



**HAL**  
open science

# Génération de récits à partir de données ambiantes

Belen Baez Miranda

► **To cite this version:**

Belen Baez Miranda. Génération de récits à partir de données ambiantes. Informatique et langage [cs.CL]. Université Grenoble Alpes, 2018. Français. NNT : 2018GREAM049 . tel-02004338

**HAL Id: tel-02004338**

**<https://theses.hal.science/tel-02004338>**

Submitted on 1 Feb 2019

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

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

## THÈSE

Pour obtenir le grade de

**DOCTEUR DE LA COMMUNAUTE UNIVERSITE  
GRENOBLE ALPES**

Spécialité : **Informatique**

Arrêté ministériel : 25 mai 2016

Présentée par

**Belén BAEZ MIRANDA**

Thèse dirigée par **François PORTET**, codirigée par **Catherine GARBAY** et co-encadrée par **Sybillle CAFFIAU**

préparée au sein du **Laboratoire d'Informatique de Grenoble**  
dans l'**École Doctorale Mathématiques, Sciences et  
Technologies de l'Information, Informatique**

## Génération de récits à partir de données ambiantes

Thèse soutenue publiquement le **3 décembre 2018**,  
devant le jury composé de :

**M. Marc CAVAZZA**

Professeur, Université de Greenwich, président

**M. Patrick GIRARD**

Professeur des Universités, Université de Poitiers, rapporteur

**M. Eddie SOULIER**

Maître de Conférences HDR, Université de Technologie de Troyes, rapporteur

**M. François PORTET**

Maître de Conférences, Grenoble-INP, LIG, directeur de thèse

**Mme. Catherine GARBAY**

Directrice de recherche CNRS HDR, LIG, co-directrice de thèse

**Mme. Sybillle CAFFIAU**

Maître de Conférences, Université Grenoble Alpes, LIG, co-encadrante de thèse







## Résumé

Le récit est un outil de communication qui permet aux individus de donner un sens au monde qui les entoure. Il constitue un élément central pour comprendre et partager leur culture, leurs connaissances et leur identité. Il véhicule une série d'événements réels ou imaginaires, provoquant un ressenti, une réaction ou même, une action. Pour cette raison, il est devenu un sujet d'intérêt pour différents domaines au-delà de la Littérature (Éducation, Marketing, Psychologie, etc.) dont les buts peuvent être très différents (persuader, réfléchir, apprendre, etc.).

Cependant, le récit reste encore peu étudié dans le contexte informatique, malgré des travaux qui visent son analyse et sa production automatique. Les algorithmes et implémentations de ce domaine se sont principalement centrées sur la production de récits fictifs et l'étude de textes littéraires. Ainsi, il n'existe pas d'approches capables de produire automatiquement des récits dont 1) la source est constituée de données non formatées et issues de la réalité et 2) et le contenu projette une perspective qui cherche à transmettre un but communicationnel particulier. Travailler avec des données brutes devient important dans le contexte du *Big Data*, qui voit leur quantité augmenter de manière exponentielle due à l'utilisation d'appareils connectés.

C'est dans ce contexte que nous présentons une approche de génération automatique de récits à partir de données ambiantes. L'objectif est de faire émerger l'expérience vécue d'une personne à partir des données produites pendant une activité humaine. Tous les domaines qui travaillent avec des données brutes pourraient bénéficier de ce travail, tels que l'Éducation ou la Santé. Il s'agit d'un effort interdisciplinaire qui inclut le Traitement Automatique de Langues, la Narratologie, les Sciences Cognitives et l'Interaction Homme-Machine.

Cette approche est basée sur des corpus et des modèles et comprend la formalisation de ce que nous appelons le *récit d'activité* ainsi qu'une démarche de génération adaptée. Elle est composée de 4 étapes : la formalisation des récits d'activité, la constitution de corpus, la construction de modèles d'activité et du récit, et la génération de texte. Chacune de ces étapes a été conçue pour surmonter des contraintes liées aux questions scientifiques posées, spécifiques à notre objectif : la manipulation de données incertaines et incomplètes, l'abstraction valide de ces données comme éléments de l'activité, la construction de modèles permettant la transposition de la réalité véhiculée par ces éléments concrets vers une perspective subjective et son rendu en langage naturel. Nous avons utilisé comme cas d'usage le ski de randonnée, un domaine d'activité où l'usage d'appareils connectés se répand de manière croissante, pour le suivi des performances mais aussi pour des raisons de sécurité, et où le partage d'expérience via le récit se pratique depuis de nombreuses années, grâce à des sites web dédiés. Les résultats obtenus sont encourageants et donnent des pistes qui ouvrent beaucoup de perspectives de recherche.



## Abstract

Stories are a communication tool that allow people to make sense of the world around them and are central to the understanding and sharing of culture, knowledge and identity. They convey a series of real or imaginary events, causing a feeling, a reaction or even an action. For this reason, the study of storytelling has become a topic of interest for a range of different areas beyond Literature (Education, Marketing, Psychology, etc.) whose goals can be very different (persuade, reflect, learn, etc.).

However, stories remain under-studied in Computer Science, despite work that focus on their analysis and automatic production. However, the algorithms and implementations of this domain have mainly focused on the production of fictional stories and on the study of literary texts. Thus, there are no approach able to automatically produce stories for which: 1) the source consists of raw real-world data; and 2) the content projects a perspective that seeks to convey a particular communicative purpose. Working with raw data becomes important in the context of *Big Data*, where its quantity increases exponentially through the use of connected devices.

It is in this context that we present an approach for the automatic generation of stories from ambient data. The objective is to bring out the lived experience of a person from the data produced during a human activity. Any areas that work with raw data could benefit from this work, such as Education or Health. It is an interdisciplinary effort that includes Natural Language Processing, Narratology, Cognitive Science and Human-Computer Interaction.

This approach is based on corpora and models and includes the formalization of what we call the *activity récit* as well as an adapted generation approach. It is composed of 4 stages: the formalization of the *activity récit*, the collection of the corpora, the construction of models of the activity and story, and the text generation. Each of these steps has been designed to overcome constraints related to scientific questions specific to our purpose: the manipulation of uncertain and incomplete data, the valid abstraction of these data as elements of the activity, the construction of models allowing the transposition from reality conveyed by these concrete elements to a subjective perspective and its rendering in natural language. As a case study, we used ski touring, an outdoor activity where the use of connected devices for monitoring performance and security is spreading more and more, and where the sharing of experiences through storytelling has been practiced for many

years thanks to dedicated websites. The results obtained are encouraging and give a wide scope for further research.

## Remerciements

L'achèvement de ce travail mené sur plusieurs années procure une grande satisfaction. En premier lieu, je voudrais remercier mes encadrants de thèse, M. François PORTET, Mme Catherine GARBAY et Mme Sybille CAFFIAU pour votre disponibilité, votre aide et pour m'avoir guidée dans mes travaux de recherche depuis mon stage. Merci également pour votre confiance, patience, passion pour la recherche ainsi que pour votre soutien durant les moments difficiles. Vous avez grandement contribué à l'aboutissement de ce travail.

Mes plus sincères remerciements vont également à M. Patrick GIRARD, Professeur à l'Université de Poitiers, et M. Eddie SOULIER, Maître de Conférences à l'Université de Troyes, qui m'ont fait l'honneur d'être rapporteurs de cette thèse, ainsi qu'à M. Marc CAVAZZA, Professeur à l'Université de Greenwich, pour avoir accepté de faire partie du jury.

J'adresse toute ma gratitude à toutes les personnes qui m'ont aidé dans la réalisation de ce travail, David BLACHON et Trung Ky NGUYEN. Je remercie particulièrement M. Frank TOLLARD, qui, grâce à son enthousiasme pour les montagnes, a fait des apports importants à cette thèse.

Je ne saurais oublier, personnels, permanents de l'équipe GETALP et IIHM et doctorants qui ont contribué à créer une atmosphère de travail agréable et conviviale. Je remercie Anubabh, Hyunyoung, Juan, Julian, Karim, Maxime, Sami, Tanguy et tous mes amis du labo, pour tous les bons moments partagés. Merci également aux collègues du Bureau 370 : Raheel, Ky, ainsi que ses anciens habitants (Alexis, Ieva, Diego...). De même, je remercie aux amis hors du labo : Sandra, Christine et la Famille Cayuela et particulièrement Andon, dont le soutien et avec qui les échanges scientifiques ont été fondamentaux.

Je remercie également l'IUT 2 de Grenoble. Un grand merci Mme Sylvie PESTY, M Benjamin LECOUTEUX, M. Didier SCHWAB, Mme MARIE-PIERRE SCHAEFFER et au Département de TechCo pour m'avoir donné l'opportunité de développer mes compétences dans le monde de l'enseignement.

Je veux remercier mes parents. Ils ont su soutenir de loin malgré qu'ils se trouvent de l'autre côté de l'Atlantique. Ils sont toujours avec moi. Ils m'ont appris à conquérir les obstacles avec courage, persévérance et le sourire.

Adriana et Brian, merci pour leur amitié et patience dans les bons et mauvais moments, bref, pour avoir toujours été là pour moi. Que notre amitié soit éternelle.

Enfin, je remercie Joseph qui a su m'encourager avec patience, joie et optimisme. Merci d'avoir cru en moi. Cette thèse et moi te devons beaucoup.

Ce travail a été possible grâce au Conseil National de la Science et de la Technologie du Mexique (CONACYT) en partenariat avec le Gouvernement Français, qui m'ont attribué une bourse de doctorat, et avec laquelle j'ai pu me consacrer à l'élaboration de la présente thèse.

# Table des matières

<b>Résumé</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Remerciements</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 État de l’art : Définition, génération et communication de récits</b>	<b>10</b>
2.1 Études de l’objet récit : la narratologie . . . . .	11
2.1.1 Pré-formalisme russe . . . . .	11
2.1.2 Le Formalisme Russe . . . . .	12
2.1.3 Le Structuralisme Français . . . . .	15
2.1.4 L’ère post-structuraliste de la narratologie . . . . .	18
2.2 Composition d’un récit textuel . . . . .	18
2.2.1 Les niveaux de description du récit . . . . .	19
2.2.2 Les composants de l’histoire . . . . .	19
2.2.2.1 Événements . . . . .	20
2.2.2.2 Existants . . . . .	21
2.2.2.3 Intrigue . . . . .	21
2.2.3 Les stratégies du discours . . . . .	22
2.3 La narratologie cognitive . . . . .	25
2.3.1 La compréhension du récit . . . . .	26
2.3.2 La production de récits . . . . .	31
2.4 Bilan . . . . .	35
<b>3 État de l’art : Génération automatique de récits</b>	<b>38</b>
3.1 La Narratologie informatique . . . . .	38
3.2 Systèmes de génération de récits . . . . .	40
3.2.1 Génération par planification . . . . .	40



3.2.1.1	<i>TALE-SPIN</i> (1977)	41
3.2.1.2	<i>Universe</i> (1985)	42
3.2.1.3	<i>StoryBook</i> (2002)	43
3.2.1.4	<i>Virtual StoryTeller</i> (2003)	46
3.2.1.5	<i>Fabulist</i> (2010)	48
3.2.2	Génération par raisonnement à partir de cas	50
3.2.2.1	<i>Minstrel</i> (1993)	51
3.2.2.2	<i>Mexica</i> (1999)	52
3.2.2.3	<i>ProtoPropp</i> (2005)	54
3.2.3	Génération de récits à partir de grammaires	56
3.2.3.1	<i>Brutus</i> (1999)	56
3.2.3.2	<i>Propper</i> (2013)	59
3.2.4	Génération de récits à base de modèles statistiques	60
3.2.5	Génération interactive de récits	62
3.2.5.1	<i>Façade</i> (2003)	62
3.2.5.2	<i>Holodeck</i> (2007) : adaptation de <i>Madame Bovary</i>	64
3.2.5.3	<i>Scheherezade-IF</i> (2015)	66
3.3	Évaluation de récits	68
3.3.1	<i>StoryEval</i> (2009)	68
3.3.2	Comparaison de plans de récit (2016)	69
3.3.3	Évaluation de la perception de la temporalité dans les récits issus de capteurs (2015)	70
3.4	Bilan	70
<b>4</b>	<b>Vers une génération automatique de récits à partir de données ambiantes</b>	<b>73</b>
4.1	La problématique	74
4.2	Choix du cas d'étude	75
4.3	Définition du <i>récit d'activité</i>	78
4.4	Approche de génération de récits d'activités	81
4.4.1	Production basée modèles d'un récit d'activité	81
4.4.2	Verrous scientifiques	84
4.4.2.1	Verrous liés à la modélisation	84
4.4.2.2	Verrous liés à la transformation des modèles	85
4.5	Démarche de conduite de thèse	86
<b>5</b>	<b>Constitution de corpus</b>	<b>88</b>
5.1	Constitution des corpus textuels	88
5.1.1	Choix de la source des corpus textuels	89
5.1.2	Corpus fréquentiel	92

5.1.3	Corpus textuel de modélisation de l'activité et de la structure textuelle du corpus annoté . . . . .	95
5.2	Constitution du Corpus <i>parallèle</i> . . . . .	97
5.2.1	Protocole de collecte de données . . . . .	99
5.2.2	Synthèse du corpus parallèle . . . . .	103
5.3	Synthèse des corpus . . . . .	106
<b>6</b>	<b>Modélisation du point de vue objectif et génération d'une séquence d'activité</b>	<b>108</b>
6.1	Extraction des connaissances sur le domaine à partir du corpus textuel de modélisation . . . . .	108
6.2	Élaboration du modèle de données . . . . .	114
6.2.1	Les ontologies basées sur la Logique de Description (DL)	114
6.2.1.1	La logique de description . . . . .	117
6.2.1.2	L'ontologie pour la structuration des données	119
6.2.2	Construction de l'ontologie basée sur la logique de description pour le ski de randonnée . . . . .	120
6.2.3	Des données captées aux données du domaine . . . . .	123
6.3	Élaboration du modèle de l'activité . . . . .	126
6.3.1	Les modèles de tâches pour exprimer le déroulement de l'activité . . . . .	127
6.3.2	Construction du modèle de tâches pour le ski de randonnée . . . . .	129
6.4	Des modèles du domaine à une séquence d'activité à rapporter	133
6.4.1	Évaluation de la perception de la séquence d'activité issue d'une séquence de tâches . . . . .	133
6.4.2	Production d'une séquence d'activité . . . . .	137
6.5	Conclusion sur la modélisation de l'activité du point de vue de son objectif et de son utilisation . . . . .	143
<b>7</b>	<b>Modélisation du point de vue subjectif et génération du récit à base de modèles</b>	<b>145</b>
7.1	Extraction des connaissances sur le monde du récit de randonnée	146
7.1.1	Analyses de fréquence . . . . .	147
7.1.2	Analyse structurelle . . . . .	151
7.1.3	Analyse du déroulement (Corps) transmis . . . . .	156
7.2	Élaboration du plan du récit . . . . .	160
7.2.1	Le modèle du plan du récit . . . . .	160
7.2.1.1	Les unités d'information du plan du récit d'activité pour le ski de randonnée . . . . .	161
7.2.1.2	Définition des <i>plots</i> pour le récit d'activité . . . . .	168

7.2.2	Construction du modèle du plan du récit . . . . .	174
7.2.2.1	Contenu selon les types de récit . . . . .	174
7.2.2.2	Structure selon les types de récit . . . . .	176
7.2.2.3	Résumé du passage de la séquence d'événe- ments valide au plan du récit . . . . .	186
7.3	Du plan du récit aux textes . . . . .	186
7.4	Conclusion de la modélisation du point de vue subjectif de l'activité : le récit . . . . .	194
<b>8</b>	<b>Conclusion</b>	<b>196</b>
8.1	Contributions de la thèse . . . . .	196
8.2	Limitations . . . . .	198
8.3	Perspectives . . . . .	199
	<b>Appendices</b>	<b>203</b>
	<b>Bibliographie</b>	<b>209</b>

# Table des figures

1.1	Présentation d'une sortie de randonnées à ski au travers deux modes différents : l'image et le récit . . . . .	5
2.1	La pyramide de Freytag . . . . .	12
2.2	Liste des 31 fonctions identifiées par Propp . . . . .	13
2.3	Schéma actanciel selon Greimas . . . . .	16
2.4	Les fonctions d'après Barthes . . . . .	17
2.5	Représentation des trois niveaux de description du récit pour lesquels il existe un consensus théorique en narratologie. . . . .	19
2.6	Organisation des composants d'un récit selon Chatman [Chatman, 1977] . . . . .	20
2.7	Illustration de la grammaire de Mandler et Johnson [Mandler et Johnson, 1977]. À gauche, texte de <i>story dog</i> , à droite abstraction de <i>story dog</i> selon la grammaire de Mandler et Johnson. les lettres <i>A T</i> et <i>C</i> indiquent respectivement <b>And</b> , <b>Then</b> et <b>Cause</b> . . . . .	27
2.8	Les 15 <i>plots units</i> primitifs décrits par Lehnert [Lehnert et Ringle, 2014]) . . . . .	29
2.9	Les <i>plots complexes</i> identifiés dans [Lehnert, 1981] . . . . .	30
2.10	Exemple d'analyse selon Lehnert[Lehnert, 1981]. En haut, récit du dépannage d'une voiture. En bas, analyse du récit selon les deux personnages. . . . .	31
2.11	Démarche mentale lors de la production d'un récit d'après [Kim, 2004] . . . . .	32
2.12	Modèle d'écriture selon [Flower et Hayes, 1981](tiré de [Grégoire et Karsenti, 2013]) . . . . .	33
2.13	Modèles proposés par [Bereiter et Scardamalia, 1987] (tiré de [Chanquoy et Alamargot, 2002]) . . . . .	34
2.14	Le cycle « Engagement and Reflection » [Sharples, 1999] . . . . .	35
3.1	Récit généré par <i>TALE-SPIN</i> . . . . .	41

3.2	Plan généré par <i>Universe</i> [Lebowitz, 1984] . . . . .	43
3.3	Architecture du système <i>StoryBook</i> [Callaway et Lester, 2002]	44
3.4	Extrait du <i>narrative stream</i> [Callaway et Lester, 2002] . . . . .	45
3.5	Exemples générés avec <i>StoryBook</i> pour l'évaluation [Callaway et Lester, 2002] . . . . .	46
3.6	Architecture du système <i>Virtual StoryTeller</i> [Theune <i>et al.</i> , 2004] . . . . .	47
3.7	Fabula générée par <i>Virtual StoryTeller</i> [Theune <i>et al.</i> , 2004] .	48
3.8	Illustration de la fabula obtenue avec l'algorithme IPOCL [Riedl et Young, 2010] . . . . .	49
3.9	Récit généré à partir de la fabula obtenue avec l'algorithme IPOCL [Riedl et Young, 2010] . . . . .	50
3.10	Récit généré par <i>Minstrel</i> . . . . .	52
3.11	Récit généré par <i>Mexica</i> . . . . .	54
3.12	Interface du système <i>ProtoPropp</i> [Gervás <i>et al.</i> , 2005] . . . . .	55
3.13	Ontologie utilisée par le système <i>ProtoPropp</i> [Gervás <i>et al.</i> , 2005] . . . . .	55
3.14	Récit généré par <i>ProtoPropp</i> avec les paramètres saisis dans la figure 3.12 [Gervás <i>et al.</i> , 2005] . . . . .	56
3.15	Récit généré par <i>Brutus</i> (issu de [Fetzer, 2002]) . . . . .	57
3.16	Architecture de <i>Brutus</i> (issue de [Friedenberg, 2008]) . . . . .	58
3.17	Exemples <i>Story actions</i> de la fonction narrative <i>Vilain</i> [Gervás, 2013, Gervás, 2016] . . . . .	59
3.18	Exemple de grammaire simple utilisant les fonctions narratives de Propp [Gervás, 2013] . . . . .	60
3.19	Arbre produit qui montre les relations entre entités saisis par l'utilisateur [McIntyre et Lapata, 2009] . . . . .	61
3.20	Récit généré avec l'approche statistique [McIntyre et Lapata, 2009] . . . . .	62
3.21	Interface du système <i>Façade</i> [Mateas et Stern, 2005] . . . . .	63
3.22	Architecture du système <i>Façade</i> [Mateas et Stern, 2005] . . . . .	64
3.23	Interaction entre l'utilisateur et "Emma Bovary" [Cavazza <i>et al.</i> , 2007] . . . . .	65
3.24	Interaction entre l'utilisateur et "Emma Bovary" [Cavazza <i>et al.</i> , 2007] . . . . .	65
3.25	<i>Ploth graph</i> de <i>Scheherezade-IF</i> autour du sujet « braquage d'une banque » [Li et Riedl, 2015] . . . . .	67
4.1	Skieurs pendant une sortie de ski de randonnée . . . . .	76
4.2	Équipements utilisés pour réaliser le ski de randonnée . . . . .	77
4.3	Données issues d'une sortie de ski . . . . .	77

4.4	Texte écrit par un skieur pour décrire une sortie vers le sommet Chamechaude ( <a href="http://www.skitour.fr/sorties/chamechaude,44954.html">http://www.skitour.fr/sorties/chamechaude,44954.html</a> , date de consultation : 14/03/18) . . . . .	78
4.5	Définition d’une histoire narrative de Mieke Bal [Bal, 2009] . . . . .	79
4.6	Les mondes du récits d’activité . . . . .	82
4.7	Modèles et instances des mondes du récits d’activités . . . . .	83
5.1	Capture d’écran du site de consultation du corpus « Text+Berg-Korpus R151v01 Écho des Alpes » du projet « Text+Berg digital » [Göhring et Volk, 2011] . . . . .	90
5.2	Fiche rapportant la sortie du 27 janvier 2015 vers le sommet Charmant Som par la route des Cottaves. ( <a href="http://www.skitour.fr/sorties/charmant-som,61537.html">http://www.skitour.fr/sorties/charmant-som,61537.html</a> ) [Date de capture : 27/07/2017] . . . . .	92
5.3	Textes appartenant au Corpus fréquentiel, tels qu’ils se trouvent sur le site <code>skitour.fr</code> . . . . .	94
5.4	Deux exemples de récits provenant du site <code>skitour.fr</code> qui décrivent deux sorties faites par le même itinéraire pour atteindre le sommet du Charmant Som (route par les Cottaves). . . . .	97
5.5	Le capteur <i>BioHarness 3</i> , la ceinture et comment elle doit être ajustée au corps. . . . .	102
5.6	Extrait de l’enregistrement no. 10 du corpus parallèle (Sortie du 30/03/2016 vers le sommet Chamechaude) . . . . .	105
6.1	Interface du logiciel d’annotation Callisto . . . . .	112
6.2	Matrice de transitions issue des annotations . . . . .	116
6.3	Architecture d’un système de représentation de connaissances basé sur la Logique de Description (issue de [Baader et Werner, 2002]) . . . . .	119
6.4	Concepts de l’ontologie . . . . .	122
6.5	Interface du logiciel <i>Protégé</i> qui montre une instance de l’ontologie générée à partir de données obtenues des capteurs . . . . .	125
6.6	Exemple de modèle de tâches édité avec K-MADe qui décrit l’activité “Envoyer un email” [Caffiau, 2009] . . . . .	128
6.7	Arborescence des tâches pour le ski de randonnée. . . . .	132
6.8	<i>Timeline</i> et cartes pour reconstruire les séquences . . . . .	134
6.9	Un participant pendant une passation . . . . .	134
6.10	Texte généré automatiquement . . . . .	135
6.11	Scores de distances et de taille en fonction du mode . . . . .	137
6.12	Interface du logiciel <i>Protégé</i> qui montre une instance d’ <i>Activity_Event</i> . . . . .	139

6.13	Automate d'un analyseur lexicale des expressions arithmétiques ( <a href="https://www.enseignement.polytechnique.fr/informatique/ARCHIVES/IF/poly/main009.html">https://www.enseignement.polytechnique.fr/informatique/ARCHIVES/IF/poly/main009.html</a> ) [Date de capture : 25/08/2018] .....	140
6.14	Traitement de validation de la séquence de l'activité . . . . .	142
7.1	Corpus de récits extraits du Corpus de modélisation . . . . .	153
7.2	Récits provenant du site <code>skitour.fr</code> décrivant la sortie faite pour atteindre le sommet du Charmant Som via la route des Cottaves ( <a href="http://www.skitour.fr/sorties/charmant-som,38284.html">http://www.skitour.fr/sorties/charmant-som, 38284.html</a> , date de consultation : 14/03/18) . . . . .	162
7.3	Les <i>plots units</i> identifiés dans le texte de la figure 7.2 . . . . .	165
7.4	Les <i>plots units</i> identifiés dans le texte de la figure 7.2 et pré- sentés dans la figure 7.3, organisés selon les rôles informatifs associés aux états affectifs. . . . .	172
7.5	Structure générale d'un récit d'activité . . . . .	177
7.6	Le plan de récit pour un récit positif sans contretemps . . . . .	181
7.7	Intégration d'étapes pendant la montée . . . . .	183
7.8	Intégration de <i>plots</i> avec difficultés et sans modification des conditions par le protagoniste . . . . .	184
7.9	Plan du récit pour les récits négatifs . . . . .	185
7.10	Correspondance entre la structure du texte généré et le plan d'un récit positif sans contretemps . . . . .	188
7.11	Correspondance entre la structure du texte généré et le plan d'un récit nuancé sans correction des conditions par le prota- goniste . . . . .	188
7.12	Correspondance entre la structure du texte généré et le plan d'un récit négatif . . . . .	189
7.13	Exemple de phrase <i>Success_Debut</i> . . . . .	192
7.14	Exemple de phrase <i>Enablement_Debut</i> . . . . .	192
7.15	Exemple de phrase <i>Success_Sommet</i> . . . . .	192
7.16	Exemple de phrase <i>Enablement_Fin</i> . . . . .	192
7.17	Exemple de phrase <i>Success_Fin</i> . . . . .	192
7.18	Récit provenant de l'enregistrement N° 5 (Date : 11/02/17, Sommet : Col de la Maurienne) . . . . .	193
7.19	Récit provenant de l'enregistrement N° 7 (Date : 18/03/16, Sommet : Col de l'Aigleton) . . . . .	193
7.20	Récit provenant de l'enregistrement N° 10 (Date : 30/03/16, Sommet : Chamechaude) . . . . .	193
7.21	Récit provenant de l'enregistrement N° 11 (Date : 22/01/17, Sommet : Grand Rocher) . . . . .	193

# Chapitre 1

## Introduction

Les récits sont omniprésents dans la communication quotidienne entre humains. Les philosophes, linguistes et spécialistes de la communication ont reconnu depuis l'Antiquité le rôle fondamental des histoires en tant que principal moyen de communication humain-humain, comme outil de la construction identitaire et comme procédé permettant de se situer dans le monde [Barthes, 1966, Propp, 1970, Genette, 2007, Fisher, 1985, Herman, 2009b]. L'avantage du récit par rapport aux autres modes de communication réside dans la disponibilité de divers moyens au niveau linguistique et structurel pour véhiculer des informations temporelles et causales, pour permettre la prise de perspective, pour exprimer des intentions et pour accentuer les informations les plus pertinentes. Il n'est donc pas surprenant que les récits écrits et oraux soient la forme principale dans laquelle les expériences sont transmises par les humains, que l'objectif soit de rapporter, d'expliquer, d'illustrer des opinions ou de transmettre des connaissances. Selon l'une des théories des sciences de la communication – le paradigme narratif [Fisher, 1985]–, les récits sont largement utilisés dans la communication technique (rapports médicales, comptes rendus d'incident...) et jouent un rôle central dans la création et le renforcement de groupes sociaux. Si le récit semble être la forme de communication la plus importante chez l'homme, il s'agit d'une forme de communication véritablement sous-employée dans le domaine de la communication homme-machine. Bien entendu, la génération, la modélisation et l'analyse d'histoires informatisées ont déjà fait l'objet de recherches ce qui a donné naissance à la narratologie informatique (*computational narratology*) [Mani, 2012]. Cependant, les progrès les plus marquants ont été réalisés dans le domaine de la créativité informatique ; domaine dans lequel les éléments à intégrer dans l'histoire sont généralement fictifs ou virtuels. Dans cette thèse, notre objectif est de rendre possible la création automatique de récits à partir de données réelles et objectives ce qui, à notre connaissance, est un objectif



qui n'a encore jamais été atteint.

La génération automatique de récits devient de plus en plus importante maintenant que le développement technologique a contribué à l'apparition d'une grande quantité de dispositifs connectés. C'est l'«Internet des Objets» (ou *Internet of Things* – IoT) qui consiste en des réseaux d'objets physiques qui peuvent se connecter entre eux, s'échanger des données et envoyer des requêtes ou des messages sur le web. Le représentant phare de ces objets est le smartphone qui permet de consulter les moteurs de recherche, d'acheter des produits en ligne, d'envoyer et de recevoir des courriels électroniques, et même d'interagir avec d'autres individus au travers des appels par vidéo, des jeux en ligne ou des réseaux sociaux dans des lieux et environnements variés. Le smartphone a notamment la particularité de s'être intégré aux activités quotidiennes des utilisateurs : au travail, à domicile, lors d'activités sportives, pour les loisirs... Au-delà de leur capacité de communication, ces objets physiques sont également truffés de capteurs permettant de recueillir en permanence des données sur le monde (et les utilisateurs) qui les entoure et contribuent ainsi à l'avènement de l'ère du *big data*. Les capteurs utilisés quotidiennement sont alors une source d'information factuelles sur l'environnement et les activités des utilisateurs.

Les informations disponibles à travers l'internet des objets peuvent être produites par l'utilisateur lui-même (comme une annonce sur un réseau social) ou extraites des données captées par les objets (comme les coordonnées GPS sur un smartphone). Par exemple, pour un sportif qui souhaite contrôler l'évolution de ses performances et de ses caractéristiques physiologiques (p.e., rythme cardiaque), des objets portatifs dotés de capteurs peuvent enregistrer à intervalles réguliers un ensemble de valeurs que le sportif peut ensuite consulter soit pour connaître une valeur à un instant donné (*Où étais-je après 10 minutes de courses ?*), soit l'évolution de ces valeurs (*Comment mon rythme cardiaque a-t-il évolué tout au long de la course ?*). Ces données peuvent ensuite être conservées ou partagées avec des personnes avec, plus ou moins, de compléments personnels en fonction de la raison du partage (médecins, entraîneurs, autres sportifs, amis...). Ce phénomène de mesure, d'analyse et de partage de ses propres informations s'étend bien au-delà des sportifs et est connu sous le terme de *quantified self* ou *QS* (mesure de soi). Le QS est un mouvement née au milieu des années 2000 qui regroupe les outils, les principes et les méthodes permettant à chacun de mesurer ses données personnelles, de les analyser et de les partager pour comprendre son corps ou son comportement ou le modifier. Ce mouvement intéresse depuis quelques années plusieurs marchés notamment celui de la publicité et des assurances.

Cependant, la quantité de données produites devient importante, d'autant plus dans le cas des données captées automatiquement. Les utilisateurs

sont donc confrontés à une quantité de données importantes stockées dans un format numérique brut ou compressé qui ne demande qu'à être exploitée pour en extraire la richesse informationnelle. Or, les systèmes (et applications) actuels ont tendance à s'appuyer sur un mode de présentation factuel pour transmettre des informations, sans adopter de perspective ni faire ressortir les liens et motivations des activités des utilisateurs pour mettre en avant des événements particuliers. Définir des méthodes permettant de transmettre ces données brutes en adoptant une perspective personnelle et en faisant ressortir ces liens sous une forme narrative cohérente est l'objectif dans lequel s'inscrit cette thèse. Bien entendu, cet objectif doit relever de nombreux défis pour y parvenir qui nécessitent encore de nombreux efforts de recherche. Ce chapitre d'introduction situe l'intérêt du récit dans le domaine de la communication et l'interprétation de données, introduit les objectifs concrets et les contributions de la thèse et présente le plan du manuscrit.

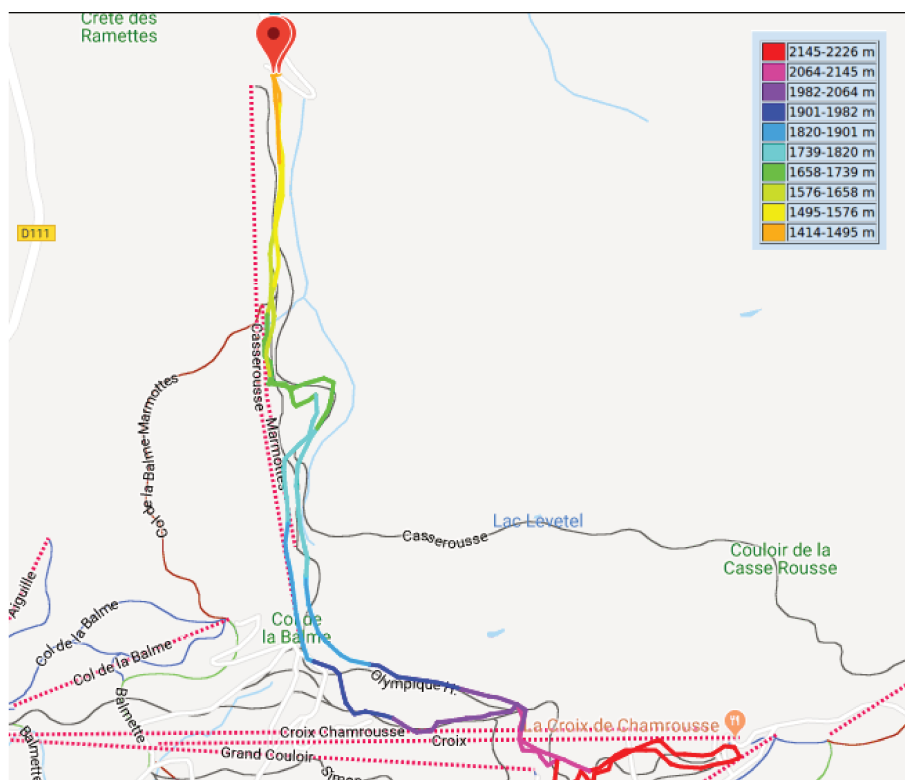
## Les stratégies pour donner un sens aux données

Différentes stratégies et traitements automatisés ont été développées pour rendre des masses de données intelligibles. La visualisation des données (ou *Dataviz*) apporte des solutions basées sur l'utilisation d'images et graphiques comme moyen de présentation et support à la prise de décisions [Cook et Thomas, 2005]. Plusieurs travaux ont été menés ces dernières années qui cherchent une transmission efficace de l'information complexe à travers des représentations visuelles simples, ainsi que l'implémentation des outils pour l'analyser ou la partager. Nous pouvons trouver dans l'état de l'art des contributions portant des rendues visuelles variées et adaptées en fonction de la nature des données présentées. Parmi eux, nous trouvons par exemple, le traitement et visualisation des jeux de données climatiques et satellitaires ([Huntington *et al.*, 2017]), des viseurs pour explorer des données génomiques [Thorvaldsson *et al.*, 2013], ou des simulations et visualisations qui aident à l'analyse des inondations [Leskens *et al.*, 2017].

À moindre échelle, la visualisation de données est également un support pour le suivi et l'analyse de données provenant d'activités humaines. Ces données permettent d'observer la performance ou l'état de la personne tout au long d'une période de temps défini. Dans la figure 1.1a nous présentons une trace GPS qui montre le déplacement d'une personne lors d'une randonnée à ski. Grâce aux couleurs, lesquelles sont associées à des valeurs indiquant l'altitude, nous pouvons comprendre de manière objective le parcours. Les

images comme celle-ci permettent de montrer une grande quantité d'information dans un espace réduit, en permettant de repérer en peu de temps des éléments ou aspects précis ou anormaux. Par exemple, en observant la figure 1.1a, nous pouvons trouver le moment où la personne est arrivée au sommet grâce à l'indication de couleur rouge.

Cependant, l'image ne transmet que des éléments du parcours qui ne peuvent être interprétés que par des spécialistes et de connaisseurs du lieu de randonnée. Par ailleurs, cette trace GPS ne transmet pas le point de vue du randonneur, ni ses difficultés ni les motivations de certains choix. C'est pourquoi, le randonneur a également écrit le récit de sa sortie qui est reproduit La figure 1.1b. Dans ce récit, nous pouvons observer que le randonneur met en valeur les points qui lui semblent intéressants et qui ont eu une influence sur la réalisation de l'activité. Par exemple, des observations ("Il y a peu du monde") ; comment cette personne se souvient de son ressenti à certain moment, soit positif ("Je me sens de bonnes jambes"), soit négatif ("j'engage une descente laborieuse malgré la quantité de neige") ; des décisions prises ("je décide de couper la grande courbe qui fait la piste") et des conséquences ("Erreur, le hors-piste est abominable"), etc. Ainsi, bien que les figures 1.1a et 1.1b présentent la même situation, l'information qu'elles véhiculent sont très différentes. L'avantage du récit sur la visualisation de données est la possibilité de partager une perspective de la réalité vécue, au travers d'une description plus personnelle, riche en éléments subjectifs qui sont difficiles à projeter sur une image, et qui cherche à partager une expérience plutôt qu'un bilan impersonnel. Par ailleurs, le support du récit est principalement le langage naturel qui contient tous les mécanismes pour transmettre les éléments narratifs soulignés pour lesquels la cognition humaine est entraînée.



(a) Parcours d'un randonneur lors d'une sortie projeté sur une carte à partir des données GPS enregistrées durant la sortie

Peu de temps pour cette sortie. On va se faire la traditionnelle croix de Chamrousse. Arrivée au parking de Casserousse, on est en limite de brouillard. Il neige un peu, il fait frais (3/4 °C). Vu le manque de temps, je décide de monter au train à la croix en remontant les pistes. Je me sens de bonnes jambes et attaque la montée à bon rythme. Arrivée à premier départ du tire-fesse intermédiaire, je décide de couper la grande courbe que fait la piste. Erreur, le hors-piste est abominable (trace qui glisse, grosses bosses). Du coup je m'épuise un peu à ne pas avancer. Au sortir de cette galère, je décide de rester sur la piste jusqu'en haut. Il y a peu de monde, je ne dérange personne. J'arrive tant bien que mal à reprendre mon rythme. Au trois quart de la montée, un peu se décolle. Je la remets rapidement et reprend mon chemin. La visibilité est abominable (4 ou 5 mètres). J'arrive cependant à rejoindre le restaurant à la croix. Je prends le temps de me couvrir et j'engage une descente laborieuse malgré la qualité de la neige. Arrivé en vu du parking la visibilité s'améliore et je peux laisser un peu glisser.

(b) Récit écrit par le randonneur

FIG. 1.1: Présentation d'une sortie de randonnées à ski au travers deux modes différents : l'image et le récit

Le domaine qui s'intéresse à la production d'une représentation des données en langage naturel est la génération automatique de textes ; plus concrètement, la sous-branche, connue sous le nom de *Data-to-text* [Reiter et Dale, 2000]. Il s'agit d'une approche qui a connu une forte poussée industrielle ces

dernières années grâce aux compagnies comme *Arria*<sup>1</sup>, *Narrative Science*<sup>2</sup> ou *Syllabs*<sup>3</sup>. L'état de l'art [Gatt et Kraemer, 2018] montre qu'il est possible d'obtenir des textes cohérents qui abstraient l'information enregistrée dans les données. Cependant, ces travaux montrent que bien que l'information textuelle communiquée soit plus efficace que les informations rendues sous forme graphique [Portet *et al.*, 2009] les textes générés ne respectent pas les propriétés d'un récit [Reiter *et al.*, 2008].

## Les récits comme solution pour transmettre l'expérience vécue

La figure 1.1 illustre bien le type d'information pertinente à transmettre sous forme de récit. En effet, les récits mettent en action des personnages, des scènes, des motivations etc. et sont donc particulièrement adaptés à la transmission d'expériences vécues à la différence par exemple de descriptions (p.e., explication de montage d'un meuble). De ce fait, cette thèse se focalise sur le récit d'activités humaines ce qui correspond non seulement aux thèmes abordés dans les récits de la vie courante [Herman, 2009b] mais également aux données majoritairement capturées par les capteurs portables.

L'activité humaine est étudiée par un grand nombre de disciplines scientifiques dans des objectifs très différents. Les matériaux utilisés pour conduire ces études sont également très divers. Dans les discipline des sciences humaines et sociales, cela peut-être les échanges vocaux [Herman, 2009b], les messages textuels [Fisher, 1985], les vidéos etc. L'arrivée des capteurs et de la numérisation rend possible l'étude à large échelle de grande quantité de traces pour lesquelles l'informatique est sollicitée mais pour laquelle les méthodes de présentations classiques ne sont pas adaptées.

Le récit textuel est donc un moyen approprié pour faciliter la compréhension d'une activité par l'humain et ce, pour des buts divers (prévention, validation d'expérience...). Nous proposons de l'utiliser comme moyen de présenter des données captées sur une activité.

## Objectif et conduite de la thèse

Cette thèse porte sur la contribution apportée à la génération automatique de récits textuels à partir de données numériques captées (données

---

<sup>1</sup>[www.arria.com](http://www.arria.com)

<sup>2</sup><http://narrativescience.com>

<sup>3</sup><https://www.syllabs.com/fr>

brutes). À la différence des travaux de recherche mentionnés plus haut, le défi soulevé dans cette thèse est le développement d'une approche capable de concilier deux mondes très différents :

- celui de la réalité physique dans laquelle l'activité enregistrée se déroule et dont les données captées contiennent des observations objectives
- celui de la subjectivité humaine constitué des intentions, des décisions et des perspectives prises par les acteurs de l'activité.

Notre objectif est de concevoir une méthode de génération de récits que nous appellerons *récits d'activité*, qui offre la possibilité d'assembler une perspective personnelle d'une activité vécue, motivée par un but communicationnel et dans un format compréhensible facilitant son partage. Cet objectif pose plusieurs questions de recherche. Les récits vont dépendre de la quantité et du contenu des données collectées. Cette dépendance implique que l'approche tienne compte de l'absence de données par exemple. Il faut ensuite s'assurer que ces données soient transmises de manière à être comprises (si des données sont liées il faut que cela soit traduits dans le récit) mais également pour satisfaire le but recherché (prévention, réaction attendues...). Cette problématique se traite à deux niveaux : sur la sélection et l'organisation des faits à inclure dans le récit et sur la génération de surface (production du texte lui-même). Nous avons abordé ces problématiques en six sous-objectifs :

**Définition formelle de récit d'activité** Étant donné que le type de texte que nous proposons n'est pas parmi ceux déjà étudiés par les théories du récit, nous proposons une définition et des reformulations plus appropriées à notre contexte de travail.

**Constitution de corpus** Nous n'avons pas pour objectif de produire des récits identiques à ceux que des auteurs humains produisent. Cependant, afin d'identifier les mécanismes mis en place pour transmettre un message, nos travaux s'appuient sur l'analyse de récits textuels existants. Pour cela, nous avons identifié un domaine applicatif qui sert de cas d'étude tout au long de nos travaux : le ski de randonnée. Cette activité a été choisie pour 3 raisons : elle s'effectue de plus en plus avec des capteurs portés, il s'agit d'une activité physique contrainte et dont le déroulé est générique (parcours pour atteindre un sommet puis retour au point de départ) mais qui implique une grande variabilité d'exécution et d'acteurs et enfin, les skieurs sont de grands communicants qui échange les récits de leurs sorties sur plusieurs forums avec des objectifs communicationnel variés (loisir, lien social, performance, danger, etc.). Ce sont ces productions écrites qui composent les corpus textuels. Cependant afin de concevoir notre approche, un corpus 'parallèle' de données captées et de récits correspondant est nécessaire.

**Abstraction des connaissances du monde objectif** À partir des observations issues de l'étude des corpus, nous proposons la création de modèles avec lesquels nous allons représenter l'activité et l'environnement. Ces modèles permettent de représenter les concepts de l'activité qui doivent être vérifiable par les experts de celle-ci.

**Abstraction des connaissances du monde subjectif** La manière de rapporter les faits nécessite l'analyse des corpus et la modélisation des éléments importants des récits ainsi que leur lieu.

**Liens et passage des mondes objectif et subjectif** En se reposant sur ces modèles, il est possible de définir la transposition du monde objectif des données vers le monde subjectif et les paramètres à prendre en compte pour réaliser celui-ci.

**Génération du texte** Finalement, à partir d'une instance de récit il faut générer un texte en langage naturel correcte et qui transmette le déroulé du récit en respectant un but communicationnel fixé.

Nous avons traité ces sous-objectifs en nous reposant sur des travaux menés dans différentes disciplines (et sous-disciplines) pour les appliquer à une problématique de la Narratologie Informatique : la génération automatique de récit à partir de données captées.

## Plan du manuscrit

Pour aborder ces sous-objectifs, nous avons structuré le manuscrit en 7 chapitres détaillés ci-dessous.

**Chapitre 2** Nous présentons une introduction à la définition du récit à travers les principales théories dont il est l'objet d'étude, notamment la Narratologie et les Sciences Cognitives. C'est à partir de celles-ci que nous allons explorer le concept du récit ainsi que ces composantes. Cette partie théorique donne les éléments fondamentaux sur lesquels repose notre travail.

**Chapitre 3** Nous exposons un bref état de l'art de la génération automatique de récits. Ils reprennent les concepts exposées dans le Chapitre 2, en ayant comme objectif la mise en œuvre informatique des stratégies narrative permettant de produire une histoire. Ce chapitre expose également les avantages et les limites de ces travaux par rapport à notre objectif.

**Chapitre 4** Nos présentons dans ce chapitre le type de récit qui constitue notre objectif : le récit d'activité. Nous y développons la problématique inhérente à la transposition du monde objectif des données vers

le monde subjectif de l'expérience vécue. Ainsi, nous proposons la définition formelle du récit qui va rendre possible cette transposition, des verrous scientifiques, et de la méthodologie de travail pour surmonter ces verrous. Par ailleurs, nous présentons le cas d'usage choisi pour tester l'approche : le ski de randonnée.

**Chapitre 5** Ce chapitre expose la constitution des corpus utilisés pour l'analyse. Trois corpus ont été acquis : un corpus de récits textuels de grande taille dont l'objectif est l'obtention des connaissances statistiques sur le langage utilisé et les éléments narratifs cités, un deuxième corpus textuel pour l'étude qualitative structurelle du récit et à des fins de modélisation et enfin, le corpus parallèle, composé de données réelles produites pendant la réalisation des sorties accompagnées de leur récits écrits par les randonneurs.

**Chapitre 6** Ce chapitre présente les modèles de l'activité. Lors de l'étude de corpus autour de l'activité, des connaissances ont été obtenues pour constituer les représentations formelles du domaine et de l'activité. Ce chapitre détaille donc les analyses effectuées, les justifications pour sélectionner l'approche par modèles, les théories utilisés ainsi que la méthodologie mise en place pour les construire. Le résultat est un modèle du domaine sous forme d'ontologie et un modèle de l'activité sous forme du modèle de tâches.

**Chapitre 7** Ce chapitre est dédié au modèle du récit. Il s'agit d'un point important de la thèse qui présente la méthode employée pour passer du monde objectif des données vers une interprétation particulière. Pour ce faire, nous présentons une analyse de corpus dont les résultats montrent comment les personnes organisent différemment leur récit selon l'interprétation du vécu et le but communicationnel. C'est à partir des connaissances acquises que nous proposons une méthode pour construire un plan de récit. Ce plan du récit est basé sur le modèle des *plots units* [Lehnert, 1981] mais adapté à notre contexte du travail. Finalement, nous présentons les trois types de plan qui projettent les trois types de vécu identifiés durant l'étude du corpus, ainsi que la génération du récit textuel final.



## Chapitre 2

# État de l'art : Définition, génération et communication de récits

L'objectif de la thèse est de produire automatiquement un récit à partir de données capturées lors d'une activité humaine. Le rôle du récit n'est pas de rendre compte de l'activité, de manière plus ou moins synthétique. Il est de communiquer ce qui a été vécu. A ce titre, il doit présenter un ensemble de dimensions narratives qu'il convient de spécifier. Avant d'introduire l'état de l'art de la génération automatique de récits (cf. Chapitre 3), nous présentons dans ce chapitre une introduction aux principales études théoriques autour des récits tant dans l'aspect narratologique que cognitif. Ces théories et modèles ont servi de base à la plupart des systèmes de génération automatique de récits qui seront décrits dans le chapitre suivant. Les différentes approches théoriques du récit en tant qu'objet sont présentées dans la section narratologie en section 2.1. La section 2.2 présente, quand à elle, les constituants consensuels d'un récit selon les théories précédemment introduites. Ces constituants sont la base de la formalisation informatique des générateurs de récits. Les approches théoriques ne se limitent pas à l'étude du contenu d'un récit, elles abordent également le récit comme un objet de transmission. Ces approches cognitives de la narratologie sont résumées en section 2.3. Ce chapitre se termine par un rapide bilan des approches présentées. Enfin, il convient de noter que cet état de l'art se concentre principalement sur les récits textuels (roman, contes, théâtre) et ne considère les récits oraux ou filmiques que de manière très ponctuelle <sup>1</sup>.

