)
- Edward L. KEENAN : The semantics of determiners. In *The Handbook of Contemporary Semantic Theory*, pages 41–63. Blackwell, 1996. (Cité à la page 19.)
- Ferene KIEFER : Les substantifs déverbaux événementiels. *Langages*, 32(131):56–63, 1998. ISSN 0458-726X. URL http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726x_1998_num_32_131_2167. (Cité aux pages 79, 142 et 213.)
- Georges KLEIBER : *Du côté de la référence verbale : les phrases habituelles*. Numéro 19 in Sciences de la communication. Peter Lang, Berne, 1987. (Cité à la page 15.)
- Brenda LACA : Indefinites, quantifiers, and pluractionals. What scope effects tell us about event pluralities. In Vogeleer SVETLANA et Liliane TASMOWSKI, éditeurs : *Non-definiteness and plurality*, Linguistik Aktuell. Benjamins, 2005. URL <http://halshs.archives-ouvertes.fr/docs/00/10/46/43/PDF/LacaIndPlurac.fin.pdf>. (Cité aux pages 16 et 17.)
- Anne LE DRAOULEC et Marie-Paule PÉRY-WOODLEY : Corpus-based identification of temporal organisation in discourse. *Proceedings of the Corpus*, pages 159–166, 2001. URL <http://ucrel.lancs.ac.uk/publications/CL2003/CL2001conference/papers/draoulec.pdf>. (Cité à la page 151.)
- Anne LE DRAOULEC et Marie-Paule PÉRY-WOODLEY : Time travel in text : temporal framing in narratives and non-narratives. *Determination of Information and Tenor in Texts : Multidisciplinary Approaches to Discourse*, pages 267–275, 2003. URL <http://w3.erss.univ-tlse2.fr:8080/index.jsp?perso=pery&subURL=articles/MAD03.pdf>. (Cité à la page 119.)
- Anne LE DRAOULEC et Marie-Paule PÉRY-WOODLEY : Encadrement temporel et relations de discours. *Langue française*, (4):45–60, 2005. ISSN 0023-8368. URL <http://www.univ-tlse2.fr/erss/textes/pagespersos/draoulec/publis/LangueFrancaise.pdf>. (Cité à la page 119.)
- Julien LEBRANCHU : Vers une analyse automatique des phénomènes itératifs en corpus. In G WILLIAMS, éditeur : *Actes des 6èmes Journées de la Linguistique de Corpus*, Lorient, France, septembre 2009. Université de Bretagne-Sud. URL <http://jlebranc.perso.info.unicaen.fr/publication/JLC2009-analyse-automatique-phenomenes-iteratifs-corpus.pdf>.