---

<sup>1</sup>voir notamment les travaux de Labov[Labov, 1972] sur ce sujet.

## 2.1 Études de l'objet récit : la narratologie

La narratologie est une discipline qui étudie les structures et techniques de la narration. Elle cherche à comprendre la composition du récit ainsi que leur fonction dans la communication humaine. Le terme « narratologie » a été adopté pour la première fois par [Todorov, 1969] qui définit la narratologie comme la science des récits. La narratologie est en constante évolution grâce aux apports d'autres disciplines comme la linguistique ou la philosophie et plus récemment les sciences cognitives.

Toutefois, malgré l'établissement de la narratologie, son sujet d'étude, le récit, reste difficile à définir. De façon très générale, un récit est une représentation d'une série d'événements temporeux et causaux [Landa et Onega, 2014]. Plusieurs propositions ont été élaborées afin d'avoir un concept plus précis qui puisse mettre en valeur la complexité du phénomène narratif. Les essais pour définir le concept du récit naissent de la nécessité de comprendre sa nature et sa valeur universelle [Perron, 2003].

### 2.1.1 Pré-formalisme russe

Les premiers travaux du domaine se sont intéressés à la structure narrative des œuvres dramatiques. Ceux-ci remontent à Aristote, qui identifie trois parties constitutives dans les drames : un début, un milieu et une fin [Aristote, 1980]. Toutefois, un des premiers essais modernes concernant l'étude de la structure des textes littéraires est proposé par Freytag [Freytag, 1900]. À partir de son étude autour de Shakespeare et des autres classiques de la littérature, il a observé que ces œuvres possédaient sept moments clés : *exposition*, *incident déclencheur*, *complication*, *climax*, *falling action*, *résolution* et *dénouement*. Ces moments définissent la structure nommée *pyramide de Freytag* (Figure 2.1). L'*exposition* présente la situation initiale. L'*incident déclencheur* est le problème qui provoque la chaîne d'événements à narrer dans l'œuvre. La *complication* correspond à une série d'obstacles à surmonter pour solutionner le problème posé. Le *climax* est l'obstacle le plus grand auquel le protagoniste doit faire face (par exemple, une lutte contre un vilain). Le *falling action* définit les actions finales des personnages. La *résolution* est la phase qui amène au dénouement.

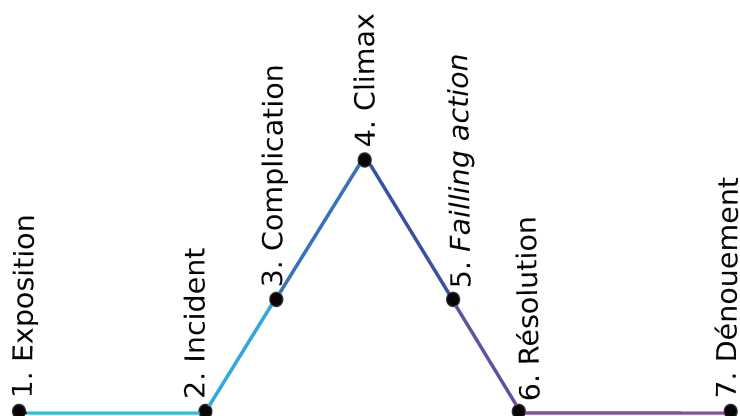


FIG. 2.1: La pyramide de Freytag

Cette structure dramatique en pyramide de Freytag se retrouve dans plusieurs œuvres de la littérature comme dans *Une vendetta* de Guy de Maupassant. L'histoire se déroule dans un petit village de Corse, où habitent une veuve et son fils (exposition). Le fils est mort tué d'un coup de couteau dans le ventre, à la suite d'une dispute (incident déclencheur). La mère devant le cadavre de son enfant lui promet la vendetta (complication). Elle entraîne sa chienne pour quelle devienne agressive pendant quelques mois (climax). Finalement, la veuve retrouve celui qui a tué son fils et elle lâche sa chienne pour qu'elle l'attaque (falling action). La chienne tue l'homme (résolution). La femme retourne chez elle et dort enfin tranquille (dénouement).

Ces types de structures de haut niveau sont utilisés pour l'analyse et la catégorisation des œuvres de fiction dramatique. Si elles permettent de structurer une trame narrative elles sont, par contre, peu adaptées à la narration de faits réels ou de trames plus complexes.

### 2.1.2 Le Formalisme Russe

Les prémices de la narratologie comme discipline visant à formaliser les œuvres littéraires sont identifiés comme étant *le Formalisme Russe*. Il s'agit d'un courant né en Russie dans les années 1914 et qui s'achève vers 1930. Ce mouvement incluait des spécialistes de l'étude de la langue et des théoriciens de la littérature. Parmi eux, Boris Eichenbaum et Victor Chklovski [Erich, 1973] cherchaient à découvrir la nature des textes littéraires à partir d'une approche formelle qui ne se concentrait que sur les structures et sur leurs éléments constitutifs, en laissant de côté d'autres aspects tels que le contexte culturel ou social [Meister, 2009].

En 1928, Vladimir Propp publia *Morphologie du conte* [Propp, 1970], où

il décrit les régularités morphologiques qui apparaissent dans plus d'une centaine de contes folkloriques russes. Pour lui, tous ces contes se sont formés à partir d'une même base narrative (un conte premier), laquelle est composée de *fonctions narratives* et de *personnages*. D'après Propp : « Par *fonction*, nous entendons l'action d'un personnage, définie du point de vue de sa signification dans le déroulement de l'intrigue » [Propp, 1970, p. 150]. Ces *fonctions* se retrouvent dans tous les contes et Propp en a identifié 31 au total listées Figure 2.2

- |  |  |
|--|--|
| 1. Éloignement ou Absence                      | 17. Le héros reçoit une marque             |
| 2. Interdiction                                | 18. Défaite du vilain                      |
| 3. Transgression                               | 19. Résolution du forfait initial          |
| 4. Interrogation                               | 20. Retour du héros                        |
| 5. Information                                 | 21. Le héros est poursuivi                 |
| 6. Tentative de tromperie                      | 22. Le héros échappe aux obstacles         |
| 7. Le héros se laisse tromper                  | 23. Arrivée incognito du héros             |
| 8. Le vilain réussit son forfait               | 24. Un vilain réclame la récompense        |
| 9. Demande au héros                            | 25. Épreuve de reconnaissance du héros     |
| 10. Acceptation de la mission par le héros     | 26. Réussite du héros                      |
| 11. Départ du héros                            | 27. Le héros est reconnu grâce à sa marque |
| 12. Mise à l'épreuve du héros par un donateur  | 28. Le faux héros/vilain est découvert     |
| 13. Le héros passe l'épreuve                   | 29. Le héros est transfiguré               |
| 14. Don : le héros reçoit un présent           | 30. Le vilain est puni                     |
| 15. Arrivée du héros à l'endroit de sa mission | 31. Mariage ou accession au trône          |
| 16. Combat du héros et du vilain               |  |

FIG. 2.2: Liste des 31 fonctions identifiées par Propp

Par exemple, la fonction d'information (5.) s'applique aussi bien quand le héros reçoit de l'information sur son ennemi que l'inverse. Une suite de *fonctions* décrivant le conte forme une ou plusieurs *séquences* caractéristiques et qui donne sa *morphologie* au conte. Par contre, certaines fonctions ne peuvent être associées à certains personnages. Par exemple, un vilain ne peut pas réaliser la fonction numéro 31 : *Mariage ou accession au trône*. Les personnages tiennent eux-même un rôle complètement défini. On trouve ainsi les personnage suivants

- *Le héros*.
- *Le faux héros* qui cherche à s'approprier les mérites du héros.
- *Le mandateur* qui désigne l'objet de la quête.
- *La princesse ou son père* qui est l'objet de la quête.
- *Le donateur* qui confie un objet magique au héros.
- *L'adjuvant* qui aide le héros à atteindre son objectif.
- *L'agresseur ou méchant* qui cherche à empêcher la quête du héros.

Le tableau 2.1 donne la décomposition du conte *Les Oies sauvages* selon le formalisme de Propp. Le résultat de cette analyse donne la séquence structurale ou *morphologie* du conte. Les travaux de Propp sont particulièrement importants car ils vont influencer les conceptions des structuralistes français des années 60 et seront reprises comme support théorique par plusieurs systèmes de génération de récits fictifs.

Conte	Fonction narrative
<i>Il était une fois, un couple qui avait laissé leur très jeune fils à la garde de sa grande sœur.</i>	2. Interdiction
<i>Mais elle relâcha sa vigilance et le garçon disparut.</i>	4. Interrogation
<i>Les oies sauvages l'avaient enlevé pendant que la grande sœur ne regardait pas.</i>	8. Manque ou méfait
<i>Poursuivant les ravisseurs, la jeune fille rencontra un four à pain, un pommier sauvage et une rivière de lait qui lui proposèrent un mets.</i>	12. Première fonction du donneur
<i>Mais elle les dédaigna.</i>	14. Réception de l'objet magique (non accepté)
<i>Elle retrouva son frère chez Baba Yaga, le saisit et s'enfuit avec lui.</i>	20. Retour du héros
<i>Cependant, Baba Yaga envoya les oies sauvages après elle.</i>	21. Poursuite
<i>Elle supplia la rivière, le pommier et le four de l'aider.</i>	22. Secours
<i>Elle accepta les mets proposés.</i>	14. Réception de l'objet magique
<i>Ils échappèrent ainsi aux oies sauvages et arrivèrent à la maison au même temps que leurs parents.</i>	

TAB. 2.1: Conte *Les Oies sauvages* analysé avec les fonctions de Propp

Le travail de Propp fut et reste populaire en narratologie mais reçut égale-

ment beaucoup de critiques. En effet, cette approche met l'accent principalement sur les fonctions en considérant les personnages comme secondaires, or certains personnages transportent du sens ne serait-ce que dans le contexte culturel et ne peuvent donc être considérés que comme un simple rouage de l'histoire. De plus, si les fonctions permettent l'analyse des contes populaires russes, Propp ne dit pas comment ces fonctions doivent s'enchaîner pour générer un conte. Enfin, cette approche n'est pas adaptée aux contes modernes et aux histoires réelles.

### 2.1.3 Le Structuralisme Français

La prochaine grande période de la narratologie est celle suivant le courant structuraliste qui s'étend entre les années 1960 et 1980. Cette période est connue comme celle de la *narratologie classique* et dont les membres les plus influents sont Claude Brémont, Algirdas Julien Greimas, Tzvetan Todorov (qui a été le premier à parler de *narratologie*), Roland Barthes et Gérard Genette. Cette période de narratologie classique s'est intéressée à construire des systèmes formels décrivant les constituants et les structures relationnelles descriptives pouvant s'appliquer à toute forme de littérature [Fludernik, 2009]. L'un des points importants de la narratologie classique est d'avoir systématiquement cherché à distinguer les événements liés à une histoire de la manière dont cette histoire est narrée, reprenant à leur compte la séparation *fabula/syuzhet* (les constituants bruts de l'histoire/la manière dont l'histoire est organisée) puis l'approfondissant.

Plusieurs définitions du récit ont été proposées durant cette période. Pour Claude Brémont, un récit est défini comme une série de séquences élémentaires dérivées d'une grammaire universelle ; chaque séquence a trois *fonctions* contenant les phases élémentaires d'un processus narratif (une éventualité à narrer dans le récit) : la fonction avec le début de l'éventualité, la fonction avec l'éventualité même et la fonction avec la conclusion de l'éventualité ou le résultat atteint [Brémont, 1973]. Todorov propose aussi une grammaire narrative universelle où le récit est une structure qui présente la transformation d'une situation en deux épisodes : le premier présente une situation stable, laquelle est perturbée par une action contraire ; le deuxième épisode porte sur le retour à l'équilibre, grâce à la réalisation d'actions opposées [Todorov, 2013].

Greimas reprend les travaux de Propp et propose le *schéma actanciel*, présenté Figure 2.3. Il substitue aux sept sphères d'action des personnages de Propp trois types d'actants organisés selon des relations contradictoires :  $\leftrightarrow$  objet, destinataire  $\leftrightarrow$  destinataire et opposant  $\leftrightarrow$  adjuvant [Greimas, 1966]. Ils forment trois axes : l'axe du désir (sujet et objet liés par une quête), l'axe du

pouvoir (opposant et adjuvant liés par une lutte pour le pouvoir) et l'axe de la transmission (destinateur et destinataire liés par un acte de communication). Par exemple, dans le conte du *petit chaperon rouge*, plusieurs rôles actanciels peuvent être attribués aux personnages : la petite fille (le sujet), sa mère (destinateur), sa grand-mère (destinataire), la galette à donner (l'objet), le loup (l'opposant) et le chasseur (l'adjuvant).

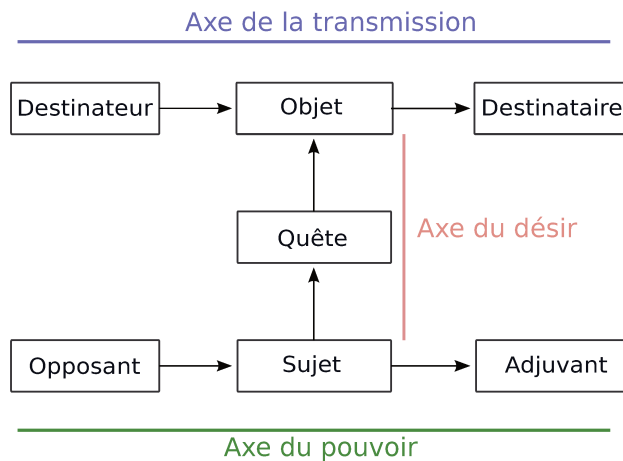


FIG. 2.3: Schéma actanciel selon Greimas

Une autre approche s'inspirant également de Propp est présentée par Roland Barthes [Barthes, 1966]. Barthes observe que les éléments d'un récit peuvent se connecter non seulement de manière séquentielle mais aussi à travers des niveaux d'abstraction, en émulant les axes syntagmatiques et paradigmatiques du langage.<sup>2</sup> Il définit le récit comme la combinaison des *fonctions* décrites sur la Figure 2.4. Pour Barthes « un récit n'est jamais fait que de fonctions ». Il distingue deux types selon son organisation : des *fonctions* (terme utilisé à nouveau, mais ici *fonction* signale particulièrement les opérations effectuées dans le récit) ; et des *indices* (éléments du récit liées à un signifié). Les fonctions et les indices sont divisés selon leur contenu. Les premiers en *cardinales* (fonctions qui portent la progression de la narration) et *catalyseurs* (fonctions qui complètent les cardinales) ; alors que pour les deuxièmes, nous trouvons les *indices* (terme employé à nouveau, mais

<sup>2</sup>En linguistique structurale, l'axe syntagmatique enchaîne les éléments d'une phrase les uns après les autres, sur un plan horizontal, afin de former un message ou une phrase. L'axe paradigmatique leur attribue une valeur à partir d'une catégorie ou d'un modèle établi. Par exemple, « Marie est une jolie fille » combine trois éléments « Marie », « est » et « une jolie fille » (axe syntagmatique), alors que nous identifions « Marie » comme un nom paradigme sujet, « est » comme verbe et « une jolie fille » comme attribut du sujet (axe paradigmatique).

maintenant pour indiquer des traits descriptifs comme les sentiments ou les sensations) et les *informants* (des repères spatio-temporels). Donc, il s'agit d'une définition structuraliste qui essaye d'organiser le récit non seulement au niveau causal, mais aussi au niveau du contenu. Par exemple, si nous revenons sur *le petit chaperon rouge*, nous pourrions trouver des *fonctions cardinales* (le petit chaperon arrivant à la maison de sa grand mère), des *catalyses* (le loup ouvrant la porte au petit chaperon). De même, nous trouvons dans ce conte des *informants* (description du loup déguisé comme la grande mère) et des *indices* (l'étonnement du petit chaperon lorsqu'elle regarde le corps de sa grande-mère).

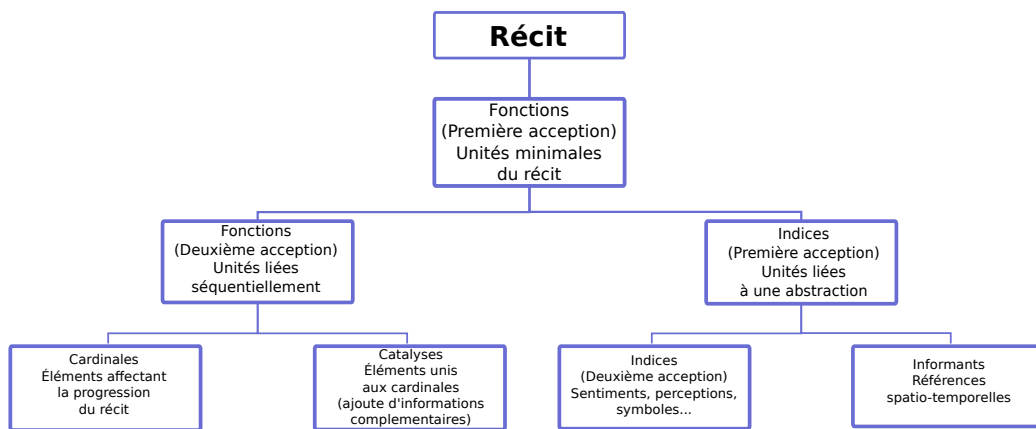


FIG. 2.4: Les fonctions d'après Barthes

Ces auteurs apportent surtout des théories axées sur la formalisation sémiotique d'une séquence d'événements narrés. D'autres théoriciens structuralistes, tels que Gérard Genette, s'intéressent à la manière, c'est-à-dire aux moyens dont l'auteur dispose pour narrer un récit (forme, voix, point de vue, etc.) sans se focaliser sur le sens [Fludernik, 2009]. Pour Genette, le récit est l'énoncé (discours écrit ou oral) à partir duquel nous avons accès aux événements racontés [Genette, 2007]. Il considère trois niveaux : l'histoire (« l'ensemble des événements racontés » [Genette, 2007, p297]), le récit, (« le discours oral ou écrit qui raconte [les événements d'une histoire] » [Genette, 2007, p297]) et finalement la narration (« l'acte réel ou fictif qui produit [le récit] c'est à dire le fait même de raconter » [Genette, 2007, p297]). L'apport de Genette en narratologie fut très important et sera synthétisé en section 2.2.

C'est ainsi que Genette et d'autres protagonistes de l'ère structuraliste vont contribuer au développement de la narratologie. Leurs travaux serviront de base à une nouvelle étape dans la définition des récits dans laquelle l'intégration d'autres domaines va enrichir l'étude du texte narratif. De nombreux systèmes de génération se basent encore sur les théories de cette époque.



### 2.1.4 L'ère post-structuraliste de la narratologie

Depuis les années 80, la narratologie est entrée dans l'ère post-structuraliste dans laquelle les approches de la narratologie se sont diversifiées en intégrant d'autres disciplines et d'autre support narratifs. En France, Jean-Michel Adam émet le postulat qu'un texte narratif ne se limite pas aux textes littéraires classiques ; le récit peut apparaître dans d'autres types de texte, comme les reportages, les journaux et même aborder des sujets appartenant à la vie quotidienne. Pour lui, le récit est un texte contenant une succession d'événements (ce qu'il appelle, ordre *chronologique*) qui sont organisés de manière à ce que nous puissions identifier une macro-structure avec laquelle nous pouvons déduire un sens, une cohésion du texte (ce qu'il appelle, ordre *configurationnel*) [Adam, 1999].

Au pays-bas, Mieke Bal reprend les concepts structuralistes et les organise de manière à ce qu'ils puissent être utilisés pour analyser et interpréter les textes narratifs de façon systématique. Elle publie ainsi *Narratology : Introduction to the Theory of Narrative* [Bal, 2009]. Une des contributions de ce livre est le raffinement des dimensions narratives. Bal suggère un modèle à trois niveaux, au lieu du binôme *histoire - discours*. Pour elle, il y a la *fabula* (l'ensemble des événements liés à l'histoire — ce que Genette appelait l'histoire), l'*histoire* (la fabula organisée d'une certaine manière — le résultat de la narration selon Genette) et le *texte* (la forme finale dans laquelle l'histoire est communiquée — ce que Genette appelait le récit).

Après avoir présenté les approches théoriques du récit en tant qu'objet, nous détaillons dans la section suivante ce que sont les constituants d'un récit. Ces constituants sont la base de la formalisation informatique des générateurs de récits.

## 2.2 Composition d'un récit textuel

Dans cette thèse, il est nécessaire de définir le récit et ses principales composantes afin de pouvoir proposer un modèle formel utilisable en informatique. Dans la section précédente, nous avons présenté des travaux qui donnent différentes façons de considérer le récit et l'acte narratif. Cependant, dans le domaine de la narratologie, certains éléments font consensus par rapport aux éléments constitutifs d'un récit. Cette section a pour but de décrire ces éléments qui seront utilisés dans la suite de ce manuscrit.

## 2.2.1 Les niveaux de description du récit

En narratologie, il existe d'abord un consensus théorique autour des niveaux de description d'un récit. Une narration (orale ou écrite) contient trois niveaux tel qu'illustré Figure 2.5. Le premier niveau, nommé *fabula*, contient tous les événements et faits de l'histoire ordonnés chronologiquement. Le deuxième niveau, nommé *sujet* (ou *histoire*) est une présentation orientée selon une logique narrative (p.ex., une intrigue) qui peut être chronologique ou non. Enfin, le troisième niveau est lié aux média utilisés pour projeter le discours (texte, image, films, etc.).

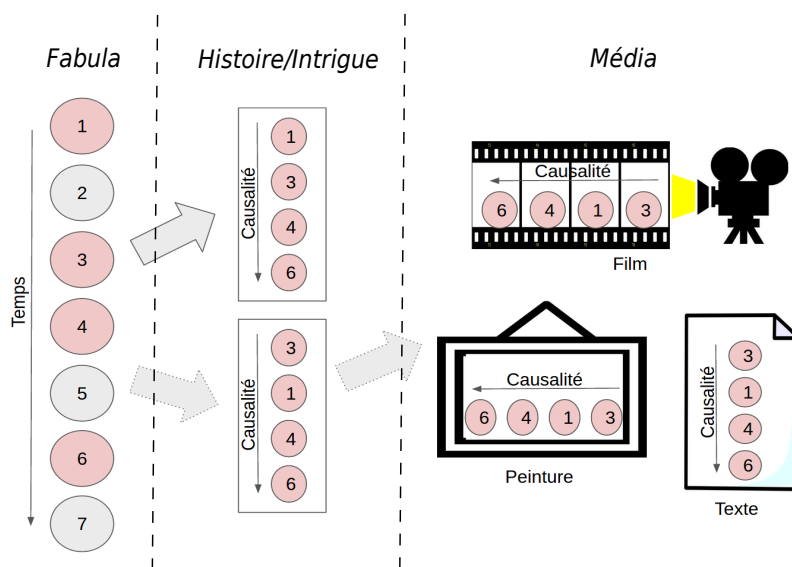


FIG. 2.5: Représentation des trois niveaux de description du récit pour lesquels il existe un consensus théorique en narratologie.

Les éléments qui composent ces niveaux de description sont décrits dans les sous sections suivantes.

## 2.2.2 Les composants de l'histoire

Il convient d'étudier plus en détail ce que contient chacun de ces niveaux de description. À cet égard la typologie de Chatman [Chatman, 1977] représentée Figure 2.6 apporte un éclairage intéressant. Cette typologie sépare les éléments liés au déroulement de l'histoire (le contenu) de ce qui relève du mode d'expression (le discours). L'histoire est composée d'événements (occupation temporelle) et d'existants (occupation spatiale) qui proviennent de

l'imaginaire de l'auteur. Le discours est quant à lui vu comme selon une certaine structure narrative (forme d'expression) et est ancré dans un artefact physique (substance de l'expression).

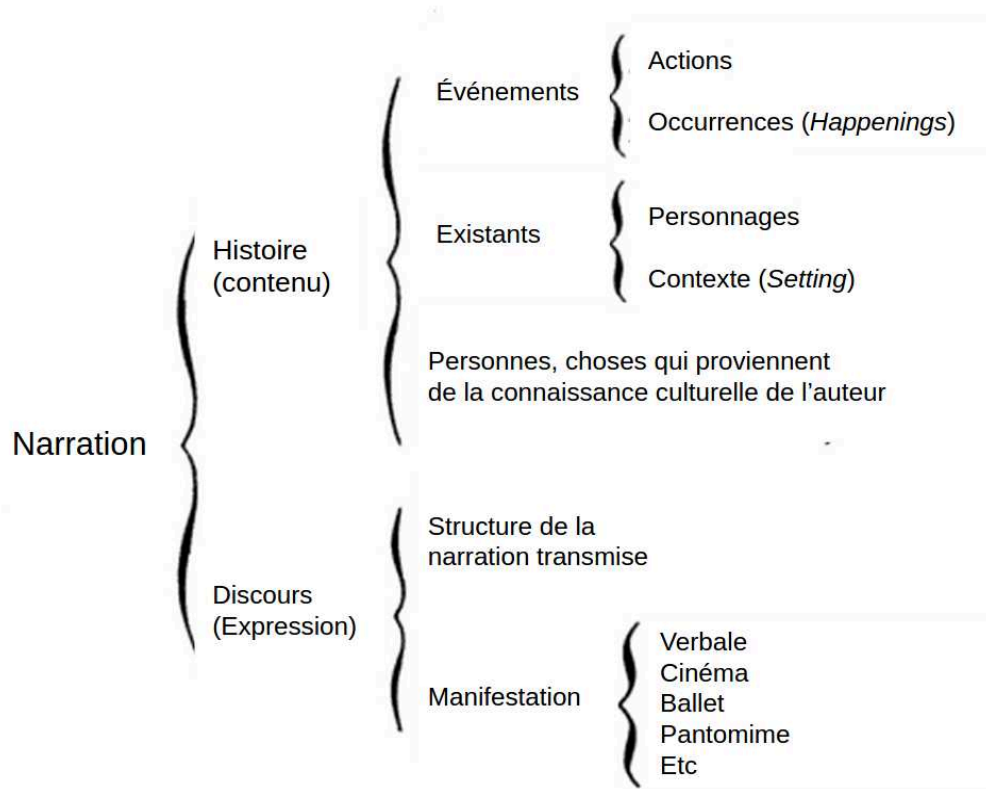


FIG. 2.6: Organisation des composants d'un récit selon Chatman [Chatman, 1977]

### 2.2.2.1 Événements

L'événement est l'élément moteur du récit : il est présent dans toutes les théories narratives. Pour [Bal, 2009], un événement est la transition d'un état vers un autre (un *état* étant défini comme un ensemble de circonstances). Par exemple, le loup soufflant sur la maison du petit cochon fait passer le cochon de l'état en sécurité à l'état en danger. Les événements peuvent être des processus (p.ex, lever du soleil) ou des mouvements (p.ex., coup de pied), qui encouragent ou impliquent un changement. Dans la même idée, Chatman fait une subdivision pour préciser la nature des événements. Ils sont classifiés en *Actions* (événements provoqués) et *Occurrences* (*Happenings* en Anglais, évé-

nements perçus ou expérimentés) [Chatman, 1977]. Les événements peuvent aussi inclure des prises de parole, des pensées et des sentiments [Chatman, 1977]. Il met aussi l'accent sur les différents niveaux d'importance des événements dans l'histoire. Ainsi, il y a des *noyaux*, c'est à dire, des événements clés pour le déroulement de l'histoire et des *satellites*, des événements qui complètent les noyaux, mais dont la présence n'est pas essentielle pour comprendre l'histoire. Cette vision est très proche de celle de Barthes [Barthes, 1966], avec les *fonctions cardinales* et les *catalyseurs*.

### 2.2.2.2 Existants

Les existants sont les entités qui se trouvent dans l'espace de l'histoire. Alors que les événements sont ancrés temporellement, les existants peuvent être vu comme des entités autour desquelles se développe l'histoire [Herman, 2009b]. Ils sont classés en deux catégories : les *personnages* et le *contexte* (ou *setting*).

Pour Chatman, les *personnages* peuvent être vus comme un ensemble de traits (caractéristiques du personnage) qui montrent une qualité personnelle, lesquels peuvent évoluer ou être modifiés au cours de l'histoire [Chatman, 1977]. Bal, quant à elle [Bal, 2009], les présente comme des acteurs dont on indique toutes les actions (humaines ou non). Ils sont les déclencheurs des événements. Les personnages peuvent être complètement fictifs, basés sur des individus de la vie réelle ou des représentations symboliques. Deux types de personnages sont présents dans l'histoire selon le degré d'importance dans le monde de l'histoire : principal (qui reste au centre de l'histoire) ou secondaire (rôle support pour les actions du personnage principal). Ceci est en phase avec les théories précédentes, notamment celle de Propp [Propp, 1970] (personnages stéréotypés) ou de Greimas [Greimas, 1966] (personnages jouant des rôles par rapport au sujet de l'histoire).

Les événements sont où les personnages participent, délimités dans un temps et espace spécifique. Les informations qui aident à situer le contenu des événements sont nommées *contexte* ou *setting*. Il s'agit d'un décor (ambiant, objets, etc) dans lequel les événements concernant des personnages se déroulent.

### 2.2.2.3 Intrigue

L'intrigue est un terme dont l'équivalent anglais *Plot* est souvent repris dans les théories et applications anglo-saxonnes. En Français, on peut définir l'intrigue comme la succession des faits et personnages essentiels permettant de saisir la *trame* de l'histoire. En reprenant les niveaux de récit cités en

section 2.2.1, l'intrigue indique un ensemble d'événements sélectionnés de la *fabula*, organisés de manière causale. Bien que la définition de l'intrigue soit particulièrement délicate, certains auteurs tels que Forster [Forster, 2016] sont connus pour leur vision tranchée de l'intrigue. Par exemple pour Forster la phrase « Le roi est mort et la reine est morte » serait une histoire composée d'événements s'enchaînant et changeant l'état du monde de l'histoire. Il est important de comprendre que quand ici Forster parle de *story* il faut comprendre *fabula*. Par contre, un texte tel que « Le roi est mort et la reine est morte de chagrin » serait une intrigue car le syntagme « de chagrin » donne une liaison causale entre les deux événements. L'intrigue est donc un arrangement de haut niveau de la *fabula* dans laquelle des interprétations sont fournies permettant d'expliquer l'histoire qui sera transmise. Plus précisément, on peut voir l'intrigue comme une réponse à la question *pourquoi les événements se sont-ils enchaînés ainsi ?* [Forster, 2016].

### 2.2.3 Les stratégies du discours

Pour construire une histoire puis un récit à partir d'une *fabula*, les éléments à narrer subissent un certain nombre de transformations. Si nous avons simplement vu une seule structuration de l'histoire par l'intrigue, l'expression de cette histoire sur le support textuel présente, elle, un grand nombre de variantes qui ont été exposées dans l'ouvrage de Gerard Genette, *Discours du récit* [Genette, 2007]. Dans cet ouvrage, Genette analyse les différentes stratégies qu'utilise un auteur pour rendre son récit. Nous résumons une partie de cet ouvrage en respectant la typologie de l'auteur.

**Voix** La voix fait référence à l'énonciation du discours et rend explicite qui parle, qui raconte. En narratologie, on distingue au moins les quatre catégories suivantes :

- *L'auteur* qui est l'entité créatrice du récit. C'est lui qui développe le monde où le récit se passe. Il se sert d'une matière source de son imagination ou de son vécu pour créer le récit. C'est ce que Chatman met en avant dans son organisation Figure 2.6 sous la forme des « connaissances culturelles de l'auteur ». L'auteur a une intention communicative qui a motivé la création du récit, tels que l'amusement, l'avertissement ou l'empathie. L'auteur peut parfois être cité dans le texte (p.ex., « l'auteur prends plaisir à entretenir le suspense »), on parle dans ce cas d'auteur impliqué (cité dans le texte mais qui reste de toute façon virtuel) [Genette, 2007, p418].

- Le *narrateur* est l'entité qui transmet le récit, c'est l'entité qui « parle ». Genette distingue le narrateur absent de l'histoire qu'il raconte (narrateur hétérodiégétique) du narrateur personnage de l'histoire (narrateur homodiégétique). Si le personnage principal est lui-même le narrateur, on parle de *narrateur autodiégétique* [Genette, 2007, p418].
- Le *narrataire* est l'entité à qui est destiné le récit et reste tout aussi virtuel que le narrateur. Il peut être intradiégétique (appartenant à l'histoire), invoqué (qui appartient au monde de l'histoire mais n'a pas de rôle dans celle narrée) ou extradiégétique [Genette, 2007, p418].
- Le *lecteur* est l'individu récepteur du récit. Il sera en charge de déchiffrer le contenu du récit. L'auteur peut avoir une connaissance abstraite des lecteurs réels et à partir de là développer le monde du récit [Genette, 2007, p418].

Par ailleurs, deux autres notions permettent de situer la voix narrative :

- *Le niveau de narration* : une narration (celle connue comme *principale*) peut contenir d'autres narrations enchâssées dans lesquelles les narrateurs peuvent changer. C'est notamment le cas dans *Les mille et une nuits* où le narrateur n'est pas toujours Shéhérazade en fonction du déroulé de l'histoire.
- *Temps de la narration* : Il s'agit de la relation entre le temps du récit et celui propre à la narration. On distingue une narration *ultérieure* (après la narration en cours), *antérieure* (avant la narration en cours), *simultanée* (au moment de la narration en cours) et *intercalée* (p.ex., alternance entre le présent et le passé).

**Ordre** L'ordre concerne la disposition des événements dans le discours. Le temps du discours est la modification de l'ordre chronologique du temps de l'histoire. Ainsi, un discours peut présenter des *anachronismes*. Il y a deux sortes d'anachronismes relevés par Genette :

- *Analepse* : mention des événements antérieurs à ceux qui sont en train d'être narrés. Les retours en arrière peuvent être utiles, par exemple, dans les récits où il y a besoin d'expliquer les antécédents d'un personnage. Par exemple, « Il commence l'ascension de la montagne de la même manière que la semaine dernière ».
- *Prolepse* : mention des événements postérieurs à ceux qui sont en train d'être narrés. Cela peut être utilisé pour expliquer le destin d'un personnage ou d'un objet mentionné dans la narration. Par exemple, « Il a pris le chemin à gauche, ce qu'il va regretter par la suite ».

Ces interruptions chronologiques peuvent être proches des événements en cours de narration. Le niveau de proximité est appelé *portée*. Tandis que la longueur de l'interruption est connue comme *amplitude*.

**Durée** Dans le discours, la durée d'un événement est définie comme la quantité de mots ou pages utilisée pour raconter un épisode. Par exemple, il peut arriver qu'une rencontre d'une durée chronologique de deux heures soit racontée sur plusieurs paragraphes, qu'elle soit compressée en une ligne ou même enlevée du discours. Le rapport entre la durée réelle d'un événement et sa narration est connu comme la *vitesse narrative*. Genette mentionne quatre types de modifications du temps du discours (TR : Temps du Récit, TH : Temps de l'Histoire) :

- Ellipse (TR = 0 ; TH = n) : Un événement de l'histoire est omis dans le discours. Cette omission peut être implicite ou explicite. Par exemple, dans « Cinq mois plus tard... », les événements des cinq mois sont explicitement ignorés dans le récit.
- Pause (TR = n, TH = 0) : La narration est interrompue pour insérer des descriptions étendues. Par exemple, les descriptions très détaillées d'un paysage comme « On est arrivé. Le paysage est magnifique, le ciel est tout bleu, les montagnes... » arrête le déroulement de l'action pour donner place à une description du lieu.
- Scène (TR = TH) : La durée d'un événement est la même que dans le discours. Les dialogues (dans un récit ou une pièce de théâtre) sont des scènes. Par exemple « Et elle a dit : "Il faut passer par là !" ».
- Sommaire (TR < TH) : Un épisode de l'histoire est résumé dans le discours. Un événement vécu par un personnage peut être présenté de façon très courte afin d'accélérer la succession de la narration. Par exemple dans « Nous allons marcher pendant plusieurs heures », ce qui s'est passé dans ces heures est compressé en peu de mots.

**Mode** La façon de narrer une histoire peut varier. Entre la reproduction fidèle de l'histoire et le résumé distancié il existe plusieurs degrés de positionnement que Genette appelle la distance narrative. Ainsi, un récit peut prendre le parti de ne présenter qu'une partie des événements qui ne seraient pas connus du narrateur ou au contraire raconter l'ensemble des événements sans se soucier de comment ceux-ci auraient pu être connus du narrateur. C'est ce que Genette appelle la focalisation.

- Distance : la façon dont le narrateur est impliqué. Si les mots du personnage sont intégrés dans la narration, nous parlons de *discours nar-*

*rativisé*. Par contre, si les mots sont cités dans la narration tels qu'ils sont dits, il s'agit d'un *discours rapporté*.

- Focalisation : la sélection, la restriction ou la présentation des événements selon un point de vue déterminé. Il existe plusieurs types de focalisation :
  - Focalisation zéro. Le narrateur est omniscient et a accès à l'ensemble des événements et des points de vue des personnages.
  - Focalisation externe. Le narrateur est un témoin externe de ce qui se passe dans la narration, par exemple dans *Les misérables* de Victor Hugo : « Il était difficile de rencontrer un passant d'un aspect plus misérable. »
  - Focalisation interne. Le point de vue appartient à l'un des personnages, par exemple dans *L'île au trésor* de Robert Louis Stevenson : « J'entrai d'un pas ferme, les bras tendus en avant. »

Nous avons abordé dans les deux sections précédentes les théories relatives à l'objet porté par la narration, et à la façon de le construire. Dans le contexte de cette thèse, nous nous intéressons également aux contraintes communicatives qui régissent la production et la réception du récit. Nous devons donc non seulement explorer le récit du point de vue structurel, mais également du point de vue cognitif. C'est pourquoi nous exposons dans la section suivante quelques travaux liés à la narratologie cognitive.

## 2.3 La narratologie cognitive

L'exploration de l'acte narratif dans la pensée a contribué à la naissance de la *narratologie cognitive* [Herman, 1999]. En effet, ce sous-domaine, héritier de la recherche post-classique de la narratologie, regroupe des théories et modèles qui permettent d'expliquer comment les éléments à raconter sont traités par les personnes. Concrètement, la narratologie cognitive cherche à comprendre les opérations mentales mises en œuvre lors de la création ou la compréhension d'un récit (par exemple, la sélection du contenu pendant la production du texte, ou la déduction des éléments implicites pendant la lecture). C'est grâce à ces processus internes que les personnes sont capables de raconter non seulement des produits de l'imagination, mais aussi leurs expériences d'une manière structurée et transparente pour les autres. Pour la narratologie cognitive le récit est un outil de communication qui donne le déroulement, les transformations et la temporalité d'une certaine situation [Herman, 2009a].



Les études de la narratologie cognitive et plus particulièrement en psychologie cognitive cherchent à comprendre comment les récits se forment dans l'esprit. Par exemple, comment caractériser l'état mental des personnages, comment un récit se structure ou comment le monde dans lequel le récit a lieu émerge. C'est ainsi que nous nous intéressons à explorer ces pistes qui permettront de guider la création d'un système de génération de récit. Les sections suivantes résument des travaux portant sur la compréhension et la production de récits.

### 2.3.1 La compréhension du récit

Un des premiers modèles concernant la compréhension des récits est basé sur la notion de *grammaire de récits*. Cette approche repose sur le postulat que les récits possèdent une structure interne hiérarchique et déterministe semblable à celle des phrases. Bien que celle-ci soit plutôt à ranger du côté du formalisme structurel, certains auteurs ont revendiqué le fait que les grammaires permettaient de distinguer les histoires des non-histoires, de servir de modèle de compréhension d'histoire en enfin que ces grammaire étaient des modèles diégétique de la mémoire [Black et Wilensky, 1979]. Dans leur modèle, [Mandler et Johnson, 1977] supposent que les lecteurs possèdent une structure abstraite qui permet d'organiser les informations rencontrées lors de la lecture d'un récit. Cette structure serait une *grammaire* formelle. La grammaire possède des règles de réécriture, qui spécifient les éléments du récit et les relations entre eux. Les extraits de règles (parmi 18) ci-dessous :

- (1) Fable -> Story AND Moral
- (2) Story -> Setting AND Event Structure
  - | State\* (AND Event\*)
- (3) Setting -> | Event\*
- ...
- | CAUSE
- (5) Event\* -> Event ((|AND Event)\*) (AND State\*)
  - | THEN
- ...

expriment qu'une histoire commence toujours par une mise en place du décor suivie d'un ensemble événements et d'états. La figure 2.7 présente un exemple de conte racontant l'aventure d'un chien, décomposé en phrases numérotés. L'arbre de droite représente la décomposition du conte où les feuilles sont les phrases du conte et les autres nœuds des éléments non terminaux de la grammaire. Dans cet arbre, nous pouvons trouver deux grandes branches. D'abord le contexte (*setting*), représenté par les phrases 1, 2 et 3 (le chien

1. Un chien tenait un morceau de viande
2. et il l'emmenait chez lui.
3. Sur le chemin il y avait un pont qui enjambait une rivière
4. En traversant, il regarda l'eau
5. et il vit son reflet
6. en pensant qu'il voyait là un autre chien tenant un morceau de viande.
7. il se mit à désirer cette autre pièce de viande.
8. Il aboya donc vers le reflet
9. et son morceau de viande tomba de sa gueule.
10. et se perdit dans la rivière
11. sans jamais pouvoir être récupéré.

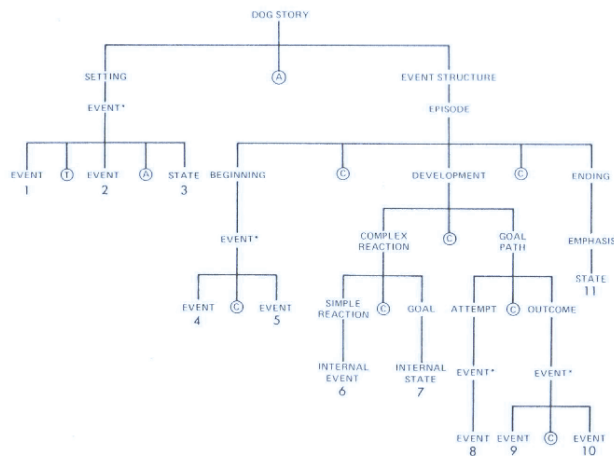


FIG. 2.7: Illustration de la grammaire de Mandler et Johnson [Mandler et Johnson, 1977]. À gauche, texte de *story dog*, à droite abstraction de *story dog* selon la grammaire de Mandler et Johnson. les lettres *A* *T* et *C* indiquent respectivement *And*, *Then* et *Cause*.

passant par le pont). Vient ensuite la structure d'événements (*event structure*), qui contient le corps de l'histoire, décomposable en un début (4 et 5, le chien regarde son reflet), un développement (6-10, le chien essaye de voler le morceau de viande à son reflet) et une fin (11, la disparition du morceau dans l'eau).

Les expériences en psychologie concernant les grammaires du récit ont permis de mesurer la familiarité des individus avec les récits et les paramètres entrant en jeu pour la tâche de compréhension [Mandler et Johnson, 1977]. Cependant d'autres études ont montré que les grammaires étaient insuffisantes pour prendre en compte la variété des connaissances nécessaires à la cohérence des récits [Black et Wilensky, 1979].

Durant la même période Schank et Abelson [Schank et Abelson, 1977] popularisèrent la notion de *script*. Un *script* est une séquence ordonnée d'actions prédéterminées ou typiques d'une situation bien connue [Schank et Abelson, 1977]. Les scripts sont censés être des représentations mentales qui permettent de prévoir, dans une situation donnée, la séquences d'actions susceptible de survenir. En compréhension de texte, les lecteurs activent le script adapté au type de texte auxquels ils sont confrontés. Ainsi le déchiffrement du contenu devient plus facile car les lecteurs s'attendent à ce qu'ils trouveront dans le texte en terme d'organisation et de présentation (type de phrases, vocabulaire, etc.). Un exemple de *script* est la représentation mentale d'un menu de restaurant ; le lecteur s'attend à trouver les noms des

plats, les ingrédients et le prix. Dans le cas d'un récit fictif, les lecteurs s'attendent à certains types de personnages (le protagoniste, l'antagoniste, des personnages drôles, etc.), une quête, un contexte fantastique où l'intrigue a un début, un milieu et une fin, tel que figuré dans la pyramide de Freytag [Freytag, 1900].

Étendant la notion de script, Herman développe le *storyworld*, un modèle mental relatif à l'univers diégétique et ce qui s'y passe. L'organisation du *storyworld* (composé de personnages et d'actions) permet au récepteur de faire des inférences à partir des situations et événements mentionnés [Herman, 2009b]. Ainsi, le contenu narratif est stocké dans une représentation mentale qui évolue et s'enrichit au fur et à mesure de la progression du récit. Le récit donne les repères nécessaires pour que le récepteur du récit soit capable de situer mentalement le récit dans un contexte plus large. C'est par cette construction du *storyworld* que le récit constituerait une ressource cognitive permettant diverses opérations cognitives, telles que l'acquisition d'un nouveau savoir [Campion, 2015] ou la résolution d'un problème.

Afin de structurer de manière souple les événements d'un récit et de prendre en compte les états mentaux des personnages, Wendy Lehnert a développé un modèle pour décrire la composition structurelle de récits, en visant une application informatique. D'après elle, lorsque les personnes sont confrontées à un récit, elles peuvent le résumer à partir de l'identification des patrons récurrents de situations clés [Lehnert, 1981]. L'identification des unités atomiques de l'intrigue, ou *plot units*, permet d'expliquer la relation causale existante entre les événements racontés. Ainsi, le récit s'analyse en identifiant les événements et les liens entre eux afin de reconnaître les patrons unitaires (*plot units*) et ainsi construire mentalement l'intrigue. D'une certaine manière, cette approche s'apparente aux scripts dans le sens où elle se fonde sur un ensemble finis de plots unitaires que le lecteur s'attend à trouver dans les récits. Néanmoins, elle en diffère fondamentalement en mettant l'accent sur les états affectifs et les réactions émotionnelles des protagonistes. Pour Lehnert, en effet, les unités de base d'un récit sont les *états affectifs* des personnages. Chaque événement ou expérience influe sur les personnages, en les affectant de manière positive ou négative, ou en modifiant leur intention. Les états affectifs sont de trois types : positif, négatif et *état mental* (réaction neutre, correspondant aux intentions ou buts du personnage) et ils sont mis en relation à travers des liens qui indiquent une certaine causalité ou conséquence, voire une simple co-occurrence. Ces liens sont au nombre de 4 : *actualisation* (*a*, un état affectif en termine un autre), *motivation* (*m*, un état affectif en provoque un autre), *terminaison* (*t*, un état affectif rend caduque un état précédent) et *équivalence* (*e*, états ou événements similaires).

Parmi l'ensemble possible de relations entre les états affectifs et les liens

qui les relie, Lehnert a identifié 15 types de *plot units* appelés *primitifs*, que l'on peut trouver dans plusieurs types de récits et qui peuvent constituer des unités plus complexes [Lehnert, 1981]. Ces unités de *plot* sont représentées Figure 2.8. Elles contiennent les états affectifs d'un personnage lorsqu'il est confronté à certains événements. Par exemple, le *plot unit* nommé « *Success* » sera reconnu si dans le texte il existe une situation mettant en relation l'intention de faire une action (une motivation M) et sa réalisation (état positif +) (« J'ai faim, j'ai donc croqué la pomme »). Lorsque le récit progresse, ces *plots units* s'intersectent, créant des *plots complexes* (Figure 7.3) et formant un graphe qui représente l'intrigue du récit.

MOTIVATION	SUCCESS	FAILURE
$\begin{matrix} M \\ M \end{matrix} \rightarrow m$	$\begin{matrix} M \\ + \end{matrix} \rightarrow a$	$\begin{matrix} M \\ - \end{matrix} \rightarrow a$
CHANGE OF MIND	LOSS	MIXED BLESSING
$\begin{matrix} M \\ M \end{matrix} \rightarrow t$	$\begin{matrix} + \\ - \end{matrix} \rightarrow t$	$\begin{matrix} + \\ - \end{matrix} \rightarrow e$
PERSEVERANCE	RESOLUTION	HIDDEN BLESSING
$\begin{matrix} M \\ M \end{matrix} \rightarrow e$	$\begin{matrix} - \\ + \end{matrix} \rightarrow t$	$\begin{matrix} - \\ + \end{matrix} \rightarrow e$
ENABLEMENT	NEGATIVE TRADE-OFF	COMPLEX POSITIVE EVENT
$\begin{matrix} + \\ M \end{matrix} \rightarrow m$	$\begin{matrix} - \\ - \end{matrix} \rightarrow t$	$\begin{matrix} + \\ + \end{matrix} \rightarrow e$
PROBLEM	POSITIVE TRADE-OFF	COMPLEX NEGATIVE EVENT
$\begin{matrix} - \\ M \end{matrix} \rightarrow m$	$\begin{matrix} + \\ + \end{matrix} \rightarrow t$	$\begin{matrix} - \\ - \end{matrix} \rightarrow e$

FIG. 2.8: Les 15 *plots units* primitifs décrits par Lehnert [Lehnert et Ringle, 2014])

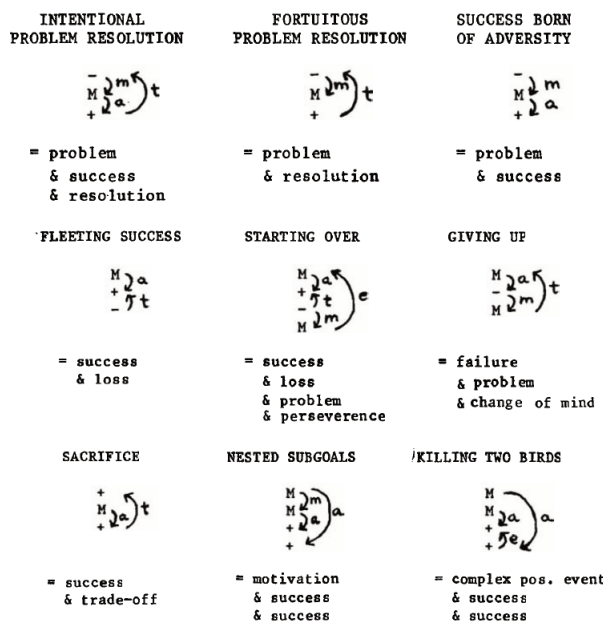


FIG. 2.9: Les *plots complexes* identifiés dans [Lehnert, 1981]

La Figure 2.10 présente un récit analysé dans [Lehnert, 1981] selon le point de vue de deux personnages. La voiture en panne et le souhait de la mettre en marche constituent le *plot unit* nommé *Problem* (encadré orange). L'appel à l'aide pour le régler permet l'interaction avec Paul (lignes noires). La réussite de l'intervention est un *Success* (encadré vert) pour les deux hommes. Finalement, le passage de l'état de la voiture de « en panne » à « en état de marche » constitue un dernier *plot unit* nommé *Resolution* (encadré violet).

(p1) John tentait de démarrer sa voiture, mais elle ne fonctionnait pas. (p2) Il demanda de l'aide à son voisin Paul. (p3) Paul régla le carburateur. (p4) John le remercia et pu prendre la voiture pour aller au travail.

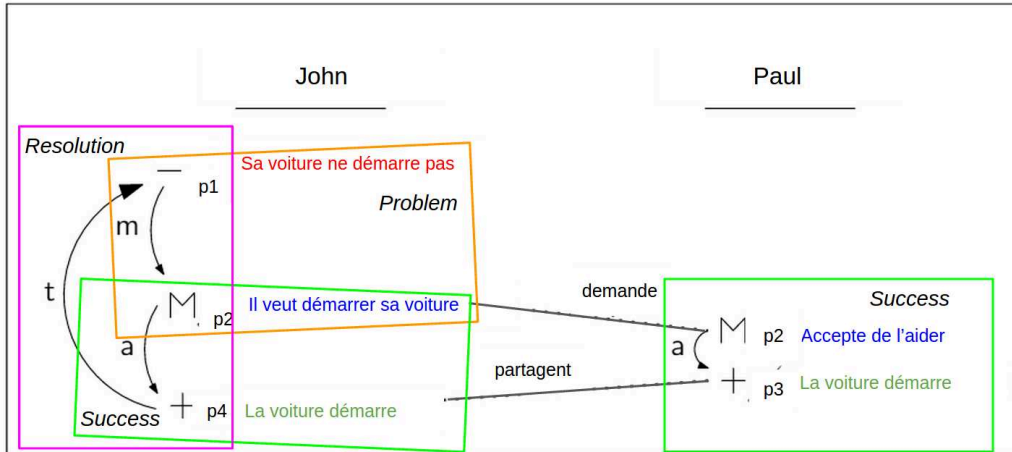


FIG. 2.10: Exemple d'analyse selon Lehnert [Lehnert, 1981]. En haut, récit du dépannage d'une voiture. En bas, analyse du récit selon les deux personnages.

L'approche de Lehnert, bien que représentant une formalisation encore vague a été utilisée dans certaines approches de génération automatique et constitue la base du modèle de *Story Intention Graph* (SIG) de David Elson [Elson, 2012].

### 2.3.2 La production de récits

Si de nombreux modèles ont été proposés pour analyser les récits, la production humaine du récit est également un domaine qui a engendré beaucoup de recherches mais reste ardu. [Kim, 2004] décrit de façon intuitive la démarche mentale derrière la conception d'un récit. Cette démarche est illustrée Figure 2.11. En premier lieu, l'auteur crée un récit contenant une séquence d'événements (*récit conçu*). Puis, il l'exprime d'une manière particulière à travers un *discours* sur un certain *media* (p.ex., livre ou un film). Le *media* servira de pont entre l'auteur et le lecteur. Le lecteur reconstruira le contenu du récit (*récit perçu*) dans son esprit à partir d'une interprétation du média qui contient le récit. Ainsi, le récit est un modèle créé dans l'esprit de l'auteur, lequel part du postulat que le lecteur le recevra tel qu'il l'a imaginé [Kim, 2004].

Quoique simplifié, ce modèle permet de situer le contexte de la thèse. En effet, pour générer un récit il convient d'identifier les décisions mises en

œuvre lors de la création, en découvrant les facteurs ou contraintes externes qui influencent ce processus.

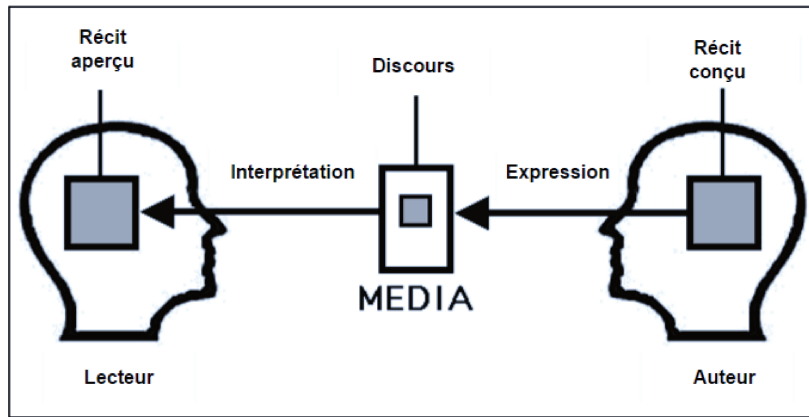


FIG. 2.11: Démarche mentale lors de la production d'un récit d'après [Kim, 2004]

Toutefois, la production d'un récit est un processus mental assez complexe pour lequel la recherche n'a pas atteint le même niveau d'approfondissement que dans le cas de la compréhension. Il n'y a aucune garantie que toutes les personnes aient la même façon de construire un texte. Des facteurs tels que la culture ou la langue d'origine peuvent influencer les processus mentaux pendant la conception d'un récit [Jasinskaja et Roßdeutscher, 2009]. Le manque de modèles sur ce sujet montre la difficulté d'établir ce qui se passe lors du processus cognitif de production d'un texte narratif. Néanmoins des travaux existent qui cherchent à mettre en évidence ces processus. La suite de cette section résume certains d'entre eux.

Flowers et Hayes ont proposé un modèle où la production de textes est vue en termes de stratégies et de processus qui sont activés par les écrivains lorsqu'ils construisent un texte [Flower et Hayes, 1981]. Ils se sont livrés à une expérience dans laquelle des professeurs et des étudiants devaient décrire à voix haute les stratégies et décisions qu'ils appliquent lorsqu'ils écrivent un texte. Les résultats ont montré une différence significative entre les personnes possédant une bonne maîtrise rédactionnelle (les experts) et celles qui ne l'ont pas (les non-experts). Ils ont observé que les écrivains avec plus d'expérience en termes de production écrite regardent l'acte de production comme un problème à résoudre, décomposable en trois grandes étapes : *planning* (génération et organisation d'idées à partir de ce qu'ils connaissent sur le texte à produire, tel que le but, le genre, le sujet, etc., en créant ainsi un schéma mental), *translation* (passer de la planification vers une première

version *brouillon*) et *révision* (modifications et corrections, si besoin, dans ce cas, l'écrivain peut revenir à l'étape de planification). Pendant toute la démarche, il existe un contrôle méta-cognitif qui aide l'écrivain à évaluer sa production durant la rédaction. Ce modèle est présenté Figure 2.12. Les deux chercheurs ont montré que l'écriture est un déploiement itératif de processus cognitifs pendant la tâche d'écriture. C'est l'interaction constante entre le contexte de production, les processus cognitifs et la mémoire à long terme de l'individu qui permet de rédiger efficacement [Flower et Hayes, 1981]. Cependant, il est difficile de savoir si ce modèle général s'applique au cas particulier du récit.

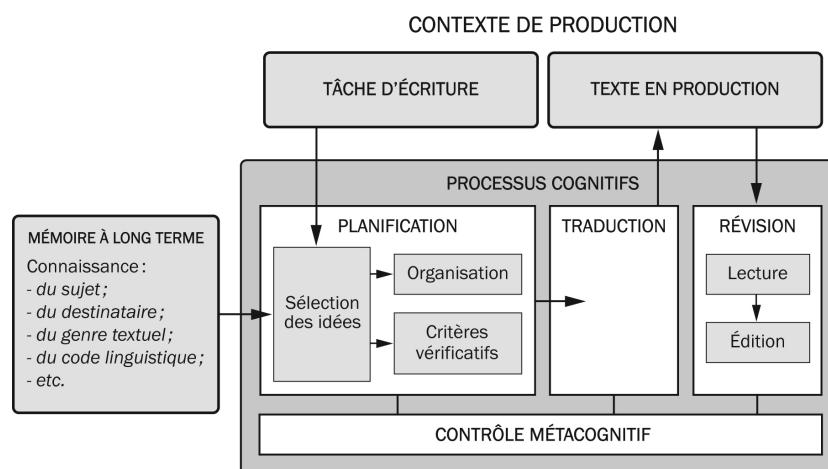


FIG. 2.12: Modèle d'écriture selon [Flower et Hayes, 1981](tiré de [Grégoire et Karsenti, 2013])

Le binôme « expert »/« non expert » a été repris par Bereiter et Scardamalia [Scardamalia et Bereiter, 1987] qui ont proposé une autre approche qui prend plus en compte les différences entre les deux types d'écrivains à travers leurs compétences lors de la production d'un texte. Cette différence provient de l'expérience que les individus ont acquise pendant leur formation scolaire. Ainsi, cette recherche voulait caractériser les différences entre ces deux types d'écrivain, à travers l'identification des stratégies mentales que chacun utilisait pendant la production. Ils ont proposé deux modèles associés à ces profils. Le premier, nommée "Savoir et raconter" (ou *Knowing Telling Strategy* en Anglais) est associé aux personnes n'ayant pas beaucoup d'expérience dans la rédaction de textes. Ce modèle montre qu'un « non-expert » s'appuie surtout sur la mémoire ; la personne prend comme matière source uniquement ses souvenirs et connaissances sur le sujet qui motive la production, sans investir du temps ou des ressources mentales en formatage



ou style. La deuxième stratégie est nommée « Transformation de connaissances » (*Knowledge Transforming strategy* en Anglais), elle est associée aux personnes les plus douées dans la rédaction de textes. Dans ce cas, le contenu du texte est obtenu après plusieurs traitements afin de l'ajuster aux buts communicatifs, en prenant plus de soin pour obtenir une sortie de meilleure qualité.

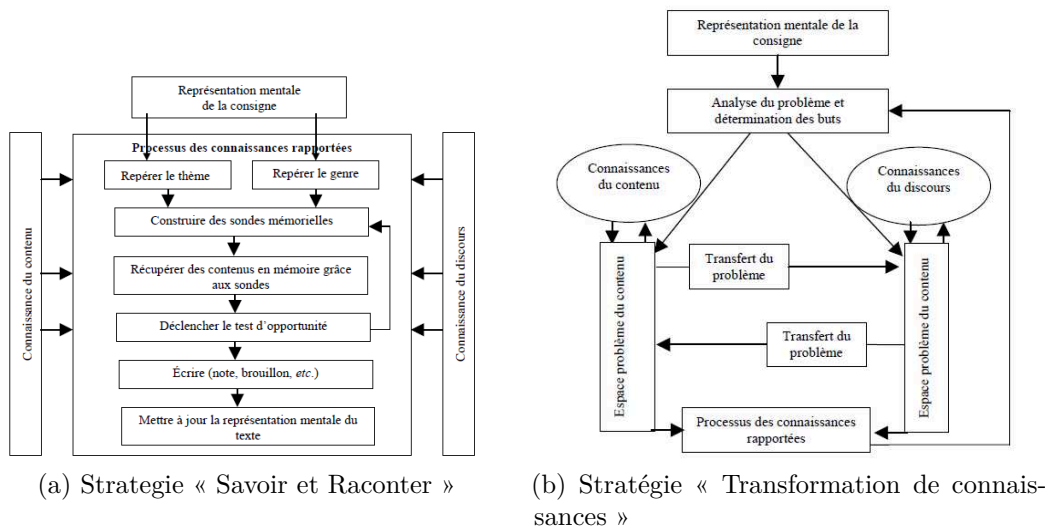


FIG. 2.13: Modèles proposés par [Bereiter et Scardamalia, 1987] (tiré de [Chanquoy et Alamargot, 2002])

Les travaux présentés proposent des modèles de production hypothétiques qui dépendent du degré d'expertise des rédacteurs. Ils observent une forte corrélation entre les compétences de rédaction et la qualité du résultat final. L'identification de facteurs externes tel que le type de texte et le public concerné sont pris en compte dans le modèle. Néanmoins, comme souligné précédemment, il n'est pas évident de transférer cette étude au cas du récit. Par exemple quelles stratégies sont sélectionnées pour raconter un événement d'une certaine manière ? Ainsi, ces études ne nous permettent pas de découvrir le mécanisme cognitif spécifique qui est mis en jeu lorsque quelqu'un élabore un récit.

Malgré la difficulté de la tâche, certaines recherches ont commencé à explorer la production du récit. Dans le champ de la créativité et du point de vue cognitif, Mike Sharples a proposé un modèle de création de textes narratifs fictifs, représenté Figure 2.14. Il signale que le processus de production/création est un cycle cognitif décomposable en deux phases qu'il appelle « engagement et réflexion » (*Engagement and Reflexion* en Anglais)

[Sharples, 1999]. Selon lui, un écrivain investit les ressources mentales qui lui permettent de projeter des concepts en un texte écrit. Dans un premier temps, l'auteur travaille sur le texte, sans se concentrer sur l'évolution du travail. Une fois fini, il peut revenir, réviser et corriger si besoin ce qu'il a produit. Le cycle commence avec une série de contraintes *internes* (par rapport au contenu, comme le type de texte, la thématique ou les connaissances à mettre en place pour créer le document) et *externes* (par rapport à l'auteur, comme les compétences propres de l'auteur vis à vis de ce type de texte) qui vont guider la création du texte. En considérant ces contraintes, l'auteur se focalise uniquement sur les idées à enchaîner afin de les mettre sur un média (étape *Engagement*). L'auteur entre ensuite dans la phase *Reflection* où il peut faire des révisions (*Reviewing*) et explorer des nouvelles idées (*Contemplation*), afin de les sélectionner pour les intégrer si besoin (*Planning*) et ainsi recommencer de nouveau le cycle, jusqu'à satisfaire les contraintes. Dans ce modèle, bien que l'étape de *Planning* spécifie une sélection de contenu, celle-ci n'est pas détaillée, montrant la difficulté d'abstraire le mécanisme mis en place et de détailler finement comment les facteurs (ou contraintes) internes ou externes influencent la création.

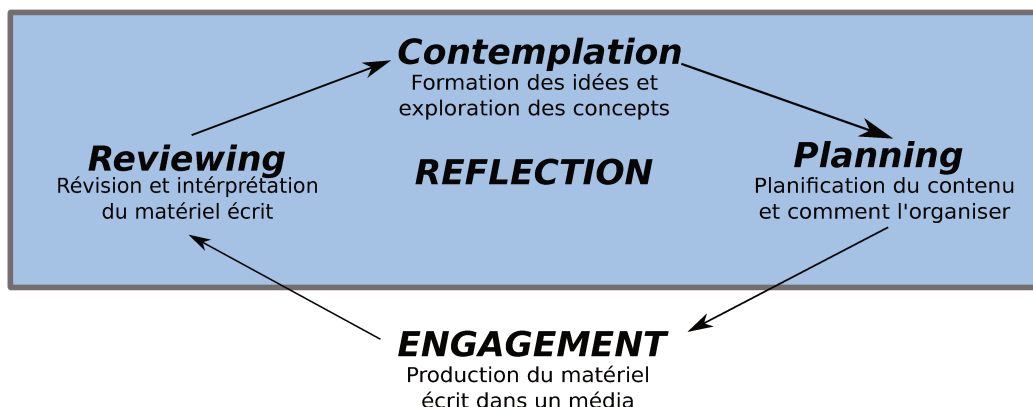


FIG. 2.14: Le cycle « Engagement and Reflection » [Sharples, 1999]

## 2.4 Bilan

Ce chapitre a fourni une introduction à la narratologie et aux notions essentielles concernant le récit. La narratologie étudie le récit en tant qu'objet littéraire et a fourni au fil des évolutions de ce domaine des définitions et structurations de plus en plus fine de ces différents éléments. En distinguant l'*acte de narration* de l'*objet* (le récit) résultant de cette narration, elle a permis de formaliser les notions intra- et extra-diététique qui entrent en jeu dans la réa-

lisation et la réception d'un récit (l'auteur, le narrateur, les personnages, les événements, l'intrigue, etc.), ainsi que les concepts qui permettent d'analyser les récits (fonctions, indice, ordre, rôle...) pour trouver des constantes (voir par exemple, Propp [Propp, 1970] et Barthes [Barthes, 1966]). Enfin, la narratologie s'est aussi intéressée au discours notamment à travers les travaux de Genette [Genette, 2007] qui mettent en lumière les stratégies d'écriture du récit utilisées par les auteurs. Ces travaux nous apprennent qu'il existe au moins deux mondes de représentation du récit : le monde de l'histoire (diégèse) et celui de la représentation de l'histoire (texte, film, discours oral) et qu'ils sont à distinguer de la notion de *fabula*. Finalement, et dans notre problématique de génération à partir de données objectives de capteurs, nous retiendrons :

- la *fabula* comme étant l'ensemble (le sur-ensemble) des événements organisés chronologiquement lié à l'histoire à transmettre ;
- l'*histoire* comme étant les événements/existants à transmettre organisés selon une certaine intrigue et certains points de vue ;
- le *récit* comme étant le texte transportant l'histoire.

La grande partie des concepts émanant de la narratologie ont été extraits d'études autour de récits fictifs, certaines théories sont d'ailleurs limitée à des genres littéraire (p.ex., Propp [Propp, 1970]). Si ces concepts sont utiles pour la formalisation et l'informatisation des récits, et nous verrons chapitre 3 que beaucoup de systèmes de génération s'inspirent des théories de la narratologie, la narratologie ne décrit pas précisément les processus de génération et de réception qui pourraient guider les choix discursif exposés. Par exemple, dans la section 2.2.3, nous avons vu que le discours altérait parfois la présentation d'un événement (soit en le décrivant en détail soit en le cachant complètement). Ce choix de l'auteur est en partie guidé par l'intrigue (pour mettre en évidence les points importants). Par contre, il n'est pas indiqué *dans quelle situation* l'auteur peut prendre cette liberté ni ce qui motive ce choix. En plus, il semble que l'utilisation de ces stratégies, ainsi que l'organisation du contenu dépend non seulement de l'intrigue, mais aussi du niveau d'expertise de l'auteur et de ses propres connaissances ([Flower et Hayes, 1981, Chatman, 1977]). Donc, le passage d'un niveau de description du récit à un autre, tel que montré dans la figure 2.5 ne se fait pas uniquement à partir des événements et existants du récit : d'autres paramètres entrent en compte, tel que le but communicatif ou le public visé par le récit.

Une autre approche de la narratologie s'intéresse au récit comme produit issu d'opérations mentales, que l'on appelle la narratologie cognitive (compréhension et production). Nous avons vu deux types de travaux, les travaux à base de grammaire où la construction et l'analyse du récit sont vus comme un

ensemble de structures hiérarchiques et déterministes, et les travaux abordant comment les personnes assimilent l'information et lui donnent un sens pendant la lecture. D'un côté, il existe une notion de contenu que les personnes, à partir de leur expérience vécue, s'attendraient à trouver (par exemple, les *storyworld* et les scripts), d'un autre, il existerait une organisation naturelle des récits auquel les lecteurs s'attendraient (grammaires de récit). Cette approche de la narratologie est plus récente et se heurte à une grande diversité de définitions et à la difficulté d'observer les opérations mentales à l'œuvre en production. Cependant, les résultats de ces recherches en cognition ont donné lieu à des modèles généraux qui essaient d'expliquer comment la pensée fonctionne quand une personne lit ou produit un texte, ainsi que certaines caractéristiques qui peuvent intervenir lors de la réalisation de ces tâches. Néanmoins, à notre connaissance, il n'existe pas une approche qui explique en détail la production d'un récit. Il n'en reste pas moins que cette discipline naissante donne quelques bases théoriques en termes de contenu et de processus mentaux qui peuvent être appliquées dans autres domaines, comme la narratologie informatique, ce que nous exposerons dans le chapitre 3.

Comme conclusion, afin d'avoir un support pour nos récits issus de données de capteur, nous avons exploré dans ce chapitre la théorie autour du récit en tant qu'objet et comme produit cognitif. Les modèles ainsi que les théories donnent les bases pour le développement de la Narratologie Informatique, discipline dans laquelle cette thèse s'inscrit. De nombreux systèmes et approches sont basés sur ces propositions pour générer leurs propres textes. Le chapitre suivant aborde l'état de l'art de ces systèmes de génération de récits, pour mettre en avant, d'un côté, comment les théories de la narratologie sont appliquées et, d'un autre côté, comment les contraintes et les choix de génération ont été techniquement abordés.

## Chapitre 3

# État de l'art : Génération automatique de récits

Le développement de l'Intelligence Artificielle et l'apport des théories sur le récit présentées au chapitre 2 ont permis l'émergence de la *Computational Narratology* [Ogata, 1998] qui concerne l'analyse et la génération automatique d'histoires et que nous nommerons *Narratologie Informatique* dans la suite de ce manuscrit. Dans ce chapitre, nous introduisons ce sous-domaine de l'Intelligence Artificielle, à travers un bref état de l'art de la génération de récits. Il est important de noter que les approches informatisées de la génération de récits ont toujours été appliquées en considérant un monde virtuel de l'histoire et non dans le cadre de données réelles de capteurs. Il est cependant essentiel de présenter ces systèmes afin d'étudier s'il est possible de s'en inspirer ou d'adapter ces approches dans le cadre de la présente thèse. Ce chapitre donne tout d'abord un aperçu de la Narratologie Informatique. Puis, différents systèmes de génération de récits qui ont marqué cette discipline sont présentés en fonction du paradigme de génération employé et en soulignant les approches théoriques sur lesquelles ils reposent. Enfin, l'évaluation de tels systèmes étant complexe, nous présentons différentes méthodes d'évaluation proposées dans la littérature. Nous terminons le chapitre avec un bilan.

### 3.1 La Narratologie informatique

La narratologie informatique est une discipline récente de l'Intelligence Artificielle qui vise à proposer des processus qui permettent de créer ou d'interpréter des narrations [Ogata, 1998, Ogata et Akimoto, 2016, Mani, 2012].

Dès les années 1980, des recherches ont été menées sur l'analyse et la génération automatique de textes narratifs fictifs [Meehan, 1977]. C'est cependant

à partir des années 1990 que cette recherche a commencé à se structurer sous le nom de *Computational Narratology* [Ogata, 1998]. Les théories de la narratologie qui ont fait émerger des structures formelles et logiques étaient adéquates pour une mise en œuvre informatisée. Ces théories ont inspiré la création d’algorithmes et de systèmes complets de production de textes narratifs [Mani, 2013].

D’après [Mani, 2013], la narratologie informatique comprend plusieurs axes de recherche :

- Axe humaniste. La narratologie informatique cherche à développer des outils et mécanismes pour modéliser et tester des théories narratives, en les appliquant sur une grande quantité de textes. Un exemple est *Story Workbench*<sup>1</sup>, un outil d’annotation semi-automatique des éléments narratifs dans les textes, particulièrement appliqué aux contes folkloriques [Finlayson, 2008]. Plus récemment, [Reagan *et al.*, 2016] se sont intéressés à la détection des principaux arcs émotionnels dans des collections de livres numériques (c’est à dire, le déroulement des émotions des personnages au fil de l’histoire) et ont trouvé 6 arcs noyaux grâce à des techniques de (*Data Mining*).
- Axe cognitif. La narratologie informatique englobe un ensemble de techniques informatiques qui visent à décrire des notions cognitives qui entrent dans la production et la compréhension des récits, ainsi que leur mise en place, afin de produire les effets narratifs étudiés dans la théorie du discours [Harrell, 2007]. Il peut s’agir par exemple de l’étude et la mise en œuvre de mécanismes pour produire du suspense chez le lecteur [O’Neill, 2013].
- Axe informatique. La narratologie informatique vise à formaliser les recherches conduites par la narratologie afin de construire des systèmes informatiques capables de produire des récits, en s’appuyant aussi sur les travaux du traitement automatique du langage naturel. Un exemple est la génération de litanies, basée sur la notion de *grammaires narratives* [Galofaro et Kubas, 2016].

La narratologie informatique appliquée au texte a des liens très forts avec le traitement automatique du langage naturel mais s’en distingue d’une part, par le fait qu’elle se focalise plutôt sur les traits des personnages, les points de vues, les émotions, les intrigues et les autres structures de haut niveau qui ont été peu abordées par le TALN, et d’autre part, par le fait que la narratologie informatique se focalise sur la création de systèmes informatiques complets pour créer, à froid ou en interaction, des récits. La difficulté de construire et d’évaluer ces systèmes pose de nouveaux défis au domaine.

---

<sup>1</sup><http://projects.csail.mit.edu/workbench/>

Ces défis ont attiré l'attention de nombreux chercheurs en informatique, créant ainsi une nouvelle communauté. Cet élan a fait naître plusieurs événements autour de la discipline, tel que *The International Conference on Computational Creativity* (<http://computationalcreativity.net> neuvième édition en 2018), *International Workshop on Computational Models of Narrative* ([narrative.csail.mit.edu](http://narrative.csail.mit.edu)), ainsi que *International Conference on Interactive Digital Storytelling*, (<http://icids.org/> onzième édition en 2018). Par ailleurs, les récents ouvrages *The Living Handbook of Narratology* et *Computational Modeling of Narrative*, ont popularisé ce domaine auprès des autres communautés informatiques.

## 3.2 Systèmes de génération de récits

L'analyse de l'état de l'art des systèmes de création automatique de récits fictifs a fait ressortir plusieurs stratégies de génération. Nous structurons donc cet état de l'art des systèmes selon la manière dont l'histoire est produite, étant donné qu'elle est l'élément structurel des récits. Les représentations informatisées de l'histoire ont pour but d'assurer la cohérence de la progression des événements à travers des liens causaux. L'un des objectifs principaux est de construire ces liens automatiquement. Bien qu'il existe une grande quantité de systèmes avec des contributions importantes, nous présenterons seulement les plus représentatifs ou reconnus par la communauté de la Narratologie Informatique. Nous proposons quatre types d'approche de construction de l'histoire : *par planification*, *par raisonnement par cas*, *par schéma et grammaire*, et *par interaction*.

### 3.2.1 Génération par planification

Une première approche de génération utilise la *planification*. Celle-ci consiste à élaborer une séquence d'actions permettant d'atteindre un but [Russell et Norvig, 2010]. Avec cette technique, les systèmes visent à fournir aux récits une structure formelle, garantissant que la succession des événements est cohérente avec le monde de l'histoire. Les liens narratifs sont inclus dans la base de connaissance sous forme de préconditions et effets qui vont orienter la recherche d'une chaîne d'actions transformant l'état initial vers l'état final. Afin d'assurer l'obtention d'une séquence complète le monde de l'histoire doit être entièrement déterministe et représentable. Ces systèmes fonctionnent donc dans un *monde fermé*, où tout ce qui peut être possible est connu en avance. La planification est une approche très utilisée pour mettre en œuvre des théories narratives qui supposent que l'intrigue est dépendante des rôles et des

actions de personnages stéréotypés tel que dans l'approche de Propp [Propp, 1970] (section 2.1.2) ou de Brémond [Bremond, 1973] (section 2.1.3).

### 3.2.1.1 *TALE-SPIN* (1977)

Il s'agit d'un des premiers travaux de la discipline. *TALE-SPIN* [Meehan, 1977] produit des récits textuels à propos d'un monde prédéfini dans une base de connaissances. Meehan implante son système en suivant le modèle de Schank et Tesler *Conceptual dependency Theory* [Schank et Tesler, 1969] pour représenter les connaissances de manière formelle et permettre d'effectuer des raisonnements. La partie fonctionnelle de *TALE-SPIN* est constituée d'un ensemble de but préconçus (par exemple, *X veut avoir Y*, *X veut que Y connaisse Z*, etc.) pour lequel un ensemble de plans peuvent être appliqués, chaque plan pouvant faire appel à un autre plan. La sortie du système est construite par planification à partir d'un état initial donné. La sortie est la trace de toutes les actions permettant de résoudre le but qui peut être vu comme une *fabula* (voir section 2.3.1).

La figure 3.1 montre une sortie du système. Le personnage est l'ours John Bear avec le problème de base : il a faim et ne veut plus être affamé. Le monde dans lequel le personnage évolue contient des myrtilles qui peuvent calmer la faim de l'ours. Ainsi, le but principal est décomposé en sous-buts, par exemple, John doit d'abord arriver à l'endroit où les fruits se trouvent puis s'en saisir et les ingérer. Il permet aussi l'intervention d'autres personnages qui peuvent aider à atteindre le but. Par exemple, si John n'avait pas su où se trouvaient les fruits mais que la base de connaissances indiquait qu'un autre personnage le savait, alors celui-ci aurait pu être intégré au récit. Ainsi, le système base l'intrigue du récit sur un ensemble de buts à atteindre. Les événements, personnages et accessoires sont choisis pour répondre à ces buts, ce qui nécessite une base de connaissances exhaustives du monde de l'histoire.

**John Bear is somewhat hungry. John Bear wants to get some berries. John Bear wants to get near the blueberries. John Bear walks from a cave entrance to the bush by going through a pass through a valley through a meadow. John Bear takes the blueberries. John Bear eats the blueberries. The blueberries are gone. John Bear is not very hungry.**

FIG. 3.1: Récit généré par *TALE-SPIN*



### 3.2.1.2 *Universe* (1985)

Le système *Universe* [Lebowitz, 1984] a été conçu pour aider les scénaristes des séries télévisées. Tout comme TALE-SPIN, le système produit un plan d'histoire à partir de fragments prédéfinis. *Universe* contient un planificateur qui sélectionne les fragments à inclure dans le plan à partir d'un but *autoriel*, c'est à dire, d'un objectif thématique de haut niveau tel que « avoir une liaison ». Pour garantir la cohérence causale, le système se sert des préconditions incluses dans les fragments. La différence avec TALE-SPIN est que l'histoire n'est pas censée avoir de fin (comme pour les séries télévisées). Par ailleurs, le genre de l'histoire n'appartient pas au domaine des contes mais à celui des mélo-drames, ce qui représente un autre niveau de complexité. Enfin, l'approche ne s'appuie pas sur une théorie cognitive mais sur l'analyse des scripts de mélodrame. Ainsi les situations sont stéréotypées et s'enchaînent de manière à respecter les canons du genre. Les connaissances pour produire les scripts sont gardés en mémoire sous forme d'une librairie de fragments d'intrigue construits manuellement. Un fort accent est mis sur la création de personnages. En effet un ensemble de personnages existe au moment de la génération qui ont chacun leurs caractéristiques et leur histoire. Les personnages sont initialisés avec des stéréotypes (avocat, gangster, etc.) qui ont chacun des traits par défauts (l'avocat a une intelligence normale mais peut avoir une faible moralité). C'est l'interconnexion entre les plans et les personnages qui va faire avancer l'intrigue tout en complétant les caractéristiques des personnages.

La figure 3.2 montre le développement d'un plan selon un but donné (« Ruiner la relation du Joshua et Fran », *\*(tell '(((churn Joshua Fran)))*) mis à la première ligne). Pour l'accomplir, le système montre une liste de fragments d'intrigue : « Dispute entre amants » (*Lovers-fight Him/Fran Her/Joshua*) ou « Être enceinte lors d'une relation » (*Pregnant-affair woman/Fran Husband/Joshua*). Ici, le fragment « Dispute entre amants » est choisi. Ce fragment demande en même temps la résolution du but « Déclencher dispute » (*Do fight Josua Fran*). De nouveau, plusieurs fragments sont appelés en tant qu'options pour continuer, (se battre à cause des enfants, de l'argent, etc. ). Le fragment « Dispute pour l'argent » (*Personal-fight A/Joshua B/Fran Subject/Money*) est choisi. Le fragment « Dispute entre amants » est fini, mais le but de l'auteur n'est pas encore résolu. Pour ce faire, un autre sous-but est requis pour l'atteindre : « Finir relation » (*Dump-Lover Joshua Fran*) qui va appeler le fragment « Se quitter » (*Break up*) pour le résoudre. Ce processus se réalise de manière itérative jusqu'à atteindre le but de l'auteur. La cohérence est garantie, mais l'histoire risque de devenir répétitive s'il n'y a pas assez de fragments. Enfin, *Universe* ne s'intéresse qu'au script

et ne génère donc pas le récit.

```
*(tell '(((churn JOSHUA FRAN)))

working on goal -- CHURN JOSHUA FRAN
Several possible plans with equal goal effect
LOVERS-FIGHT HIM/FRAN HER/JOSHUA
LOVERS-FIGHT HIM/JOSHUA HER/FRAN
JOB-PROBLEM P1/FRAN P2/JOSHUA JOB/BUREAUCRAT
JOB-PROBLEM P1/JOSHUA P2/FRAN JOB/SLEAZY-LAWYER
PREGNANT-AFFAIR WOMAN/FRAN HUSBAND/JOSHUA
ACCIDENT-BREAKUP P1/FRAN P2/JOSHUA
STEAL-CHILD HUSBAND/JOSHUA HER/FRAN
COLLEAGUE-AFFAIR WIFE/FRAN HUSBAND/JOSHUA
AVALANCHE-ACCIDENT HIM/FRAN HER/JOSHUA
AVALANCHE-ACCIDENT HIM/JOSHUA HER/FRAN
-- picking one
-- using plan LOVERS-FIGHT HIM/FRAN HER/JOSHUA

working on goal -- DO-FIGHT JOSHUA FRAN
Several possible plans with equal goal effect
PERSONAL-FIGHT A/JOSHUA B/FRAN SUBJECT/IN-LAWS
PERSONAL-FIGHT A/JOSHUA B/FRAN SUBJECT/MONEY
PERSONAL-FIGHT A/JOSHUA B/FRAN SUBJECT/SECRETS
PERSONAL-FIGHT A/JOSHUA B/FRAN SUBJECT/FLIRTING
PERSONAL-FIGHT A/JOSHUA B/FRAN SUBJECT/KIDS
-- picking one
-- using plan PERSONAL-FIGHT A/JOSHUA B/FRAN SUBJECT/MONEY

>>> JOSHUA and FRAN fight about: MONEY
working on goal -- DUMP-LOVER JOSHUA FRAN
-- using plan BREAK-UP DUMPER/JOSHUA DUMPED/FRAN

>>> JOSHUA tells FRAN he doesn't love her
```

FIG. 3.2: Plan généré par *Universe* [Lebowitz, 1984]

### 3.2.1.3 *StoryBook* (2002)

*StoryBook* [Callaway et Lester, 2002] est un système de génération qui met en place un ensemble de mécanismes permettant la génération de récits de haute qualité linguistique. *StoryBook* est capable de produire des récits longs et variés autour d'une seule histoire (*Le Petit Chaperon rouge*). L'architecture présentée figure 3.3) prend en entrée un ensemble de spécifications et produit un texte complètement formé [Callaway et Lester, 2002]. *StoryBook* prend comme données d'entrée des spécifications, comme les personnages et les événements à mettre dans le récit. En fait, chaque histoire demande un grand nombre de connaissances à saisir dans la *story ontology*. La fabula consiste en un grand nombre d'instances de personnages, d'événements et d'objets très détaillés qui sont également insérés dans la base de connaissances. Un exemple d'opérateurs produisant une fabula est donné ci-dessous :

```
(NewNarrative MyLRRH-Narrative001 Narrator001)
(AddActor Woodman001 Woodman Person Male)
```

```

(AddActor Wife001 Wife Person Female)
(AddLocation Cottage001 Cottage)
(AddLocation Forest001 Forest)
(AddActor Little-Red-Riding-Hood001 Little-Red-Riding-Hood
Person Female "Little Red Riding Hood")
...

```

L'intrigue (*narrative stream*) de l'histoire est produite par un planificateur (*Narrative Planner*), qui représente un plan sous la forme d'un automate dont un chemin atteignant l'état final est considéré comme le plan narratif. Un *narrative stream* est présenté figure 3.4. D'après Callaway et Lester [Callaway et Lester, 2002] l'histoire n'est ni plus ni moins que des éléments extraits de la fabula et organisés par ordre chronologique. Une fois le plan obtenu, il sera traité par les autres modules qui vont mettre en place des opérations pour le passage en texte naturel (*Narrative organizer*, en charge par exemple du choix lexical, *Sentence planner*, pour la création des structures abstraites des phrases). Enfin, le plan est transmis au module de révision (enrichissement des phrases, fusion de phrases courtes), puis la structure linguistique profonde qui en résulte est transformée en texte en utilisant le générateur FUF/SURGE [Elhadad et Robin, 1996].

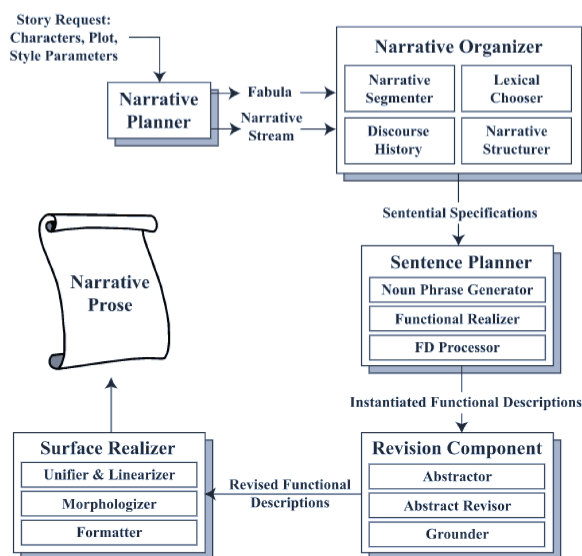


FIG. 3.3: Architecture du système *StoryBook* [Callaway et Lester, 2002]

Dans *StoryBook*, les efforts ont été portés sur le traitement linguistique et le rendu narratif du récit plutôt que sur l'intrigue (de fait, *StoryBook* ne génère qu'une seule histoire, celle du Petit Chaperon Rouge). Les auteurs n'intègrent pas de théorie narrative, arguant du fait que les théories

de l'époque étaient impropres à l'informatisation. Ils se réfèrent cependant à Propp [Propp, 1970] et Bal [Bal, 2009] pour souligner l'importance de distinguer les notions du *fabula*, de *sujet* et d'*intrigue*. De plus, la partie narration utilise des concepts mis en avant par Genette [Genette, 2007]. Par exemple, *StoryBook* considère aussi le narrateur comme paramètre dans le *narrative stream*, aspect important d'après Genette (section 2.2.3) et que nous pouvons observer dans la figure 3.4, deuxième ligne, «*narrator-mode narrator001 third-person disembodied*». Ce paramètre de *StoryBook* correspond à la focalisation zéro, en indiquant qu'il n'est pas prise la perspective d'un personnage en particulier. Toutefois, les aspects communicatifs (but, lecteur) ne sont pas abordés. Par ailleurs, comme les auteurs partent du principe que le plan ne présente pas d'inconsistances, il n'existe pas non plus un mécanisme de vérification du plan de l'histoire.

```
((narration-mode historical mixed complex ascii narrated english)
(narrator-mode narrator001 third-person disembodied)
(new-scene scene001)
(new-actor woodman001)
(new-actor mother001)
(new-actor little-red-riding-hood001)
;;; "Once upon a time there was a woodman and his wife."
(actor-property exist-being woodman001)
(refinement and-along-with woodman001 wife001)
(refinement belonging-to wife001 woodman001)
(specification exist-being process-step-type once-upon-a-time)
;;; "The woodman and his wife lived in a pretty cottage."
(prop-relationship living-in woodman001 cottage001)
(refinement and-along-with woodman001 wife001)
(refinement belonging-to wife001 woodman001)
(detail cottage001 pretty-appearance)
```

FIG. 3.4: Extrait du *narrative stream* [Callaway et Lester, 2002]

Les auteurs s'étant concentrés sur la qualité linguistique, ils ont pu faire une évaluation sur la qualité du résultat. 20 étudiants ont été recrutés pour évaluer différents aspects des contes générés. 9 mesures dont le style, la logique de l'histoire, la diction, la lisibilité etc. étaient évalués sur une échelle ordinaire (A à F). 10 textes ont été générés à partir de 2 *fabula* et différents paramètres de narration. Des extraits de ces textes sont montrés figure 3.5. Les résultats de l'évaluation ont montré une appréciation favorable, surtout au niveau de la lisibilité, grâce à une intégration adéquate de pronoms et à la fusion de phrases (agrégation). Par contre, tout comme les systèmes précédents, il s'agit d'un système complexe à mettre en œuvre et coûteux en termes de ressources. L'originalité de celui-ci est de mettre l'accent sur l'aspect linguistique, alors que les éléments de l'histoire restent les mêmes.

<p>Little Red Riding Hood had not gone far. Little Red Riding Hood met the wolf.  “Hello,” said the wolf. The wolf was the cunning looking creature. The wolf said, “Where is Little Red Riding Hood going?”  “Little Red Riding Hood is going to Little Red Riding Hood’s grandmother’s house,” said Little Red Riding Hood.</p> <p style="text-align: center;"><b>Pronoms non traités  (Module «Narrative Organizer»)</b></p>	<p>She had not gone far. She met a wolf.  “Hello,” greeted the wolf. The wolf was a cunning looking creature. He asked, “Where are you going?”  “I am going to my grandmother’s house,” she replied.</p> <p style="text-align: center;"><b>Phrases simples et enchaînées  sans connecteurs  (Module «Revision component»)</b></p>
<p>Little Red Riding Hood had not gone far when Little Red Riding Hood met the wolf.  “Hello,” greeted the wolf, who was the cunning looking creature. The wolf asked, “Where is Little Red Riding Hood going?”  “Little Red Riding Hood is going to Little Red Riding Hood’s grandmother’s house,” replied Little Red Riding Hood.</p> <p style="text-align: center;"><b>Sans traitement</b></p>	<p>She had not gone far when she met a wolf.  “Hello,” greeted the wolf, who was a cunning looking creature. He asked, “Where are you going?”  “I am going to my grandmother’s house,” she replied.</p> <p style="text-align: center;"><b>Résultat obtenu lors d'appliquer  l'ensemble de modules</b></p>

FIG. 3.5: Exemples générés avec *StoryBook* pour l'évaluation [Callaway et Lester, 2002]

### 3.2.1.4 *Virtual StoryTeller* (2003)

Il s'agit d'un *framework* multi-agent pour faire émerger des contes à travers la simulation d'agents semi-autonomes et de leurs interactions [Theune *et al.*, 2004]. Ces agents utilisent des plans pour atteindre leur but personnel. En réaction à des événements, des objets ou d'autres agents, ils peuvent produire des émotions. Les auteurs ne se réclament pas d'une théorie narrative classique mais de l'approche *emergent narrative* [Aylett, 1999] qui tranche avec les scripts et plans prédéfinis afin de laisser plus de liberté dans l'organisation du plan de l'histoire.

La génération est décomposable en deux niveaux 3.6. Le premier correspond au niveau du récit (*Story level*). Dans ce niveau, les personnages créent la *fabula* en réalisant des actions dans un monde virtuel. Ce contexte nécessiterait de définir exhaustivement le monde virtuel, mais ce système est suffisamment souple pour ajouter lui-même des informations manquantes nécessaires à la construction du récit par l'intermédiaire d'opérateurs d'inclusion. Cette insertion à la volée d'éléments manquants s'inspire de l'improvisation théâtrale<sup>2</sup>. La *fabula* naît des buts des agents et de leur interactions et se représente par la chaîne causale des événements expérimentés par les

<sup>2</sup>Dans le théâtre d'improvisation, les acteurs créent une scène et une intrigue simplement à partir d'une thématique délivrée sur l'instant et sans préparation.

agents/personnages. Ces agents sont guidés par un autre agent dit *Director*, dont le but est de garantir la création d’une intrigue cohérente (par exemple, ajouter des informations si besoin).

Une fois la *fabula* réalisée, le narrateur associe les éléments de la fabula à une réalisation textuelle en néerlandais. Le *Narrator* est un agent intéressant car il contient les trois modules classiques de la Génération Automatique de textes (*Planification du document* où l’intrigue est structurée, *microplanification* et *réalisation de surface*). Finalement, l’histoire est narrée par l’agent *Presenter*, qui est un agent incarné qui joue le rôle de narrateur à travers la synthèse vocale du conte généré.

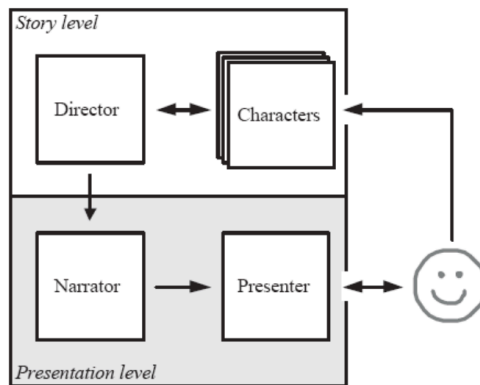


FIG. 3.6: Architecture du système *Virtual StoryTeller* [Theune et al., 2004]

Un exemple de *fabula* du *Virtual StoryTeller* est présenté dans la figure 3.7. Elle représente l’histoire d’un pirate qui veut devenir riche. Ceci motive le but de partir à la recherche d’un trésor. Les événements qui ont mené à la réalisation du but sont représentés en bas de la figure. Chaque événement est lié par un lien rhétorique issu de la *Rhetorical structure theory (RST)* [Mann et Thompson, 1988]. La *fabula* est donc une structure d’événements organisés en discours, ce qui facilite le travail de la génération de texte qui suit. *Virtual StoryTeller* montre qu’il est possible de créer des contes émergents sans nécessiter de description exhaustive du monde virtuel de l’histoire. Il faut cependant souligner que le traitement du plan de l’histoire est très orienté vers la production textuelle et se contente d’utiliser la *fabula*.