- Julien LEBRANCHU et Yann MATHET : Vers une prise en charge approfondie des phénomènes itératifs par TimeML. In *Actes de TALN 2011*, Montpellier, France, 2011. URL <http://jlebranc.perso.info.unicaen.fr/publication/taln2011-timeml-phenomenes-iteratifs.pdf>.
- Kiyong LEE, James PUSTEJOVSKY, Harry BUNT, Branimir BOGURAEV et Nancy IDE : Language resource management - Semantic annotation framework (SemAF) - Part 1 : Time and events. Rapport technique, International Organization for Standardization, Geneva, Switzerland, 2007. URL http://lirics.loria.fr/doc_pub/SemAFCD24617-1Rev12.pdf. (Cité aux pages 194 et 200.)
- Florence LEFEUVRE et David NICOLAS : La phrase nominale existentielle et la distinction aspectuelle télélique/atélique. *Revue de Sémantique et Pragmatique*, 14(14):101–119, 2004. URL http://halshs.archives-ouvertes.fr/docs/00/13/80/90/PDF/Telicite_et_PNexistentielleFLDN.pdf. (Cité à la page 213.)
- François LÉVY, Gérard BECHER, Patrice ENJALBERT, Estelle FIEVÉ, Laurent GOSSELIN et Gérard LIGOZAT : Rapport Technique du projet OGRE (Ordres de Grandeur et Répétition). Rapport technique, CTAN, 2005. URL <http://arxiv.org/pdf/cs/0610004>. (Cité aux pages 3 et 49.)
- Jung-Hae LIM : *La fréquence et son expression en français*. Honoré Champion, 2002. (Cité aux pages 20, 53 et 139.)
- Laurent MASCHERIN : *Analyse morphosémantique de l'aspectuo-temporalité en français. Le cas du préfixe RE-*. Thèse de doctorat, Université de Nancy 2, 305 pages, 2007. URL <http://tel.archives-ouvertes.fr/docs/00/28/06/60/PDF/2007NAN21011.pdf>. (Cité à la page 53.)
- Yann MATHET : Une approche cognitive de l'itération et sa modélisation objet. *RTE*, 2007. (Cité aux pages 4, 5, 69 et 274.)
- Yann MATHET : *Aspects de l'itération : une approche pluridisciplinaire*, chapitre Une approche objet et sa modélisation cognitive. 2011 – à paraître. (Cité aux pages 69 et 167.)
- Henning NOLKE : *Linguistique modulaire : de la forme au sens*. Peeters, 1994. (Cité à la page 63.)
- Barbara H PARTEE : *Noun Phrase Interpretation and Type-Shifting Principles*, volume 8 de *Groningen-Amsterdam Studies in Semantics*, chapitre 5, pages 115–143. Foris Publications, 1987. (Cité à la page 19.)
- Cédric PERSON : *Traitement de la temporalité du récit*. Thèse de doctorat, Université de Caen, 182 pages, 2004. (Cité aux pages 3, 7, 39, 159, 170, 225 et 239.)
- Marcel PROUST : *À la recherche du temps perdu*. Gallimard, 1913. (Cité aux pages iii et 5.)

- James PUSTEJOVSKY, José CASTAÑO, Robert INGRÍA, Roser SAURÍ, Rob GAIZAUSKAS, Andrea SETZER et Graham KATZ : TimeML : A Specification Language for Temporal and Event Expressions. In *Proceedings of the International Workshop of Computational Semantics (IWCS-2003)*, Tilburg, The Netherlands, 2003. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.85.5610&rep=rep1&type=pdf>. (Cité à la page 193.)
- Hans REICHENBACH : *Elements of symbolic logic*. Free Press, New York, 1947. (Cité à la page 30.)
- Laure SARDA : Fonctionnement des cadres spatiaux dans les résumés de films. In Michel CHAROLLES et Marie-Paule PÉRY-WOODLEY, éditeurs : *Langue Française*, volume 148, pages 61–79, 2005. URL http://halshs.archives-ouvertes.fr/docs/00/06/78/16/PDF/Cadres_spatiauxLF2005.pdf. (Cité à la page 119.)
- Roser SAURÍ, Jessica LITTMAN, Robert GAIZAUSKAS, Andrea SETZER et James PUSTEJOVSKY : *TIMEML Annotation Guidelines, Version 1.2.1*, 2006. URL http://www.timeml.org/site/publications/timeMLdocs/annguide_1.2.1.pdf. (Cité à la page 198.)
- Jung Ho SHIN, Young S HAN et Key-Sun CHOI : A HMM Part-of-Speech Tagger for korean with Wordphrasal Relations. In Ruslan MITKOV et Nicolas NICOLOV, éditeurs : *Recent Advances in Natural Language Processing : Selected Papers from RANLP'95*, volume 136 de *Current Issues in Linguistic Theory (CILT)*, pages 439–449, Amsterdam/Philadelphia, 1997. John Benjamins. (Cité à la page 194.)
- Elise TERRAN : *Le cadrage temporel en français*. Thèse de doctorat, Université Paris III - Sorbonne Nouvelle, 2002. (Cité aux pages 42 et 43.)
- Lucia M. TOVENA et Marta DONAZZAN : On ways of repeating. *Recherches linguistiques de Vincennes*, 2008. URL <http://rlv.revues.org/1724>.
- Veerle VAN GEENHOVEN : *For*-adverbials, frequentative aspect, and pluractionality. *Natural Language Semantics*, 12:135–190, 2004. ISSN 0925-854X. URL <http://dx.doi.org/10.1023/B:NALS.0000031389.36427.af>. (Cité aux pages 16 et 17.)
- Veerle VAN GEENHOVEN : Atelicity, pluractionality, and adverbial quantification. In Henk VERKUYL, Henriette SWART et Angeliek HOUT, éditeurs : *Perspectives on Aspect*, volume 32 de *Studies in Theoretical Psycholinguistics*, pages 107–124. Springer Netherlands, 2005. ISBN 978-1-4020-3232-5. URL http://dx.doi.org/10.1007/1-4020-3232-3_6. (Cité aux pages 16, 17 et 18.)
- Zeno VENDLER : *Linguistics in Philosophy*. Cornell U. P, 1967. (Cité aux pages 35, 125, 146, 194, 217 et 218.)
- Co VET : *Temps, aspects et adverbes de temps en français contemporain. Essai de sémantique formelle*. Genève : Librairie Droz, 1980. (Cité à la page 30.)
- Denis VIGIER : Les adverbiaux praxéologiques détachés en position initiale et leur portée. *Verbum*, XXVII(3):293–312, 2005. URL <http://halshs.archives-ouvertes.fr>.

[fr/halshs-00349248/PDF/Vigier_D._2005_Les_adverbiaux_praxeologiques_detaches_en_position_initiale_et_leur_portee_Verbum_XXVII.pdf](http://halshs-00349248/PDF/Vigier_D._2005_Les_adverbiaux_praxeologiques_detaches_en_position_initiale_et_leur_portee_Verbum_XXVII.pdf). (Cité à la page 119.)

Antoine WIDLÖCHER : *Analyse macro-sémantique des structures rhétoriques du discours - Cadre théorique et modèle opératoire*. Thèse de doctorat, Université de Caen Basse-Normandie, Caen, 396 pages, 2008. (Cité aux pages 91, 120 et 154.)

Antoine WIDLÖCHER et Yann MATHET : La plate-forme Glozz : environnement d'annotation et d'exploration de corpus. In *Actes de TALN 2009*, Senlis, France, 2009. URL http://www-lipn.univ-paris13.fr/taln09/pdf/TALN_120.pdf. (Cité à la page 92.)



TABLE DES MATIÈRES

INTRODUCTION	1
Les phénomènes itératifs	3
L'itération : quelle manifestation textuelle?	4
La temporalité des phénomènes itératifs	6
Plan de la thèse	8
I LA QUESTION DE LA TEMPORALITÉ ET DES ITÉRATIONS	11
1 LES PHÉNOMÈNES ITÉRATIFS DANS LA LITTÉRATURE	15
1.1 Linguistique d'« inspiration » formelle	15
1.1.1 Les phrases habituelles	15
1.1.2 La pluriactionnalité	16
1.2 Linguistique formelle	18
1.3 Linguistique descriptive	20
1.3.1 La fréquence et son expression	20
1.3.2 Les phrases à subordinées temporelles	21
2 DES MODÈLES DE LA TEMPORALITÉ	25
2.1 Modèles référentiels	25
2.1.1 DRT : Discourse Representation Theory	25
2.1.2 SDRT : Segmented Discourse Representation Theory	29
2.2 Modèles aspecto-temporels	30
2.2.1 Reichenbach : un modèle temporel de la référence	30
2.2.2 Le modèle de la Sémantique de la Temporalité de L. Gosselin	32
2.2.3 Le modèle de J.-P. Desclés et de Z. Guentchéva	40
2.3 Modèles discursifs	42
2.3.1 Les cadres du discours	42
2.3.2 Les cadres temporels	43
2.3.3 La dynamique énonciative de surface	44
II DE LA SDT AUX REPRÉSENTATIONS DISCURSIVES	47
3 L'ITÉRATION DANS LE MODÈLE SDT	51
3.1 Inscription temporelle de l'itération	51
3.2 Sources de l'itération	53
3.3 Calcul des visées aspectuelles	55
4 UNE FORMALISATION ALGÈBRIQUE	59
4.1 Les expressions calendaires itératives	60