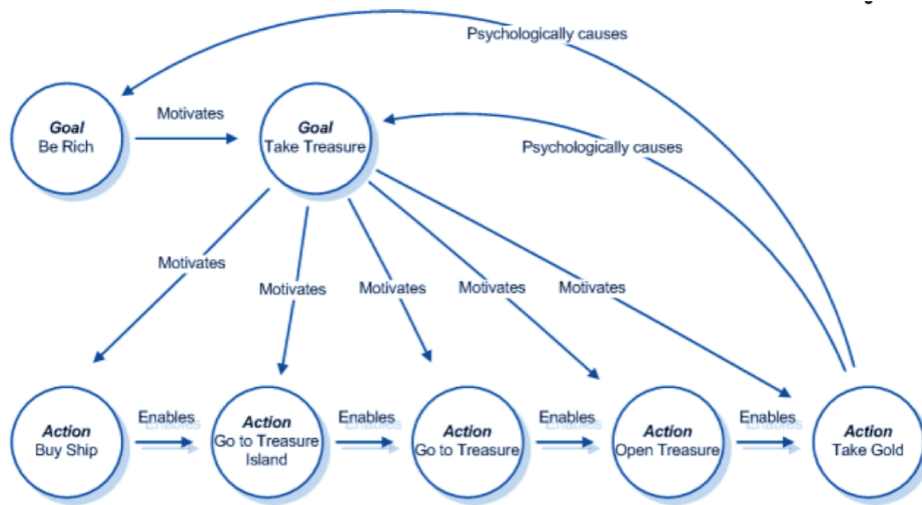


FIG. 3.7: Fabula générée par *Virtual StoryTeller* [Theune *et al.*, 2004]

### 3.2.1.5 *Fabulist* (2010)

*Fabulist* [Riedl et Young, 2010] est un système de génération automatique qui utilise des liens causaux pour créer des récits avec plusieurs personnages. Les auteurs approchent la construction du conte par planification car un récit peut être vu comme une série d'actions qu'un personnage réalise pour atteindre un état final à partir d'un état initial. Son architecture présente trois étapes : la génération de la *fabula* (ensemble d'actions possibles), la *génération du discours* (plan issu de la *fabula*) et la représentation finale sous forme textuelle, (génération par *template*). L'aspect le plus remarquable est la création de la *fabula*. Elle émerge des actions des personnages en fonction de leur intention. C'est-à-dire que la création du plan prend non seulement en compte les différents buts des personnage (qui entraîne une modification de l'état du monde) mais également les intentions de ceux-ci lors d'événements (qui guident les choix des personnages). Pour cette planification les auteurs ont proposé l'algorithme *Intent-driven Partial Order Causal Link* (IPOCL) dérivé de l'algorithme *Partial Order Planning*. IPOCL se base sur un ensemble d'actions prédéfinies pour modifier l'état du monde. Par exemple, l'action ci-dessous, indique que l'apparition d'un monstre effrayant devant un personne implique que ce personnage veut la mort du monstre.

```

Action : appear-threatening (?monster, ?char, ?place)
actors : ?monster
happening : t
constraints : monster(?monster), character(?char), place(?place)
precondition : at(?monster, ?place), at(?char, ?place), scary(?monster)
               , ?monster ?char
effect : intends(?char, ¬alive(?monster))
  
```

Cette prise en compte des intentions permet de motiver les actions des personnages. Elle permet aussi de rendre plus explicite les histoires et sous-histoire dans un même récit. Ceci est observable dans la fabula illustrée figure 3.8. Le but de l'histoire est que Jasmine épouse le Roi. Pour cela un autre personnage intervient : Aladin. Chaque couleur correspond au *thread* narratif d'un personnage central. Ainsi l'histoire concernant le roi et son objectif d'épouser Jasmine est en rouge. En orange on trouve la sous-histoire d'Aladin et son périple pour trouver la lampe du Génie. Le seul but d'Aladin est de ramener la lampe à Jafar et non que Jafar épouse Jasmine. Cette fabula forme ainsi différents flux narratifs qui permettent d'exposer une histoire selon différents points de vue.

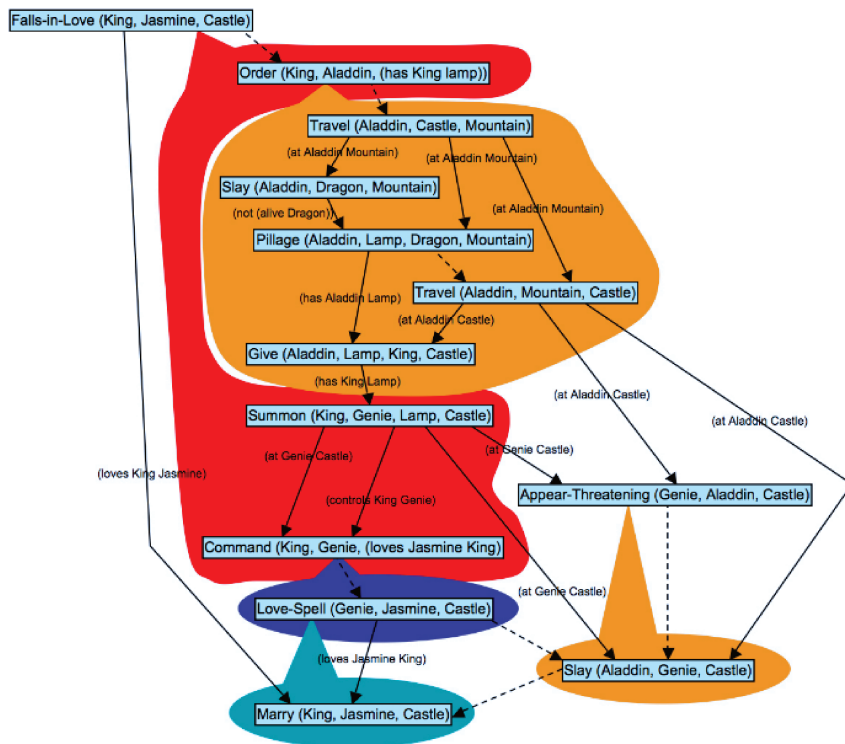


FIG. 3.8: Illustration de la fabula obtenue avec l'algorithme IPOCL [Riedl et Young, 2010]



There is a woman named Jasmine. There is a king named Jafar. This is a story about how King Jafar becomes married to Jasmine. There is a magic genie. This is also a story about how the genie dies.

There is a magic lamp. There is a dragon. The dragon has the magic lamp. The genie is confined within the magic lamp.

King Jafar is not married. Jasmine is very beautiful. King Jafar sees Jasmine and instantly falls in love with her. King Jafar wants to marry Jasmine. There is a brave knight named Aladdin. Aladdin is loyal to the death to King Jafar. King Jafar orders Aladdin to get the magic lamp for him. Aladdin wants King Jafar to have the magic lamp. Aladdin travels from the castle to the mountains. Aladdin slays the dragon. The dragon is dead. Aladdin takes the magic lamp from the dead body of the dragon. Aladdin travels from the mountains to the castle. Aladdin hands the magic lamp to King Jafar. The genie is in the magic lamp. King Jafar rubs the magic lamp and summons the genie out of it. The genie is not confined within the magic lamp. King Jafar controls the genie with the magic lamp. King Jafar uses the magic lamp to command the genie to make Jasmine love him. The genie wants Jasmine to be in love with King Jafar. The genie casts a spell on Jasmine making her fall in love with King Jafar. Jasmine is madly in love with King Jafar. Jasmine wants to marry King Jafar. The genie has a frightening appearance. The genie appears threatening to Aladdin. Aladdin wants the genie to die. Aladdin slays the genie. King Jafar and Jasmine wed in an extravagant ceremony.

The genie is dead. King Jafar and Jasmine are married. The end.

FIG. 3.9: Récit généré à partir de la fabula obtenue avec l’algorithme IPOCL [Riedl et Young, 2010]

La transformation du conte obtenu à partir de la fabula est présentée figure 3.9. Bien que le texte ne soit pas d’une grande qualité, il fait bien ressortir les intentions des personnages qui permettent d’un côté de garantir une cohérence du récit et de l’autre de justifier les actions des personnages. De plus, la prise en compte des intentions de différents personnages permet d’avoir différentes focalisations et ainsi de raconter une même histoire différemment.

*Fabulist* est donc un générateur de récits se focalisant sur la génération de plusieurs flux narratifs. L’approche adoptée nécessite une connaissance exhaustive de l’univers virtuel sur lequel il repose.

### 3.2.2 Génération par raisonnement à partir de cas

La génération de récits a également été problématisée comme un raisonnement à partir de cas (RàPC). Cette approche s’appuie sur l’hypothèse que les individus en face d’une nouvelle situation se servent des expériences vécues et les utilisent par analogie pour affronter cette nouvelle situation. Ce type de raisonnement trouve ses racines dans les travaux de Schank et Abelson [Schank et Abelson, 1977] sur les scripts ainsi que sur celui des mémoires dynamiques [Schank, 1982]. Il est généralement constitué d’une base de cas (mémoire) qui décrivent des situations et leurs solutions. Lorsqu’un nouveau cas arrive un moteur de recherche trouve un ou des cas similaires et propose une solution (qui peut être construite à partir de plusieurs cas). Dans le cas

de la narratologie informatique, le raisonnement par cas est une approche qui cherche à imiter les opérations cognitives mises en œuvre pour produire de nouveaux récits à partir de fragments de récits stockés en mémoire. La réutilisation du matériel est une stratégie connue dans la narratologie informatique (nous pouvons la trouver dans *Universe*). Dans la suite, nous allons montrer les principaux systèmes qui se servent du RàPC pour produire des récits.

### 3.2.2.1 *Minstrel* (1993)

*Minstrel* [Turner, 1993] génère de longs récits sur le roi Arthur. Le système s'intéresse à créer un récit avec une morale, en suivant un but thématique tout comme *Brutus* et *Universe*. L'objectif de Turner était de modéliser le processus d'un auteur afin de le reproduire sur une machine. Il fut grandement influencé par Schank et sa notion de mémoire dynamique [Schank, 1982] ainsi que par les *scripts* abordés dans la section 2.3.2 d'où son choix de raisonnement à partir de cas.

*Minstrel* génère un récit à partir d'un thème. S'il existe un fragment de récit correspondant au thème celui-ci peut être sélectionné. Ces fragments impliquent des descriptions de personnages et de scènes. Le processus de création est géré par un agenda. Le but de "créer un récit" est partitionné en sous-but : sélectionner un thème, illustrer un thème, appliquer des buts dramatiques, etc. Chaque sous-but est inséré dans une file de priorité et le prochain but à satisfaire est sélectionné et résolu jusqu'à ce que la file de buts soit vide. Ce processus peut ajouter de nouvelle scène à l'histoire et ajouter de nouveaux buts d'auteur à la file.

Le raisonnement à partir de cas est mis en œuvre à travers des TRAMs (*Transform-Recall-Adapt Methods*) qui peuvent modifier un fragment si aucun fragment correspondant au but n'est trouvé. Par exemple, si le thème est "la mort du chevalier", et si le seul cas disponible est le fragment d'"un chevalier combattant et tuant un troll, combat au cours duquel le chevalier est blessé", alors aucun fragment ne peut être appliqué. Dans cette situation, *Minstrel* peut relaxer une contrainte sur le thème, par exemple remplacer la mort par un accident ou une blessure et ré-appliquer la recherche. Dans ce cas le fragment de la base sera trouvé. Puis le système pourra essayer de transformer le fragment en réinsérant la contrainte relaxée afin de faire en sorte que le chevalier meure dans le combat. Ce nouveau fragment sera inséré dans la base et pourra servir de cas.

Un exemple de texte généré est présenté figure 3.10. Il raconte l'histoire de Jennifer qui se transforme en dragon pour se venger d'une autre dame en lui faisant peur. Jennifer est tuée par son amant, le chevalier, qui voulait

impressionner le roi, sans savoir de qui il s'agissait. La morale *Deception is a weapon difficult to aim* met en évidence que, lorsque l'on met en œuvre une supercherie, il est difficile d'en garder le contrôle.

#### The Vengeful Princess

Once upon a time there was a lady of the court named Jennifer. Jennifer loved a knight named Grunfeld. Grunfeld loved Jennifer. Jennifer wanted revenge on a lady of the court named Darlene because she had the berries which she picked in the woods and Jennifer wanted to have the berries. Jennifer wanted to scare Darlene. Jennifer wanted a dragon to move towards Darlene so that Darlene believed it would eat her. Jennifer wanted to appear to be a dragon so that a dragon would move towards Darlene. Jennifer drank a magic potion. Jennifer transformed into a dragon. A dragon moved towards Darlene. A dragon was near Darlene.

Grunfeld wanted to impress the king. Grunfeld wanted to move towards the woods so that he could fight a dragon. Grunfeld moved towards the woods. Grunfeld was near the woods. Grunfeld fought a dragon. The dragon died. The dragon was Jennifer. Jennifer wanted to live. Jennifer tried to drink a magic potion but failed. Grunfeld was filled with grief.

Moral: Deception is a weapon difficult to aim.

FIG. 3.10: Récit généré par *Minstrel*.

Tout comme les systèmes précédents, *Minstrel* est un générateur à base de connaissances qui, s'il ne nécessite pas une description complète du monde de l'histoire, doit inclure les opérateurs de transformation des fragments. Ces opérateurs ne sont ni triviaux ni indépendants de l'histoire produite. En effet, si "se tuer avec une épée" est une transformation crédible de "se blesser avec une épée" il n'en va pas de même du fragment "se blesser avec une agrafe". Éviter ce type de transformation nécessite un ensemble bien trop important de connaissances du monde.

#### 3.2.2.2 *Mexica* (1999)

Tout comme *Minstrel*, *Mexica* cherche à reproduire le modèle cognitif de la création de récits. Ce système a comme base la recherche cognitive

de Sharples (section 2.3.2 ) [Pérez y Pérez, 1999, Sharples, 1999]. *Mexica* génère des récits sur les anciens habitants de la ville de Mexico. Dans *Mexica*, un récit est composé d'une séquence d'actions que les personnages peuvent réaliser nommées *Story actions* (qui ont des préconditions et des post-conditions), lesquelles ont des effets (sous forme de réactions émotionnelles des personnages). Dans un fichier texte, l'utilisateur donne comme entrée les actions, ainsi qu'un ensemble de récits nommés *previous stories* [Pérez y Pérez, 2007], contenant également des *story actions*. *Mexica* abstrait le contenu des *previous stories* pour construire des schémas nommés *Story World Context* (SWC) afin de créer la base de connaissances du système (ou la *fabula*). Les SWC contiennent trois éléments : une *story action*, son effet et l'action qui doit découler de cet effet.

Le récit est généré en utilisant un modèle d'engagement et réflexion (*engagement-reflection*) [Pérez y Pérez, 1999]. Durant la phase d'engagement, le système produit de nouvelles données selon le contenu présent et les contraintes rhétoriques (il n'y a pas de buts explicites ni d'état ou de structure d'histoire figée). Dans le mode réflexion, le système détecte les impasses générées lors de l'engagement, satisfait la cohérence et évalue la nouveauté et l'intérêt de l'histoire en cours.

Brièvement, durant la phase d'*engagement*, le système utilise la première *story action* de l'utilisateur et la base de connaissances. Ensuite, *Mexica* cherche dans la base de connaissances un SWC contenant la même *story action* de l'utilisateur. Si c'est le cas, le système ajoute la *story action* suivante qui se trouve dans le SWC récupéré de la base de connaissances. Le processus est réalisé itérativement et le résultat est une séquence de *story action*. La phase de *reflection*) reçoit ce qui a été produit dans l'étape précédente et regarde s'il y a des inconsistances, notamment des *story actions* qui se répètent.

La figure 3.11 montre un récit produit par *Mexica*. Elle raconte l'histoire d'une princesse qui aide un guerrier blessé. À la différence d'autres approches, telles que la planification, la composition du récit est faite à partir des réactions émotionnelles des personnages (effets) qui servent de liens causaux et fournissent la cohérence du récit. Par exemple, la princesse guérit les blessures du guerrier (*story action*). L'effet résultant est une réaction émotionnelle (le guerrier se sent redevable et remercie la princesse). Ainsi, l'action déclenchée est de donner un cadeau qui exprime ce remerciement. Dans l'exemple, le guerrier donne du cacao, en créant ainsi un lien émotionnel entre les deux personnages. Un aspect remarquable de ce système est la génération guidée par les émotions des personnages, similaire à la formalisation proposée par [Lehnert, 1981], où le récit émerge à partir des réactions (positives ou négatives) aux événements. Toutefois, et de manière similaire à *Minstrel*, la

créativité des récits dépend du nombre de cas gardés en mémoire. De plus, le système demande un ensemble de ressources non négligeable pour la base de connaissances. En particulier, l'utilisateur doit théoriquement savoir indiquer de manière adéquate les *previous stories* et les *story actions*, sinon il y a un risque de création erronée ou incomplète de récits.

```
Jaguar_knight was an inhabitant of the
Great Tenochtitlan. Princess was an
inhabitant of the Great Tenochtitlan.
Jaguar_knight was walking when
Ehecatl (god of the wind) blew and
an old tree collapsed injuring badly
Jaguar_knight. Princess went in
search of some medical plants and
cured Jaguar_knight. As a result
Jaguar_knight was very grateful to
Princess. Jaguar_knight rewarded
Princess with some cacauatl (cacao
beans) and quetzalli (quetzal)
feathers.
```

FIG. 3.11: Récit généré par *Mexica*

### 3.2.2.3 *ProtoPropp* (2005)

*ProtoPropp* est un autre système qui emploie le raisonnement à partir de cas [Gervás *et al.*, 2005] pour construire automatiquement des contes de fées. Il prend comme base la théorie formulée par Vladimir Propp [Propp, 1970] (section 2.1.2) et se fonde sur la notion de fonction narrative. La génération commence avec la sélection des fonctions narratives à partir d'une interface graphique représentée figure 3.12. Ceci permet d'obtenir une description partielle du récit devant être généré. Puis, un module contenant des structures narratives en mémoire est consulté afin de trouver des cas (c'est à dire, de séquences partielles) similaires contenant les fonctions données au début du traitement. Ces cas gardés en mémoire sont des contes annotés et analysés préalablement pour identifier les fonctions narratives qui les composent. Si un cas est trouvé, il est adapté en suivant la description initiale. Pour gérer les cas, *ProtoPropp* emploie les techniques du web sémantique. Une ontologie du domaine représentée figure 3.13 décrit les fonction narratives et contient les instances des cas d'où sont extraits les fonctions narratives et les relations existantes entre elles pour garantir la cohérence du récit. Une fois le processus d'adaptation terminé, l'histoire est transmise à un module de génération

de textes.

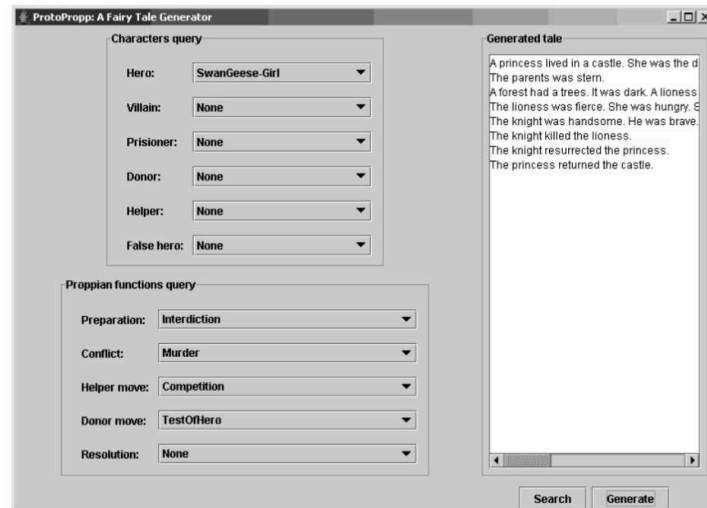


FIG. 3.12: Interface du système *ProtoPropp* [Gervás *et al.*, 2005]

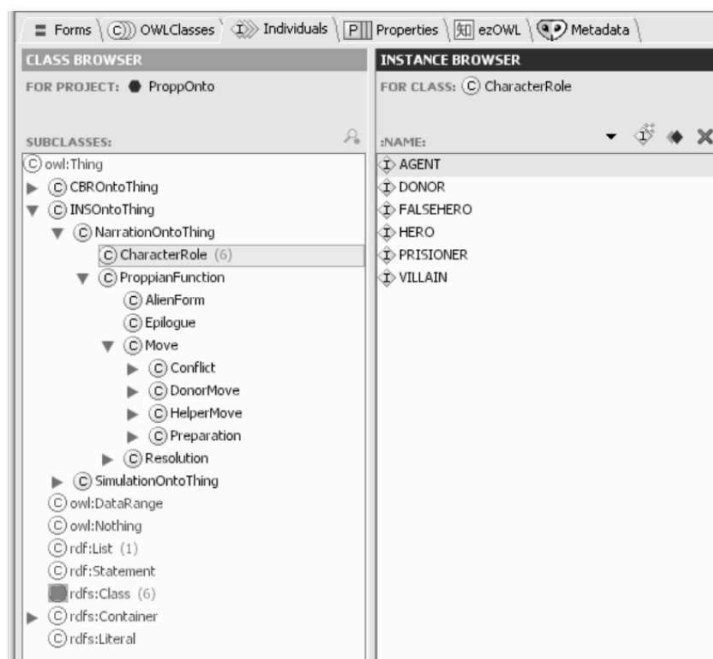


FIG. 3.13: Ontologie utilisée par le système *ProtoPropp* [Gervás *et al.*, 2005]

Once upon a time there was a princess. The princess said not to go outside. The princess went outside. The princess heard about the lioness. The lioness scared the princess. The lioness kidnapped the princess. The knight departed. The knight and the lioness fought. The knight won the fight. The knight solved the problem of the princess. The knight returned. A big treasure to the knight.

FIG. 3.14: Récit généré par *ProtoPropp* avec les paramètres saisis dans la figure 3.12 [Gervás *et al.*, 2005]

Une évaluation avec des participants a été mise en place pour juger les récits selon leur qualité linguistique, leur cohérence et leur degré d'intérêt et d'originalité [Peinado et Gervás, 2006]. Un des textes générés et évalués se trouve figure 3.14. Les résultats indiquent que le système est capable de produire des récits acceptables. Au niveau de la cohérence et de l'intérêt, les évaluations sont plutôt favorables. Par contre, les sorties du système ne présentent pas une qualité adéquate en termes d'explicitation des liens causaux. L'aspect novateur est ici l'utilisation de techniques du web sémantique pour établir des relations causales entre les fonctions narratives. Cependant, les inférences dépendent en grande partie des annotations faites sur le corpus de cas. Par ailleurs, la structure du récit reste figée étant donné qu'elle est basée sur des contes de fées. Autrement dit, bien que les fonctions soient modifiables, il y a aura toujours un début, un conflit et une résolution tel que le décrit la pyramide de Freytag (section 2.1.1).

### 3.2.3 Génération de récits à partir de grammaires

Une autre possibilité pour générer des récits est de modéliser la structure d'un récit à travers une grammaire (section 2.3.1).

#### 3.2.3.1 *Brutus* (1999)

*Brutus* génère des récits autour du concept de trahison [Bringsjord et Ferrucci, 1999, Bringsjord et Ferrucci, 2000]. C'est un système qui a suscité un grand intérêt en raison de la qualité des textes qu'il générait. Les auteurs ont déclaré ne pas s'appuyer sur une théorie particulière mais il utilise de fait beaucoup de situations et de plans prédéfinis à la manière de Universe ou TALE-SPIN.

Une sortie du système est présentée figure 3.15. Ce récit relate la trahison subie par l'étudiant Dave de la part d'un professeur. Ce professeur lui avait assuré qu'il allait signer le passage de sa soutenance mais refuse finalement le jour de sa soutenance.

Dave wanted desperately to be a doctor. But he needed the signatures of three people on the first page of his dissertation, the priceless inscriptions which, together, would certify that he had passed his defense. One of the signatures had to come from Professor Hart, and Hart had often said — to others and to himself — that he was honored to help Dave secure his well-earned dream.

Well before the defense, Dave gave Hart a penultimate copy of his thesis. The professor read it and told Dave that it was absolutely first-rate, and that he would gladly sign it at the defense. They even shook hands in Hart’s book-lined office. Dave noticed that Hart’s eyes were bright and trustful, and his bearing paternal.

At the defense, Dave thought that he eloquently summarized Chapter 3 of his dissertation. There were two questions, one from Professor Rogers and one from Dr. Teer; Dave answered both, apparently to everyone’s satisfaction. There were no further objections.

Professor Rogers signed. He slid the tome to Teer; she too signed, and then slid it in front of Hart. Hart didn’t move.

“Ed?” Rogers said.

Hart still sat motionless. Dave felt slightly dizzy.

“Ed, are you going to sign?”

Later, Hart sat alone in his office, in his big leather chair, saddened by Dave’s failure. He tried to think of ways he could help Dave achieve his dream.

FIG. 3.15: Récit généré par *Brutus* (issu de [Fetzer, 2002])

Brutus possède une architecture fonctionnelle assez complexe qui est représentée figure 3.16. Pour générer une histoire, après avoir choisi un thème (ici la trahison) le système doit trouver un traître et un trahi. Dans ce but, le système cherche dans la base de connaissances les personnages qui correspondent au thème et qui peuvent se retrouver dans les mêmes activités. La base de connaissances est un ensemble de définitions en logique de première ordre (FLEX et Prolog).

L'intrigue (plot) est ensuite générée par le développement des plans de chaque personnage. Ainsi le traître Hart suit le plan suivant :

```
instance evilGoal is a goal
  agent is hart and
  plan is {lie101, refuse_to_sign101} and
  success is status of strivers_defense is failed.
```

Les événements tels que `refuse_to_sign101` ou `strivers_defense` sont eux même des plans avec des préconditions, des actions et des effets, tous réalisés à la main pour l’histoire en train d’être générée.

Une fois les événements et personnages définis en mémoire, le récit est structuré à l’aide d’une grammaire de récit du type suivant :



1. Story → Setting + Goals\_and\_plans + Betrayers\_evil\_action + betrayed's\_state
2. Goals\_and\_plans → Betrayed's\_goal + Betrayers\_promise + Betrayers\_goal
3. Setting → setting\_description(betrayal\_location,pov,betrayed)
4. Betrayed's\_goal → personal\_goal\_sentence(betrayed)
- ...

Chaque élément entre parenthèse correspond aux données transmises au générateur de textes pour produire un fragment du récit textuel final.

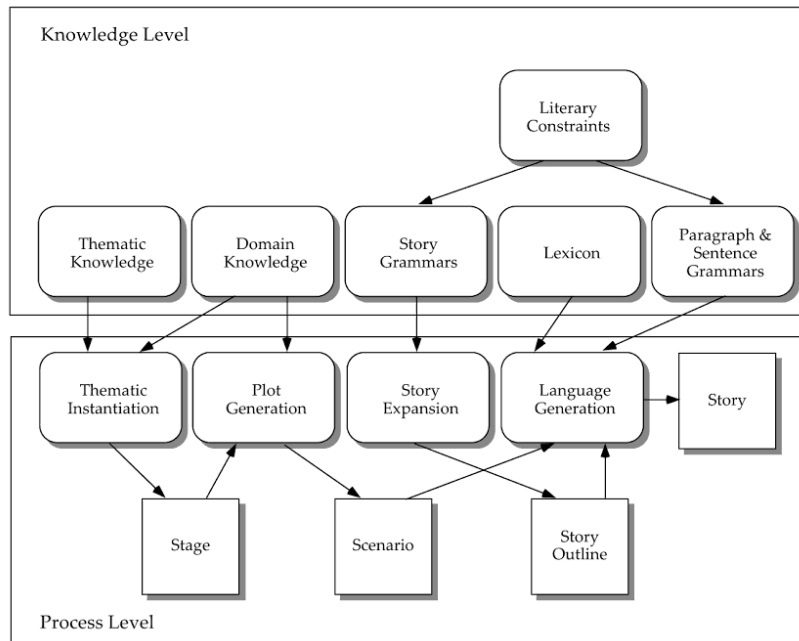


FIG. 3.16: Architecture de *Brutus* (issue de [Friedenberg, 2008])

Si Brutus permet de réaliser des sorties d'une très bonne qualité, le système nécessite un ensemble de connaissances très précises et non génériques. Ceci en fait un système extrêmement coûteux en connaissances à intégrer dans le système pour générer une seule histoire.

### 3.2.3.2 *Propper* (2013)

Dans *Propper* [Gervás, 2013, Gervás, 2016], le formalisme de Propp (section 2.1.2) a été revisité et adapté en termes informatiques. Il en résulte une grammaire de récits (à la différence de *ProtoPropp* cf. section 3.2.2.3) qui permet de guider la génération du plan du récit. Dans ce système, l'intrigue et la fabula sont obtenus par deux modules. Le *plot driver generator* génère une séquence de fonctions narratives prises parmi les 31 définies par Propp [Propp, 1970] (cf. Figure 2.2) et valides pour le monde du conte. Le *fabula generator* va analyser la séquence et instancier les fonctions narratives en des actions (*story action*) (et des personnages). Les *story actions* sont des instances des fonctions narratives sous forme de prédicats, et possèdent des pré-conditions et des post-conditions. Les fonctions sont instanciées de manière itérative jusqu'à ce que toutes les actions de la séquence satisfassent les post- et pré-conditions. Un troisième module *flow generator* que nous ne décrirons pas gère le déroulé textuel du récit. L'approche par grammaire est intéressante car il existe une garantie de cohérence grâce à la mise en place de pré-conditions entre les *story actions*. Par contre, cette grammaire est spécifique au monde des contes de fées.

Character function	<i>villainy</i>
Preconditions	married H Y <b>hero H</b> <b>villain X</b>
Action	makes_disappear X Y
Postconditions	<b>victim Y</b> sundered H Y

FIG. 3.17: Exemples *Story actions* de la fonction narrative *Vilain* [Gervás, 2013, Gervás, 2016]

Plusieurs méthodes ont été testées pour les différents modules. Pour le module *plot driver generator* une grammaire de plot a été utilisée. la figure 3.18 donne un extrait de la grammaire utilisée dans [Gervás, 2013].

```

...
FOLKTALE      = COMPLICATION DONOR COURSE_OF_ACTION CLOSURE
FOLKTALE      = COMPLICATION COURSE_OF_ACTION
COMPLICATION  = TRIGGER
COMPLICATION  = TRIGGER MEDIATION
TRIGGER       = A
TRIGGER       = a
MEDIATION     = C
COURSE_OF_ACTION = TASK
...

```

FIG. 3.18: Exemple de grammaire simple utilisant les fonctions narratives de Propp [Gervás, 2013]

Une évaluation a été faite de la génération de plot [Gervás, 2013]. Considérant des plots de référence, la mesure compte le ratio de fonctions narratives mal positionnées par rapport à la référence. Bien que la méthode par grammaire n’obtienne pas 100% de bons résultats, la conclusion de Gervás est que la grammaire permet des variations valides des contes qui ne sont pas capturées par les plots de références.

Plus récemment, le système a été étendu pour permettre l’exploitation de schèmes thématiques définissant des classes d’intrigues comme ”quête” ou ”vengeance” [Gervás *et al.*, 2015], de manière similaire à *Minstrel*. Un schème narratif est ici compris comme un vocabulaire qui caractérise différents types de thèmes et donne des pistes sur les fonctions narratives à inclure pour réaliser un récit avec ce thème. Par contre, d’après les auteurs, il semble difficile d’avoir une structure canonique pour chaque type d’intrigue ; les récits peuvent varier au niveau des fonctions, et par conséquent ne peuvent pas toujours s’ajuster à une description thématique donnée.

### 3.2.4 Génération de récits à base de modèles statistiques

Les systèmes de génération de récits des sections précédentes sont extrêmement dépendants des bases de connaissances utilisées pour créer les récits. Ceci implique que les connaissances doivent souvent être saisies bien avant la génération proprement dite que cela soit pour les éléments de l’histoire ou pour sa structuration (p.ex., fragments d’histoire). Par conséquent la construction de ces systèmes demande une quantité considérable de temps et d’effort. C’est pourquoi depuis quelques années des approches basées sur des corpus et à base de modèles statistiques sont apparues [McIntyre et Lapata, 2009, Harrison *et al.*, 2017, Roemmele, 2016a, Jain *et al.*, 2017]. Par exemple, McIntyre et Lapata [McIntyre et Lapata, 2009] proposent un système pour

généraliser automatiquement des contes de fées à partir d'un corpus de contes de fées. Avec cette approche, la formalisation informatique des théories narratologiques ou cognitives n'est plus nécessaire, car la composition du récit émerge du corpus. Le corpus est constitué de 437 contes de fées édités par Andrew Lang<sup>3</sup>. Le corpus est traité pour extraire les mots avec leur contexte d'apparition. Les verbes sont transformés en prédicats pour remplir une base de connaissances qui servira à guider la création de l'histoire à partir d'un sujet.

Le sujet de l'histoire est saisi par l'utilisateur (sous forme d'une phrase simple ou de mots isolés). Le sujet contient les *entités* et l'action qui serviront comme personnages et action centrale. Le système sélectionne dans la base de connaissances les relations les plus proches du sujet et forme la structure du récit : un arbre hiérarchique où le niveau le plus haut est le sujet fourni et chaque feuille représente un événement classé. La profondeur de l'arbre indique la longueur du récit (si l'arbre a une profondeur de trois, le récit aura trois phrases). Un exemple d'arbre se trouve figure 3.19 représentant le sujet « le chien poursuit le canard ». Ici nous observons les prédicats trouvés pour l'entité « chien » en relation avec l'entité « canard » à partir du prédicat « aboyer(chien,canard) ».

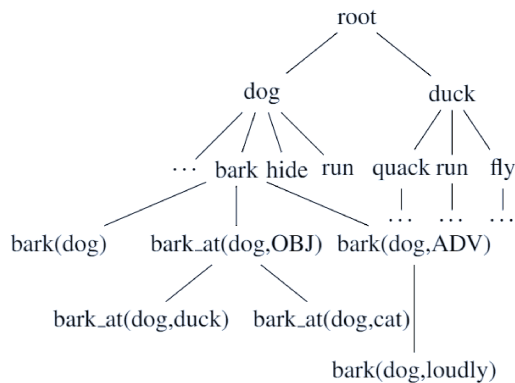


FIG. 3.19: Arbre produit qui montre les relations entre entités saisies par l'utilisateur [McIntyre et Lapata, 2009]

---

<sup>3</sup>Écrivain écossais (1844-1912) fameux pour l'édition et la publication d'une collection de contes de fées folkloriques.

The family has the baby. The baby is to seat the lady at the back. The baby sees the lady in the family. The family marries a lady for the triumph. The family quickly wishes the lady vanishes.

FIG. 3.20: Récit généré avec l'approche statistique [McIntyre et Lapata, 2009]

Un exemple de récit généré est présenté dans la figure 3.20. Le résultat montre des textes cohérents selon une évaluation réalisée avec des participants [McIntyre et Lapata, 2009]. L'approche proposée permet théoriquement de travailler sans investir de ressources et de temps dans la modélisation. Cependant, il existe des limites à surmonter. Au niveau des connaissances, il est implicitement demandé de travailler avec des domaines dont les récits sont fait de phrases simples afin de faciliter l'apprentissage statistique. Le traitement structurel est très pauvre, se contentant de sélectionner exclusivement les événements les plus fréquents. Enfin, en l'absence d'une théorie narrative mise en place, il est difficile d'affirmer que ces récits possèdent une vraie intrigue. Finalement, la qualité linguistique est basse au niveau des phrases, particulièrement la causalité n'est pas explicitée sous forme de connecteur. D'après ces travaux, il apparaît donc qu'il existe un compromis entre la rapidité de développement et la qualité narrative des réalisations.

### 3.2.5 Génération interactive de récits

À la différence des systèmes précédents où le rôle de l'utilisateur se limite à apporter des données d'entrée, l'approche interactive intègre l'utilisateur comme auteur ou co-auteur du récit. Ainsi, la cohérence du récit ne provient pas d'un mécanisme automatique, mais de l'individu qui décide de comment le récit va progresser, ce qui donne l'opportunité de créer des intrigues plus variées et plus créatives. L'utilisateur va pouvoir, par exemple, modifier le comportement des éléments de la narration, comme les personnages. Actuellement, il s'agit d'une tendance qui suscite l'intérêt du monde de l'industrie numérique, qui cherche constamment des moyens d'offrir à ses clients des expériences à la fois plus attrayantes et plus motivantes (par exemple, l'industrie du jeu vidéo). Dans la suite, nous présentons des exemples de systèmes qui fonctionnent en interaction avec l'utilisateur.

#### 3.2.5.1 *Façade* (2003)

*Façade* [Mateas et Stern, 2005] (<http://www.interactivestory.net/>) est un jeu interactif se jouant en vision subjective. Il met en scène le joueur

rendant visite à un couple d'amis – *Grace* et *Trip* – en crise. L'utilisateur doit jouer le rôle de médiateur entre les deux personnages, soit pour aider à résoudre les conflits, soit pour influencer une éventuelle séparation. Pour ce faire, l'utilisateur doit prendre des décisions qui vont affecter les réactions des personnages et par conséquent, la progression de la situation. La figure 3.21 montre la vue du joueur avec *Grace* au premier plan ainsi que le texte entré par le joueur afin de faire progresser l'histoire.



FIG. 3.21: Interface du système *Façade* [Mateas et Stern, 2005]

*Façade* est reconnu pour son architecture capable de traiter des interactions avec l'utilisateur en temps réel (fig. 3.22). *Façade* intègre un directeur du drame (*drama management*, similaire à celui de *Virtual StoryTeller*), des agents semi-autonomes (c'est à dire, les personnages), un monde de l'histoire et un module de compréhension de la langue pour interpréter les entrées de l'utilisateur. Lorsque l'utilisateur interagit avec le système en prenant une décision, le *directeur dramatique* sélectionne l'unité narrative (nommé *Beat*) appropriée selon cette décision. Le choix des *beats* est guidé par la pyramide de Freytag afin d'assurer une tension dramatique [Freytag, 1900]. Les *beats* (d'une durée d'environ 60 secondes) contiennent un ensemble de comportements encodés, ainsi que des dialogues pour chacun des personnages principaux autour d'un sujet (par, exemple, un secret ou un problème). Chaque *beat* possède des pré-conditions et effets qui vont aider le *directeur dramatique* à sélectionner le prochain *beat*, en suivant une structure dramatique prédéfinie en mémoire inspirée de [Freytag, 1900] (section 2.1.1). Ainsi, l'intrigue émerge à partir des interactions du joueur avec le système via des dialogues textuels [Mateas et Stern, 2005].

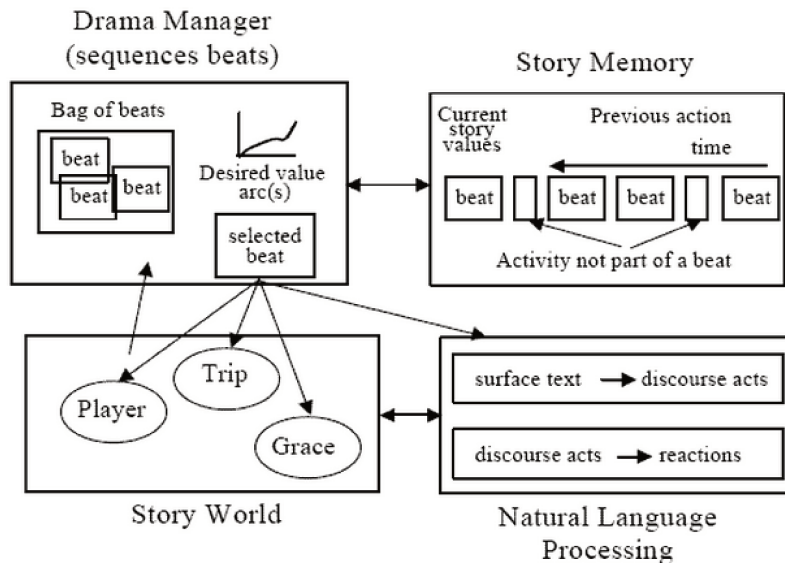


FIG. 3.22: Architecture du système *Façade* [Mateas et Stern, 2005]

La principale contrainte est le besoin d'avoir une quantité considérable de *beats* préalable, afin de couvrir l'ensemble des variations de séquences narratives, désavantage propre aux systèmes qui travaillent avec des unités pré-chargées (*Universe*, *Minstrel* ou *Mexica*).

### 3.2.5.2 *Holodeck* (2007) : adaptation de *Madame Bovary*

*Holodeck* [Cavazza *et al.*, 2007] est un système interactif en 3D se déroulant dans l'univers de *Madame Bovary*. Une vue du système et de ses éléments fonctionnels est montrée Figure 3.23. La construction de l'histoire est assurée par planification. Le mécanisme interactif de narration (*Interactive Storytelling Engine*) est en charge de créer la séquence d'actions de l'histoire et d'assurer la cohérence vis à vis de l'utilisateur. Chaque personnage est doté d'un planificateur qui contient les actions (nommées *story actions*) qu'ils peuvent réaliser. Lorsque l'utilisateur interagit, le mécanisme de narration cherche dans les plans des personnages quelles actions pourraient être appliquées. Ainsi, l'intrigue naît de l'interaction directe entre l'individu et les personnages.

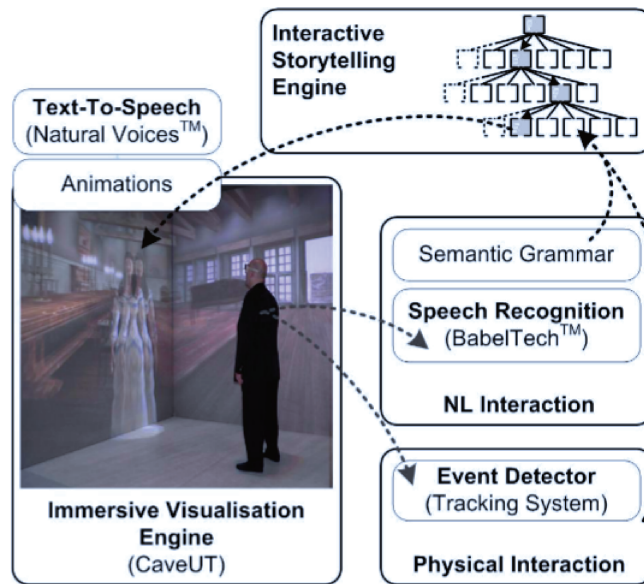


FIG. 3.23: Interaction entre l'utilisateur et "Emma Bovary" [Cavazza *et al.*, 2007]

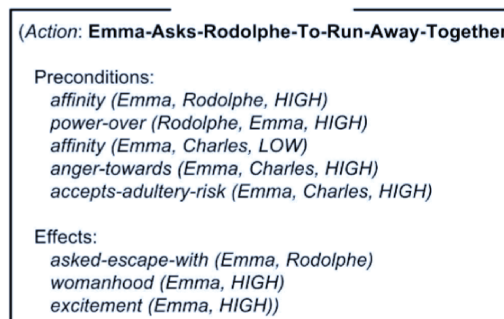


FIG. 3.24: Interaction entre l'utilisateur et "Emma Bovary" [Cavazza *et al.*, 2007]

Un exemple de *story action* est affiché figure 3.24. Ici, *Emma* demande à Rodolphe de s'en aller. L'action est déclenchée par l'état émotionnel de *Emma* sous forme de *pré-conditions*, par exemple la forte attirance d'*Emma* vers Rodolphe (action *affinity (Emma, Rodolphe, HIGH)*), ainsi que son peu d'intérêt pour *Charles* (action *affinity (Emma, Charles, LOW)*). Cette action a pour effet de demander de fuir tout en provoquant une forte excitation chez *Emma* (*excitement (Emma, Rodolph, HIGH)*). Les effets peuvent forcer les actions qui vont modifier l'état mental des personnages, selon les choix de



l'utilisateur. Ainsi, l'utilisateur a une grande marge de manœuvre dans un monde réaliste mais qui reste restreinte à ce que la planification autorise.

### 3.2.5.3 *Scheherezade-IF* (2015)

*Scheherezade* [Li et Riedl, 2015] est un système de génération qui utilise des corpus d'histoire pour construire sa base de connaissances à la manière de [McIntyre et Lapata, 2009]. La différence est que le corpus est plus volumineux et composé d'histoires simples de la vie quotidienne. Pour pouvoir utiliser le système, les auteurs ont fait appel au *crowdsourcing* afin de rédiger des textes annotés sémantiquement. Si l'utilisateur entre un thème nouveau, celui-ci est envoyé au système de *crowdsourcing* et plusieurs histoires sont rédigées sur le thème. Après analyse des récits, les événements narratifs les plus importants (consensuels) de l'intrigue sont extraits grâce au regroupement de phrases (par *clustering*). Les événements contigus sont ensuite liés deux à deux pour les ordonner temporellement et former la structure *plot graph* dont un exemple est donné figure 3.25. Le *plot graph* est la représentation de tous les événements extraits des récits humains. Le processus d'apprentissage du *plot graph* commence avec l'analyse textuelle de phrases dans les récits collectés (une phrase simple et avec un seul verbe par événement). L'analyse organise ces phrases qui sont similaires pour identifier les événements qui apparaissent fréquemment dans l'ensemble des textes, ceux optionnels et ceux qui sont incompatibles entre eux. Il en résulte que le *plot graph* constitue non pas un seul plot mais un plot dont plusieurs variations (dues aux spécificités de chaque récit sur lequel le plot est construit) sont possibles. Ainsi le *plot graph* contient des chemins d'histoire alternatives et mutuellement exclusives.

Lorsque le système est en mode de création, le choix des chemins alternatifs est résolu par l'interaction utilisateur. Les alternatives concernant les personnages non joueurs sont résolus par tirage aléatoire.

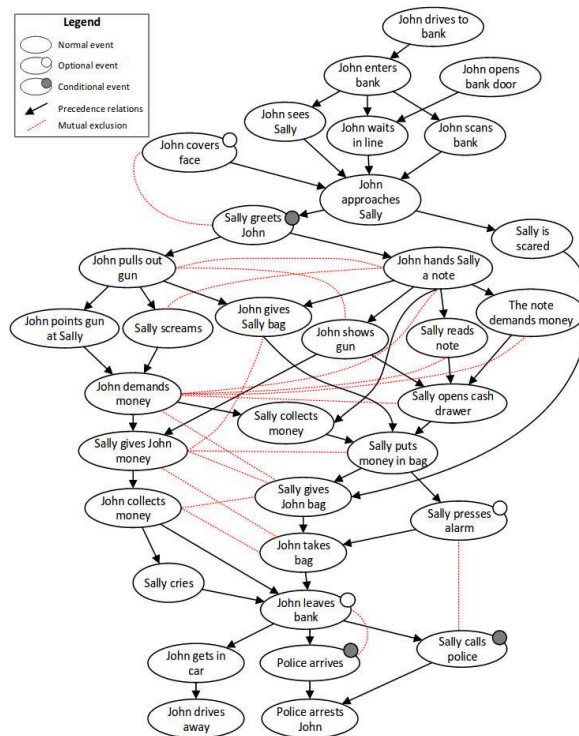


FIG. 3.25: *Ploth graph* de *Scheherazade-IF* autour du sujet « braquage d’une banque » [Li et Riedl, 2015]

L’approche *Scheherazade-IF* est très originale. Elle consiste à contourner le problème de la création de récits en utilisant les capacités d’un ensemble d’auteurs. Il est cependant évident que cette approche contourne le problème de l’acquisition de connaissances d’une manière qui ne la rend pas moins consommatrice en temps et en coût et demeure donc inimaginable à grande échelle. Cependant, il est possible d’imaginer que les textes de la base de données puissent être collectés automatiquement sur le web lorsque les technologies le permettront. Concernant la construction de l’histoire, l’utilisateur est en charge de sélectionner l’état suivant pour faire progresser l’intrigue. Cette interaction permet d’obtenir de la variation dans les récits à partir d’un même *Ploth graph*, mais qui reste limitée aux actions présentes dans celui-ci et limite donc la créativité. Cette approche est donc intéressante au niveau de la collecte d’informations mais également par le fait qu’à notre connaissance, elle est la première à représenter dans une seule structure (*Ploth graph*) l’unification de récits d’auteurs différents.

## 3.3 Évaluation de récits

L'évaluation de récits est, comme dans le plupart des tâches du TALN, une activité difficile et qui n'a pas encore de méthodologie établie dans la communauté. En effet, si l'on considère la complexité des systèmes, le fait que plusieurs couches d'abstraction doivent être évaluées (personnage, intrigue, style, intérêt, etc), qu'il n'existe pas une seule façon de narrer une histoire et enfin que les systèmes sont soit orientés créativité soit orientés narration, on voit qu'il est difficile de concevoir et construire une méthode d'évaluation générale et applicable à tous. Néanmoins, nous avons pu observer que plusieurs systèmes ont mis en place leur propre méthode d'évaluation, comme *Proto-Propp* ou *Propper*. Il est donc possible d'obtenir une évaluation à partir de l'observation des constituants (par exemple, la construction de l'intrigue) ou du résultat obtenu (la qualité linguistique du récit généré). Ainsi, plusieurs méthodes ont été proposées dans la littérature au cours des années. Cette section donne une description succincte des principaux travaux proposant des méthodologies d'évaluation.

### 3.3.1 *StoryEval* (2009)

*StoryEval* propose une observation empirique des récits, qu'ils soient interactifs ou non [Rowe *et al.*, 2009]. L'approche résulte de l'adaptation et de l'intégration des méthodologies employées par les sciences cognitives, l'interaction homme-machine et le traitement automatique des langues. La motivation des auteurs est de trouver un ensemble de techniques qui puissent évaluer la complexité des récits générés ainsi que les composants des systèmes de génération. *StoryEval* contient quatre outils pour mesurer la qualité des histoires, qui concernent plusieurs niveaux.

Le premier outil de *StoryEval* (*Narrative metrics*) évalue le mécanisme qui produit le récit à partir de la mesure de traits spécifiques trouvés dans les récits. Autrement dit, il cherche à rendre compte de comment la narration est présentée (style, grammaire, etc.). Une deuxième mesure d'ordre cognitivo-affectif (*Cognitive-affective studies*) explore la réponse des utilisateurs lors d'évaluations avec des participants. Le troisième outil (*Director-centric studies*) évalue la performance et l'effectivité de l'agent en charge de surveiller la construction du récit (*director agent*). Finalement, le quatrième outil (*Extrinsic Narrative Evaluation*) cherche à évaluer si le récit atteint effectivement le but pour lequel il a été conçu. *StoryEval* est présenté comme un *framework* d'évaluation qui offre la possibilité d'évaluer un récit à partir de plusieurs perspectives. Toutefois, l'article qui le présente [Rowe *et al.*, 2009], s'il liste

bien les points à évaluer par rapport à la littérature, reste très vague sur les mesures concrètes permettant de répondre aux points soulevés. Par ailleurs, l'étude se concentrant sur des systèmes de génération qui travaillent sur des domaines virtuels ou à partir de textes, ne met pas en avant d'autres points importants dans notre cas tels que la relation entre la quantité et la qualité d'information disponible et la construction de l'histoire.

### 3.3.2 Comparaison de plans de récit (2016)

L'un des aspects important pour l'évaluation d'un récit est d'être capable de mesurer la qualité du plan du récit. [Amos-Binks *et al.*, 2016] proposent une méthodologie d'évaluation de plans par comparaison afin d'identifier et de spécifier les différences majeures entre deux plans correspondant à deux solutions différentes pour un même récit. Les plans considérés sont des *Conflict Partial Order Causal Link (CPOCL)* [Ware *et al.*, 2014] dans lesquels des "plans classiques" sont augmentés des buts des personnages, des cadres intentionnels et des conflits. Sont également modélisés le plan de l'histoire, le problème à résoudre par le plan ainsi que la définition de la solution, étant donné la possibilité d'avoir plusieurs solutions pour un seul problème. La comparaison entre plans vise à mesurer des différences à un niveau sémantique plutôt que syntaxique comme cela est classiquement fait dans la littérature. Pour ce faire, une méthode de résumé de plan (*story plan summarization model*) est mise en œuvre pour capturer les éléments les plus importants et les intentions du personnage. La structure résultante est appelée *Important-Step Intention-Frame*. Elle constitue une représentation sémantique concise du plan, sur laquelle fonder la comparaison. La distance sémantique entre deux plans  $P1$  et  $P2$  est calculée par la distance de Jaccard  $\delta(P1, P2)$  entre les événements importants  $E$  et les cadres intentionnels  $J$  des résumés  $\psi_1$  et  $\psi_2$  de  $P1$  et  $P2$ .

$$\delta(P1, P2) = 1 - \frac{1}{2} \left( \frac{|E(\psi_1) \cap E(\psi_2)|}{|E(\psi_1) \cup E(\psi_2)|} + \frac{|J(\psi_1) \cap J(\psi_2)|}{|J(\psi_1) \cup J(\psi_2)|} \right)$$

Comme souligné par les auteurs, CPOCL est un cadre très contraint qui oblige les connaissances du monde de l'histoire à une description très précise. Cependant, cette approche est une des premières à apporter un tel niveau de formalisation pour l'évaluation des plans d'histoire.

### 3.3.3 Évaluation de la perception de la temporalité dans les récits issus de capteurs (2015)

Un autre aspect important à être mesuré est la perception de la temporalité dans les plans du récit. [Baez *et al.*, 2015] a réalisé une évaluation où il est demandé de reconstruire la trame en ordre chronologique d'un ensemble de récits textuels générés à partir des annotations basés sur un modèle de tâches. L'objectif était de savoir si le modèle pourrait porter une séquence d'événements temporellement cohérente pour raconter une activité humaine. Les participants, 18 francophones, devaient d'abord de lire les textes présentés (le texte originale et la version générée avec des annotations), et après placer les événements trouvés en ordre chronologique dans une ligne de temps.

La distance entre les réponses du participant et le plan du récit de référence a été observé au travers de la taux d'erreur de mots (*Word Error Rate*), unité de mesure utilisé dans les systèmes de reconnaissance de la parole. Les résultats ont montré que la reconnaissance de l'ordre chronologique dans les textes générés était légèrement meilleure en comparaison avec les textes originaux. D'abord, ils portaient moins de détails à comparaison des textes originaux, où il y a une quantité majeure d'information, ce qui fait semblant qu'il y a plus d'événements. Par ailleurs, dans les textes générés montraient un ordre chronologique plus explicite, ce qui a aidé à percevoir plus facilement la structure et ainsi reconstruire la séquence.

## 3.4 Bilan

Nous avons présenté dans ce chapitre l'état de l'art concernant la génération automatique de récits. Cette tâche fait partie de ce que l'on appelle la narratologie informatique plus connue sous le nom de *computational narratology*. Nous avons pu voir que de nombreuses approches ont utilisé comme formalisation une adaptation des théories de la narratologie et reposent sur une base de connaissances conséquente. Cependant, il peut être noté que les approches statistiques prennent une place grandissante par leur capacité d'exploiter les nombreuses ressources disponibles sur le web, grâce aussi à l'évolution des méthodes formelles pour représenter les plans des récits.

Nous avons observé qu'il existe plusieurs façons de représenter et de générer des récits automatiquement. Toutes ces approches, sans exception, se placent dans le cadre d'un monde fictif dans lequel les connaissances sont considérées comme disponibles et délimitant l'ensemble du monde de l'histoire. À partir de connaissances de base, les systèmes construisent un plan d'histoire qui respecte certaines contraintes. Nous avons classé ces systèmes

selon la méthodes de construction du récit : par planification, par raisonnement à partir de cas, par grammaire générative de récits, par méthodes statistiques et par interaction avec l'utilisateur. Malgré cette diversité d'approche, les systèmes exposés montrent des aspects récurrents que nous devons examiner vis à vis des objectifs de la thèse. Nous pouvons résumer cette analyse en quelques points :

**Domaine de connaissances** Dans les systèmes présentés, la connaissances est soit directement insérée manuellement sous forme structurée soit extraite à partir de récits textuels. À notre connaissance, il n'existe pas de système générant des textes respectant les contraintes du récit à partir de données capteurs incertaines et incomplètes.

**Monde fermé** Les bases de connaissances des systèmes présentés sont considérées comme contenant l'ensemble des informations de l'histoire à générer. À part dans le raisonnement à partir de cas (avec les limitations exposées section 3.2.2.1), de nouvelles connaissances ne sont pas censées être insérées dynamiquement dans la base. Par ailleurs, une grande part de de la structuration de l'histoire repose sur la connaissance *a priori* des intentions et des buts des personnages ainsi que des pré-conditions. Ces informations sont utilisées pour établir le plan du récit. Dans notre cas, les informations issues de capteurs sont incomplètes (et incertaines), ce qui signifie qu'il n'y a pas de garantie que certaines actions soient observables et qu'elles puissent être utilisées comme conditions lors de la construction d'un plan. Par ailleurs, l'intention des personnages n'est pas directement observable (ni leur rôle) et doit donc être modélisée comme une variable cachée.

**Théorie narrative** La plupart des systèmes font directement ou indirectement référence soit à des théories de la narratologie (p.ex., Propp [Propp, 1970]), soit à celles de la cognition (p.ex., Schank [Schank et Tesler, 1969]), soit à aucune de celles-ci (p.ex. approches purement statistiques). Ce support théorique influence le développement et l'évaluation des systèmes qui se focalise généralement soit sur la structure narrative, soit sur le processus cognitif. Il serait intéressant dans cette thèse d'aborder ces deux aspects, c'est-à-dire de respecter une structure narrative dont la composition est justifiée d'un point de vue cognitif.

**Intrigue et plan de récit** On peut constater de cet état de l'art que les notions de *Plot*, intrigue et plan de récit ne sont pas toujours compatibles entre les différents travaux. Par exemple, certains *plot* s'instancient comme des fabula alors que d'autres ne servent qu'à structurer les éléments cruciaux du récit ou à assurer la cohérence. Certains plans de récit sont composés d'opérateurs (*story action*) avec préconditions et

intentions alors que d'autres sont le résultat d'une organisation déduite ou supportée de manière statistique. Dans notre cas, il sera nécessaire de clairement définir ces notions. De plus, les éléments du plan du récit ne peuvent pas être assemblés selon les méthodes précitées car nous ne disposons ni d'un modèle d'histoire ni d'un corpus linguistique d'entrée étant donné que l'information d'entrée provient de capteurs. Il conviendra cependant d'identifier la connaissance nécessaire pour supporter une structuration cohérente.

**Intégration des aspects communicatifs** Les systèmes se sont principalement intéressés à la création automatique de récit. Par contre, l'aspect destinataire (*à qui* ce récit est-il destiné?) et l'aspect but communicatif (pour quel *but* ce récit est-il généré?) sont assez peu présents dans les travaux de l'état de l'art. Ces questions seront soulevées dans ce manuscrit. Il s'agira en particulier de savoir comment moduler et adapter la génération du récit pour tenir compte de manière explicite de l'intention communicative de l'auteur et du profil du destinataire.

L'évaluation de ces systèmes représente un défi relevé par un faible nombre de travaux. Si les systèmes fictifs peuvent être évalués à partir de la théorie sur laquelle ils reposent ou sur corpus, dans notre cas, le traitement des données capteurs ainsi que la non observabilité des informations couplée à l'impossibilité de créer des actions fictives pour satisfaire une contrainte ajoutent de la complexité à l'évaluation de notre approche. Tous ces défis seront abordés selon la démarche que nous adopterons pour générer des récits à partir de données non textuelles qui est présentée dans le chapitre suivant.

## Chapitre 4

# Vers une génération automatique de récits à partir de données ambiantes

L'état de l'art présenté dans les chapitres précédents montre les contributions majeures du domaine portant sur la production de récits fictifs pour lesquelles les informations nécessaires tant au niveau de l'histoire (déroulement, personnages, etc.) qu'au niveau de la forme (structure du texte, vocabulaire, etc.) sont généralement identifiées avant la génération. Dans le cadre de cette thèse nous cherchons à générer des récits à partir d'observations provenant de la réalisation d'une activité humaine réelle. L'objectif est de faire "émerger" l'histoire à partir d'un ensemble de données capturant de l'activité humaine plutôt que de générer une simple description textuelle de cet ensemble. Les activités humaines étant variables et parfois non observables, il n'est pas possible d'appliquer directement les approches de génération de récits de l'état de l'art. Ainsi, il est nécessaire de proposer une démarche de génération de récits spécialement conçue pour travailler avec des données réelles et sans connaissance exhaustive et *a priori* de l'ensemble des éléments de l'histoire.

Les récits que nous cherchons à produire dans cette thèse sont particuliers. D'une part, ils sont basés sur une expérience réelle (comme peut l'être un article de journal) mais en plus, cette histoire est obtenue (inférée) à partir de sources objectives tels que des capteurs de données ambiantes. La particularité des sources de données de notre approche de génération nous a amené à définir un type de récit spécifique. Nous l'avons nommé *récit d'activité*. Un récit d'activité est un récit basé sur des données captées lors d'une activité humaine. Notre approche de génération répond alors à deux problématiques principales : *comment construire un plan d'histoire à partir des données captées qui respecte les canons narratifs et qui représente une activité cohérente*



et comment transformer ce plan en un récit transmettant un message à un lecteur.

## 4.1 La problématique

Comme présenté dans le chapitre 3, la Narratologie Informatique cherche à analyser et à générer automatiquement des textes narratifs. Pour ce faire, les auteurs appliquent les théories narratologiques issues de l'étude du récit (cf. chapitre 2). Ces travaux nécessitent alors l'adaptation et la précision des concepts narratologiques. Toutefois, les particularités liées à la présentation et au partage d'une activité humaine sous forme de récit textuel ne sont pas étudiées par la Narratologie Informatique. Dans cette thèse nous aborderons la problématique en considérant les quatre dimensions suivantes.

**Données de capteur en entrée.** Les informations données en entrée aux systèmes de génération automatiques ou semi-automatiques sont des informations qui appartiennent au monde de l'histoire (dont les informations ont un sens vis à vis de l'histoire). Dans notre cas, travailler avec des données observées par capteur implique de passer par une étape de transformation des signaux du monde physique vers des informations propres à être intégrées dans la fabula. Autrement dit, la fabula doit être construite à partir des données avant de savoir quelle forme l'histoire prendra.

**Connaissances incomplètes et incertaines.** Les systèmes de génération de récits fictifs fonctionnent en *monde fermé*. Par *monde fermé*, nous faisons référence au fait que la base de connaissances est soit complète (seul existe ce qui est dans la base) soit complétée à travers des critères qui respectent le monde de l'histoire (e.g., section 3.2.2). Un autre parti pris important est que toutes les informations utiles à l'histoire sont observables et donc connues ; cela comprend : les sentiments des personnages, l'état du monde où l'histoire se passe (lieux, objets), tous les événements ordonnés chronologiquement, etc. Dans le cadre de cette thèse, les récits doivent être générés dans une perspective de *monde ouvert*. Le *monde ouvert* implique que nous dépendons uniquement de l'information captée et de connaissances *a priori* sur l'activité. Nous n'avons pas accès, par exemple, aux opinions des protagonistes. De plus, les informations provenant de capteurs, il faut prendre en compte le fait que beaucoup d'informations ne sont pas observable et que celles inférées sont incertaines.

**Informations pertinentes pour les récits d'activité réelle.** Les systèmes générateurs de récits fictifs s'appuient sur l'étude d'œuvres narratives provenant de l'imagination des auteurs. Leurs approches se focalisent sur le développement de mécanismes qui permettent d'imiter le processus créatif

pour produire un nouveau récit, suscitant des émotions et ouvrant un univers imaginaire. Dans le cas de récits basées sur une activité humaine réelle, l'objectif est différent. Il s'agit de relater une activité réelle et de transmettre une expérience vécue. Il s'agit aussi de renseigner le lecteur du récit, susceptible de s'engager à son tour dans cette activité, sur son intérêt, son accessibilité ou au contraire les risques qu'elle peut entraîner. En conséquence, les éléments narratifs des récits fictifs produits par les systèmes de génération ne se trouvent pas nécessairement dans les récits de l'activité humaine : ainsi, les protagonistes de l'histoire, s'ils sont présents dans les deux cas, ne vont pas jouer de rôles stéréotypés (il n'y aura pas de "gentil" ou de "méchant").

**Relations entre les événements de l'histoire.** Les relations entre les événements des récits fictifs sont des relations causales et temporelles qui sont établies afin de construire l'intrigue. Les analyses des récits fictifs ont permis d'aboutir à une liste exhaustive d'intrigues et de plans (ie : organisation d'événements) qui permettent de les exposer. Grâce à cela, des approches comme la planification ont été utilisées comme solution pour générer des plans d'événements. Dans le cas de récits d'activités réelles, les événements et leur organisation peuvent varier d'un récit à l'autre (y compris pour la même activité). Par exemple, une personne faisant la même activité deux jours différents peut agir différemment (changer la routine, omettre une action, etc.). Son ressenti, son expérience, peuvent varier profondément, selon le contexte du moment. Enfin, son intention de communication peut varier, impliquant par exemple une pure volonté pédagogique de détailler l'activité réalisée, ou au contraire l'envie de partager une émotion ressentie, comme le sentiment d'une réussite.

## 4.2 Choix du cas d'étude

Nous avons sélectionné le ski de randonnée comme cas d'usage pour la thèse. Il s'agit d'une activité sportive réalisée au pleine nature et dans un contexte non balisé qui consiste à gravir une montagne sans l'aide des montées mécaniques et à redescendre à ski dans un terrain non balisé (figure 4.1). Cette activité est réalisée individuellement ou en groupe et permet la découverte de paysages et d'itinéraires. Sa pratique demande une maîtrise technique et une bonne forme physique pour faire face à un certain nombre de difficultés : longueur de l'itinéraire, raideur de la pente à la montée ou à la descente, qualité de la neige...Sa seconde particularité réside dans l'importance des prises de décisions : se déroulant en itinéraire non balisé, sans sentier, l'activité nécessite un excellent sens de l'orientation, et une capacité à utiliser les outils de navigation (carte, GPS, etc). Enfin elle se déroule dans

un milieu très changeant : la qualité de la neige varie beaucoup (glace, neige dure, croûte, neige lourde, poudreuse...), selon l'orientation, la température ou les chutes de neige précédentes, modifiant les difficultés de la randonnée et en augmentant éventuellement les risques (avalanches en particulier). C'est pour l'ensemble de ces raisons que le partage de connaissances autour de l'activité est extrêmement populaire, et il faut noter que le récit joue ici un véritable rôle de "passeur d'expérience", support clé pour transmettre et comprendre comment réagir face à telle ou telle situation.



FIG. 4.1: Skieurs pendant une sortie de ski de randonnée

Nous avons choisi cette activité comme cas d'usage car elle permet d'explorer les quatre dimensions de la problématique que nous avons identifiées dans la section précédente.

**Données captées en entrée.** Un avantage de cette activité est l'utilisation par les skieurs d'équipements portant des capteurs, comme illustré dans la figure 4.2. Les usages de ces dispositifs électroniques sont variés : aide à l'orientation, enregistrement de l'itinéraire réalisé, mesure de la performance et des paramètres biométriques, aide à la détection des victimes d'avalanches... Les données de ces capteurs doivent être interprétées pour être présentées aux skieurs. La figure 4.3 présente des représentations graphiques des données biométriques (Figure 4.3a) et des données GPS (Figure 4.3b). La présentation sous forme de texte demandera une interprétation de ces valeurs numériques. Cependant, la Figure 4.3a met en évidence la variation des valeurs fournies par un même capteur ainsi que la présence de valeurs aberrantes. La collecte dépend énormément de la performance ou de la bonne manipulation des dispositifs, et il existe donc un risque d'avoir des données incomplètes ou incorrectes (e.g. un point géographique mal identifié). Enfin, les données sont obtenues uniquement au moment de la réalisation de l'activité et de l'utilisation des dispositifs. Donc, il n'est pas possible d'avoir des informations hors de la réalisation de la sortie (par exemple, il est impossible d'avoir des connaissances sur ce qui a motivé le choix de l'itinéraire).

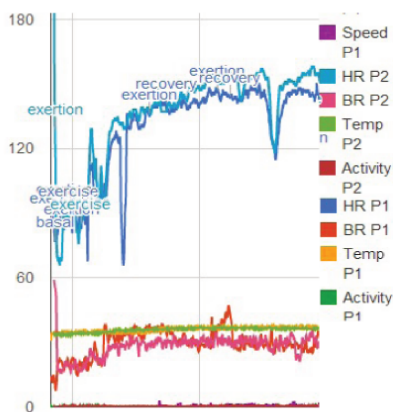


(a) Vêtements et articles divers

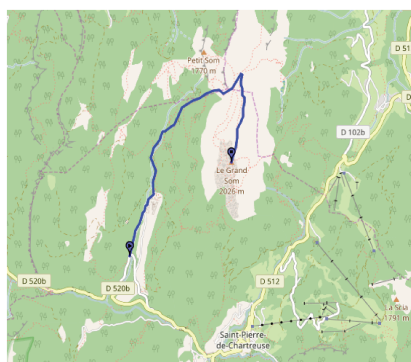


(b) Dispositifs électroniques fréquemment utilisés par les skieurs

FIG. 4.2: Équipements utilisés pour réaliser le ski de randonnée



(a) Données biométriques



(b) Carte géographique sur laquelle est affichée un itinéraire de sortie de randonnée

FIG. 4.3: Données issues d'une sortie de ski

**Informations pertinentes pour les récits d'activité réelle.** Le ski de randonnée est une activité populaire et avec des pratiquants passionnés qui aiment partager leurs expériences et conseils. Ce partage est réalisé via des modes différents dont l'utilisation de sites web dédiés (comme le site skitour<sup>1</sup>). Ces sites permettent aux randonneurs de partager des informations selon différents moyens dont des récits textuels dans lesquels ils exposent leur point de vue sur leur expérience. Par exemple, la figure 4.4 présente un récit écrit par un randonneur. Ce récit raconte une expérience personnelle en incluant plusieurs types d'informations : factuelles (par exemple : *Partis à 7h20 du parking; il y a 20 cm [de neige] de plus qui sont tombé dans*

<sup>1</sup><http://http://www.skitour.fr/>

la nuit) ; et subjectives (par exemple, *génial !* . Grâce à ces expressions, le texte devient plus vivant, ce qui permet de transmettre certains aspects que l'auteur considère importants, comme le niveau de danger auquel il s'est exposé ou sa réaction à quelque chose d'inattendu.

*Partis à 7h20 du parking, il y a au moins une dizaine de personnes devant moi, dont un pote qui a démarré 10 min plus tôt. Nickel, la trace est faite sous le tire-cul et presque jusqu'en haut (cote 2030m), où je retrouve mon pote que je n'ai jamais rattrapé. Lui redescend, (Boulot oblige), moi j'ai envie de monter au sommet et de voir le soleil se lever. Du coup, je trace les 50 derniers mètres, tjs un peu dans les cailloux à cet endroit là. Une petite plaquounette poudreuse part sous mes skis au niveau du petit ressaut rocheux avant d'arriver sur les câbles. Le deuxième ressaut au niveau des câbles est bien colmaté, on doit faire tout juste 4 pas sur le caillou, le reste est en neige ! Lever de soleil au sommet, derrière Belledonne, génial ! Je ne suis pas descendu sous la falaise, au droit des râteliers, par le petit couloir, ça me parassait un peu trop craignos, à la vue de l'accumulation présente tout le long et de la petite plaque partie à la montée. Je suis donc passé à gauche des râteliers, dans un grand champ de neige vierge (Attention, ça devait être soufflé auparavant, j'ai touché l'éboulis une fois ou deux), puis j'ai rejoins le champignon qui, pour une fois, n'était pas ravagé par les millions de skieurs qui montent la haut (Vu ce qui montait, ce soir c'est mort, il va falloir farfouiner pour faire sa trace). Par rapport à hier (cf. sortie David Zijp), il y a 20 cm de plus qui sont tombé dans la nuit. On se croirait au canada, bien que je n'ai jamais skié la bas :-)* Un Chamechaude de haut niveau.

FIG. 4.4: Texte écrit par un skieur pour décrire une sortie vers le sommet Chamechaude ( <http://www.skitour.fr/sorties/chamechaude,44954.html>, date de consultation : 14/03/18)

**Relations entre les événements de l'histoire.** Il s'agit d'une activité contrainte en actions (monter vers un sommet et descendre), mais riche en nombre d'exécutions qui varient en grande partie à cause des conditions du terrain, de la météo et, même, des skieurs. En ce qui concerne l'intention de communication, on note dans le récit donné en exemple de nombreux détails sur les conditions rencontrées, souvent utilisés pour valoriser la performance du skieur.

### 4.3 Définition du *récit d'activité*

Pour tirer parti des contributions de la narratologie et appliquer la génération de récits à des données de capteurs nous proposons une démarche qui intègre les éléments narratologiques les plus adéquats pour raconter une activité humaine, dans l'objectif de transmettre une expérience vécue, et en

considérant que nous allons travailler avec des données incomplètes. Avant de concevoir cette démarche, nous devons définir le type de récit que nous envisageons. Nous proposons la définition des récits que nous cherchons à produire sous le nom de *récit d'activité*, précisé à partir de [Baez et al., 2014].

D'après Brémond, un récit est un discours sur une succession d'événements dans une unité d'action qui est pertinente pour l'humain [Brémond, 1973]. Barthes [Barthes, 1966] précise que la séquence d'événements et d'actions dans un récit est produite par des liens causaux. D'après Genette [Genette, 2007], un récit caractérise le contenu d'une *histoire narrative*. Plusieurs définitions d'une histoire narrative (en anglais *Narrative*) existent dans la littérature, Mieke Bal [Bal, 2009] la définit comme ceci :

"A *narrative* text is a text in which an agent relates ('tells') a story in a particular medium, [...]. A *story* is a fabula that is presented in a certain manner. A *fabula* is a series of logically and chronologically related events that are caused or experienced by actors. *Actors* are agents that perform actions. [...] To act is defined here as to cause or to experience an *event*."

FIG. 4.5: Définition d'une histoire narrative de Mieke Bal [Bal, 2009]

En appliquant cette définition avec les caractéristiques du récit tels qu'Adam [Adam, 1999] les expose, nous posons les définitions de narrateur (ie d'agents, définition 4.1), d'événements (définition 4.3), de fabula (définition 4.8) et d'acteurs (définition 4.4). De plus, nous introduisons dans ces définitions la notion de destinataire. En effet, dans sa définition d'un récit, Brémond précise que la succession d'événements présents dans le discours est "*pertinente pour l'humain*". En conséquence, nous intégrons également le destinataire (définition 4.6) dans notre définition du récit d'activité (définition 4.2). Enfin, la prise en compte du lecteur et la définition de *narrative* par [Bal, 2009], qui suggère que quelque chose guide la manière de présenter les événements de l'histoire (" *A story is a fabula that is presented in a **certain manner***" nous ont fait ajouter la notion de but communicationnel (définition 4.5) dans la définition du récit d'activité (définition 4.2). Comme le montrent [Hovy, 1988] et [Reiter et Dale, 2000] le but communicationnel est un concept important qui permet de spécifier 1) pourquoi un récit doit d'être généré et 2) les opérations de production nécessaires pour atteindre ce but. L'intégration d'un but communicationnel est donc un concept particulièrement important dans le cas de la génération de récits car il guide la construction de celui-ci.

**Définition 4.1 *Narrateur* ( $\mathcal{N}$ ) :** Entité en charge de transmettre l’histoire au destinataire  $\mathcal{D}$ .

Dans les récit d’activité  $\mathcal{RA}$ , nous considérons que le rôle du narrateur est de transmettre l’expérience vécue par les acteurs en suivant une certaine perspective. Cette définition est inspirée des propos de [Genette, 2007], lequel signale que le narrateur souhaitera montrer plus ou moins d’information selon son degré d’implication dans le récit (par exemple, s’il participe ou s’il est seulement un observateur). Il est important d’indiquer qu’à la différence de [Genette, 2007], où trois types de focalisation sont définis (voir section 2.2.3), le narrateur dans nos récits d’activité reste limité à la focalisation externe, puisqu’il s’agit de transmettre ce qui a été fait, et non ses sentiments ou pensées (focalisation interne). Bien que l’objectif soit d’exprimer sous forme de récit une expérience réelle vécue, nous ne cherchons pas à l’exprimer en prenant la place de celui qui l’a vécue. Les récits d’activité sont une plateforme pour partager les expériences mais toujours en prenant une distance vis à vis de la personne qui les a générées. De cette manière, le récit transmet l’association entre le contexte environnemental de l’activité et un vécu physique, décisionnel et émotionnel. Ainsi, nous pourrions percevoir le récit d’activité comme une entité contenant un ensemble de connaissances et faits passés pendant l’activité en question, sans inclure une introspection ou une exposition intime de celui qui le raconte.

**Définition 4.2 *Récit d’activité* ( $\mathcal{RA}$ ).** Un récit d’activité présente une séquence d’événements  $\mathcal{E}$  de manière à communiquer l’expérience vécue d’un acteur  $\mathcal{A}$  lors de la réalisation d’une activité humaine en suivant un but communicationnel  $\mathcal{B}$  et pour un destinataire  $\mathcal{D}$  particulier. Les événements présentés reflètent uniquement ce qui a été observé, décidé ou ressenti pendant l’exécution de l’activité.

Ainsi, nous représentons le récit d’activité  $\mathcal{RA}$  comme une tuple de quatre composants :

$$\mathcal{RA} = \langle \mathcal{E}, \mathcal{A}, \mathcal{B}, \mathcal{D} \rangle \quad (4.1)$$

**Définition 4.3 *Événement* ( $\mathcal{E}$ ) :** C’est l’unité du récit d’activité  $\mathcal{RA}$  représentant une transition entre deux états. Cette transition est causée par une action qu’un (ou plusieurs) acteur  $\mathcal{A}$  a réalisée, observée ou expérimentée dans certaines circonstances, délimitées en temps et dans l’espace. Dans un récit, un événement montre des changements que l’acteur réalise (ou subit) pendant l’activité humaine et qui traduisent sa progression.

**Définition 4.4 *Acteur* ( $\mathcal{A}$ ):** Entité qui réalise ou est témoin des événements de l’activité humaine. Il peut avoir des objectifs à accomplir pendant

la réalisation de l'activité. Un acteur peut être soit une entité vivante (p.ex., animal) soit un phénomène physique (p.ex., le soleil fait fondre la neige) et peut correspondre à plusieurs individus (p.ex., le groupe décide de partir).

**Définition 4.5 *But communicationnel* ( $\mathcal{B}$ ):** Effet que l'auteur souhaite atteindre chez un lecteur lorsqu'il partage l'expérience de l'acteur. C'est le but communicationnel qui guide la construction du plan du récit (définition 4.7).

**Définition 4.6 *Destinataire* ( $\mathcal{D}$ ):** Individu(s) à qui le récit d'activité sera transmis. Dans le cas de récits textuels, le destinataire est le lecteur visé.

**Définition 4.7 *Plan du récit* ( $\mathcal{PR}$ ):** Un plan du récit est un ensemble d'événements sélectionnés de la fabula (définition 4.8) et leur organisation pour satisfaire un but communicationnel  $\mathcal{B}$  tout en respectant un ordre cohérent vis à vis des relations causales entre les événements. Cette notion peut être rapprochée de celle d'intrigue (telle que présentée dans la section 2.2.2.3). Dans la suite du manuscrit, nous ne considérerons que la notion de plan du récit en prenant comme point de vue sur l'intrigue qu'elle est une partie du plan du récit (la séquence des événements)

**Définition 4.8 *Fabula* ( $\mathcal{F}$ ):** Séquence d'événements vécus et expérimentés par les acteurs de l'activité, organisés chronologiquement.

## 4.4 Approche de génération de récits d'activités

Nous présentons dans cette section la démarche proposée pour la génération de récits d'activités, tenant compte de la problématique spécifique à ce type de récit dans ses quatre dimensions et des définitions proposées. Nous exposons ensuite les verrous scientifiques originaux soulevés par cette démarche.

### 4.4.1 Production basée modèles d'un récit d'activité

La démarche que nous proposons cherche à réunir deux mondes : un monde *objectif*, dans lequel l'activité se passe dans la vie réelle et un monde *subjectif*, dans lequel l'expérience vécue de l'activité est retranscrite pour être transmise. La figure 4.6 représente ces deux mondes. Le monde objectif de l'activité est constitué des données numériques (*monde des données*)



capturées et de leur interprétation en tant que déroulement temporel d'une activité (*monde de l'activité*) réalisée par un acteur. Cette interprétation implique les connaissances d'un expert sur la nature des données et sur le monde objectif de l'activité. Le monde subjectif implique un auteur dont le but est de construire le récit de l'expérience vécue, en vue de le transmettre à un lecteur. Pour cela, l'auteur constitue l'histoire (*monde du récit*) et la rend compréhensible sous forme textuelle (*monde du langage*). Dans notre approche, l'auteur sera une machine. Pour cette raison, nous avons choisi d'appliquer certaines restrictions à la transcription subjective de l'activité vécue. Dans le cas d'un récit d'activité produit par un auteur humain (a fortiori si l'auteur est l'acteur), celui-ci intègre des aspects comme les pensées ou les émotions. À partir des données objectives, ces informations ne peuvent qu'être fictives. Cependant, elles permettent au lecteur de mieux se projeter (elles rendent le récit plus "vivant"). Nous avons fait le choix de limiter les avis subjectifs à des avis sur des conditions environnementales (avis partagés par la communauté). Par exemple, si la température est de  $-20^{\circ}\text{C}$  et que le vent se lève, il est possible de considérer un point de vue subjectif négatif sur les conditions météorologiques. De même, au contraire des récits d'activité produits par les skieurs, nous avons choisi d'exclure du monde du langage les tournures personnelles telles que l'utilisation des pronoms "je" et "nous" car notre approche est d'utiliser un narrateur extérieur au récit.

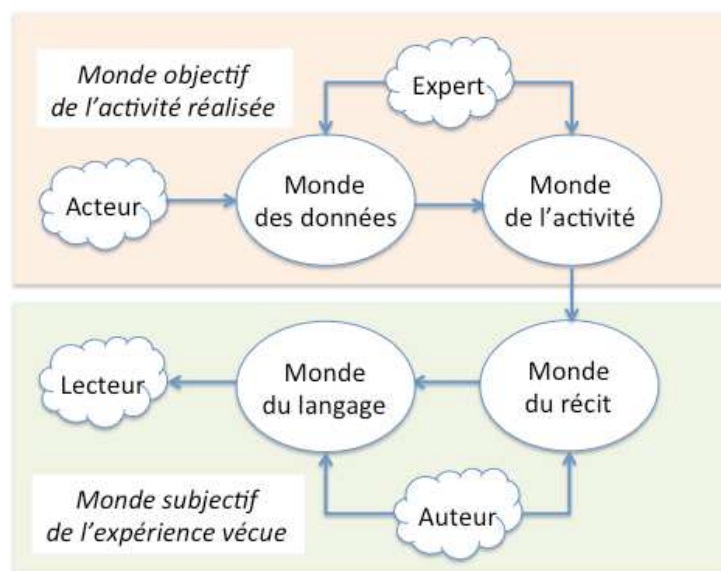


FIG. 4.6: Les mondes du récits d'activité

Ces deux mondes représentent deux points de vue différents sur une ac-

tivité humaine. Ils mettent en jeu des connaissances et sont régis par des contraintes différentes. Afin de mettre cela en évidence, et de mieux le formaliser, nous proposons de nous appuyer sur des modèles (figure 4.7). Ceux-ci permettent de représenter et d'organiser les connaissances en respectant les spécificités des points de vue. De plus, lorsque le modèle est appuyé sur une notation formelle, il peut être utilisé comme support pour le traitement automatique. Le monde des données est alors composé d'un modèle de données (données sémantiquement liées au ski de randonnée) et d'instances de données (données produites lors de la réalisation d'une sortie particulière). La même distinction entre connaissances générales (sous forme de modèles) et connaissances liées à une sortie particulière (niveau instance) s'applique au monde de l'activité, au monde du récit et au monde du langage.

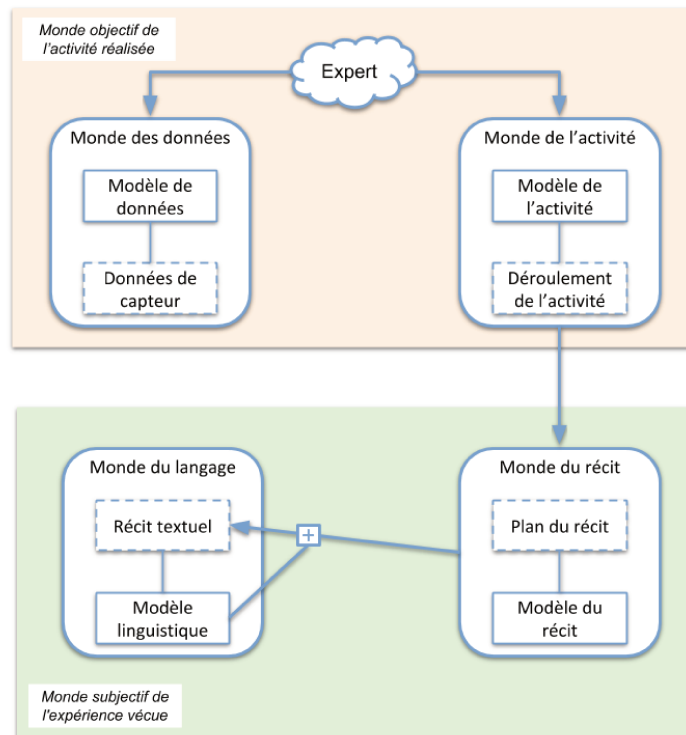


FIG. 4.7: Modèles et instances des mondes du récits d'activités

L'utilisation des modèles des données, de l'activité, du récit et de la linguistique nous permet de définir des "passages" entre les différents mondes du récit d'activités :

- des données vers un déroulement d'événements (interprétation des données)

- d'un déroulement d'événements à un plan du récit (transposition du monde de l'activité réalisée au monde de l'expérience vécue)
- du plan du récit au texte (génération "de surface" en langage naturel)

Les passages sont définis au niveau modèle par des transformations (génération ou tissage) qui sont ensuite applicables à partir d'un jeu de données acquises lors d'une randonnée à ski.

## 4.4.2 Verrous scientifiques

La mise en place de cette approche de génération de récit repose sur deux étapes :

- définir les modèles
- définir les transformations entre les modèles

### 4.4.2.1 Verrous liés à la modélisation

La réalisation de modèles constitue l'activité de modélisation. Celle-ci a pour but d'exprimer des connaissances en suivant une notation afin de construire un modèle. Dans la suite nous distinguerons les notations des modèles eux-mêmes afin de préciser où se situent les contributions de ce travail. Outre les enjeux de modélisation propres à chacun des deux mondes (monde objectif de l'activité, monde subjectif de l'expérience vécue), un enjeu particulier est celui de leurs relations mutuelles : comment concevoir des modèles d'activité cohérents vis-à-vis des données et soutenant une transposition sous forme de récit ? Comment concevoir des modèles de récit cohérents vis-à-vis de l'activité réalisée et soutenant une transcription linguistique de l'expérience vécue ?

**Élaboration des modèles de l'activité humaine** Les modèles de données et de l'activité représentent les connaissances sur l'activité réalisée (le ski de randonnée dans notre cas). Le premier verrou est la modélisation des données et de l'activité par les experts en modélisation. Celle-ci n'est pas facile [Amaro *et al.*, 2014], en particulier en raison du "gap sémantique" entre les données capteurs et leur interprétation en termes d'activité : il s'agit d'aboutir à des modèles reflétant les composants de l'activité, leurs propriétés et leurs relations, contenant tous les concepts nécessaires à l'activité de ski de randonnée et soutenant sa transcription sous forme de récit.

**Définition du modèle du récit d'activité** Comme nous l'avons dit précédemment, le récit d'activité est un type de récit pour lequel tous les

concepts des autres récits ne s'appliquent pas nécessairement. Le second verrou porte sur la définition d'un modèle du récit pour le récit d'activité. Dans cette thèse, nous chercherons à modéliser en priorité les aspects liés au plan du récit afin de nous focaliser sur le déroulement vécu. L'enjeu est la transposition réalisée par ce modèle entre le monde de l'activité réalisée et celui de l'expérience vécue. Il est aussi de soutenir la transcription linguistique.

**Élaboration du modèle du récit d'activité** De la même manière que pour les modèles de données et d'activité, la définition de la notation pour exprimer le récit d'activité ne suffit pas à la réalisation d'un modèle. Comme montré dans la section 2.2.1, les plans de récit d'activité sont construits en fonction du but communicationnel. Les buts communicationnels sont dépendants du domaine et s'ils guident le processus de création du plan du récit, ils ne sont pas explicites. Ainsi, un problème est l'identification des buts communicationnel et des choix de création du modèle de récit d'activité en fonction de ces buts.

#### 4.4.2.2 Verrous liés à la transformation des modèles

Une fois les modèles réalisés, il faut pouvoir définir les transformations qui permettent d'aboutir au récit d'activité textuel. Les étapes suivantes sont à considérer.

**Lier les données objectives (numériques et captées) aux modèles** La quantité de données produites dans l'activité peut être considérable, ce qui demande des traitements pour reconnaître quelles sont celles qui sont utiles pour la production du récit et comment les regrouper et les interpréter pour leur donner une sémantique et les associer aux concepts du modèle d'activité.

#### **Transformation des modèles de l'activité en modèles de l'expérience**

Le passage du monde objectif au monde subjectif demande une sélection et une restructuration des données d'activité à inclure dans le plan du récit. Cette transformation implique de lier les concepts entre les deux mondes mais également d'établir des règles de transformation pour passer de l'un à l'autre (et donc de connaître les paramètres à prendre en compte pour pouvoir appliquer ces règles). En particulier, l'ordre de présentation des événements dans un récit répond à un besoin d'intrigue qu'il faut chercher à faire émerger des données captées.

**Enrichir les modèles avec des informations subjectives** L'enregistrement des données capteurs ne collecte pas l'ensemble du vécu de l'acteur. Il existe des informations hors de la portée des dispositifs électroniques,

comme les pensées ou les opinions. Identifier des moyens pour produire (même de manière limitée pour les raisons citées plus haut) un point de vue subjectif sur les données objectives demande d'ajouter des connaissances à celles exprimées dans les modèles de l'activité.

**Choix du modèle linguistique** Lorsqu'une personne lit un récit racontant la réalisation de l'activité humaine, elle a des attentes. Les satisfaire peut faciliter le partage de l'expérience vécue. Pour cela, le mécanisme de présentation textuelle du récit d'activité doit considérer le type de lecteur visé afin de produire un texte qui essaie de satisfaire ses besoins et limitations (le choix d'un vocabulaire trop technique par exemple peut induire des difficultés de compréhension).

## 4.5 Démarche de conduite de thèse

Notre approche de génération de récit d'activité nécessite l'expression de connaissances dans plusieurs domaines : celui des données, de l'activité (ici, le ski de randonnée) et de la narratologie. Ces connaissances sont exprimées dans des modèles. Différentes méthodes permettent à des experts d'acquérir des connaissances sur le domaine de l'activité auprès des acteurs eux-mêmes. Des techniques d'entretiens (comme la technique des incidents critiques...) permettent d'identifier des procédures par un dialogue entre l'expert et l'acteur de l'activité. L'activité alors recueillie est celle interprétée par l'acteur qui en fait un récit (partiel) oral en répondant aux questions. Les processus de création de récits à l'oral et à l'écrit étant différents [Nonnon, 2000], nous avons choisi d'initier les modélisations à partir de ressources textuelles. Nous nous reposons sur des textes écrits par des randonneurs après leurs sorties. Ceux-ci sont soit extraits du site web *skitour.fr* soit produits à notre demande par des randonneurs. De plus, tout comme les travaux présentés dans le chapitre 2, nous nous appuyons sur l'analyse de corpus textuels pour définir le modèle du récit d'activité. Trois corpus ont donc été constitués (voir chapitre 5) et sont utilisés pour la réalisation des modèles des données, de l'activité et du récit. Ils ont également permis de concevoir les modèles de génération de textes. Par ailleurs, cette thèse nécessite de lier des données brutes de capteurs à des récits concernant ces données. Cependant, à notre connaissance, il n'existe aucun corpus disponibles de données capteurs et de récits alignés. L'acquisition d'un tel corpus fait donc partie intégrante de la démarche de cette thèse. Ainsi un des corpus a été constitué de données captées lors de randonnées à ski et de récits correspondants, produits par leurs acteurs.

Tous les travaux réalisés sur la modélisation sont basés sur l'analyse des

différents corpus. Différentes itérations ont été nécessaires pour aboutir à des modèles utilisables dans notre approche. Les premières itérations concernant les modèles de données et de l'activité avaient pour objectif la production de modèles cohérents vis à vis de l'activité (des retours informels de rando-neurs nous ont permis de les structurer). Puis, des itérations ont eu lieu pour les rendre conformes l'un par rapport à l'autre et par rapport aux données qui pouvaient être collectées via des capteurs (la construction des ces modèles est détaillée dans le chapitre 6). Les itérations apportées à la notation d'un modèle du récit ont été réalisées pour adapter une notation existante à notre type de récit et, dans un second temps, pour permettre l'identification des liens entre les concepts du modèle du récit et les données objectives de l'activité (Chapitre 7). Chacune des étapes de modélisation et de transformation a donc été réalisée par analyses successives de récits produits par des acteurs de l'activité et de connaissances apportées par la capture de données d'activités. Enfin, nous avons mené conjointement la résolution "théorique" des verrous soulevés avec le développement logiciel pour la mise en œuvre de notre approche (Chapitre 8). Ces deux types d'études en parallèle nous ont amenés à apporter des modifications supplémentaires dans les modèles et les transformations afin de préciser les notions modélisées pour l'automatisation du traitement.

# Chapitre 5

## Constitution de corpus

Comme nous l’avons présenté dans le chapitre 4, notre démarche s’appuie sur l’étude conjointe des connaissances contenues dans des données et dans des récits de sortie (textes écrits par des skieurs pour raconter leur expérience). Pour permettre l’acquisition de connaissances nous avons constitué trois corpus. Ceux-ci sont respectivement utilisés pour les tâches :

1. Identifier les événements et acteurs des récits de ski de randonnées existants ainsi que les différents déroulement possibles d’une sortie.
2. Identifier les caractéristiques linguistique et communicationnelle du récit d’activité lorsqu’il est utilisé pour raconter une expérience de ski de randonnée (vocabulaire, structure...)
3. Associer les informations extractibles des capteurs et de ressources externes aux événements rapportés dans un récit d’activités afin de choisir les algorithmes d’inférence.

Afin de baser les analyses des deux premières tâches sur des données les plus réalistes possibles, les deux premiers corpus sont extraits de sites web utilisés par les skieurs eux-mêmes sur des sorties passées. La première section présente la constitution de ces deux corpus textuels. La seconde section est dédiée au troisième corpus, appelé “corpus parallèle”, pour la troisième tâche qui contient deux types de données (données de capteurs et récits textuels) que nous avons du créer à partir d’une expérimentation humaine.

### 5.1 Constitution des corpus textuels

L’utilisation de récits textuels produits par des skieurs (sans demande de notre part) comme source de données est un choix fait pour deux raisons. La première est que nous voulions générer des récits textuels en nous basant sur

ce que l'humain produit (démarche courante dans les approches de génération de texte). La seconde est d'acquérir des connaissances sur l'activité plus générale. En effet, la collecte peut être fait selon d'autres méthodes, comme l'observation *in situ*, la réalisation d'entretiens... Cependant, ces méthodes donnent une vue de l'activité propre à un skieur particulier (il faut donc dépenser beaucoup de temps à acquérir des sources venant de skieurs différents) et présentée dans le but de l'étude (but communicationnel faussé). Plusieurs sites permettent aux skieurs d'échanger des informations sur leurs sorties. Nous avons utilisé le site de *skitour.fr* pour construire nos corpus. Nous exposons dans la première section les raisons de ce choix, puis dans les deux sections qui suivront les caractéristiques des corpus textuels.

### 5.1.1 Choix de la source des corpus textuels

Les corpus textuels ont pour objectif d'être utilisés pour permettre l'identification des concepts du domaine de ski de randonnées et des concepts de récits d'activité du point de vue des skieurs eux-même. Pour cela, les corpus doivent :

- être constitués de récits produits par les acteurs de l'activité, en français
- porter sur des randonnées à ski (lien avec le domaine)

Afin de faciliter l'analyse de ces corpus, notamment pour l'identification des concepts, nous avons fait le choix de ne considérer que les récits portant sur des sorties réalisées dans les Alpes. En effet, les skieurs étant des passionnés, ils utilisent des points de repères (comme "la bergerie" ou des lieux-dits) que les autres skieurs de la même région connaissent. Afin de pouvoir nous appuyer sur les experts du domaine à Grenoble (pratiquant principalement dans les Alpes) nous avons choisi de limiter la zone géographique des randonnées considérées aux massifs proches de Grenoble (Chartreuse, Vercors, Belledonne, Écrins)

Il existe un corpus textuel en partie en Français accessible sur le ski dans les Alpes. Ce corpus, réalisé par l'université de Zurich est appelé « Text+Berg digital » [Göhring et Volk, 2011]. Il s'agit d'un corpus parallèle français-allemand de plus de 4 millions de mots. Le sous-corpus français est nommé « Text+Berg-Korpus R151v01 Écho des Alpes ». La Figure 5.1 montre le résultat d'une recherche du mot 'ski' dans l'interface d'interrogation du site web et renvoie toutes les occurrences contenant ce mot. Il est composé de textes annotés appartenant au Club Alpin Suisse (CAS)<sup>1</sup>. Il présente des informations autour de la culture, de l'environnement et des activités effectuées

---

<sup>1</sup><https://cqpweb.linguistik.uzh.ch/>



dans les Alpes Suisses. La plate-forme en ligne permet de faire de recherches à partir d'un mot, afin de récupérer les collocations pour observer le contexte où il apparaît.