4.2	L'ontologie temporelle	61
4.3	La relation circonstancielle itérative	62
4.3.1	Déterminer les séries à mettre en relations	62
4.3.2	Calculer la relation circonstancielle	63
4.4	Les adverbes fréquentatifs	66
5	L'ITÉRATION SELON UNE APPROCHE OBJET	69
5.1	Hypothèse cognitive	69
5.2	Modélisation objet	72
5.2.1	L'itération	73
5.2.2	Les itérateurs	74
5.2.3	Le modèle itératif et ses constituants	75
5.2.4	Les sélections	76
5.3	Lien avec la <i>SdT</i>	77
5.4	Discussion	78
III LES PHÉNOMÈNES ITÉRATIFS EN CORPUS		83
6	OBSERVATION DES PHÉNOMÈNES ITÉRATIFS	87
6.1	Constitution et description du corpus [ITER]	87
6.1.1	Les articles du journal [LE MONDE]	88
6.1.2	Le sous-corpus [ROMAN]	89
6.2	Annotation manuelle du corpus	90
6.2.1	Méthode d'annotation	90
6.2.2	Outil d'aide à la tâche d'annotation	91
6.3	Premiers constats sur les phénomènes itératifs dans le corpus [ITER]	92
6.3.1	Les structures itératives	92
6.3.2	Les substantifs événementiels	99
6.3.3	Les déclencheurs	100
6.4	Organisation textuelle des structures itératives	102
6.4.1	La position initiale des déclencheurs	103
6.4.2	La reprise événementielle	104
6.4.3	La distance entre les structures itératives	105
6.4.4	Exemples de configurations textuelles	107
7	REPRÉSENTATION DES STRUCTURES ITÉRATIVES	113
7.1	Organisation du discours	113
7.1.1	Cadre du discours	113
7.1.2	Cadre et structure itérative	116
7.2	Unité - Relation - Schéma	120
7.2.1	Les unités discursives	121
7.2.2	Les relations discursives	122
7.2.3	Les schémas discursifs	123

7.3	Représentation textuelle des structures itératives	125
7.3.1	Les procès	125
7.3.2	Les compléments circonstanciels de temps	126
7.3.3	Les déclencheurs des structures itératives	127
7.3.4	Les structures itératives	128
7.4	Exemples d'annotation	132
8	ANALYSEUR DES STRUCTURES ITÉRATIVES	135
8.1	Introduction	135
8.2	Traitements de base pour une analyse en TAL	137
8.2.1	Segmentation : du texte aux mots	137
8.2.2	Segmentation en phrases	138
8.3	Repérage des constituants	139
8.3.1	Traitement des procès	139
8.3.2	Traitement du temps	142
8.3.3	Traitement de l'aspect	145
8.3.4	Traitement des déclencheurs	147
8.4	Analyse des structures itératives	150
8.4.1	Méthode d'analyse	150
8.4.2	Du point de vue de l'implémentation	153
IV	LES PHÉNOMÈNES ITÉRATIFS ET LE CALCUL ASPECTUEL	157
9	LE CALCUL ASPECTUO-TEMPOREL	161
9.1	Principes du calcul de la représentation	161
9.2	Des procès semelfactifs aux procès itératifs	163
9.3	La temporalité absolue des procès modèles	166
9.4	Les circonstanciels de temps	167
9.5	La dimension textuelle	170
9.5.1	La portée des connecteurs temporels	171
9.5.2	La récursivité	172
9.5.3	La sélection	174
10	LE SYSTÈME <i>caste</i>	179
10.1	Description des informations aspectuo-temporelles	180
10.2	Principes de fonctionnement de <i>CAsTe</i>	183
10.3	Résultats commentés	186
11	UN FORMALISME XML POUR LES STRUCTURES ITÉRATIVES	193
11.1	Qu'est-ce que TIMEML	193
11.1.1	Les entités de TIMEML	194
11.1.2	Les relations de TIMEML	196
11.2	Les phénomènes itératifs dans TIMEML	197
11.2.1	Les structures itératives	197

11.2.2	Les déclencheurs	198
11.3	TIMEML ++ : traitement des phénomènes itératifs	200
11.3.1	Les structures itératives	200
11.3.2	Les déclencheurs	201
CONCLUSION		209
ANNEXES		215
A	LEXIQUE DES TYPOLOGIES DE PROCÈS	217
A.1	Classification	219
A.2	Exemple avec <i>Rouler</i>	220
A.3	Démarche & résultat	221
B	RÈGLES DE CALCUL ASPECTUO-TEMPOREL	225
B.1	Les phénomènes itératifs	226
B.2	Les temps morphologiques	226
B.3	Les déclencheurs	233
B.4	La subordination temporelle	238
B.5	Les relations circonstancielles	239
B.6	Les circonstants de durée de type [depuis+durée]	243
C	EXTRAIT DE <i>la fille aux yeux d'or</i>	245
D	SYNTAXE ET SÉMANTIQUE XML	247
D.1	DTD pour le calcul-aspectuel	247
D.2	DTD de TIMEML pour les itérations	249
RÉFÉRENCES BIBLIOGRAPHIQUES		253
Bibliographie		255
Table des matières		263
Table des figures		267
Liste des tableaux		270
Liste des règles		271

TABLE DES FIGURES

FIGURE 1	Représentation schématique d'une itération.	6
FIGURE 2	Représentation de la construction d'une DRS	26
FIGURE 3	Illustration d'une DRS sur la pluralité.	28
FIGURE 4	Illustration d'une DRS sur la quantification.	29
FIGURE 5	Exemple de SDRS avec les relations de discours.	30
FIGURE 6	Représentation du temps dans le modèle de Reichenbach.	31
FIGURE 7	Représentation des intervalles de phrases simples	33
FIGURE 8	Représentation des intervalles de phrases complexes	34
FIGURE 9	Liste des relations temporelles au sein de la <i>SdT</i>	35
FIGURE 10	Échelle de saillance relative pour le calcul du degré de saillance des intervalles.	38
FIGURE 11	Exemple de structure aspectuo-temporelle globale	40
FIGURE 12	Représentation de l'itération dans la GAC	41
FIGURE 13	Représentation de l'itération dans la GAC	42
FIGURE 14	Structure aspectuo-temporelle d'une itération simple	52
FIGURE 15	Structure aspectuo-temporelle d'un conflit	55
FIGURE 16	Exemple d'application du calcul des visées aspectuelles	57
FIGURE 17	Opération de restriction de deux séries	60
FIGURE 18	Représentation d'une expression calendaire itérative.	61
FIGURE 19	Représentation d'un procès type.	64
FIGURE 20	Représentation d'une série d'itérés : CC/Procès	64
FIGURE 21	Représentation d'une série d'itérés : Procès/CC	65
FIGURE 22	Représentation des configurations avec un CC_{unit}	65
FIGURE 23	MOPI : processus de construction d'une itération	70
FIGURE 24	MOPI : espaces mentaux	71
FIGURE 25	MOPI : schématisation « objet » d'une itération	73
FIGURE 26	MOPI : diagramme UML	74
FIGURE 27	MOPI : schématisation du phénomène de récursivité d'une ité- ration	76
FIGURE 28	MOPI : schématisation « objet » d'une sélection	77
FIGURE 29	Représentation au sein de MOPI d'exemples d'itérations fon- dées sur un substantif événementiel.	80
FIGURE 30	Interface de GLOZZ permettant l'annotation et l'exploration de corpus.	92
FIGURE 31	Répartition des itérations dans le corpus [ITER] selon leur taille.	96