Your query "ski" returned 799 matches in 269 different texts (in 7,454,070 words [4,577 texts]; frequency: 107.19 instances per million words)

lc << >> > | Show Page: 1 | KWIC View | Show in corpus order

No	Filename	Solution 751 to 799	Page 16 / 16
751	<a href="#">1920_fr_a28</a>	section ait eu un résultat magique ? Peut-être ! A côté du <b>ski</b> qui tend à devenir le « maître de l' heure » ,	
752	<a href="#">1917_fr_a43</a>	Toutes deux se dérobent . Dans le premier cas , le <b>ski</b> comme le pied la pénètre et s' appuie sur la couche sous-jacente	
753	<a href="#">1924_fr_a223</a>	à peu , notamment , lorsqu' on aura trouvé un modèle de <b>ski</b> donnant les avantages du drib , tout en supprimant les inconvénients	
754	<a href="#">1921_fr_a44</a>	Concours de skis de SI-Cergues du 20 février . Les champions du <b>ski</b> à Gstaad . N° 774. Portrait du lieutenant-col . Messner , président	
755	<a href="#">1899_fr_a1</a>	pas un grimpeur , pas un elubiste ! UNE PARTIE DK <b>SKI</b> AU MONT-ROSE Quant à nous , impérissable sera le souvenir que laisseront	
756	<a href="#">1922_fr_a124</a>	publie la vue d' un original tremplin pour le saut en <b>ski</b> . Ce tremplin , le plus élevé connu , se trouve à	
757	<a href="#">1904_fr_a37</a>	tin ) , W. Baumann . Le S.A.C. et le sport du <b>ski</b> , Dr C. Frey . Encore quelques remarques du Dr A. Nilssen	
758	<a href="#">1922_fr_a186</a>	avaient , depuis longtemps , préconisé cette mesure . Courses estivales de <b>ski</b> au Col de la Jungfrau Ces courses ont eu lieu avec grand	
759	<a href="#">1906_fr_a71</a>	plaisirs impatientement attendus . Vive la luge , le patin , le <b>ski</b> ! Vive la neige ! Ils s' en donnent à cœur joie	
760	<a href="#">1921_fr_a86</a>	l' occasion de nos courses de premier printemps , les fervents du <b>ski</b> se sont groupés en une association spéciale , « le Ski-Club Moléson	
761	<a href="#">1924_fr_a223</a>	la montagne des skieurs déjà au courant de la technique habituelle du <b>ski</b> . Ils n' ont donc aucune raison d' organiser des concours tels	
762	<a href="#">1917_fr_a76</a>	Nul ne pouvait mieux qu' eux faire comprendre au public que le <b>ski</b> est aussi bien à la fois une science qu' un art .	
763	<a href="#">1916_fr_a28</a>	deux amis et moi , d' aller faire une course en <b>ski</b> . Le communiqué de la station météorologique du 5 février mentionnait un	
764	<a href="#">1916_fr_a27</a>	plus élevés de la Suisse où l'on puisse monter si facilement en <b>ski</b> . Le vent soufflait violemment et nous cherchâmes un abri dans les	
765	<a href="#">1913_fr_a54</a>	Ski . Non encordés . A voulu continuer malgré faiblesse . <b>Ski</b> . Départ temps douteux . » » t > » » »	
766	<a href="#">1906_fr_a84</a>	Mahtix . L' hiver en Norvège . Karl Roll . Excursion en <b>ski</b> au Snehaetta . Henry Hokk . Le développement des courses de ski	

FIG. 5.1: Capture d'écran du site de consultation du corpus « Text+Berg-Korpus R151v01 Écho des Alpes » du projet « Text+Berg digital » [Göhring et Volk, 2011]

Les textes de ce corpus ne concernent pas uniquement l'activité de ski de randonnée. En outre, en plus des récits, il est constitué de données textuelles variables (rapports, contes...), sans possibilité d'extraire un sous-corpus constitué uniquement de récits d'activité. Enfin, certains des textes sont anciens. Or, tant le domaine d'activité que les modes d'expression des skieurs évoluent au cours du temps. Pour ces trois raisons nous avons choisi de constituer de nouveaux corpus composés uniquement de récits d'activité produits en français, par des skieurs de randonnée ayant fait des sorties dans les Alpes.

Nous avons cherché des sources spécialisées sur Internet, en regardant s'ils offraient des textes liés à notre cas d'usage. Il se trouve que les sites spécialisés en ski de randonnée prolifèrent sur le net. Nous pouvons y trouver des itinéraires, des conseils ainsi que des forums sur lesquels les gens peuvent exprimer leur retour d'expérience. Certains sites ne proposent pas aux skieurs d'exprimer leur avis sous la forme d'un texte libre (comme [www.espace-skiderando.com](#) et dans une moindre mesure [camptocamp.org](#)), or c'est dans ces espaces 'libres' que les skieurs produisent des récits de leur sortie. Nous avons donc choisi d'extraire nos corpus textuels du site [skitour.fr](#) dont l'interface exposée Figure 5.2 détaille les différentes informations dispo-

nibles pour une sortie.

*Skitour.fr* est le site de référence de la communauté de skieurs francophones. Les sorties sont décrites avec des informations sur l'itinéraire (fichier .gpx, points de repère, durée du parcours, variantes), sur les conditions (date, participants, météo, altitude) mais aussi avec des informations sur l'expérience vécue comme des remarques sur certains points de l'itinéraire, des liens avec d'autres sorties relatées ou un résumé de la sortie produit par le skieur.

Plusieurs skieurs profitent de cet espace pour rendre public le récit de leur expérience pendant la randonnée. Bien que les indications techniques aident à comprendre les conditions pendant la sortie, exprimer le ressenti participe au partage de l'activité. Le récit du vécu est un point de vue appartenant au skieur qui permet le suivi du parcours, en même temps que la mise en valeur de certains aspects particuliers pour le skieur. Dans le contexte du ski de randonnée, ce point de vue peut aider à communiquer certains dangers observés qui ont provoqué une réaction ou une décision (par exemple, abandonner la route prévue). Grâce à ce point de vue, il est possible aussi de savoir le niveau d'expertise requis pour faire ce parcours, selon la description des manœuvres effectuées. Finalement, il expose le ressenti sur la sortie dans son ensemble (satisfaction, points positifs de l'itinéraire...). Ainsi, il s'agit d'un commentaire enrichi où la personne partage son interprétation personnelle de l'activité avec d'autres. Et pour ce ressenti, l'outil le plus utilisé est le récit. Ces récits constituent donc un ensemble de données idéal pour notre cas.

En raison de la quantité des récits disponibles (le site propose environ 5000 itinéraires en France), nous avons limité la recherche des récits aux sommets populaires autour Grenoble. Ce choix favorise l'accès à des experts pouvant nous aider à identifier les itinéraires et la signification du vocabulaire appartenant à la communauté. Il facilite également la collecte de données capteurs sur place et le repérage dans les données numériques des endroits, accidents ou points les plus fréquemment mentionnés dans le corpus.

**SKITOUR**  
cent pour cent ski rando

ACCUEIL | ACTU | TOPO GUIDES | MATERIEL | ANNONCES | PARTIR | RESSOURCES | PHOTOS | FORUMS | CONNEXION

Accueil > Topo Guide > Sommets > Chartreuse > Charmant Som > par les Cottaves > Sortie du 27 janvier 2015

Compléter/Modifier le topo [?]

**Charmant Som, par les Cottaves** 1867 m

**Données techniques**

Massif : Chartreuse  
Secteur : Chartreuse  
Orientation : E  
Dénivelé : 780 m

**Difficulté de montée [?] :** F  
Difficulté ski [?] : 2.3 E1  
Pente : Facile

**No de jours :** 1  
Type : Aller-Retour

**Cartes :** 233MCT

**Météo**  
RAS

**Informations pratiques**

Par fort enneigement, il peut être compliqué de se garer aux Cottaves (pas de place), se garer alors sur la route du col, à l'embranchement de la route des Cottaves, et monter au mieux dans la forêt pour rejoindre l'itinéraire normal vers 1200.

**Variantes :**

- par Mauboucheit (D+ 800m - S(2) 2) : A 1150 prendre la piste à droite et la suivre jusqu'à 1400, ou elle s'arrête au dessus des barres. Monter alors O dans des petites arêtes roides (pénible) jusqu'au sommet de Mauboucheit (1113). Longer la crête vers le Sud-Ouest sur 30 mètres pour trouver un passage qui permet de descendre sur le col de Mauboucheit, ou l'on retrouve l'itinéraire classique.
- descende par les Nivoles (D+ 780m - S(4) 2,3) : Descendez par l'itinéraire classique jusqu'au col de Mauboucheit puis basculez N pour rejoindre une belle combe sous les falaises de Mauboucheit. On rejoint le bas de la voie des dalles.

Sortie(s) GPS associée(s) : 21.01.17 28.12.12 30.01.19

**Conditions récentes**

- Plus de sortie les 15 derniers jours
- Afficher toutes les sorties liées
- Voir les conditions d'enneigement du massif

**Charmant Som, par les Cottaves [?]** : Sortie du mardi 27 janvier 2015 par alexis0

Participants : Lulu55, Emile

Membres associés : David Zip (06 01), Alex (05 01), Sabine (08 01), Raphaël Joubert (09 01), nicola (08 01)

**Commentaire de la sortie**

Prévu	Actuel	Observations
Brouillard principalement, mers	Clair	
Colonne	Côte	1807-1830 S 13h00
Conditions d'accès	Pneu/rapaz	1830-1850 SE 15cm de fraîche sur environ 60
Etat de la route	Forêt	1500-1200 SE Poudre lourde
Éloigné à partir de l'intersection pour les Cottaves	Élois 10 et 15cm de fraîche sur tout l'itinéraire, hormis sur l'arête, de bonnes conditions donc même si on a du mal à se tacher car on est encore à l'affût du moindre caillou.	
Altitude de parking :	Élois 10 et 15cm de fraîche sur tout l'itinéraire, hormis sur l'arête, de bonnes conditions donc même si on a du mal à se tacher car on est encore à l'affût du moindre caillou.	
1200m	Altitude de chausage (montée) : 1000m	
	Altitude de déchausage (descente) : 1000m	
	Activité avalancheuse observée : RAS	
	Sévérité [?]	★★★★

**Résumé de la sortie/Remarques diverses**

Itinéraire suivi : Topo - Dénivelé réel effectif : 800m

Les petites ballades en Chartreuse commencent à me manquer... Avec sa belle forêt, ses petits étangs mais surtout sa belle neige, il est vrai que la Chartreuse a ses atouts. Résultats aujourd'hui un petit Charmant Som sur ce pas de montagne de rochers un peu avec les OVA... et plus tard dans la semaine on se signera sûrement pour une autre rando, sûrement plus longue, dans ce massif !

Ça commence à être bon en Chartreuse : vive l'hiver !

**Photos**

**Commentaires**

- Par Jubb05 le mardi 27 janvier 2015 à 19:51  
Lâchez vous pas trop quand même, y'en a qui bossent. 😊
- Par alexis0 le mardi 27 janvier 2015 à 19:52  
Avec ce qui va tomber dans les prochains jours tu pourras en profiter ce week-end 😊
- Par Jubb05 le mardi 27 janvier 2015 à 20:09  
Si t'as encore peur des cailloux, faut aller tout droit... Pour rayer le ski dans le sens de la carre... 😊
- Par alexis0 le mardi 27 janvier 2015 à 20:23  
C'est exactement ce que je me dis en descente 😊
- Par Joubert le jeudi 29 janvier 2015 à 05:49  
Fait le 28 avec tour par la combe NO.

Seuls les membres peuvent poster des commentaires sur cette sortie

Bilan technique de la sortie (Information sur le sommet, itinéraire et trace GPS mise sur une carte géographique)

Des remarques ou variantes proposées

Indication des conditions récentes (Dernières sorties et conditions d'enneigement)

Météo, conditions d'accès et conditions de neige

Zone de texte libre, éventuellement accompagnée de photos

Commentaires des autres membres de la communauté par rapport à la sortie

FIG. 5.2: Fiche rapportant la sortie du 27 janvier 2015 vers le sommet Charmant Som par la route des Cottaves. (<http://www.skitour.fr/sorties/charmant-som,61537.html>) [Date de capture : 27/07/2017]

### 5.1.2 Corpus fréquentiel

Le corpus fréquentiel a pour objectif d'apporter une source de connaissances au niveau langagier, c'est à dire, comment un avis est transmis aux autres individus au travers de mots et phrases. Vu la quantité de textes appartenant au Corpus textuel de modélisation, celui-ci ne peut pas assurer une représentativité suffisante pour recueillir ces informations. Ainsi, il a été nécessaire de constituer un deuxième corpus textuel. Les textes qui le com-

posent ont été extraits de la même section de la fiche que ceux du Corpus textuel de modélisation. Néanmoins, vu l’objectif établi pour la nouvelle collecte, il a fallu mettre en place une approche différente où la récupération de textes soit effectuée plus efficacement. Ainsi, au lieu de les sélectionner manuellement et afin d’avoir une représentativité adéquate, nous avons décidé de mettre en place une extraction automatique à partir du site `skitour.fr`.

Pour ce faire, nous avons utilisé la librairie *open-source Jsoup*<sup>2</sup>. Développée en Java, elle permet d’analyser la structure d’un site web, de manière à manipuler ou à récupérer son contenu. Nous avons développé un script qui permet d’explorer les sorties proposées, de trouver la section concernée et de collecter le texte souhaité, en prenant soin de filtrer des entrées vides, c’est à dire des parcours où il n’y a pas de commentaire en texte libre.

Vu la quantité de sorties disponibles sur le site, nous avons limité l’extraction. Pour ce corpus, nous avons récupéré uniquement des sorties faites vers des sommets des départements français de l’Isère et de Haute Savoie sur la période 2009 - 2016.

Au contraire des textes appartenant au Corpus textuel de modélisation, ici il n’a pas été établi un critère spécifique pour la sélection. Ce corpus est destiné à être utilisé dans le cadre de recherches quantitatives. Cela veut dire que le seul objectif visé est d’avoir assez de matériel pour observer une connaissance du langage propre à l’activité. Ainsi, ce corpus de grande échelle ne considère pas, par exemple, la taille ou le nombre de mots utilisés dans chaque texte. Cela implique également la possibilité de trouver différents types de textes dans le corpus.

Par exemple, la figure 5.3 présente deux textes inclus dans le Corpus fréquentiel. Le texte de la Figure 5.3a est un commentaire concernant une sortie vers le sommet du Pic de la Belle Étoile (Massif de Belledonne). Il s’agit d’un avis très bref sur la qualité de la neige.

Le corpus fréquentiel apporte des connaissances au niveau lexical, grammatical et sur la cohésion interne du récit. Au niveau lexical, nous pourrions observer la fréquence d’usage de certains mots pour décrire un événement ou un fait, et identifier des termes propres au domaine. Concernant l’aspect grammatical, nous pourrions noter l’emploi des pronoms personnels ou des temps de conjugaison, par exemple. Ce corpus sert aussi pour analyser la syntaxe utilisée par les skieurs, par exemple s’ils décrivent leur ressenti avec des phrases simples (comprenant un seul verbe) ou complexes (comprenant plusieurs verbes conjugués). Finalement, nous pourrions observer la cohésion interne du texte, observable via l’usage des connecteurs. Pour ces objectifs, le premier texte de la figure 5.3a apporte des connaissances au même titre

---

<sup>2</sup><https://github.com/jhy/jsoup/>

que celui de la Figure 5.3b qui lui, est un récit. Dans le texte de la figure 5.3a nous pouvons observer les différents termes utilisés pour la description. Les mots “croulée” et “transfo” servent comme adjectifs qui indiquent des états particuliers de la neige. Ainsi, nous notons qu’il existe un certain type de vocabulaire et d’expressions techniques propres au domaine, qui aident à mieux comprendre les conditions de certaines sorties. Dans le récit de la figure 5.3b, nous pouvons remarquer en plus l’usage des connecteurs pour indiquer la progression du parcours, tels que ‘après’ ou ‘puis’. Il est possible aussi d’observer la composition des phrases, notamment l’emploi des pronoms personnels (comme ‘nous’) et des verbes employés au présent. Enfin, il est possible d’observer que l’auteur de cette sortie utilise différents types de structures de phrases : sans verbe (comme “Première descente”) et avec verbe (“nous remontons au col de Bellefont”). Un autre aspect important est la mention des noms propres qui font référence aux sommets et points remarquables fréquemment visités par les skieurs lorsqu’ils font une sortie, par exemple “col de Bellefont”. Finalement, nous pouvons observer que le style dans lequel la sortie a été rédigée est assez informel et comporte même des fautes de frappe et d’orthographe et des mots incomplets.

*La neige dans tous ses états. Globalement croulée, bonne poudreuse en versant nord et ouest abrités du vent. Transfo en dessous de 2000m pour la descente (13h00)*

(a) Sortie du 09 mars 2008 vers le sommet pic de la Belle Étoile de Belledonne (<http://www.skitour.fr/sorties/pic-de-la-belle-etoile,11679.html>, date de consultation : 14/03/18)

*Partis du col de Marcieu ou il y avait déjà quelques voitures, nous suivons les nombreuses traces jusqu’à l’Aulp du Sueil. Puis direction les Lances de Malissard dans une poudre légère à souhait. Première descente du sommet le long des barres rocheses à l’ombre. De la nous remontons au col de Bellefont et rejoignons les lances à pied par les arêtes. éme descente tout aussi bonne. Après cel la descente du couloir en virgule n’est pas fabuleuse, surtout qu’il y a des bouloches dans la deuxième partie du couloir...*

(b) Sortie du 21 janvier 2009 vers le sommet Lances de Malissard (<http://www.skitour.fr/sorties/lances-de-malissard,16674.html>, date de consultation : 14/03/18)

FIG. 5.3: Textes appartenant au Corpus fréquentiel, tels qu’ils se trouvent sur le site `skitour.fr`

Nous avons extrait un total de 2 840 textes contenant 16 592 phrases et 307 456 mots. Le tableau 5.1 présente une synthèse du contenu de ce corpus.

Sommet	Massif	Nbre de Textes	Nbre de Mots	Nbre de Phrases
Aiguille de la Grande Sassièr	Alpes grées	41	7344	368
Chamechaude	Chartreuse	643	52880	2832
Charmant Som	Chartreuse	514	57347	3014
Croix de Belledonne	Belledonne	191	25964	1385
Grand Moretan	Belledonne	77	8912	486
Grand pic de Belledonne	Belledonne	34	5329	289
Grand Som	Chartreuse	382	38339	2173
Grande Moucherolle	Vercors	15	1029	52
Lances de Malissard	Chartreuse	273	30269	1646
La Pyramide	Belledonne	18	2119	130
La Sure	Vercors	11	783	44
Pic Bayle	Grandes Rousses	37	4025	190
Pic de la Belle Etoile	Belledonne	492	51241	2913
Rochers du Midi	Chartreuse	33	4036	195
Soeur Agathe	Vercors	18	1261	62
Traversée de Belledonne	Belledonne	44	13769	655
Traversée de la Chartreuse	Chartreuse	6	493	27
Traversée du Vercors	Vercors	11	2316	131

TAB. 5.1: Synthèse des textes du Corpus fréquentiel

### 5.1.3 Corpus textuel de modélisation de l'activité et de la structure textuelle du corpus annoté

L'objectif principal de ce corpus est d'être le support à une analyse pour identifier la manière avec laquelle les skieurs décrivent le parcours réalisé et expriment leurs impressions sur le déroulement. Pour que ce corpus ne soit constitué **que** de récits dans lesquels les skieurs donnent des informations tant sur l'itinéraire que sur son ressenti, il a été nécessaire de sélectionner les textes du corpus un par un.

Par ce mode de constitution, le nombre de textes du corpus est limité, c'est pourquoi nous avons cherché à le rendre le plus consistant possible sur le domaine. Nous avons choisi de ne considérer que les sorties vers le Charmant Som (alt. 1,867 m), Chamechaude (alt. 2,082 m) (tous deux situés dans le massif de la Chartreuse) et le Grand Eulier (alt. 2,232 m) (situé dans la chaîne de Belledonne) sur la période 2011-2014. Les récits sélectionnés respectent quatre critères :

- Description la plus complète possible du parcours (en incluant le départ, l'itinéraire, la descente et le retour au point du départ )
- Mention des conditions environnementales qui aident à visualiser le parcours décrit (qu'elles soient bonnes ou non)
- Mention des repères géographiques qui aident à identifier les points visités par le ou les skieurs

- Récit cohérent (suivi causal et/ou séquentiel du déroulement du récit)

Ainsi, les textes écartés du corpus comprennent :

- Les commentaires composés d'un bilan général du parcours, mais sans détails. Principalement des textes composés de quelques lignes dans lesquelles le skieur mentionne brièvement si l'itinéraire a été suivi ou non, si le sommet a été atteint et son avis global (positif ou négatif).
- Les commentaires visant à signaler un aspect en particulier, laissant le reste de la sortie caché (un passage, le sommet, une rencontre, la météo, une anecdote, etc.)
- Les comparaisons entre plusieurs sorties réalisées.

La figure 5.4 présente deux textes correspondant au même itinéraire de randonnée. La figure 5.4a montre un texte qui ne répond pas aux critères souhaités pour notre corpus. Nous pouvons observer ici que le skieur décrit en quelques lignes ses impressions générales, sans aucune trame narrative. Cette personne indique la météo, ainsi que des notes personnelles sans une relation directe avec la sortie, sans donner d'indication sur l'itinéraire. Ainsi, nous ne pouvons pas le considérer comme un récit à inclure dans notre corpus, vu le manque d'information sur la progression de l'activité et le ressenti du skieur.

La figure 5.4b montre un récit du corpus collecté. Ici, il est possible de suivre le déroulement de la randonnée à travers des changements d'état lors des déplacements. On trouve une présentation du contexte : des descriptions factuelles qui permettent aux personnes de se repérer et de visualiser les conditions météo et celles du terrain (un exemple d'indication météorologique est surligné en magenta). Il présente également des actions qui se répercutent sur la réalisation de la randonnée, comme une prise de décision par rapport à l'itinéraire ou une contrainte technique (un exemple est surligné en cyan). Finalement, il contient également des impressions personnelles sur les conditions pendant la montée ou la descente et sur les éléments non attendus, comme la rencontre avec la faune du lieu (la rencontre avec des chamois, surlignée en jaune).

*beau temps, première sortie de l'hiver en rando, première en chartreuse et c'était que du bonheur ! La montée par les cottaves évite bien la foule de marcheurs, fondeurs et lugeurs qui montent par la route. Par contre ça doit être chaud quand la neige est dure ou trop avalancheuse ; la montée est bien raide....enfin il me semble.*

(a) Texte écarté (<http://www.skitour.fr/sorties/charmant-som,53658.html>, date de consultation : 14/03/18)

*On n'a tout de même bien fait de ne pas partir pour le grand som comme prévu à la base. Vu le risque de formation de plaque, on préfère s'arrêter aux cottaves et rester un maximum dans la forêt. **Bon le -19°C au parking** nous fait hésiter mais en même temps on n'est pas venu jusqu'ici pour rien. On met un gros rythme d'entrée de jeu histoire de se réchauffer les doigts. Une fois les premiers onglés passés, la montée se passe sans encombre. On se cale derrière le dernier rocher avant l'attaque finale pour se camoufler avec les gros gants, le casque, la doudoune. **Comme ça on n'a plus qu'à enlever les peaux au sommet.** Pari réussi, malgré le vent et les -30 °C en ressenti au sommet **malgré le vent et les -30 °C en ressenti au sommet,** pas d'onglé!!! La descente est franchement pas terrible jusqu'à l'attaque de la forêt (neige croutée) mais franchement excellente sous les arbres où la neige n'a pas vu le soleil!!! **Les trois chamois qu'on a croisés on l'air d'être du même avis ;-)***

(b) Texte conservé (<http://www.skitour.fr/sorties/charmant-som,38284.html>, date de consultation : 14/03/18)

FIG. 5.4: Deux exemples de récits provenant du site `skitour.fr` qui décrivent deux sorties faites par le même itinéraire pour atteindre le sommet du Charmant Som (route par les Cottaves).

Le corpus textuel de modélisation a été réalisé en deux temps. Une première version, Corpus1v1, comportait 11 récits qui ont servi à une première analyse (v1 des modèles de données et d'activité - voir le chapitre 6). Cinq textes supplémentaires ont été ajoutés dans une seconde analyse (Corpus1v2) visant à évaluer les modèles issus de la première analyse. La table 5.2 présente une synthèse du corpus.

## 5.2 Constitution du Corpus *parallèle*

Dans le cadre de cette thèse, il est nécessaire d'observer comment l'activité se déroule en termes de données vis à vis des informations présentées dans les récits. Nous avons donc besoin d'un corpus constitué, pour une même sortie, des données capteurs et du récit d'activité produit par le skieur. Comme nous l'avons présenté dans la section 5.1.1, le skieur peut rendre accessible



N° de texte	ID	Auteur	Route	Sommet	Nbre de Mots	Nbre de Phrases	Date de la sortie
1	35813.xml	Bernard T..	Route du Charmant Som	Charmant Som	112	7	08/12/2011
2	37304.xml	tiamus	Route par les Cottaves	Charmant Som	115	11	10/01/2012
3	38025.xml	viervier	Combe de l'If en boucle	Charmant Som	470	32	02/02/2012
4	38284.xml	tiamus	Route par les Cottaves	Charmant Som	170	10	07/02/2012
5	38809.xml	Antonin17	Route du Charmant Som	Charmant Som	91	7	18/02/2012
6	44954.xml	Frédéric Grimbert	Versant Ouest	Chamechaude	273	12	16/01/2013
7	44987.xml	Toz	Combe de l'If en boucle	Charmant Som	363	28	16/01/2013
8	44999.xml	taramont	Boucle par Canaple	Charmant Som	207	3	16/01/2013
9	45189.xml	frederic bertrumé	Boucle par couloir NO et Bergerie	Charmant Som	276	25	21/01/2013
10	45545.xml	TiBougnat	Combe de l'If en boucle	Charmant Som	251	21	21/01/2013
11	45715.xml	Toz	Couloir de l'Oursière	Grand Eulier	303	21	30/01/2013
12	46259.xml	padre1501	Boucle par couloir NO et Bergerie	Charmant Som	511	27	09/02/2013
13	48511.xml	el-Manu	Boucle de Chamechine	Charmant Som	391	13	21/03/2013
14	53945.xml	Philoupi	Boucle par Canaple	Charmant Som	156	16	31/01/2014
15	54085.xml	sale_gosse	Boucle par couloir NO et Bergerie	Charmant Som	284	12	02/02/2014
16	54085.xml	Papusa	Route par les Cottaves	Charmant Som	141	6	11/02/2014

TAB. 5.2: Bilan du Corpus textuel de modélisation

son ressenti sur `skitour.fr` mais le site ne lui permet pas de télécharger et stocker des données collectées par les dispositifs connectés (autres que des données gpx). Par conséquent, nous ne pouvons pas récupérer de données associant données et récits via une extraction de données sur ce site web.

### 5.2.1 Protocole de collecte de données

Nous avons donc mené une collecte de ces deux types de données. Chaque entrée du corpus contient des enregistrements faits par des dispositifs dotés de capteurs portés par les skieurs pendant une sortie de ski de randonnée et un récit narrant la sortie, écrit par ceux qui ont porté les capteurs. Ce corpus est utilisé pour les étapes de la démarche nécessitant l'analyse du lien entre les données capteurs et textuelles, c'est à dire pour :

- observer comment l'activité se déroule
- servir comme référence pour expliquer des aspects dans les données (par exemple, expliquer certains manques dans les données)
- essayer d'identifier des étapes du déroulement et conditions dans les enregistrements (par exemple, trouver où un skieur a fait une pause).
- tester la démarche de génération avec les données collectées, en ayant les récits parallèles comme référents textuels.

Nous avons remis à chaque participant des dispositifs équipés de capteurs, un questionnaire et un guide à consulter pour la collecte. Celui-ci devait être consulté la veille du départ et contient des informations concernant les capteurs (démarrage, manipulation...) ainsi que les consignes pour chaque étape de la collecte afin de garantir la bonne qualité de données récoltées.

Le jour de la sortie, juste avant le début de la sortie, le participant doit allumer les capteurs et les placer sur lui. Une fois les capteurs démarrés, le participant peut commencer la randonnée. Aucune manipulation des dispositifs n'est nécessaire pendant la sortie. Les capteurs doivent rester allumés tout au long de la randonnée. Une fois la randonnée terminée, le participant doit arrêter l'enregistrement et éteindre les capteurs (les instructions sont aussi dans le guide).

Pour finaliser la collecte, il est demandé au participant de compléter un questionnaire le jour même de la sortie. Celui-ci sert à recueillir des informations plus précises autour de la sortie, telles que la date et le lieu du parcours, les points d'intérêt et les problèmes rencontrés (pendant la randonnée ou par rapport aux capteurs). De même, il est demandé au participant d'écrire le récit narrant la sortie enregistrée.

Deux dispositifs de capture de données aux caractéristiques complémentaires ont été utilisés : un dispositif spécialisé, le *Zephyr BioHarness 3*<sup>3</sup> et un Smartphone équipé de l'application *RecordMe* [Blachon *et al.*, 2014].

Nous présentons dans la suite une description détaillée des capteurs utilisés pour la collecte de données. D'abord, nous avons ceux qui se trouvent dans le portable. Ils servent pour reconnaître des éléments de contexte ainsi que la réalisation de l'activité. Voici la liste avec sa respective fonction :

**Accéléromètre** Mesure l'orientation du téléphone ainsi que le déplacement du même

**Gyroscope** Détecte des changements d'accélération de rotation

**Antenne-relais** En tant qu'émetteur et récepteur d'ondes radioélectriques, permet la liaison avec une station appartenant à un opérateur de télécommunication qui couvre une zone géographique

**Wi-Fi** Connecte le portable à une réseau d'haute débit permettant l'accès à Internet

**GPS** Détecte la géolocalisation de l'appareil

**Bluetooth** Permet la transmission des données à une distance courte

**Magnétoscope** Mesure le champ magnétique

**Baromètre** Détecte la pression atmosphérique

**Capteur de proximité** Indique la présence d'objets proches à l'appareil

**Luminosité** Détecte l'intensité de la lumière sur le portable

L'accéléromètre et le gyroscope servent pour reconnaître des déplacements. L'antenne-relais et le GPS aident à indiquer des lieux visités ou passés lors des déplacements. Le *bluetooth* détecte la présence de personnes. Finalement, les capteurs restants donnent des informations autour du contexte au moment d'utiliser l'appareil. La table 5.3 montre une comparaison de fonctionnement des capteurs pour reconnaissance d'éléments de contexte sur un smartphone (voir [Blachon, 2016]).

---

<sup>3</sup><https://www.zephyranywhere.com>

	Disponibilité	Échantillonnage	Puis. Instantanée (mW)
Accéléromètre	Toujours	Régulier 50 Hz	5
Gyroscope	Modéré	Régulier 50 Hz	30
Antenne-relais	Toujours	Événementiel	5-10
Wi-Fi	Souvent	Événementiel	15-50
GPS	Souvent	Régulier 1 Hz	380-430
<i>Bluetooth</i>	Souvent	Événementiel 1 Hz	-
Luminosité	Modérée	Régulier 50 Hz	3
Magnétomètre	Souvent	Régulier 50 Hz	12
Baromètre	Modérée	Régulier 50 Hz	1
Proximité	Souvent	Régulier 50 Hz	7

TAB. 5.3: Comparaison des sources suivant des critères objectifs ([Blachon, 2016])

Le *Zephyr BioHarness 3* (figure 5.5) est composé d'un ensemble de capteurs intégré dans une ceinture qui sert comme support pour l'attacher au corps. Ce capteur enregistre les données physiologique des individus pendant une activité sportive en intérieur ou en extérieur (il supporte des températures entre  $-10^\circ$  et  $50^\circ$ ). Les données enregistrées sont :

- Fréquence cardiaque
- Respiration
- Accélération linéaire
- Posture
- Température corporelle

Les spécifications sont présentés dans le tableau 5.4.

Mesure	
Fréquence cardiaque	25-240 BPM
Fréquence respiratoire	4-70 BPM
Température	10-60°C
Accélération	+16g
Cycles de charge	300
Range de transmission	Jusqu'à 100 m <sup>2</sup>
Fréquence	2.4-2.4835GHz
Puissance de sortie	+10dBm

TAB. 5.4: Spécifications des capteurs intégrées dans le capteur *BioHarness 3*



FIG. 5.5: Le capteur *BioHarness 3*, la ceinture et comment elle doit être ajustée au corps.

Pour compléter les données biométriques avec des informations environnementales (la situation géographique pendant l'activité), nous avons utilisé des smartphones Nexus 4 et Samsung Galaxy S3 équipés de l'application *RecordMe* [Blachon *et al.*, 2014]. Cette application permet d'enregistrer en continu des données à partir des différents capteurs du portable, en respectant la vie privée de l'utilisateur. *RecordMe* permet l'enregistrement :

- des valeurs de l'accéléromètre
- de paramètres acoustiques
- de paramètres de la centrale inertielle (accélération et vitesse angulaire)
- de la luminosité
- de la pression atmosphérique
- de la localisation géographique

Enfin, le questionnaire est composé de deux parties. La première demande au skieur (qui portaient les capteurs) de rédiger un récit de la sortie. La seconde est composée de 25 questions sur l'activité :

- 1 question sur la date
- 3 questions sur l'itinéraire suivi
- 1 question sur l'expérience du skieur sur cette itinéraire
- 4 questions sur les conditions météorologiques (à différents moments clés de la sortie)
- 3 questions sur les conditions sociales (groupe) de la sortie
- 2 questions sur des événements inattendus éventuels
- 3 questions sur le skieur
- 8 questions sur l'utilisation des capteurs (en particulier les problèmes rencontrés)

## 5.2.2 Synthèse du corpus parallèle

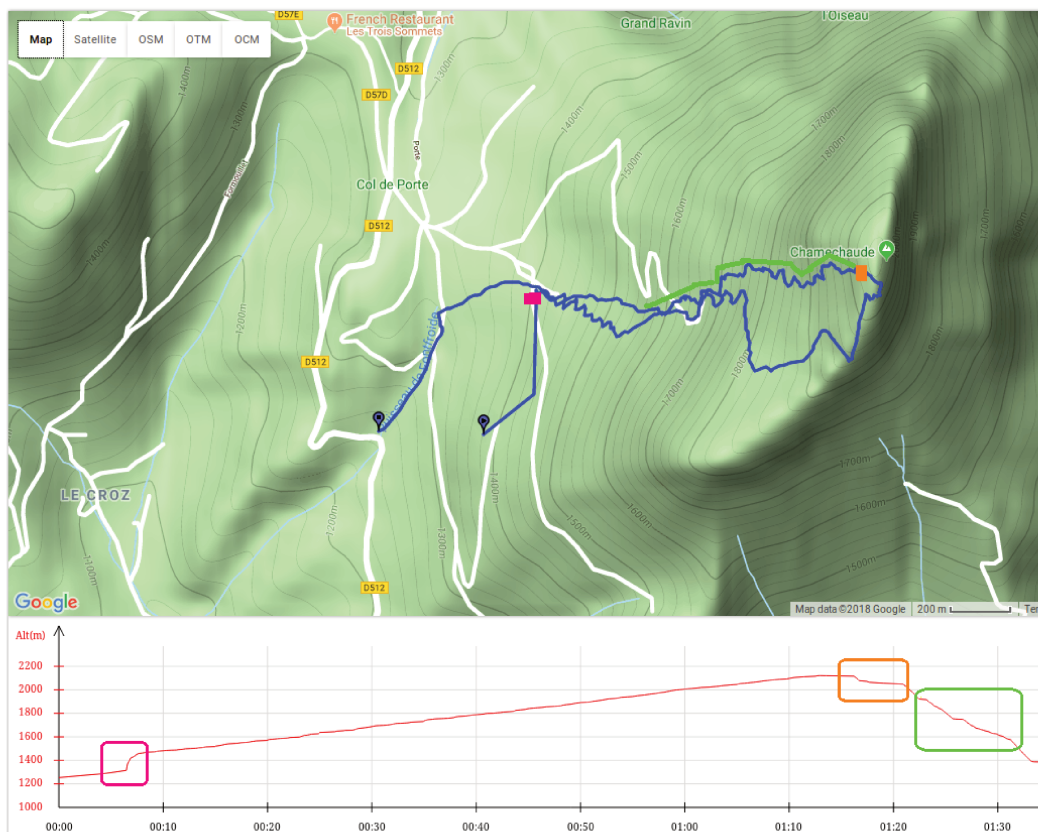
Nous avons collecté un total de 12 entrées (enregistrements et récits correspondant) sur la période 2013-2017. Elles ont été produites par trois participants sur des sommets proches de Grenoble. La table 5.5 présente la synthèse du contenu du corpus.

La figure 5.6 présente un exemple de récit produit par un participant et daté du 30/03/2016 (Enregistrement no. 10). La mise en relation avec les données captées (comme sur la figure 5.6a) permet d'expliquer certaines valeurs.

Par exemple, pour cette sortie, les données concernant le déplacement (affichées dans la figure 5.6a à partir du fichier .gpx) contiennent des trous ou des anomalies. Ici, la première, en couleur magenta, indique un moment où la progression du participant est perturbée. La deuxième, en couleur orange, signale aussi une perturbation lors de la descente, avec un arrêt pendant quelques minutes. Finalement, la troisième, de couleur verte, montre que les conditions de descente sont difficiles, et des pauses nécessaires. Ce récit parallèle (5.6b) peut être utilisé pour essayer d'expliquer ces irrégularités et nous permettre de proposer un moyen de les traiter dans notre approche. Dans le récit, nous pouvons observer qu'en effet, il y a trois moments décrits dans la sortie qui peuvent expliquer les anomalies dans le déplacement. D'abord, le participant décrit qu'au début de la montée il a trouvé une langue de névé qui lui permet de monter avec les skis malgré le manque de neige ( « Le départ est peu enneigé mais on trouve encore une langue de névé pour monter ski aux pieds » ) et sa mention semble expliquer la perturbation de la montée. Concernant la deuxième irrégularité, le récit indique qu'effectivement, le début de la descente a été interrompu à cause du manque de neige ( « Une langue de terre/cailloux barre la descente et l'on doit déchausser » ). Finalement, le récit indique que la descente a été effectuée dans un terrain difficile ( « La neige est molle et lourde ce qui m'impose quelques pauses » ), ce qui correspond à la troisième irrégularité observée.

N° d'enregistrement	Durée	Auteur	Itinéraire	Sommet	Nbre de mots dans le récit	Nbre de phrases le récit	Date de la sortie
1	7h50	C.	Boucle par le Bacheux depuis la Lavoire	Col de Merlet	672	43	10/03/2013
2	7h50	D.	Boucle par le Bacheux depuis la Lavoire	Col de Merlet	385	20	10/03/2013
3	2h30	C.	par le lac Achard	Croix de Chamrousse	432	25	17/03/2013
4	3h10	C.	par le Col de Porte	Chamechaude	362	19	24/03/2013
5	2h	F.	par Val Cenis	Col de la Maurienne	243	18	11/02/2016
6	2h15	F.	par le col de Porte	Chamechaude	78	7	11/03/2016
7	6h14	F.		Col de l'Aigleton	50	6	18/03/2016
8	2h	F.	par le Col de Porte	Chamechaude	246	17	23/03/2016
9	2h30	F.	par les pistes depuis Casserousse	Chamrousse	213	18	25/03/2016
10	2h10	F.	par le Col de Porte	Chamechaude	203	12	30/03/2016
11	6h	C.	Depuis le hameau de la Coche	Grand Rocher	411	26	22/01/2017
12	6h	D.	Depuis le hameau de la Coche	Grand Rocher	202	9	22/01/2017

TAB. 5.5: Bilan du Corpus Parallèle



(a) Extrait des données enregistrées par le participant montrant son déplacement pendant la sortie et récit correspondant

Un chamechaude de l'après midi. J'ai décidé de monter tranquillement. **Le départ est peu enneigé mais on trouve encore une langue de névé pour monter ski aux pieds.** Le ciel est couvert et le vent se fait de plus en plus présent au cours de la montée. La trace de montée est bonne et ne glisse pas bien qu'elle soit parfois bien raide. La course se fait classiquement jusqu'à la première crête que l'on atteint skis aux pieds. Le vent est assez fort sur la crête et me couche vers la crête. Après un dépeutage rapide, je décide de descendre le long de la barre. **Une langue de terre/cailloux barre la descente et l'on doit déchausser.** Entre le vent et la pente, ce n'est pas bien confortable, mais on s'en sort. Je rejoins les grandes pentes sous le paravalanche. **La neige est molle et lourde ce qui m'impose quelques poses.** J'arrive à la forêt où les conditions ne sont pas trop mauvaises. On rejoint la piste qui est relativement skiable. On doit déchausser un peu avant l'arrivée pour joindre deux pans de neige.

(b) Récit associé

FIG. 5.6: Extrait de l'enregistrement no. 10 du corpus parallèle (Sortie du 30/03/2016 vers le sommet Chamechaude)



## 5.3 Synthèse des corpus

Dans ce chapitre nous avons présenté la constitution et le contenu des corpus. Ces collections représentent la source de connaissances pour la démarche que nous proposons. Pour produire les modèles à utiliser dans notre approche, nous avons constitué trois corpus autour de notre cas d'étude. Le tableau 5.6 en fait une synthèse. Tous les corpus sont composés d'au moins un récit textuel (le corpus parallèle contient en plus des données capteurs). Ces récits montrent la combinaison des concepts du domaine avec des éléments subjectifs, ainsi que la manière dans laquelle ceux-ci sont placés dans le texte (corpus textuel de modélisation). De même, ces récits exposent la manière dans laquelle l'expérience vécue est exprimée en mots (corpus fréquentiel). Finalement, grâce au corpus parallèle, il a été possible d'observer le type de données obtenus pendant l'activité. Le premier corpus, nommé *Corpus fréquentiel*, est composé de 2867 textes qui permettent de connaître le vocabulaire de l'activité. Le deuxième corpus, nommée *Corpus textuel de modélisation de l'activité et de la structure textuelle*, contient 16 textes racontant comment une sortie de ski de randonnée s'est passée. Avec celui-ci, nous pouvons récupérer des informations sur le domaine (concepts et déroulement d'une randonnée à ski) mais aussi sur la manière d'exprimer le ressenti du skieur (ordre des événements, cohésion entre phrases, etc.). Finalement, nous avons construit un corpus *parallèle* composé de 12 enregistrements de données produits par des skieurs de randonnée portant des capteurs (un capteur biométrique et un smartphone) associés aux récits correspondants. Ce dernier corpus vise à observer le passage du monde objectif au monde subjectif de l'activité et permet d'avoir de vraies données pour tester la démarche.

Nom du corpus	Objectif	Données	Nbre d'entrée
Corpus textuel de modélisation	<ul style="list-style-type: none"> <li>- identifier les concepts et actions caractéristiques pour les skieurs</li> <li>- identifier les structures des récits produits</li> </ul>	récits issus de <code>skitour.fr</code>	16 textes
Corpus fréquentiel	<ul style="list-style-type: none"> <li>- identifier les éléments lexicaux et syntaxiques pour le domaine</li> </ul>	textes issus de <code>skitour.fr</code>	2 869
Corpus parallèle	<ul style="list-style-type: none"> <li>- connaître le lien entre données capteurs et événements du récit</li> <li>- servir de source de test pour l'implémentation</li> </ul>	données capteurs, récit libre et informations subjectives sur certains points de la sortie	12 textes + données capteurs

TAB. 5.6: Synthèse des corpus constitués

## Chapitre 6

# Modélisation du point de vue objectif et génération d'une séquence d'activité

Le chapitre précédent présente la constitution des trois corpus que nous avons utilisé pour notre étude. Ceux-ci contiennent des informations sur l'expérience vécue pendant une sortie de randonnée à ski. Pour modéliser les connaissances sur le domaine, c'est à dire les données et l'activité de randonnée, nous avons analysé le corpus textuel de modélisation. Les analyses réalisées sur ce corpus permettent l'identification des concepts et actions associés à la randonnée (du point de vue des skieurs). Ce chapitre présente la construction des modèles de données (section 6.2) et d'activité (section 6.3) à partir de l'analyse du corpus textuel de modélisation (section 6.1) ainsi que leur utilisation dans notre démarche de génération de récit (peuplement et construction d'une séquence d'activité avec des données issues de données collectées via des capteurs) (section 6.4).

### 6.1 Extraction des connaissances sur le domaine à partir du corpus textuel de modélisation

Le corpus textuel de modélisation a été constitué en deux temps. La première version (corpus textuel de modélisation, v1) contient 11 récits et la seconde version (corpus textuel de modélisation, v2) en contient 16 (les 11 de corpus textuel de modélisation v1 + 5 autres). Les analyses sur le corpus textuel de modélisation v1 nous ont permis d'établir les premières

versions des analyses alors que corpus textuel de modélisation v2 a été utilisé pour en vérifier la couverture. Les analyses et les connaissances acquises pour chacune sont synthétisées dans le tableau 6.5. Nous précisons dans la suite quelle version a fourni les données présentées.

Extraire des connaissances sur l'activité à partir des expériences relatées par les acteurs de l'activité est une tâche réalisée dans le cadre de la conception de systèmes interactifs. Nous avons utilisé sur les textes du corpus textuel de modélisation v1 la méthode proposée par [Sebillotte, 1991] pour l'analyse d'entretiens retranscrits. Cette méthode permet d'identifier <sup>1</sup> :

- les objectifs
- les plans (listes des sous-objectifs pour atteindre un objectif)
- les actes (ou tâches)
- les liens entre les actes (causalité, précédence...)
- les paramètres qui conditionnent les actes. Pour le ski de randonnées, ces paramètres sont par exemple les acteurs ou les conditions météorologiques.
- les exceptions par rapport au déroulement normal

L'utilisation de cette méthode sur des récits libres (c'est à dire dont la structure et le contenu ne sont pas conditionnés par les objectifs de l'interviewer) nous a permis d'identifier 7 actes qui sont listés dans le Tableau 6.1.

Dans leurs récits, lorsque les skieurs décrivent une action (un acte), ils apportent de l'information pour aider à mieux comprendre ce qu'ils ont vécu ou fait. Les extraits reportés dans le Tableau 6.2 montre que ces descriptions peuvent inclure des informations ponctuelles, comme la position géographique ou les objets utilisés. Par exemple, les peaux<sup>2</sup> pour les skis. C'est à partir de ces informations que nous avons constitué les *paramètres* des actes. Nous avons construit des catégories en regroupant les informations ayant des caractéristiques (ou traits) en commun. Par exemple, nous avons créé l'objet *météo* en réunissant tous les types d'informations associés à ce concept. Dans notre exemple, nous avons inclus *la température*, *la pluie*, *la chute de neige*, *la présence du vent* et *l'ensoleillement*. Treize paramètres décrivant l'environnement ou les conditions de réalisation des actes ont été ainsi identifiés et sont reportés dans le Tableau 6.3.

Le tableau 6.2 met en évidence que pour chaque acte, le temps (sous forme de durée ou ponctuelle), la localisation et le(s) acteur(s) sont des paramètres indiqués par les auteurs de récits. En dehors de ces paramètres communs à

---

<sup>1</sup>La terminologie employée dans ce manuscrit n'est pas celle de [Sebillotte, 1991], elle a été adaptée au vocabulaire employé dans ces travaux

<sup>2</sup>les peaux (ou peaux de phoque) désignent le tissu collé à la semelle du ski pour qu'il ne glisse que dans un sens, permettant alors de remonter la pente.