FIGURE 32	Répartition des sélections dans le corpus [ITER] selon leur taille.	96
FIGURE 33	Répartition des déclencheurs par type dans le corpus [ITER].	100
FIGURE 34	Analyse partielle des cadres de discours de l'extrait précédent.	116
FIGURE 35	Représentation des univers de discours sur un texte itératif.	118
FIGURE 36	Représentation du conflit de subordination entre deux cadres.	119
FIGURE 37	Représentation du non respect du principe de vériconditionnalité.	120
FIGURE 38	Le modèle unités, relations et schémas	121
FIGURE 39	Le modèle U-R-S : les unités	122
FIGURE 40	Le modèle U-R-S : les relations	122
FIGURE 41	Le modèle U-R-S : les schémas	123
FIGURE 42	Le modèle U-R-S : la propriété de récursivité	124
FIGURE 43	Le modèle U-R-S : la propriété de discontinuité	124
FIGURE 44	Représentation du modèle d'annotation des différents procès de l'énoncé.	126
FIGURE 45	Modèle d'annotation des compléments circonstanciels de temps.	127
FIGURE 46	Modèle d'annotation du procès et des déclencheurs.	127
FIGURE 47	Représentation du modèle d'annotation d'un déclencheur issu d'un conflit.	128
FIGURE 48	Modèle d'annotation des structures itératives.	129
FIGURE 49	Représentation des différentes configurations associées à une structure itérative.	130
FIGURE 50	Représentation du modèle d'annotation d'une sélection.	131
FIGURE 51	Représentation du modèle d'annotation d'une sélection sans déclencheurs.	131
FIGURE 52	Représentation du modèle d'annotation sur un extrait de <i>La fille aux yeux d'or</i> de Balzac.	132
FIGURE 53	Exemple de structures de trait des circonstanciels semelfactif et itératif.	143
FIGURE 54	Structure de trait des subordonnées temporelles.	144
FIGURE 55	Exemple de structures de trait des déclencheurs.	147
FIGURE 56	Exemple de grammaire CMDL	154
FIGURE 57	Exemple d'annotation de ADSI	155
FIGURE 58	Schéma récapitulatif du principe du calcul temporel au sein de la <i>SdT</i> .	162
FIGURE 59	Représentation du chronogramme de l'exemple 82 au passé simple.	164
FIGURE 60	Représentation de la temporalité des procès modèles	166
FIGURE 61	Représentation atemporelle des procès modèles	167
FIGURE 62	Exemple de chronogramme de la portée des connecteurs.	172
FIGURE 63	Exemple de chronogramme de deux itérations imbriquées.	174
FIGURE 64	Exemple de chronogramme d'une sélection	176

FIGURE 65	Schéma du fonctionnement de notre calculateur de contraintes temporelles	183
FIGURE 66	Schéma des interactions	184
FIGURE 67	Échelle de saillance relative pour le calcul du degré de saillance des intervalles d'après Gosselin (1996).	185
FIGURE 68	Illustration du chronogramme du premier extrait de Flaubert. .	188
FIGURE 69	Illustration du chronogramme du second extrait de Flaubert. .	191
FIGURE 70	Représentation des relations entre les entités formalisant une itération dans le standard TIMEML.	200
FIGURE 71	Capture d'écran de l'application de constitution du lexique des typologies de procès.	222
FIGURE 72	Chronogramme d'un procès itératif au présent.	227
FIGURE 73	Chronogramme d'un procès itératif à l'imparfait.	228
FIGURE 74	Chronogramme d'un procès itératif au passé simple.	228
FIGURE 75	Chronogramme d'un procès itératif au futur.	229
FIGURE 76	Chronogramme d'un procès itératif au passé composé.	230
FIGURE 77	Chronogramme d'un procès itératif au plus-que-parfait.	231

LISTE DES TABLEAUX

TABLEAU 1	Nombres d'articles par secteur pour le sous-corpus [LMD].	89
TABLEAU 2	Taille du corpus	90
TABLEAU 3	Quantité de structures itératives dans le corpus [ITER].	93
TABLEAU 4	Quantité de structures itératives dans le corpus [ITER].	93
TABLEAU 5	Proportions des propositions incluses dans des structures itératives.	95
TABLEAU 6	Nombres d'articles contenant des structures itératives par secteur pour le sous-corpus [LMD].	95
TABLEAU 7	Proportion des itérations dans les sous-corpus [ROMAN] et [LMD] selon leur taille.	96
TABLEAU 8	Proportion des sélections dans les sous-corpus [ROMAN] et [LMD] selon leur taille.	96
TABLEAU 9	Quantité et proportion des structures itératives fondées sur un substantif événementiel.	100
TABLEAU 10	Proportion des déclencheurs par type.	101
TABLEAU 11	Proportion des déclencheurs de type « sélecteur ».	101
TABLEAU 12	Répartition des déclencheurs en position initiale selon le type et le corpus.	104
TABLEAU 14	Ontologie de Vendler	146
TABLEAU 15	Couple des temps morphologiques favorisant permettant une lecture itérative d'une subordonnée temporelle.	149
TABLEAU 16	Ontologie de Vendler	217
TABLEAU 17	Définition des types de procès proposée par Laurent Gosselin.	218
TABLEAU 18	Tableau de correspondance entre classe et type de procès.	220
TABLEAU 19	Types de procès pour chaque type d'association du verbe <i>rouler</i>	221
TABLEAU 20	Répartition des verbes annotés selon leur catégorie.	223

LISTE DES RÈGLES

1	Règle des circonstants temporels.	168
2	Cas de la portée d'un circonstanciel semelfactif.	169
3	Cas de la portée entre un circonstanciel et un déclencheur calendaire.	169
4	Cas de la portée entre deux circonstanciels.	170
5	Règle sur la portée des connecteurs temporels.	171
6	Règle d'extraction des itérés de l'itération.	175
7	Règle d'identité des intervalles de procès associées à l'itération et à la sélection.	176
8	Règle de construction des intervalles de la série.	226
9	Règle de construction des intervalles du procès.	226
10	Cas des itérations au présent.	227
11	Cas des itérations à l'imparfait.	227
12	Cas des itérations au passé simple.	228
13	Cas des itérations au futur.	229
14	Cas des itérations au passé composé.	230
15	Cas des itérations au plus-que-parfait.	231
16	Cas des itérations à l'infinitif passé.	232
17	Cas des itérations avec un auxiliaire de visée aspectuelle.	233
18	Cas des déclencheurs de type calendaire.	235
19	Cas des déclencheurs de type quantificationnel.	235
20	Cas des déclencheurs de type événementiel.	236
21	Cas des déclencheurs de type présuppositionnel.	237
22	Cas des déclencheurs de type itératifs.	238
23	Règle des subordonnées temporelles.	238
24	Relation d'accessibilité avec IC.	240
25	Relation de recouvrement avec IC.	240
26	Relation de coïncidence avec IC.	240
27	Relation de coïncidence sur la borne gauche avec IC.	241
28	Relation de coïncidence sur la borne droite avec IC.	241
29	Relation d'accessibilité inversée avec IC.	242
30	Relation d'antériorité avec IC.	242
31	Relation de postériorité avec IC.	242
32	Règle des circonstants temporels de type [depuis + durée].	243

Cette thèse a été composée avec L^AT_EX₂ε et la mise en page a été réalisée avec le style `classicthesis`, disponible sur le CTAN. Compilé le 4 janvier 2012.