<b>Actes trouvés dans le corpus textuel de modélisation</b>	
<b>Texte du corpus</b>	<b>Actes</b>
<i>Partis à 7h20 du parking</i> (44954.txt)	Aller au point du départ
<i>Montée à bon rythme au sommet du Charmant Som</i> (38025.txt)	Avancer
<i>Pause thé</i> (54085.txt)	Faire une pause
<i>Tonio a même voulu embrasser un sapin qui lui a foutu une gifle</i> (46259.txt)	Subir un incident
<i>Il fait déjà bien chaud avec 1°C au thermomètre de la voiture</i> (54494.txt)	Évaluer les conditions de randonnée
<i>Lassés, on décide de forcer le passage skis sur le sac, et finalement à pied ça ne va pas si mal</i> (38025.txt)	Prendre une décision
<i>On retrouve les randonneurs de la voie normale et notre ami Alain et sa compagne de Skalp</i> (45545.txt)	Faire des rencontres

TAB. 6.1: Extraits de textes illustrant les actes identifiés par l'analyse du corpus textuel de modélisation v1

tous les actes, les actes *Faire une pause*, *Subir un incident*, *Évaluer les conditions de randonnée*, *Prendre une décision* et *Faire des rencontres* contiennent des ensembles de paramètres distincts les uns des autres. *Aller au point de départ* et *Avancer* contiennent les mêmes paramètres mais la localisation d'*Aller au point de départ* est l'un des points de départ des randonnées (connus par les experts du domaine). Comme notre corpus textuel de modélisation contient uniquement des récits de l'ascension du sommet Chamechaude, nous pouvons différencier *Aller au point de départ* de *Avancer* car les différents points de départs possibles sont connus et limités. En conséquence, nous pouvons extraire des textes, les actes et leurs paramètres. Cette possibilité nous a permis de nous adapter plus facilement aux styles littéraires des skieurs. En effet, pour aider à mieux visualiser les événements racontés, les skieurs ont l'habitude d'employer des mots et expressions subjectifs, particulièrement des figures de style tels que les métaphores, l'exagération, l'appel aux sens, et des références

<b>Paramètres des Actes trouvés dans le corpus textuel de modélisation</b>	
<b>Actes</b>	<b>Paramètres</b>
<i>Aller au point de départ</i>	Mouvement, Localisation départ, Niveau de difficulté, Humain, Localisation arrivée, Équipement, Durée
<i>Avancer</i>	Mouvement, Localisation départ, Niveau de difficulté, Humain, Localisation arrivée, Équipement, Durée
<i>Faire une pause</i>	Pause, Humain, Localisation, Durée
<i>Subir un incident</i>	Humain, Problème, Durée, Localisation
<i>Évaluer les conditions de randonnées</i>	Humain, Localisation, Temps, Météo, Condition Physiologique, Terrain
<i>Prendre une décision</i>	Humain, Décision, Temps, Localisation
<i>Faire des rencontres</i>	Êtres vivants rencontrés, Localisation, Temps, Durée

TAB. 6.2: Extraits de textes illustrant les actes identifiés par l'analyse du corpus textuel de modélisation

culturelles. Pour extraire les actes à partir des paramètres de réalisation de ceux-ci, une annotation manuelle des textes du corpus textuel de modélisation v2 a été mise en place. Nous avons construit des séquences d'actes à partir de l'identification des paramètres. Pour cela, nous avons utilisé l'outil d'annotation Callisto [Day *et al.*, 2004]. Le logiciel fonctionne avec un document DTD contenant des balises *span* pour colorer des zones de texte avec leurs attributs. Dans notre cas, le fichier DTD contient parmi les étiquettes (tags), les 9 étiquettes qui permettent d'identifier les actes avec leurs paramètres (table 6.6). Nous pouvons observer un exemple d'étiquetage de l'objet "Déplacement" sur Callisto dans la figure 6.1. Le résultat est l'obtention d'un fichier .xml contenant l'ensemble des objets étiquetés en ordre chronologique (tels qu'ils apparaissent dans le texte).

Paramètres	Attributs
Niveau de difficulté	Difficulté de ski, Pente
Localisation	Élément naturel, Construction, Direction, Route, Piste
Mouvement	Avancer, Descendre, Monter
Humains	Groupe, Personne
Terrain	Neige, Trace
Conditions Physiologiques	Fatigue, Effort
Pause	Discussion, Repos, Observation, Attente, Repas
Décision	Changement de parcours, Arrêt
Météo	Vent, Pluie, Chute de neige, Ensoleillement, Température
Êtres vivants	Humains, Animaux
Équipement	Ski, Chaussures, Lunettes, GPS, Peaux de phoque
Problème	Accident, Désorientation, Problème technique, Route bloquée, Problème Physique
Temps	Durée, Moment, Instant

TAB. 6.3: Paramètres de réalisation des actes

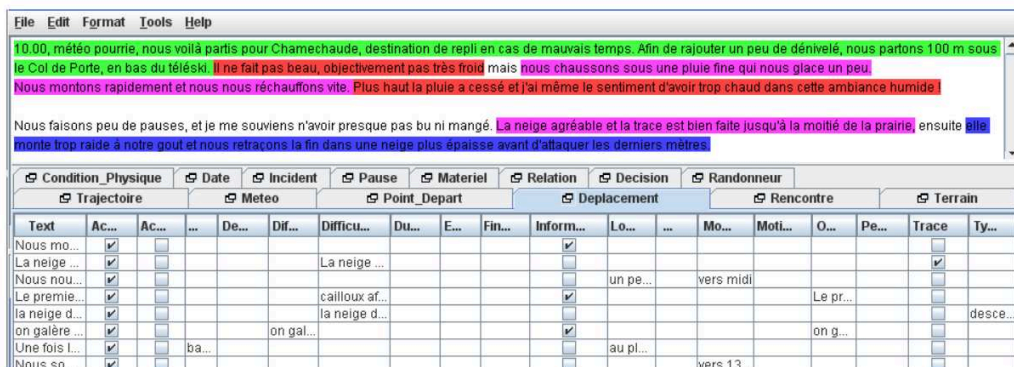


FIG. 6.1: Interface du logiciel d'annotation Callisto

Au total, nous avons obtenu 238 instances d’acte (identifiées par leurs paramètres) sur l’analyse du corpus textuel de modélisation v2. Le tableau 6.4 montre les fréquences d’instances détectées pour les actes. Nous pouvons constater que les connaissances liées au mouvement (les déplacements et les pauses) sont les plus fréquemment rapportées.

<b>Actes</b>	<b>Paramètres</b>	<b>Fréquence</b>	<b>Pourcentage</b>
Aller au point du départ	Point_Départ	11	4,6%
Avancer	Déplacement	111	46,6 %
Faire une pause	Pause	30	12,6 %
Subir un incident	Incident	1	0,4 %
Évaluer les conditions de randonnée	Terrain	36	15,5 %
	Météo	13	5,46 %
	Condition_Physique	1	0,4 %
Prendre une décision	Décision	14	5,88 %
Faire des rencontres	Rencontre	17	7,14 %

TAB. 6.4: Les annotations obtenues et leur fréquence d’apparition

À partir de cette analyse, nous avons également construit une matrice de transitions (Figure 6.2). Pour des raisons de lisibilité la figure 6.2 n’en présente que les probabilités supérieures à 0.07. Celle-ci présente la probabilité que certaines tâches se succèdent. Nous avons remarqué que les transitions avec un poids important correspondent à celles qui forment une paire avec le “déplacement”. L’observation du terrain (0.61), de la météo (0.77), la prise de décision (0.71), la mention des rencontres (0.65) et les pauses (0.47) sont toujours suivis d’un déplacement. Cela montre encore une fois que la mention du mouvement est un facteur essentiel qui guide le déroulement des événements et qui constitue donc le fil conducteur du récit.



Corpus	Analyse	Connaissances acquises
corpus textuel de modélisation v1 (11 récits)	Analyse textuelle selon [Sebillotte, 1991]	7 actes 13 paramètres (et leurs attributs)
corpus textuel de modélisation v2 (16 récits)	Analyse textuelle par annotation	16 séquences d’actes, fréquence des actes relatés et matrice de transitions

TAB. 6.5: Les analyses sur le corpus textuel de modélisation

## 6.2 Élaboration du modèle de données

Le modèle de données contient les concepts permettant de représenter l’activité indépendamment de son déroulement. C’est également dans ce modèle qu’est exprimé le lien entre le niveau sémantique et le niveau numériques des “valeurs captées”. Nous avons choisi d’exprimer le modèle de données par une ontologie basée sur la Logique de Description (DL). La section 6.2.1 présente ce modèle et la section 6.2.2 décrit la modélisation des données pour construire l’ontologie du domaine de ski de randonnée.

### 6.2.1 Les ontologies basées sur la Logique de Description (DL)

Les données de capteurs représentent des informations qui permettent de connaître objectivement le déroulement d’une activité. Toutefois, travailler avec cette source de connaissance demande de prendre en compte leur incomplétude et le manque de certitude dans les enregistrements (en raison des erreurs de captation dues au bruits et artefacts par exemple). Par exemple, des capteurs d’accélération n’informent que de manière indirecte sur l’activité des personnes. De plus certaines informations, telles que la qualité de la neige ou les raisons d’une décision ne peuvent être acquises par des capteurs simples. Ces caractéristiques propres aux données issues de capteurs posent un problème pour une modélisation du domaine qui doit répondre à la question : *Comment attribuer un sens à un ensemble de données non complet ?* Nous avons étudié les modèles existants pour représenter des connaissances qui soient capables de gérer cette incomplétude. Les approches travaillant sur l’hypothèse d’un *monde ouvert* permettent de ne pas considérer comme fausse une information non connue [Gil, 2005] et donc de faire des raisonnements sur l’information existante ou non dans la base de connaissances. Les

Acte	Étiquette	Attributs
Avancer	Déplacement	Difficulté, Localisation, Trace, Pente, Durée_Déplacement, Moment_Déplacement, Mode_Déplacement, Difficulte_Ski, Etat_Acteur, But_Déplacement, Opinion_Acteur, Motivation_Déplacement, Début_Déplacement, Fin_Déplacement, Acteur_Principal, Acteur_Secondaire
Aller au point de départ	Point_Départ	Localisation_Départ, Moment_Départ, Acteur_Principal, Acteur_Secondaire
Faire des rencontres	Rencontre	Moment_Rencontre, Personnes_Rencontre, Lieu_Rencontre, Durée_Rencontre, Acteur_Principal, Acteur_Secondaire
Evaluer les conditions de randonnée	Terrain	Moment_Terrain, Fragilité_Neige, Epaisseur_Neige, Risque_Avalanche, Cailloux, Arbres, Roche, Animal, Acteur_Principal, Acteur_Secondaire
	Condition_Physique	Tension, Etat, Acteur_Principal, Acteur_Secondaire
	Météo	Moment_Meteo, Vent, Pluie, Chute_Neige, Ensoillement, Temperature, Acteur_Principal, Acteur_Secondaire
Faire une pause	Pause	Moment, Durée, Debut, Fin, Lieu, Cause, Acteur_Principal, Acteur_Secondaire
Prendre une décision	Décision	Décision_Prise, Nouveau_Objectif, Localisation_Décision, Raison_Décision, Moment_Décision, Acteur_Principal, Acteur_Secondaire
Subir un incident	Incident	

TAB. 6.6: Étiquettes employées pour obtenir les séquences d'actes par l'annotation sur Callisto

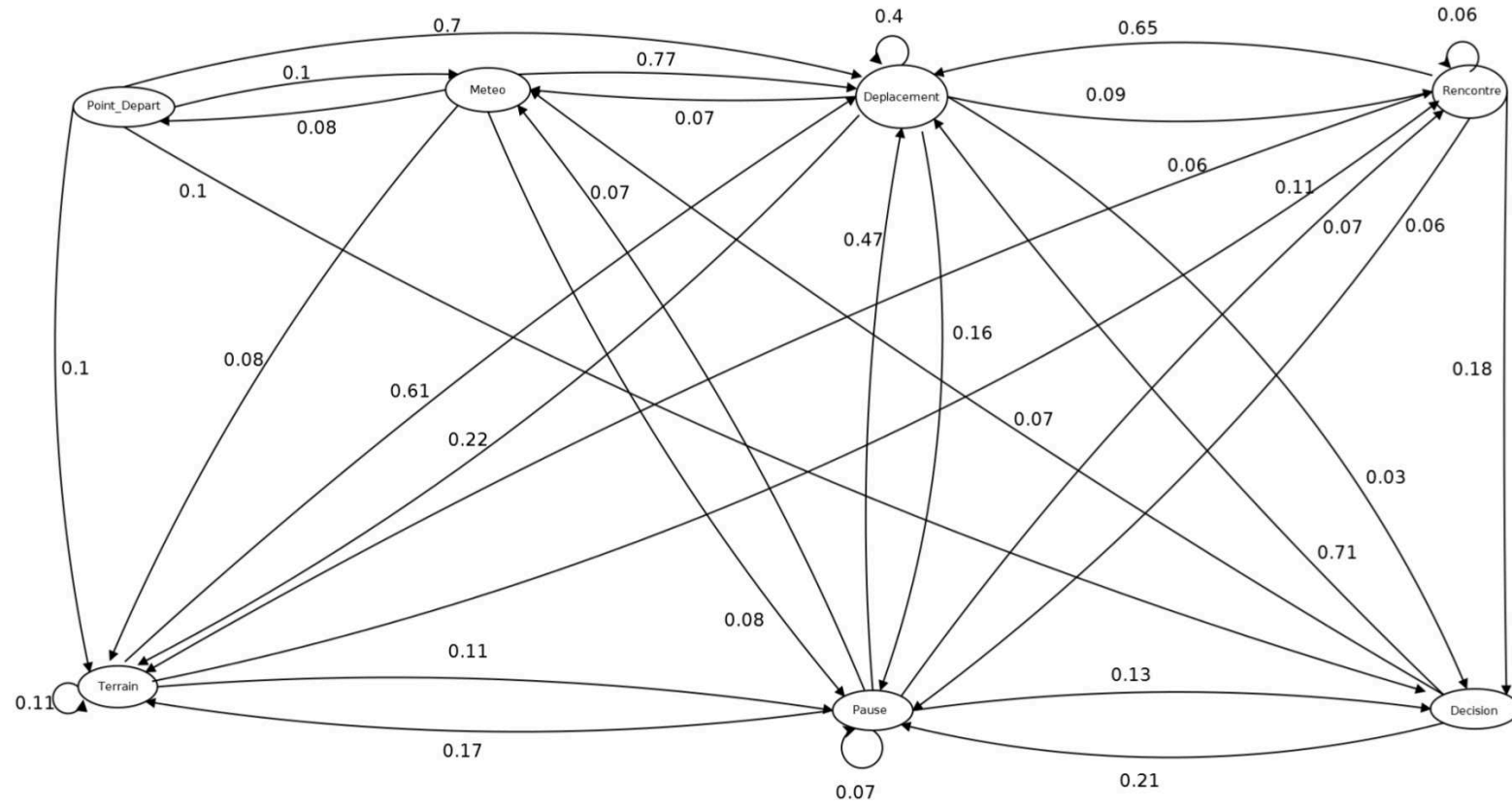


FIG. 6.2: Matrice de transitions issue des annotations

ontologies basées sur la logique de description [Rudolph, 2011] sont une de ces approches (comme OWL DL). Ce type de représentation est capable d'organiser l'information de manière structurée (sous forme d'une taxonomie de concepts) et ainsi de faire des raisonnements complexes à partir des informations connues et inconnues. De même, elles permettent la mise en place d'un vocabulaire personnalisé et réutilisable, ainsi que la définition des relations sémantiques entre les concepts.

### 6.2.1.1 La logique de description

La logique de description (*Description Logic* – DL) travaille sous l'hypothèse d'un *monde ouvert* [Rudolph, 2011]. Elle est capable de traiter des informations incomplètes, faire des inférences et faire évoluer la représentation du domaine si besoin. Ainsi, un système en DL recevant une requête est capable de faire des raisonnements complexes nécessitant plusieurs interprétations possibles [Fournier-Viger, 2005].

La DL possède une famille de langages qui permettent de représenter les connaissances d'un domaine particulier d'une manière formelle et structurée [Baader et Werner, 2002]. Chaque modèle est organisé en trois types d'entités : *concepts* (descriptions génériques), *individus* (instances des concepts) et *rôles* (propriétés qui permettent la relation entre *concepts* et entre *individus*). Sacco et Tzitzikas [Sacco et Tzitzikas, 2009] définissent un langage à travers une signature :

**Définition 6.1** Soit  $S = (O, C_a, R_a, C_{str})$  une signature où  $O$  est un ensemble d'individus,  $C_a = \{C_1, C_2, \dots\}$  un ensemble de concepts,  $R_a = \{R_1, R_2, \dots\}$ , un ensemble de propriétés qui permet la relation entre termes et  $C_{str}$  un ensemble de constructeurs permettant de créer des concepts plus complexes.

Dans le cas du ski de randonnée une signature pourrait être composée des éléments :

- Individus ( $O$ ) : *Jean, Charmant Som, Route D512, Cottaves, chute de neige...*
- Concepts ( $C_a$ ) : *Skieur, Route, Parking, Mouvement, Avancer, Manger, Neige, Personne, Auberge, Cailloux, Accident, Lac, Sommet, Météo, Tempête...*
- Constructeurs ( $C_{str}$ ) : *L'événement Avancement a-Mouvement, a-Route, a-Skieur*
- Rôle ( $R_a$ ) : *est-localisé, a-activité, a-acteur,...*

et un mécanisme d'interprétation tel que :  $I$  est une interprétation d'une signature  $S$  avec  $I = \langle \Delta^I, \cdot^I \rangle$  où :

- $\Delta^I$  est un ensemble non vide, le domaine de l'interprétation
- $\cdot^I$  est une fonction affectant :
  - un élément  $a_i^I \in \Delta^I$  à chaque individu  $a_i \in O$
  - un sous-ensemble  $C_i^I \subseteq \Delta^I$  à chaque concept atomique  $C_i \in C_a$
  - et une relation  $R_i^I \subseteq \Delta^I \times \Delta^I$  à chaque rôle atomique  $R_i \in R_a$

Une base de connaissances peut donc être définie à partir d'un langage de description et d'une signature comme ceci :

**Définition 6.2** Soit un langage de description  $\mathcal{L}$  et une signature  $S$ . Une base de connaissances  $\Sigma$  dans  $\mathcal{L}$  est un couple  $\Sigma = \langle T, A \rangle$  telle que :

- $T$  (ou *TBox*) est un ensemble fini contenant les informations terminologiques ou générales du domaine, représenté par un ensemble d'axiomes de la forme  $C_i \sqsubseteq C_j$ , ce qui veut dire que chaque instance de  $C_i$  est nécessairement une instance de  $C_j$ .  $C_i \equiv C_j$  est équivalent à  $C_i \sqsubseteq C_j$  et  $C_j \sqsubseteq C_i$ . Les formules de  $T$  sont appelées axiomes terminologiques
- $A$  (ou *ABox*) est un ensemble contenant des assertions de la forme  $o : C$  ou  $(o_i, o_j) : R$ , où  $C$  est un concept,  $R$  est un rôle, et  $o_i, o_j$  appartiennent à un ensemble d'individus  $O$ , ce qui veut dire "l'individu  $o$  appartient à  $C$ " et "il y a une relation  $R$  entre  $o_i$  et  $o_j$ ". Les formules de  $A$  sont appelées des assertions.

Pour être mise en place la logique de description nécessite l'utilisation d'un langage basé sur la logique de première ordre qui permet d'exprimer les connaissances du domaine [Baader et Werner, 2002]. À partir des connaissances exprimées dans la ABox et la TBox, les individus de la ABox peuvent être identifiés comme des instances de concepts de la TBox afin de se reposer sur les connaissances du niveau TBox ou, le cas échéant, enrichir le modèle de connaissance. La figure 6.3 présente l'architecture d'un système de représentation de connaissances basé sur la logique de description [Baader et Werner, 2002].

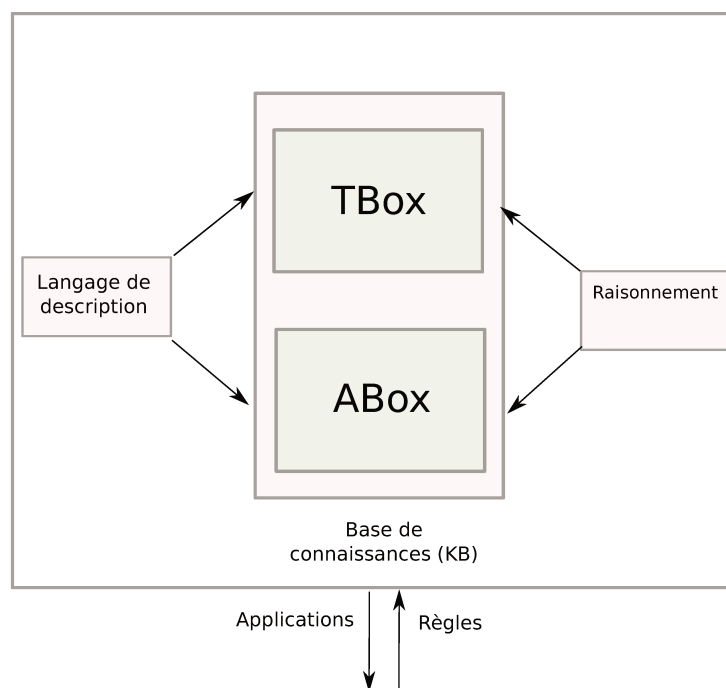


FIG. 6.3: Architecture d'un système de représentation de connaissances basé sur la Logique de Description (issue de [Baader et Werner, 2002])

### 6.2.1.2 L'ontologie pour la structuration des données

La logique de description aide à garantir l'attribution d'un sens aux données en prenant en compte leur incomplétude. L'utilisation d'une ontologie permet de structurer les connaissances sur les concepts par un ensemble de primitives qui aident à modéliser un domaine de connaissances ou un discours [Gruber, 2008]. L'utilisation d'ontologies pour exprimer l'ensemble des concepts du domaine et du récit (Chapitre 7) facilite les échanges entre ces deux modes. Une ontologie exprime de manière explicite les concepts, leur définition (ie leurs propriétés) et les relations existantes entre eux. Staab et Studer [Staab et Studer, 2004, Ahson et Ilyas, 2010] donnent la définition suivante :

**Définition 6.3** Soit une ontologie une tuple à 4 composants :

$$O = \langle C, R, I, A \rangle \quad (6.1)$$

où  $C$  est l'ensemble de concepts connus aussi comme *Classes*,  $R$  sont les relations (ou propriétés),  $I$  est l'ensemble des instances et  $A$  est l'ensemble d'axiomes (ou restrictions de rôle)

Dans le cas du ski de randonnée, une ontologie comprend les éléments suivants :

- Concepts (*C*) : *Skieur, Route, Parking, Mouvement, Avancer, Manger, Neige, Personne, Auberge, Cailloux, Accident, Lac, Sommet, Météo, Tempête...*
- Relations (*R*) : *est-localisé, a-activité, a-acteur,...*
- Instances (*I*) : *Jean, Charmant Som, Route D512, Cottaves, chute de neige...*
- Axiomes (*A*) : *Mouvement est une activité humaine, Tempête est un type de météo, Skieur est une personne...*

Il existe plusieurs langages pour formaliser les ontologies. Les RDF (*Resource Description Framework*) par exemple, expriment les connaissances sous forme de triplets sujet-prédicat-objet, mais sans aucune possibilité de faire des inférences. Le langage d'ontologies Web OWL, par contre, permet l'utilisation de connecteurs logiques et la description explicite et complexe des propriétés qui vont lier les concepts. Pour exprimer les données du domaine du ski de randonnée en tenant compte des incertitudes sur les données capteurs et en favorisant la structuration avec une description détaillée de différents niveaux de connaissances (concepts comme propriétés), nous avons choisi le langage OWL DL développé par W3C [McGuinness *et al.*, 2004, Lacy, 2005] . Il permet d'exprimer les connaissances d'un domaine en utilisant les termes de la DL décrite préalablement.

## 6.2.2 Construction de l'ontologie basée sur la logique de description pour le ski de randonnée

La construction de notre ontologie a été faite en 3 étapes :

**Identification des relations entre les éléments du domaine** Dans la section 6.1 nous avons présenté les 13 paramètres et leurs attributs identifiés par l'analyse des textes du corpus textuel de modélisation v2. En plus de la relation entre paramètres et attributs, une organisation entre les paramètres eux-mêmes permet d'identifier les liens de classification (au sens de la conception objet). Enfin, les dernières relations modélisées dans notre modèle de données visent à exprimer des liens "temporels". Les seuls liens de ce type considérés dans le modèle de données expriment que les valeurs des deux éléments sont affectées dans la même fenêtre temporelle. Par exemple, lorsqu'un skieur observe les conditions météorologiques, les éléments du domaine *météo, lieu, moment* sont liés. Ces liens sont représentés sous le concept d'*événement logique*. Nous avons identifié 5 événements :

- Activity\_Event : fenêtres temporelles définies par l'activité d'un protagoniste
- Physiology\_Event : fenêtres temporelles définies par l'état physique d'un protagoniste
- Condition\_Event : fenêtres temporelles définies par des conditions environnementales (autres que météorologiques)
- Weather\_Event : fenêtres temporelles définies par des conditions météorologiques
- Decision\_Event : fenêtres temporelles définies par une prise de décision

**Vérification des concepts par un expert du domaine** Les concepts et relations sont présentés à un expert de l'activité (ski de randonnée) pour contrôler les éléments de l'ontologie (classification et concepts/attributs). En même temps que le contrôle de ces relations, l'expert "remonte" les concepts de ski de randonnée manquants dans le modèle. Cette démarche a permis d'aboutir à l'ensemble des concepts présentés dans la figure 6.4.

**Utilisation de *Protégé* pour éditer l'ontologie** Le résultat de l'étape précédente est édité à l'aide du logiciel *Protégé* [Gennari *et al.*, 2003]. Cet outil d'édition permet de construire des ontologies à partir de l'organisation hiérarchique et de la définition des relations en utilisant le langage OWL DL. L'équivalence entre termes utilisés par la Logique de Description et *Protégé* est montrée dans le tableau 6.7. *Protégé* est doté d'un moteur d'inférence qui permet de vérifier la cohérence de l'ontologie et permet la création des individus pour peupler l'ontologie à partir des relations préalablement définies. Le résultat final est une structure de connaissances liées les unes aux autres. L'ontologie sur *Protégé* (Figure 6.5) est organisée en *classes*, groupes d'objets qui contiennent des objets avec des caractéristiques similaires. Les classes de notre ontologies sont les événements et les entités. Elles sont liées entre elles à travers les liens appelés *object\_property*. Par exemple, pour connaître les conditions météorologiques pendant un événement concernant la météo, on utilise *has\_meteo* pour mettre en relation la sous-classe *weather\_event* avec la classe *météo weather* appartenant à *entity*. Les valeurs des attributs sont exprimées par les *data property*. Elles peuvent être de divers types (entiers, chaînes de caractères, etc.). L'ontologie obtenue est composée de 96 *classes*, 11 *object property* et 28 *data property*.



Context	Difficulty Level	Difficulty_Mounting	Chapel Shelter Station
		Difficulty_Ski	
		Slope	
	Location	Bridge	
		Building	
		Bus_Stop	
	Weather	Corridor	
		Direction	
		Parking	
		Route	
		Track	
		Humidity	
		Snowfall	
		Sunshine	
		Temperature	
Wind			
Human_Activity	Movement	Advance	
		Descent	
	Static	Mount	
		Discussion	
		Earl	
Human_Physiology	Barrel	Observation	
		Rest	
		Waiting	
		Digestion	
		Drowsiness	
	Problem	Exercise	
		Exertion	
	Problem	Recovery	
		Accident	
		Disorientation	
Material_Problem			
Path_Blocked			
Physical_Problem			

Living	Animal	Group Person	
	Human		
	Plant		
Non_Living	Natural_Element	Lake	Equipment
		Peak	
		Saddle	
	Object	Spring	
		Glasses	
		GPS	
		Helmet	
		Ski_pole	
		Rope	
		Shoes	
		Ski	
		Seal_skin	
Crampon			
Material	Rock		
	Snow		
	Stones		

FIG. 6.4: Concepts de l'ontologie

Logique de description	<i>Protégé</i>
Concepts	Classes
Rôles	Propriétés
Individus	Instances

TAB. 6.7: Équivalences entre termes de la Logique de Description et l’Ontologie *Protégé*

### 6.2.3 Des données captées aux données du domaine

L’une des problématiques de notre approche est de donner une sémantique aux valeurs captées. Pour cela, des algorithmes dédiés à chaque capteur permettent d’extraire des informations symboliques des données numériques enregistrées (signaux). Par exemple, les données numériques de positions GPS sont transformées en déplacement montant ou descendant allant d’un point A à un point B (*Activity\_Event*). Ce passage à des données symboliques est réalisé en amont de l’application de notre démarche [Baez *et al.*, 2015] et ce mécanisme de transformation est appelé *abstraction*. Pour chaque abstraction une instance correspondant au concept est créée automatiquement dans l’ontologie à laquelle sont ajoutées les valeurs des attributs en fonction des traitements opérés sur les sources de données. La liaison entre le modèle et les données obtenues est faite en JAVA. Tous les noms des classes et propriétés de l’ontologie sont stockés dans le fichier *story elements.java*. Ce fichier est utilisé pour associer les abstractions issues des données aux concepts de l’ontologie. Par ailleurs *story elements.java*, contient également les paramètres et seuils des algorithmes d’abstraction. Par exemple si dans les données GPS (fichier .gpx), il y a une valeur de vitesse égale à 0, alors une instance de l’objet “Pause” est créée. La librairie JAVA *Carados*<sup>3</sup> est utilisée pour construire et manipuler l’ontologie (inclure des classes et des instances) de manière conforme à *OWL*.

La figure 6.5 présente une capture d’écran de *Protégé* montrant une instance générée à partir des données GPS. Dans la partie haute à gauche de l’interface, *Entity* et *Event* organisent les concepts dont les instances sont présentées dans la partie basse (instance appartenant à la sous-classe sélectionnée (dans la partie haute)). La partie centrale de l’interface présente les informations sur la classe à laquelle appartient l’instance (sur la figure *Acti-*

<sup>3</sup>*Carados* est une librairie construite par Dave Westwater et utilisée comme package utilitaire pour le projet *BabyTalk*[Portet *et al.*, 2009]

*vity\_Event*, sous-classe de *Event*). Enfin, à droite, le nom (génééré automatiquement à partir du nom de la sous-classe), les assertions et les propriétés de l'instance sont visibles. Les *object property*, comme *has\_actor* indiquent les relations avec les autres concepts d'*Entity* alors que les *data property*, indiquent les attributs de l'instance dont les valeurs sont issues des données enregistrées (comme la valeur 2927.0 pour l'altitude).

The screenshot displays the Protégé software interface for an ontology named 'Skitouring'. The main window is divided into several panes:

- Class hierarchy:** Shows a tree structure of classes. 'Activity\_Event' is highlighted in blue. Annotations point to 'Thinking' as a 'Sous-classes dans Entity' and 'Event' as 'Sous-classes dans Event'.
- Instances:** A list of instances for 'Activity\_Event', including 'activity\_event\_0' through 'activity\_event\_17'. 'activity\_event\_0' is selected, with an annotation pointing to the list as 'Liste d'instances issues de données'.
- Types:** Shows 'Activity\_Event' as the selected class.
- Property assertions:**
  - Object property assertions:** 'takes\_place\_at parking', 'has\_activity advance', and 'has\_actor person\_1'. An annotation points to these as 'Object property (Relations concept-concept)'.
  - Data property assertions:** 'Start\_Altitude 2927.0f', 'Start\_Time "2017-01-22T07:59:19"^^xsd:dateTime', 'Average\_Speed 5.5912707E-8f', 'Denivelee 8.0f', 'Travel\_Mode "stop"^^xsd:string', 'End\_Altitude 2919.0f', and 'End\_Time "2017-01-22T08:16:00"^^xsd:dateTime'. An annotation points to these as 'Data property (relation concept-valeur)'.

At the top right, a dropdown menu shows 'Instance de la sous-classe sélectionnée'. At the bottom, a status bar indicates 'No Reasoner set. Select a reasoner from the Reasoner menu' and 'Show Inferences' is checked.

FIG. 6.5: Interface du logiciel *Protégé* qui montre une instance de l'ontologie générée à partir de données obtenues des capteurs

## 6.3 Élaboration du modèle de l'activité

La modélisation des connaissances de l'activité sous forme d'ontologie ne permet pas de modéliser directement la manière dont les concepts interagissent entre eux pendant la réalisation de l'activité. Par exemple, si les instances des concepts *prendre une pause* et *prendre une décision* se suivent, sachant que certaines instances ne sont pas toutes découvertes (cf. incomplétude des données), l'ontologie ne contient pas la connaissance ni les mécanismes d'inférence permettant de vérifier si cette succession fait sens par rapport à l'activité de ski de randonnée. De plus, l'ontologie ne permet pas d'inférer les relations les événements (par exemple, savoir à quel moment un certain vécu, comme le passage à travers un ruisseau, a eu lieu : au début, pendant la montée ou durant la descente). Par ailleurs, il est nécessaire dans plusieurs domaines de connaître et d'exprimer le comportement des individus dans une situation particulière, par exemple pour étudier l'utilisation d'un produit ou d'un service informatique. Pour ce faire, on s'appuie sur des modèles dont l'objectif est de décrire ce que les personnes font, les contraintes et les procédures de l'activité. Parmi ces modèles, les *modèles de tâches* sont utilisés dans le domaine de la conception des interfaces interactives (IHM). Ils ont été retenus en raison de leur aptitude à représenter de manière générique et non ambiguë un ensemble très complet de propriétés de l'activité humaine, selon une perspective orientée vers les buts et les formes de déroulement de l'activité. La présence d'outils pour éditer les modèles, simuler des déroulements particuliers d'activité et vérifier leur cohérence à également joué en faveur de ce choix.

Il existent des approches de génération de récits fictifs qui utilisent des formalismes similaires à celui du modèle de tâches pour structurer le récit, notamment ceux utilisés en planification. Le plus reconnu est l'utilisation des *Hierarchical Task model* (HTN) dans la génération par interaction. Le HTN est un formalisme où chaque tâche est décomposée en sous-tâche jusqu'au actions élémentaires, lesquelles sont les feuilles du modèle. Comme ces modèles supportent une planification partielle, il sont adaptés à des intrigues qui sont construites au fur et à mesure de l'interaction des utilisateurs avec le système interactif. La tâche racine correspond au but de l'histoire qui sera achevée lorsque tous les niveaux inférieurs seront accomplis. Nous avons présenté dans l'état de l'art (Chapitre 3) comment [Cavazza *et al.*, 2007] les a utilisé pour schématiser les actions des personnages. Grâce à cette approche, il est possible d'intégrer l'expression des émotions des personnages à travers l'inclusion d'opérateurs contenant des états mentaux qui répondent à des préconditions ou des effets. Cependant, dans notre cas, il n'y a pas d'inter-

activité lors de la construction du récit et le modèle que nous cherchons à construire tient plus du domaine que du monde du récit.

La section 6.3.1 présente ce modèle qui exprime l'ensemble des déroulements d'une activité. Nous avons construit le modèle pour le ski de randonnée à partir des analyses sur les actes (identifications, fréquences et séquences) (section 6.3.2).

### 6.3.1 Les modèles de tâches pour exprimer le déroulement de l'activité

De manière analogue à la conception des IHM, nous avons besoin d'un référent formel pour identifier dans les données comment les acteurs ont réalisé leur activité. Le modèle de exprime une activité sous forme de qu'un acteur accomplit pour atteindre un but. Plusieurs notations permettent d'exprimer les modèles de en utilisant les mêmes concepts [Limbourg et Vanderdonckt, 2004]. Parmi ceux-ci, le modèle *K-MAD* (*Kernel of Model for Activity Description*) est le résultat des recherches menées pour proposer un noyau pour la modélisation de contenant tous les concepts pour décrire une activité sous une forme aussi générique que possible et de manière non ambiguë [Lucquiaud, 2005]. Dans K-MAD, une tâche est définie comme étant une *Activité nécessaire pour atteindre un but*. Quatre ensemble de caractéristiques définissent à la fois la tâche (que nous pouvons assimiler aux actes identifiés par l'analyse de corpus) et les relations qu'elle a avec les autres au cours du déroulement [Caffiau *et al.*, 2010] :

- *les informations de la tâche* : les attributs de la tâche comprenant le nom, le but assigné, sa durée et sa place dans l'arborescence
- *les caractéristiques de la tâche* : l'exécutant (qui réalise la tâche : utilisateur, système, interactif, abstrait), la modalité (tâche cognitive ou motrice), la fréquence (si elle se répète ou non ) et le degré d'importance (ce qui aide à déterminer sa priorité d'exécution dans le modèle : peu importante, importante, très importante)
- *l'ordonnancement local de la tâche* : les propriétés qui indiquent l'exécution de la tâche indépendamment de sa place dans l'arbre ou de ses sous-tâches, en indiquant sa nécessité (si elle est obligatoire ou non) son interruptibilité (si la tâche peut être arrêtée à tout moment), son itérativité (si elle peut être répétée), les acteurs concernés, les pré-conditions (conditions pour être déclenchée)
- *l'ordonnancement global de la tâche* : les liens temporels d'exécution avec les autres tâches qui décrivent une procédure pour atteindre un but commun (séquentiel, alternatif, parallèle et sans ordre).

Dans le cadre de ces travaux, nous n'utilisons pas l'ensemble de ces caractéristiques. Nous ne distinguons donc pas ce qui relève de la description de la tâche (Identifiant) de ce qui se rapporte à l'ordonnancement (optionnalité, itération, opérateur d'ordonnancement). Ce modèle permet également d'inclure des objets qui interagissent ou à utiliser pour réaliser une tâche. Un objet est une ressource nécessaire pour accomplir la tâche. Chaque objet possède un nom (texte) et des attributs qui sont définis à un niveau abstrait (par un ensemble de caractéristiques) et à un niveau concret (ensemble de valeurs attribuées aux caractéristiques). Nous utilisons les objets pour faire un lien symbolique avec les données du domaine.

Un logiciel nommé *K-MADe* [Baron *et al.*, 2006] permet l'édition de modèles de tâches, ainsi que la simulation d'un déroulement particulier de l'activité (*scénarios*). La figure 6.6 présente un exemple de modèle de tâches édité avec l'outil K-MADe (exemple issu de [Caffiau, 2009]). La tâche mère (*Envoyer un nouvel email* au niveau "Racine") est décrite avec l'exécutant du type "abstrait", car elle est constituée de plusieurs sous-tâches. Le but de la tâche mère, et donc du modèle, est d'envoyer un email. Pour atteindre ce but, la procédure est séquentiellement (opérateur de "Séquentiel") de *Créer un nouvel email*, de *Éditer* et enfin de *Envoyer email*. Ces trois sous-tâches sont réalisées dans le cadre d'une interaction entre l'acteur et le système interactif (tâches d'exécutant "Interactif").

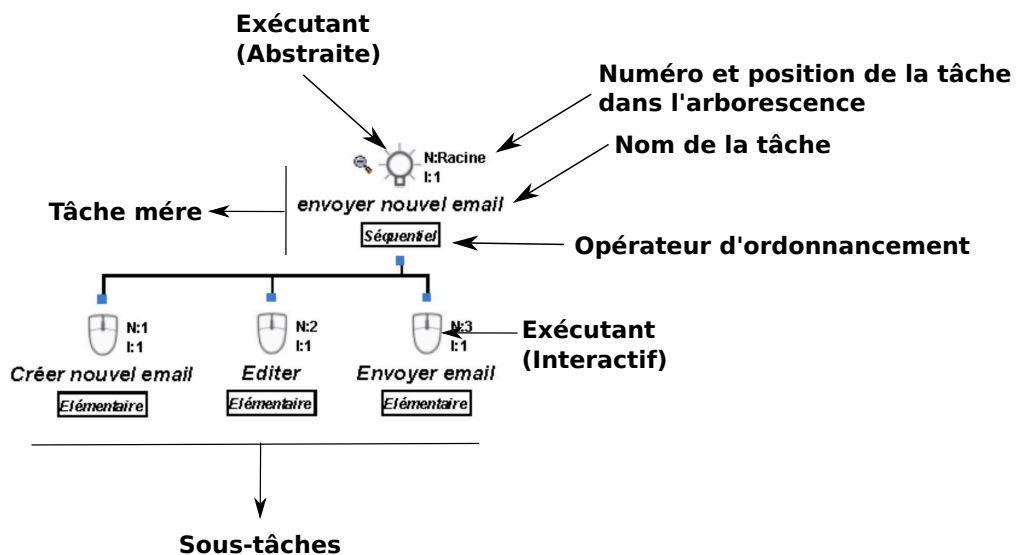


FIG. 6.6: Exemple de modèle de tâches édité avec K-MADe qui décrit l'activité "Envoyer un email" [Caffiau, 2009]

### 6.3.2 Construction du modèle de tâches pour le ski de randonnée

La construction du modèle de tâches de ski de randonnée a été faite en 5 étapes :

- 1. Définition des tâches concrètes** Les 7 actes que nous avons identifiés (cf Tableau 6.1) constituent les tâches concrètes de l'arbre. Les fréquences d'apparition des actes obtenues par l'analyse du corpus annoté permettent d'observer que les déroulements variaient beaucoup en termes de présence des tâches. Par exemple, dans un récit A, il y aura des tâches que le récit B ne mentionne pas. Donc, pour constituer le modèle contenant tous les déroulements possibles, nous utilisons les caractéristiques d'optionnalité des tâches. Pour exprimer la *nécessité* nous différencions les tâches *obligatoires* (réalisées à toutes les randonnées comme *Aller au point du départ*) et les tâches *optionnelles* dont l'apparition dans le récit est due à un déroulement particulier (*incident* qui a nécessité une *Pause* par exemple). L'étude des séquences d'actes des récits a permis également d'identifier les tâches réalisées une seule fois dans l'activité (comme *Aller au point du départ*) ainsi que celles réalisées itérativement (comme *Avancer*). En dehors de ces caractéristiques (nom, optionnalité et itération), nous n'avons utilisé aucun des autres attributs des tâches K-MAD pour les tâches concrètes. Le tableau 6.8 synthétise les caractéristiques des tâches concrètes pour le ski de randonnée.



Description de tâches en termes	
Nom de la tâche	Nécessité, itération
<i>Aller au point du départ</i>	Obligatoire, fait une seule fois
<i>Avancer</i>	Obligatoire, faite itérativement
<i>Faire une pause</i>	Optionnelle, faite itérativement
<i>Subir un incident</i>	Optionnelle, faite itérativement
<i>Évaluer les conditions de randonnée</i>	Optionnelle, faite itérativement
<i>Prendre une décision</i>	Optionnelle, faite itérativement
<i>Faire des rencontres</i>	Optionnelle, faite itérativement

TAB. 6.8: Caractéristiques de tâches concrètes

**2. Construction de l'arborescence** Les tâches concrètes composent le niveau le plus bas de l'arbre. L'étude de la matrice de transitions (Figure 6.2) nous permet de constater que les tâches *évaluer les conditions de randonnée* et *prendre une décision* sont mentionnées fréquemment ensemble et dans le même ordre, *prendre une décision* apparaissant seulement lorsqu'il est nécessaire d'interrompre ou modifier le déroulement prévu. Cette relation signifie qu'elles constituent une procédure comprenant une évaluation des conditions et une prise de décision éventuelle. Cette procédure a été regroupée dans la tâche *Analyser la situation courante*. *Faire une pause* et *Subir un incident* sont deux manières d'exprimer une interruption momentanée (*Interrompre le trajet*) dans le parcours (tâche *Avancer*). Les interruptions et *avancer* se rapportant aux déplacements, elles sont intégrées à la tâche *Faire le trajet*. Le niveau le plus haut de l'arborescence a été obtenu en regroupant l'ensemble des tâches liées au suivi de l'itinéraire (*Suivre l'itinéraire*) que nous avons placé en séquence avec l'accès au parcours (*Aller au point de départ*).

**3. Édition du modèle avec K-MADe** L'opérationnalisation du modèle de tâches pour le cas d'usage a été fait avec K-MADe. Grâce à ses fonctionnalités, nous avons édité et visualisé les caractéristiques et informations des tâches, les objets que chacune manipule ainsi que la structure hiérarchique.

- 4. Vérification par rapport aux séquences d'actes** K-MADe permet de jouer des *scénarios* (séquences de tâches) pour vérifier leur cohérence vis à vis des contraintes exprimées dans le modèle de tâches. Nous avons vérifié que les déroulements extraits de récits (corpus textuel de modélisation v2) étaient inclus dans le modèle de tâches. Pour cela, nous avons créé un ensemble de scripts qui permettent de transformer automatiquement le contenu d'un fichier *Callipso* en un fichier xml selon le format des scénario K-MADe. Bien que nous ayons construit le modèle de tâches de l'activité en utilisant les données extraites de l'analyse des textes du corpus textuel de modélisation, nous avons eu le cas de séquences invalides. Par exemple, nous avons identifié que la première tâche obligatoire de l'activité est *Aller au point du départ*, donc l'absence de cette tâche dans une séquence invalide la séquence dans son ensemble. Or, dans certains récit les auteurs ont fait une ellipse sur cette tâche, n'indiquant le point de départ que dans les données factuelles demandées dans le site. Tous les cas où la séquence était détectée comme invalide vis à vis du modèle ont été étudiés un à un afin de vérifier que la cause n'était pas due à la modélisation de l'activité. Toutes provenaient des ellipses faites par les auteurs.
- 5. Vérification par un expert du domaine** Finalement, nous avons consulté un expert du ski de randonnée. Nous avons utilisé la fonctionnalité de simulation pour avoir l'avis de l'expert sur le modèle réalisé. Grâce à son expertise, une tâche a été ajoutée, celle-ci termine l'activité (*Rentrer*). Du point de vue de la modélisation, cette tâche pouvait être vue comme un cas particulier de la tâche *Avancer* (pour laquelle le *Déplacement* menait au point de départ) mais pour l'expert, ces tâches ont une sémantique différente car le but de la tâche est uniquement de pouvoir rentrer chez soi alors qu'*Avancer* répond au but d'atteindre un objectif dans l'itinéraire (sur le parcours). Ces deux tâches ayant des buts différents, elles doivent bien apparaître comme des tâches différentes dans le modèle de tâches.

L'arborescence du modèle final est présentée Figure 6.7. À chaque tâche, des objets sont associés (basés sur les paramètres du tableau 6.3) dont les valeurs sont affectées lorsque les tâches sont réalisées (valeurs dépendantes du scénario). Les protagonistes sont eux représentés par des concepts *Acteurs* et directement liés aux tâches.