RÉSUMÉ

Cette thèse s'inscrit dans le domaine du traitement automatique des langues, et concerne l'étude des phénomènes itératifs en langue. Il s'agit notamment de rendre compte de la sémantique de procès exprimant une répétition (d'événements ou d'états) au sein d'un même texte, tant du point de vue de l'aspect que du temps. Nous avons mené nos travaux dans le cadre d'un travail d'équipe pluridisciplinaire couvrant des orientations linguistiques, logiques et informatiques. Ces travaux reposent sur une double appréhension commune d'un énoncé itératif : une entité événementielle en quelque sorte générique qui est présentée comme répétée dans le temps, d'une part ; l'itération proprement dite, conçue comme ensemble des occurrences du modèle, d'autre part. Nous nous focalisons pour notre part sur deux facettes : l'inscription discursive, en nous reposant sur une modélisation objet de la construction du sens des itérations (Mathet, 2007) ; et le calcul aspectuo-temporel au travers de la *Sémantique de la temporalité (SdT)* (Gosselin, 2011). Nous menons en premier lieu une étude de corpus, sur des articles de journaux et des œuvres littéraires, qui vise à caractériser l'inscription discursive des phénomènes itératifs. Nous exposons sur cette base leurs propriétés discursives dans la perspective de circonscrire automatiquement les itérations, c'est-à-dire repérer la borne initiale et finale du segment textuel correspondant. Nous établissons également un modèle d'annotation pour cette automatisation permettant *in fine* d'instancier les itérations au sein du modèle objet. Dans un second temps, nous abordons les phénomènes itératifs sous l'angle du calcul aspectuo-temporel. Fondé sur les travaux de Laurent Gosselin sur le calcul aspectuel des procès semelfactifs, nous étudions les ajustements nécessaires pour l'appliquer aux procès itératifs, en intégrant la notion de double visée aspectuelle issue de la *SdT*, dans la perspective de l'automatisation du calcul. Enfin, nous complétons ces deux aspects de notre travail par l'intégration de nos résultats au sein de TIMEML, un formalisme XML de structuration des connaissances temporelles.

MOTS-CLÉS : Traitement Automatique des Langues, sémantique temporelle, temps et aspect, itérations en langue, inscription discursive

ABSTRACT

The study of iterative phenomena in French language: Discursive inscription and aspectual calculation, towards an automatic processing.

This thesis deals with Natural Language Processing (NLP). It studies the iterative phenomena in French, and it covers the semantics of event repetitions within the same text, both in term of aspect and time. Our study is based on a teamwork focusing on multidisciplinary approach to the phenomena iterative guidance covering language, logic and science computer. This work is based on two common concepts: on one hand, a temporal entity called typical event that is presented as repeated over time, and on the other, we have the iteration itself, constructed as set of all occurrences of the model event. This thesis focuses on two aspects: the discursive inscription, basing on an object modeling for the construction of the meaning of iterations (Mathet, 2007); and the aspectual calculation through the semantics of temporality (Gosselin, 2011). We conduct, as a first step, a study corpus of newspaper articles and literature, which aims to characterize the discursive inscription of iterations. On this basis we expose their properties with a view towards automatically indentifying the iterations, *i.e.* matching the beginning and the closing of a textual segment. We establish also a model of annotation for this automation in order to instantiate iterations in the object model. In a second step, we consider the aspectual and temporal calculation of iterations. Based on the work of Gosselin on the aspectual calculation of semelfactive events, we study the adjustments in order to apply it to the iterative events, integrating the concept of two simultaneously independent referred aspectual, in view of the implementation process. Finally, we complete these two aspects of our work through the integration of our results in TIMEML, an XML formalism for events and temporal expressions in natural language.

KEYWORDS: Natural Language Processing, temporal semantic, iteration in french language, discursive inscription, aspect and time