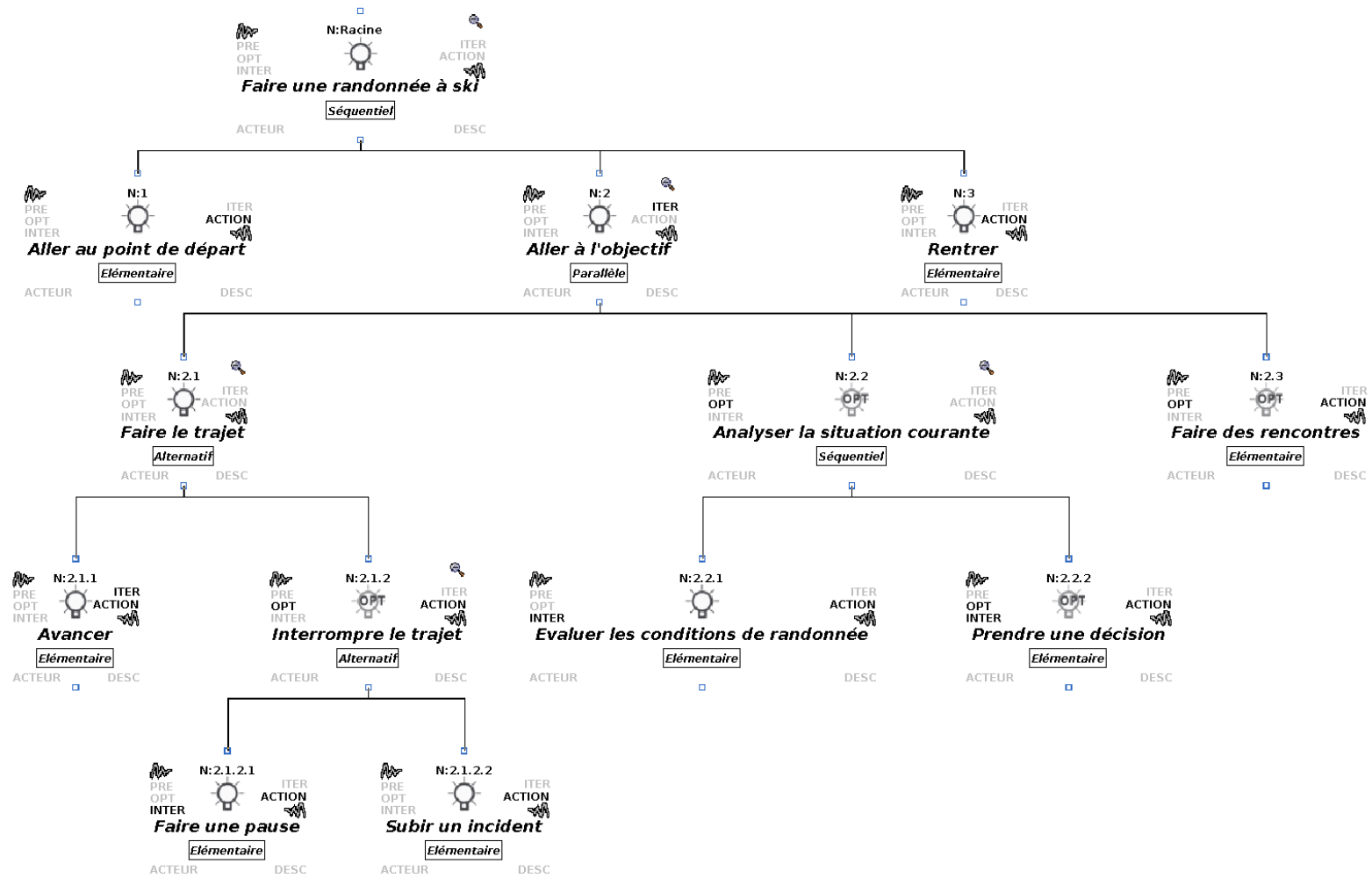


FIG. 6.7: Arborescence des tâches pour le ski de randonnée.

## 6.4 Des modèles du domaine à une séquence d'activité à rapporter

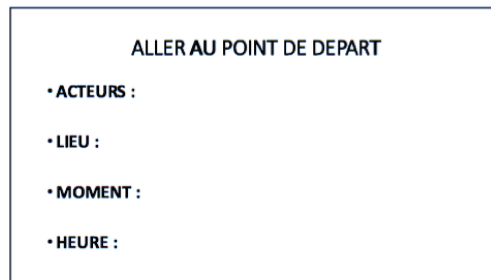
Les modèles de données et de l'activité représentent les connaissances sur l'activité et le lien entre les données captées lors de l'activité avec leur sémantique pour les acteurs. Pour pouvoir être décrit de manière à être compréhensible par le lecteur, il faut que le déroulement de l'activité soit cohérent vis à vis de ces connaissances. Nous formulons l'hypothèse que pour cela il faut :

- que les concepts, entités présents dans le récit soient du domaine
- que les événements suivent l'ordre et les contraintes temporelles de l'activité

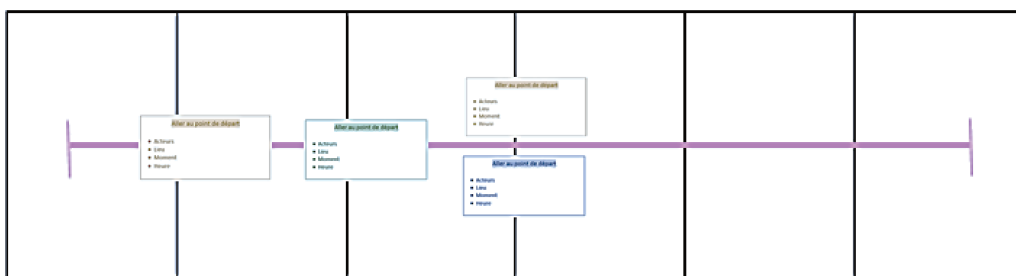
Afin de vérifier cette hypothèse, nous avons mené une évaluation de la perception d'une séquence d'activité (ie déroulement d'une sortie) décrite textuellement par des concepts du domaine à partir d'une séquence de tâches. La séquence était conforme au modèle de tâches et les concepts utilisés pour la décrire conforme au modèle de données (section 6.4.1). Les résultats obtenus lors de cette évaluation ont montré que les lecteurs perçoivent correctement les déroulements des sorties à partir de textes produits avec les connaissances de ces deux modèles. La section 6.4.2 présente la procédure de génération d'une séquence d'événements à partir de l'ontologie et du modèle de tâches que nous avons défini.

### 6.4.1 Évaluation de la perception de la séquence d'activité issue d'une séquence de tâches

Nous avons mis en place une évaluation afin d'étudier si une séquence de l'activité basée sur le modèle de tâches (c'est à dire dont les événements sont conformes aux tâches du modèle et dont l'ordre de séquence respecte les contraintes exprimées dans le modèle de tâches), lorsqu'elle était présentée sous forme textuelle à partir des concepts du modèle de données, permettait au lecteur de se représenter le déroulement de la sortie ([Baez *et al.*, 2015]). Pour ce faire, nous avons présenté des séquences d'activité sous forme textuelle et nous avons demandé à 18 personnes francophones (12 hommes et 6 femmes), âgées de 19 à 38 ans et habitants dans la région grenobloise, de reconstruire sur une *timeline* le déroulement de l'activité rapportée (cf Figure 6.8). Les éléments à placer sur la *timeline* étaient les tâches du modèle de tâches (sous forme de cartes 6.8). Les participants avaient la possibilité d'y ajouter toutes les informations qu'ils souhaitaient (cf Figure 6.9).



(a) Cartes



(b) *Timeline*

FIG. 6.8: *Timeline* et cartes pour reconstruire les séquences



FIG. 6.9: Un participant pendant une passation

Chaque participant a eu trois textes à analyser les uns après les autres, tous basés sur l'analyse de récits réels (issus du corpus textuel de modélisation v2). Les récits choisis ont été sélectionnés (par les évaluateurs) en fonction de leur taille, la complexité du déroulement décrit, la clarté de la description et la qualité linguistique ainsi que le nombre des protagonistes de la sortie. Chaque passation a duré environ 25 minutes. Les textes à analyser étaient

soit présentés dans leur version originale (tel que trouvée sur le site skitour.fr), soit issus d'une génération automatique de texte à partir des connaissances des modèles (Figure 6.10), sans que les participants ne soient informés du mode de production du texte. Le tableau 6.9 présente la répartition des textes attribués aux 18 participants.

Ils montent pour aller sommet du Charmant Som. Ils observent par rapport à l'ensoleillement qu'il y a de brume. Ils avancent le couloir NO. Ils rencontrent 2 skieurs au sommet. Les deux skieurs décident de faire demi-tour en raison de être un peu intimidés par le couloir vierge. Ils descendent. Ils observent par rapport à l'épaisseur de la neige que le fond dur et couche fraîche. Ils s'arrêtent à cause de profiter du lieu. Ils descendent jusqu'à 1450m. Ils s'arrêtent à cause de remonter le Pré Batard. Ils observent par rapport à l'ensoleillement qu'il y a de brume. Ils avancent pour aller la Combe de l'If. Ils avancent pour aller jusqu'au Nord pour avoir l'avantage d'offrir de belles vues de face du couloir. Ils s'arrêtent pour manger. Ils observent par rapport à l'ensoleillement qu'il y a de brume. Ils descendent. Ils s'arrêtent à cause de chausser. Ils observent par rapport à l'épaisseur de la neige que grosse poudre. Ils avancent sur une plaque dure en ski. Ils affrontent un décrochage de ski. Ils avancent. Ils décident de se diriger au jugé vers le habert en raison de ne pas avoir plus guère. Ils s'arrêtent à cause de remettre les peaux pour remonter au Collet. Ils observent par rapport à l'épaisseur de la neige que poudre épaisse. Ils décident de forcer le passage skis sur le sac puis à pieds en raison de être un passage difficile à trouver. Ils s'arrêtent à cause de rechausser. Ils avancent pour aller le Collet pendant 1h30 pendant 500m. Ils descendent pour aller aux Révols.

FIG. 6.10: Texte généré automatiquement

<b>Répartition des textes</b>			
<b>Participant</b>	<b>Texte 1</b>	<b>Texte 2</b>	<b>Texte 3</b>
<i>1 (mod 6)</i>	37304-gen	38809-gen	54494-original
<i>2 (mod 6)</i>	37304-original	38809-original	54494-gen
<i>3 (mod 6)</i>	37304-original	38809-gen	54494-original
<i>4 (mod 6)</i>	37304-original	38809-gen	54494-gen
<i>5 (mod 6)</i>	37304-gen	38809-original	54494-original
<i>6 (mod 6)</i>	37304-gen	38809-original	54494-gen

TAB. 6.9: Répartition des textes par participant

La distance entre les réponses du participant et les récits utilisés comme référence a été calculée à travers le “taux d’erreur de mots” (*Word Error Rate* – WER). Il s’agit d’une mesure standard du TALN notamment en reconnaissance automatique de la parole pour évaluer l’erreur (la distance d’édition) entre le mot prononcé et le mot transcrit. Ainsi la distance entre un phrase de référence et sa transcription est calculée par l’équation du WER suivante :

$$WER = \frac{S + D + I}{N} \quad (6.2)$$

où

- $S$  est le nombre de substitutions (mots de la référence remplacés par d’autres mots dans la transcription)
- $D$  est le nombre d’omissions (mots de la références non trouvés et non substitués dans la transcription)
- $I$  est le nombre d’insertions (mots dans la transcription non trouvés dans la référence et non substitués)
- $N$  est le nombre de mots total de mots prononcés

Le résultat du calcul de la distance pour les textes générés (*computer*) et les textes écrits (*human*) est présenté dans la figure 6.11. Une ANOVA a été réalisée sur les valeurs de distance. Elle montre un effet du texte, c’est à dire, un effet selon le mode de production du texte ( $F(1,18) = 7,583$ ,  $p=0,0131$ ). Ces résultats ont montré également que la distance dépend du participant ( $F(17,18)=2.281$ ,  $p=0.0457$ ). Les résultats montrent que la perception de la séquence a été meilleure avec des textes générés automatiquement qu’avec ceux issus du corpus. Les textes générés présentent un ordre chronologique plus explicite, ce qui peut expliquer que les participants ont été plus capables de percevoir la structure chronologique des événements.

Un autre résultat observé est la taille des séquences décrites (sur les *timelines*) en nombre d’événement en fonction du type de textes (Figure 6.11). Une différence sur la taille des séquences identifiées a été trouvée selon le type de récit lu ( $F(1,48)=5.604$ ,  $p=0.022$ ). La taille de la séquence spécifiée est plus importante lorsqu’elle est construite à partir d’un texte généré.

Ainsi, les résultats semblent indiquer que les textes générés introduisent moins d’erreurs de compréhension du déroulement que les textes produits par les skieurs. Les textes humains entrelacent plusieurs informations qui pourraient rendre plus difficile la compréhension de la séquence ou, au contraire, présentent des omissions qui seraient aussi un problème pour capturer correctement le déroulement.

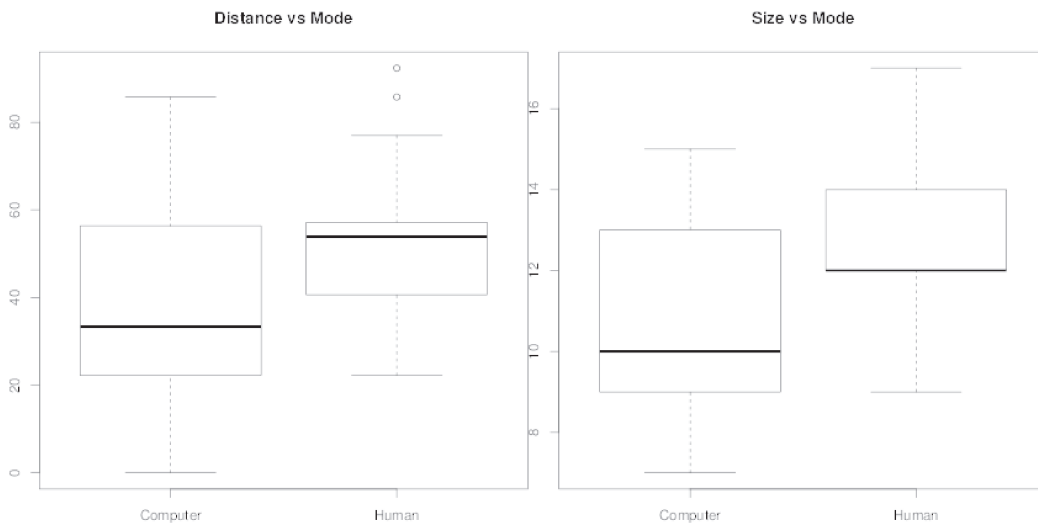


FIG. 6.11: Scores de distances et de taille en fonction du mode

Ainsi, cette étude a montré que l'utilisation d'une séquence de tâches (issue du modèle de tâches) comme support à une activité présentée dans un texte permet aux lecteurs de comprendre les étapes d'une sortie au moins aussi bien que le permettent les récits humains.

## 6.4.2 Production d'une séquence d'activité

Les données issues de la réalisation d'une activité sont des données organisées dans le temps (il est possible de savoir quand une donnée a été produite et si c'était avant, pendant ou après une autre). Pour passer de cette vue orientée vers les évolutions des données aux cours du temps en une vue centrée sur les actions (et faisant intervenir plusieurs données), c'est-à-dire organiser les données de manière cohérente non plus uniquement vis à vis du temps mais également vis à vis de l'activité, nous proposons d'utiliser les connaissances exprimées dans le modèle de tâches.

Sanou et al. [Sanou *et al.*, 2008] ont montré comment des événements (produits sur une interface) pouvaient être utilisés pour vérifier que le déroulement de l'activité était conforme à celui attendu (exprimé dans un modèle de tâches). Nous nous appuyons sur le même principe pour chercher à établir une séquence d'actes (avec leurs paramètres) qui soit conforme à l'activité à partir des données captées. Cependant, les événements que nous captions n'étant pas issus de l'utilisation d'une interface (monde fermé) notre approche nécessite un traitement et l'utilisation de l'ontologie pour pouvoir avoir des données à un niveau d'abstraction permettant une interprétation sémantique. L'ontologie permet de construire des instances d'événements avec des valeurs



obtenues par traitement des données captées (voir la section 6.2.2). Dans notre ontologie 5 types d'événements peuvent être instanciés (Tableau 6.10). Pour une sortie réalisée, l'ontologie sera donc peuplée d'événements de ces types là avec des propriétés complétées donnant des informations temporelles (les propriétés *Start\_Time* et *End\_Time*) et de liens avec les entités (comme *has\_actor*). La Figure 6.12 présente l'instance *Activity\_Event\_13* (une pause réalisée de 12 :40 :12 à 12 :57 :45 près d'une fontaine) qui fait suite à l'instance d'événement *Activity\_Event\_10*. Ces événements sont alors associés à des tâches du modèle de tâches selon les règles basées sur la sémantique des événements ou des valeurs attribuées à leurs propriétés (Tableau 6.10).

Description de tâches en termes	
Événements de l'ontologie	tâches concrètes du modèle de tâches
<i>Activity_Event</i>	si <i>Travel_Mode</i> !=stop alors Avancer si <i>Travel_Mode</i> =stop alors Faire une pause ou Subir un incident
<i>Condition_Event</i>	Évaluer les conditions de randonnée
<i>Decision_Event</i>	Prendre une décision
<i>Physiology_Event</i>	Évaluer les conditions de randonnée
<i>Weather_Event</i>	Évaluer les conditions de randonnée

TAB. 6.10: Caractéristiques de tâches concrètes

À partir de ce traitement, il est possible d'aboutir à une séquence de tâches concrètes qui, présentée dans le format des fichiers de scénarios de K-MADe, est jouable dans l'outil de simulation (en suivant la même procédure que celle décrite pour la vérification du modèle, section 6.3.2). Deux cas se présentent alors :

- la simulation réussit à jouer la séquence. La séquence d'événements telle que construite avec l'ontologie correspond à un déroulement permis par la modélisation de l'activité.
- la simulation ne réussit pas à jouer la séquence. Il faut trouver une séquence valide vis à vis de l'activité qui prend en compte la séquence d'événements telle que construite avec l'ontologie.

Lorsque la simulation échoue, il faut chercher à construire une séquence d'événements valide en nous reposant sur les événements à notre disposition.

The screenshot displays the Protégé interface for the class `Activity_Event`. On the left, a class hierarchy tree shows the following structure:

- owl:Thing
  - Entity
    - Abstract\_Entity
      - Context
        - Difficulty\_Level
        - Location
        - Weather
          - Human\_Activity
            - Movement
              - Advance
                - Descent
                  - Mount
                    - Static
                      - Human\_Physiology
                        - Objective
                          - Problem
                            - Ski\_Touring
                              - Time
                                - Physics\_Entity
                                  - Living
                                    - Non\_Living
                                      - Event
  - Activity\_Event** (selected)
  - Condition\_Event
  - Decision\_Event
  - Difference\_Event
  - Location\_Event
  - Physiology\_Event
  - Weather\_Event

The main panel shows the description of `Activity_Event` with the following types and annotations:

          - Types:** Activity\_Event
          - Annotations:** Same Individual As, Different Individuals

The right panel displays property assertions and data for the instance:

          - Object property assertions:**
            - has\_end\_natural\_element shelter\_12
            - has\_actor person\_1
            - has\_activity advance
            - previous\_event activity\_event\_10
            - has\_start\_natural\_element spring\_11
            - has\_end\_natural\_element spring\_11
          - Data property assertions:**
            - Denivelee -10.0f
            - Travel\_Mode "stop"^^xsd:string
            - Start\_Time "2017-01-22T12:40:12"^^xsd:dateTime
            - Average\_Speed 2.3494055E-8f
            - End\_Altitude 1597.0f
            - Start\_Altitude 1607.0f
            - End\_Time "2017-01-22T12:57:45"^^xsd:dateTime

FIG. 6.12: Interface du logiciel Protégé qui montre une instance d' *Activity\_Event*

Avec notre approche de construction de la séquence d'événements jouée lors de la simulation, celle-ci peut échouer lorsque, à un moment donné, la tâche jouée n'est pas celle attendue. C'est le cas si les tâches obligatoires n'ont pas d'événements dans la séquence, ou si plusieurs événements représentent une tâche non itérative. Il faut donc, soit ajouter des événements qui correspondent aux tâches obligatoires, soit supprimer des événements. Dans le cas où deux événements d'une tâche non itérative sont séquentiellement présents dans la séquence, nous faisons le choix de les fusionner. Dans le cas où ils sont distants ou qu'il faut ajouter un événement correspondant à une tâche obligatoire, il faut considérer tous les cas possibles.

Pour vérifier la cohérence, une implémentation a été faite sous forme d'automates à états finis. Les automates sont des machines abstraites capables de reconnaître si une chaîne appartient ou non à un langage régulier donné. La figure 6.13 présente un analyseur lexical sous forme d'automate. Il sert pour identifier l'appartenance d'une chaîne de caractères qui correspond à une expression arithmétique. Pour ce faire, il doit être capable de reconnaître une séquence contenant des blancs, des lettres ou chiffres et les opérateurs appropriés (identifiés dans l'image comme *autre*) dans l'ordre fixé par l'automate. Cet analyseur accepte la séquence " 24 +" car elle commence avec un espace (un blanc), suivie de deux chiffres et d'un opérateur. Par contre, la séquence "+24 " sera rejetée vu qu'elle commence avec un opérateur.

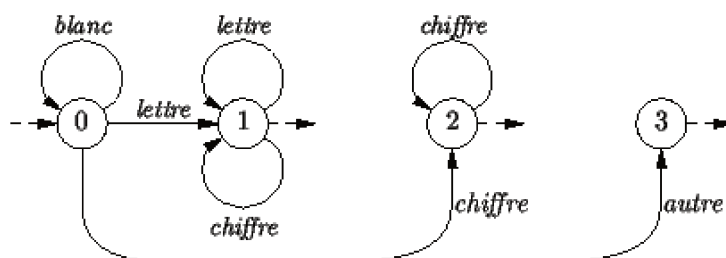


FIG. 6.13: Automate d'un analyseur lexical des expressions arithmétiques (<https://www.enseignement.polytechnique.fr/informatique/ARCHIVES/IF/poly/main009.html>) [Date de capture : 25/08/2018]

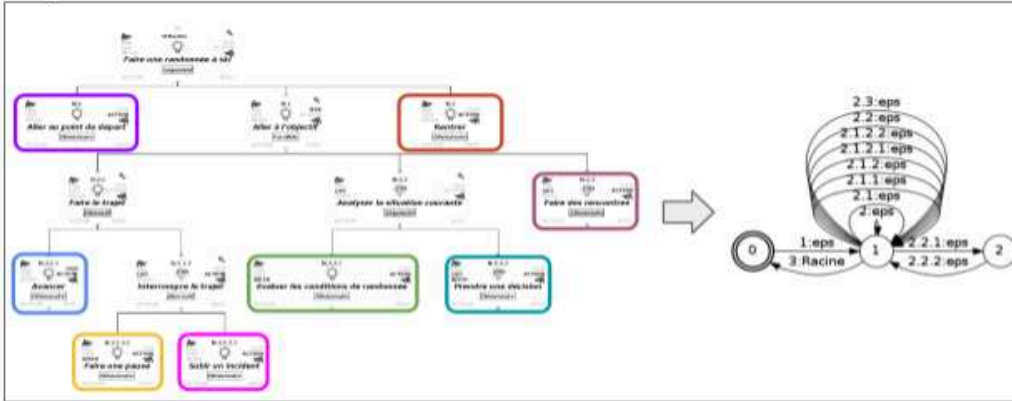
Dans notre contexte, c'est la séquence d'activité qui peut être vue comme la chaîne de caractères, alors que le modèle de tâches (exprimé sous forme d'automate) correspond au langage régulier. La démarche pour valider la séquence obtenue a été réalisée par composition d'automates et implantée en utilisant la librairie *OpenFST*<sup>4</sup>. Nous la présentons dans la Figure 6.14. Le premier pas (1) est la transformation du modèle de tâches vers un automate

<sup>4</sup><http://www.openfst.org/twiki/bin/view/FST/WebHome>

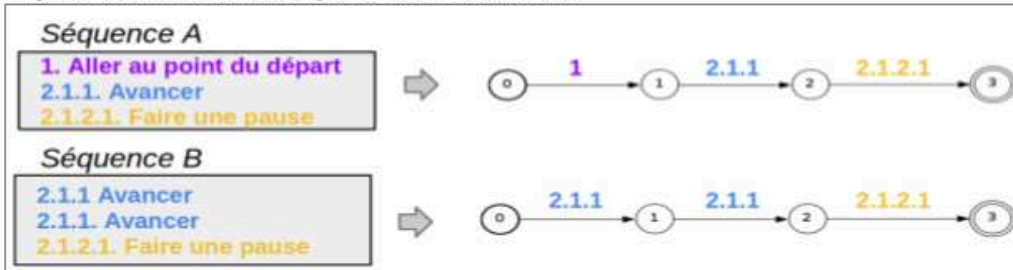
à états finis qui conserve la hiérarchie. Le deuxième pas (2) est la transformation de la séquence abstraite dans les données sous forme d'automates. La figure 6.14 (2) présente deux exemples : la *séquence A* et la *séquence B* contenant comme symboles le numéro attribué à chaque tâche d'après le niveau où elle se trouve dans le modèle (par exemple, le 1 est pour "Aller au point du départ"). Le troisième pas (3) est la composition de l'automate du modèle et celui de la séquence. Il y a deux résultats possibles : 1) si la séquence est acceptée par l'automate du modèle, cela veut dire qu'il s'agit d'une exécution valide de l'activité (*Séquence A*) ou 2) la séquence est rejetée en raison d'une inconsistance vis à vis du modèle de tâches (*Séquence B*).

Afin d'identifier les tâches pouvant manquer dans la séquence, un *mécanisme de recherche* de la séquence valide la plus proche de la séquence originale a été mis en place. Il consiste à composer l'automate du modèle de tâches avec l'automate d'un *modèle d'édition*. Le *modèle d'édition* est très utilisé dans le Traitement Automatique de Langues, particulièrement dans la Reconnaissance et Synthèse de la Parole. Il permet d'identifier les différences d'une chaîne en la comparant avec une autre prise comme référence. Trois variations sont comprises : *insertion*, *omission* et *substitution*. Dans cette thèse, nous avons développé le modèle d'édition avec les omissions possibles (c'est à dire, des tâches manquantes). Le résultat de la composition du modèle d'édition et de la séquence rejetée est le treillis de tous les chemins valides possibles. Dans notre approche nous sélectionnons le chemin le plus court, c'est à dire, celui avec le moins d'omissions. Dans l'exemple de la figure 6.14 (4), nous pouvons observer qu'après l'exécution de la méthode de recherche, la *Séquence B* n'a pas été acceptée car elle ne comportait pas la tâche "Aller au point du départ", laquelle doit toujours apparaître au début des exécutions. Dans ce cas, plusieurs choix peuvent être considérés. Soit la tâche doit être reconstruite à partir d'autres données connues, soit une ellipse doit être générée explicite ou implicite. Nous reviendrons sur cet aspect dans la conclusion du mémoire.

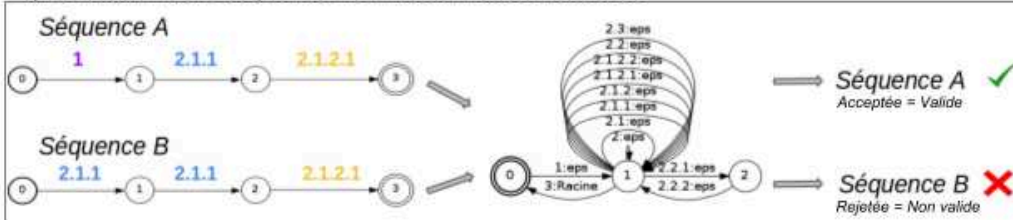
### 1) Transformation du modèle de tâches vers un automate



### 2) Transformation de la séquence vers un automate



### 3) Validation de la séquence vis à vis du Modèle de tâches



### 4) Obtention de la séquence avec le chemin le plus court quand la séquence a été rejetée

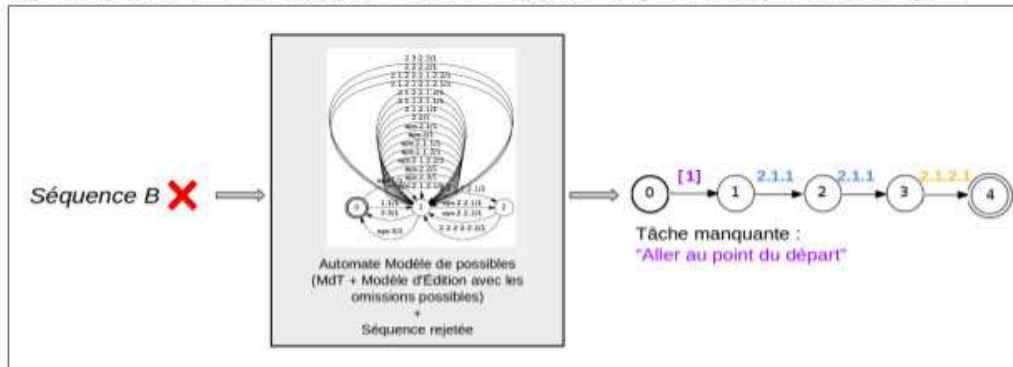


FIG. 6.14: Traitement de validation de la séquence de l'activité

## 6.5 Conclusion sur la modélisation de l'activité du point de vue de son objectif et de son utilisation

L'approche par modèles que nous avons adoptée vise à surmonter les verrous scientifiques concernant l'interprétation des données brutes issues d'activités humaines. Ainsi, nous présentons la formalisation d'une activité humaine observée à partir de capteurs à travers deux modèles qui représentent l'activité selon deux axes : le premier concerne la connaissance propre aux éléments du domaine, alors que le deuxième concerne la réalisation de l'activité. Il s'agit d'un développement nécessaire et innovant par rapport à l'état de l'art puisqu'à la différence des systèmes de génération à partir de données fictives, nous devons nous adapter aux seules connaissances que nous pouvons obtenir à l'aide de capteurs (observateur objectif de l'activité).

Les modèles ont été construits à partir d'analyses menées sur un corpus constitué de 16 récits de randonnées. Ce choix de sources implique des désavantages dans l'élaboration des modèles. L'un d'entre eux a été la difficulté d'identification des éléments du domaine à modéliser. Parfois, les skieurs, en cherchant à retransmettre leur expérience de manière vivante, expriment leurs opinions et sentiments. Cela représentait une contrainte car ce sont des aspects qui sont difficiles à obtenir par des capteurs et que, par conséquent nous avons décidé de ne pas inclure dans les récits que nous envisageons. Les capacités des capteurs sont limitées à l'enregistrement des données objectives et observables de celui qui les porte.

Le résultat des transformations des données dirigées par les modèles est l'obtention d'une séquence contenant le déroulement de l'activité enregistrée, dont les éléments inférés sont ordonnés chronologiquement. Nous avons montré que la production d'un texte reposant sur les seules connaissances exprimées dans les modèles de données et de tâches ne gênait pas la compréhension du déroulement de la sortie par les lecteurs. Cette séquence sera donc la base des récits dans les étapes postérieures. Cependant, en nous appuyant sur les récits pour produire nos modèles, il existe un risque d'interprétation erronée ou incomplète dans notre traitement. Les skieurs peuvent juger qu'une information n'a pas d'intérêt vis à vis du récit qu'ils font et donc ne pas en faire mention. Toutefois, nous avons minimisé ce risque grâce au support de l'expert du domaine, qui a été présent tout au long de la démarche de modélisation. Par exemple, comme nous l'avons mentionné dans la section 6.3.2, lors de l'évaluation du modèle de tâches par un expert du domaine, nous avons ajouté la tâche *Rentrer*. Pour l'expert, elle fait partie intégrante de l'activi-

té (puisqu'il faut rejoindre son moyen de transport) mais pour les auteurs des récits du corpus textuel de modélisation v2, elle était systématiquement omise (ellipse) soit par ce qu'elle n'illustrait aucun point intéressant vis à vis de l'information qu'ils souhaitaient transmettre u récit soit par ce que cette tâche est implicite dans les récits. Cette ajout illustre la distinction entre le déroulement conforme vis à vis de l'activité (lié au domaine) et le déroulement relaté dans les récits (faits d'intérêt pour l'auteur). Le passage d'une séquence du déroulement de la sortie que nous obtenons par l'utilisation des modèles de données et de tâches à un récit (dont le plan nécessite de choisir les événements à présenter et l'ordre dans lequel cela doit être fait) nécessite d'autres connaissances et transformations de données que nous abordons dans le chapitre suivant.

# Chapitre 7

## Modélisation du point de vue subjectif et génération du récit à base de modèles

Nous avons présenté dans le chapitre précédent comment nous obtenions une séquence d'actions valide vis à vis de l'activité à partir des données captées. Le présent chapitre vise à exposer la construction des plans de récits (voir définition 4.7) d'activité et la génération de phrases basées sur ces plans. Les plans du récit permettent de transposer les événements du monde de l'activité au monde du récit.

Comme nous l'avons évoqué au chapitre 4, l'enjeu est de faire émerger le récit à partir d'un ensemble de données capturant l'activité humaine. Ce récit doit rendre compte des éléments objectifs de cette activité, et en transmettre une description cohérente, mais aussi rendre compte du ressenti du protagoniste, et transmettre au lecteur une perspective communicationnelle, passant ainsi du monde objectif de l'activité vers le monde subjectif de l'expérience.

Les éléments objectifs et subjectifs qui fondent l'activité ne sont pas tous observables et l'on ne dispose pas d'une connaissance exhaustive de toutes les circonstances permettant le récit de l'activité. Les conditions environnementales ont une influence considérable sur la réalisation de l'activité, rendant son déroulé, le ressenti du protagoniste ou son intention de communiquer extrêmement fluctuants. Pour ces raisons, les approches à base de planification ne peuvent pas être employées car on ne dispose pas d'un modèle ou d'un référent exprimant l'influence du but communicationnel sur la construction d'un récit. Les lois de construction du récit doivent donc être repensées.

Nous proposons dans cette thèse de bâtir le récit à partir d'une base modulaire d'unités de représentation spécifiant les caractéristiques du déroulé et du ressenti de l'activité. Différents récits, correspondants à des déroulés, des ressentis et des buts communicationnels différents, seront obtenus par l'or-



ganisation et la sélection de ces unités de représentation. Une méthodologie flexible, travaillant en deux temps, est mise en place pour construire le récit : constitution d'un fil narratif par une structure d'unités reflétant le déroulé de l'activité, expression du ressenti par sélection d'unités complémentaires venant enrichir ce fil narratif.

Une approche à base de modèles, issue d'une adaptation du modèle de Lehnert, est proposée dans ce but. D'une grande modularité, l'intérêt de ce modèle est de conférer une structure narrative aux événements centrée sur la notion d'état affectif et de relations causales entre états, tout en préservant le caractère émergent de la construction du récit.

Quoique non spécifique au ski de randonnée, le travail est ancré dans l'acquisition de connaissances sur ces récits, à partir de l'analyse des corpus de modélisation, afin de donner une base concrète et vérifiable aux propositions effectuées.

Nous présentons d'abord l'analyse des corpus faite pour acquérir de la connaissance sur la réalisation d'un récit textuel (section 7.1) puis comment ces connaissances nous permettent de définir des plans de récit d'activité (section 7.2) et enfin, à partir d'un plan, comment générer des phrases pour constituer le récit textuel (section 7.3).

## **7.1 Extraction des connaissances sur le monde du récit de randonnée**

Pour présenter des données objectives sous forme de récit, nous avons analysé les corpus textuel et fréquentiel dans le but d'acquérir des connaissances sur comment les randonneurs racontent leurs sorties aux autres randonneurs (le vocabulaire, la structure, les points jugés comme importants). Pour cela nous avons réalisés deux types d'analyse, une analyse de fréquence (mots et collocations) et une analyse structurelle des récits (voir tableau 7.1).

| Corpus             | Analyse  | Connaissances  |
|--------------------|--|--|
| Corpus fréquentiel | Analyse des fréquences de mots<br>Analyse des fréquences de collocations | vocabulaire<br>expression  |
| Corpus textuel     | Analyse structurelle   | structures globales<br>type et contenu du corps<br>contenu de l'introduction<br>contenu de la conclusion |

TAB. 7.1: Les analyses sur les corpus pour la connaissance sur le récit

### 7.1.1 Analyses de fréquence

Nous avons analysé le langage que les skieurs utilisent pour raconter leurs sorties. Par exemple, comment les personnes expriment les déplacements, ou s'ils utilisent une terminologie spécifique pour parler d'un certain fait. Pour cette analyse, nous utilisons le Corpus fréquentiel comme source (section 5.1.2). Concernant les outils d'analyse, nous avons employé la librairie Python NLTK. Nous avons mis en place des scripts pour obtenir les mots ainsi que leur fréquence et les collocations (bigrammes et trigrammes).

Pour obtenir les 20 mots les plus fréquents (Tableau 4), le script analyse les 307 456 mots du corpus fréquentiel, supprime les mots vides<sup>1</sup> et indexe les 166 434 mots non vides pour compter le nombre total d'utilisation.

7 des 20 mots les plus fréquents sont directement liés au domaine du ski de randonnée (*Domaine* dans le tableau 4). 7 autres mots traduisent le ressenti des narrateurs vis à vis d'éléments de l'activité (*Ressenti* dans le tableau 4). Ces mots sont principalement des adjectifs et des adverbes. Ces deux groupes montrent l'importance de la description des conditions et de l'environnement dans le récit d'activité pour les randonneurs.

Une explication possible est que les randonneurs n'ont pas besoin de s'exprimer longuement sur les actions réalisées car la complexité et l'intérêt de la sortie résident dans les caractéristiques de l'environnement dans lequel elle se déroule.

Cette première analyse de fréquence nous montre également que les mots avec une charge négative ne sont pas fréquents, ce qui nous fait penser que

<sup>1</sup>Un mot vide est un mot tellement commun dans la langue française qu'il n'apporte aucune connaissance. Par exemple les mots "de", "en", "le" sont des mots vides.

les randonneurs partagent plutôt des aspects positifs sur leurs sorties.

| <b>Mot</b> | <b>Fréquence</b> | <b>Normalisé</b> | <b>Catégorie</b> |
|------------|------------------|------------------|------------------|
| neige      | 2456             | 0,2456           | Domaine          |
| bien       | 2104             | 0,2104           | Ressenti         |
| descente   | 1951             | 0,1951           | Domaine          |
| plus       | 1441             | 0,1441           | Ressenti         |
| peu        | 1384             | 0,1384           | Ressenti         |
| col        | 1320             | 0,1320           | Domaine          |
| très       | 1306             | 0,1306           | Ressenti         |
| sommet     | 1253             | 0,1253           | Domaine          |
| montée     | 1085             | 0,1085           | Domaine          |
| fait       | 1037             | 0,1037           |                  |
| sortie     | 1004             | 0,1004           | Domaine          |
| bonne      | 994              | 0,0994           | Ressenti         |
| belle      | 976              | 0,0976           | Ressenti         |
| trace      | 923              | 0,0923           | Domaine          |
| bon        | 906              | 0,0906           | Ressenti         |
| tout       | 870              | 0,0870           |                  |
| même       | 828              | 0,0828           |                  |
| puis       | 816              | 0,0816           |                  |
| encore     | 783              | 0,0783           |                  |

TAB. 7.2: Les 20 mots les plus fréquents du Corpus fréquentiel

Les fréquences de collocations représentent les fréquences de la présence collocalisée de deux mots (bigrammes) ou trois mots (trigrammes). Les scripts de calcul de ces fréquences procèdent de la même manière que celui du calcul des fréquences de mots mais en indexant tous les couples (ou triplets) de mots détectés. Les tableaux 7.3 et 7.4 présentent les 20 bigrammes et les 20 trigrammes les plus fréquents dans les textes du corpus fréquentiel.

| <b>Bigramme</b> | <b>Fréquence</b> | <b>Intérêt</b>               |
|-----------------|------------------|------------------------------|
| belle étoile    | 291              | Navigation                   |
| quand même      | 257              |                              |
| grand som       | 223              | Navigation                   |
| charmant som    | 188              | Navigation                   |
| neige dure      | 176              | Conditions Environnementales |
| neige bien      | 162              | Conditions Environnementales |
| bonne neige     | 158              | Conditions Environnementales |
| petit som       | 154              | Navigation                   |
| très bonne      | 129              |                              |
| dent pra        | 123              | Navigation                   |
| belle sortie    | 122              |                              |
| même si         | 121              |                              |
| col porte       | 118              | Navigation                   |
| neige très      | 113              | Conditions Environnementales |
| très bon        | 111              |                              |
| petite sortie   | 107              |                              |
| bon ski         | 99               | Conditions Environnementales |
| belle journée   | 97               |                              |
| très belle      | 97               |                              |
| peu plus        | 95               | Conditions Environnementales |

TAB. 7.3: Les 20 bigrammes les plus fréquents du Corpus fréquentiel

| <b>Trigramme</b>       | <b>Fréquence</b> | <b>Intérêt</b>               |
|------------------------|------------------|------------------------------|
| pic belle étoile       | 44               | Navigation                   |
| très bon ski           | 29               |                              |
| très belle sortie      | 26               |                              |
| bien belle journée     | 22               | Conditions Environnementales |
| très bonne neige       | 21               | Conditions Environnementales |
| toujours aussi beau    | 21               | Conditions Environnementales |
| grande lance domene    | 21               | Navigation                   |
| sommet grand som       | 18               | Navigation                   |
| vallon belle étoile    | 17               | Navigation                   |
| sommet charmant som    | 17               | Navigation                   |
| quand même bien        | 17               |                              |
| refuge combe madame    | 17               | Navigation                   |
| sommet belle étoile    | 16               | Navigation                   |
| très peu monde         | 16               | Conditions Environnementales |
| neige bien dure        | 15               | Navigation                   |
| très bonnes conditions | 15               | Conditions Environnementales |
| très bonne sortie      | 15               |                              |
| bonne petite sortie    | 14               | Conditions Environnementales |
| sud belle étoile       | 14               | Navigation                   |
| montée grand som       | 14               | Navigation                   |

TAB. 7.4: Les 20 trigrammes les plus fréquents du Corpus fréquentiel

En plus des observations déjà faites sur les fréquences de mots, ces deux analyses de collocation nous permettent d'identifier deux types d'information sur l'environnement. Le premier type porte sur les points de repère (*Navigation* dans les tableaux 7.3 et 7.4). Ceux-ci sont utilisés par les randonneurs pour connaître l'itinéraire suivi. Dans cette catégories d'information, les fréquences des noms de sommet mais également des mots qui s'y réfèrent

(comme *jusqu'au sommet*) confirment que l'arrivée au sommet est un point important qui représente le succès du parcours. Le second type d'information porte sur les conditions environnementales (météo, qualité de la neige...). Un aspect à noter dans ce second type est la combinaison des adjectifs et ad-  
verbes (*très bonne*) comme mécanisme pour exprimer le degré de satisfaction du narrateur.

## 7.1.2 Analyse structurelle

Nous avons analysé chaque récit d'activité du corpus textuel pour identifier les éléments de structure :

- Introduction : section au début du récit qui apporte des informations pour aider le lecteur à comprendre le contexte de la sortie.
- Corps : section qui présente l'évolution des faits.
- Conclusion : section en fin de récit qui dresse un bilan de la sortie.

Le tableau 7.5 présente le nombre d'apparitions de chacun de ces éléments de structure dans les 16 récits d'activité du corpus textuel.

| Élément de structure | Nombre de présence (/16) |
|----------------------|--------------------------|
| Introduction         | 11                       |
| Corps                | 16                       |
| Conclusion           | 11                       |

TAB. 7.5: Proportion de présence des éléments structurels

Dans un second temps, nous avons analysé les introductions, les corps et les conclusions pour identifier leurs contenus. Trois types de contenu apparaissent dans l'introduction (tableau 7.6), deux types de corps (7.7) et cinq types de contenu peuvent être présents dans la conclusion (tableau 7.8).

L'introduction peut présenter des commentaires généraux sur les conditions ou donner des informations sur la préparation de la randonnée, telles que la sélection du sommet ("Superbe journée pour un doublée Chamechaude et Charmant Som" [récit 53945.xml]) ou une brève description de la route ("Partis tard, vers 11h depuis les téléskis par la route qui monte au col." [récit 38809.xml]) ; elle peut aussi donner la motivation ou l'état d'esprit des randonneurs ("une journée comme celle là, on pouvait pas la gâcher..." [récit 44987.xml]).

| Contenu                  | Description   | Fréquence |
|--------------------------|---|-----------|
| Description              | Commentaires par rapport aux conditions météorologiques ou de terrain   | 6         |
| Préparation de la sortie | Décisions prises d’après les conditions avant de commencer la sortie, combien de personnes vont faire la randonnée, etc | 9         |
| Impressions personnelles | Observations subjectives ou commentaires dont l’information ne peut pas être mesurée                                    | 10        |

TAB. 7.6: Contenu de l’introduction

Le corps porte le déroulement de l’activité réalisée pendant la sortie. Nous avons identifié deux types de corps, selon la façon dont les skieurs expriment le parcours réalisé : le premier a tendance à donner une description impersonnelle de l’activité (Figure 7.1a) , alors que le second rapporte l’activité mais de manière plus subjective et engagée (Figure 7.1b). Par exemple, dans la figure 7.1a nous pouvons observer que le récit se limite à faire des références autour des conditions de la sortie, notamment des observations sur l’état de la neige, aspect fondamental pour le ski de randonnée (“Au niveau d’Orgeval la neige a déjà commencé a bien transformer”). La figure 7.1b présente l’expérience vécue plus en détail, en ajoutant des commentaires qui dénotent des impressions (“on traîne pas et on enquille la NW, belle ambiance”) ou des commentaires qui enrichissent le récit (“puis on lâche les watts dans la directe, dommage qu’il y ait autant de raclette dans les passage étroits, ça passe partout ce truc... ”). Bien qu’il n’y ait pas de séparation nette entre ces deux types, chaque corps a tendance à orienter son style vers un de ces deux types. Les moyens utilisés sont l’inclusion de commentaires, l’utilisation de pronoms personnels comme “je” ou l’insertion de contenu extra-diégétique employé pour illustrer ou compléter le déroulement (comme des références vers des sorties précédentes dans le même endroit afin de comparer les conditions environnementales).

*Montée dans les sapins tranquille, la neige est damée par les nombreux passages. Il y a du monde aujourd'hui aussi bien en ski qu'en raquettes ou même a pied et en ski de fond. Au niveau d'Orgeval la neige a déjà commencé a bien transformer. Montée sur l'arrête sud sans difficulté. Pour redescendre c'est quasi une neige de printemps, assez lourde.*

(a) Sortie du samedi 18 février 2012 vers le sommet Charmant Som (<http://www.skitour.fr/sorties/charmant-som,38809.html>, date de consultation : 30/08/18)

*Montée peinarde au Charmant, dégun, on a du faire 17 pauses dont un premier casse dalle aux Bergeries. Alors qu'on se croyait seuls au monde on est rejoints au sommet par un groupe de 4 puis de 2, mais c'est place Grenette ce truc alors on traîne pas et on enquille la NW, belle ambiance, bonne poudre en choisissant bien. On croise la collective du Gucem en bas, remontée du Pré Batard jusqu'à Chamechine, casse dalle 2 le retour et go dans le bouzin. Bonne neige, le goulet du bas nécessite un gratonnage mi herbe mi rocher de 5 à 8m mais pas pire, puis c'est tranquille dans la foret dans une neige un peu transfo, parfois un peu lourde mais pas pire, ça va pas durer longtemps. Stop pour revenir à l'épingle (merciiii), on remonte au col de Porte pour voir les copains ;- ) on les voit qui se lancent dans la NNW qu'on avait aussi envisagé ce matin mais c'est une couenne et on avait la journée. On se dit que ça doit pas être terrible et une fois les zozos sortis des difficultés on file et on les nargue par sms. Réponse : c'est mort bon blablabla, ah bon ? M\*rde, on lâche quand même l'affaire et on se casse mais arrivés au Sappey on craque, on n'a pas de volonté quoi et on retourne au col de porte. La montée est avalée en un peu plus d'une heure, pas mal pour des mecs crevés et le sale gosse avec son split, puis on lâche les watts dans la directe, dommage qu'il y ait autant de raclette dans les passage étroits, ça passe partout ce truc... retour dans la foret d'abord bon puis de plus en plus lourd voire bien bien loooooourd jusqu'en bas, ça c'est fait ! un 3.3, un 4.3/5.1 et un 4.2 pour finir, en ce jour de "terrain miné" c'est cadeau surtout quand on a une pensée aux copains qui sont allés buter au... refuge Agnel ;-), et le sale gosse est content de sa journée, c'est pas beau ça !!?*

(b) Sortie du 21 mars 2013 vers le sommet Charmant Som Étoile de Belledonne (<http://www.skitour.fr/sorties/charmant-som,48511.html>, date de consultation : 30/08/18)

FIG. 7.1: Corps de récits extraits du Corpus de modélisation



| Contenu                     | Description   | Fréquence |
|-----------------------------|---|-----------|
| Un récit plutôt impersonnel | Commentaires par rapport au déroulement de l'activité, tel que déplacements, observations ou événements météorologiques | 6         |
| Un récit plutôt personnel   | Impressions ou ressentis qui accompagnent la description du parcours  | 10        |

TAB. 7.7: Contenu du corps

La conclusion peut permettre de revenir sur un point particulier (“Les trois chamois qu’on a croisés on l’air d’être du même avis ;-)” [38284.xml]) de la sortie ou faire un point global en présentant un résumé avec les points les plus remarquables ou intéressants de la sortie (“Beau tour très complet du Charmant Som. on visite toutes les faces, y compris les plus sauvages du versant N” [38025.xml]). De même, la conclusion est un espace où les impressions personnelles (“Un bon bol d’air et des jolis paysages, cela fait du bien” [3513.xml]), des comparaisons avec d’autres sorties (“Par rapport à hier (cf. sortie David Zijp), il y a 20 cm de plus qui sont tombé dans la nuit. On se croirait au canada, bien que je n’ai jamais skié la bas :-) Un Chamechaude de haut niveau.” [44954.xml]), ainsi que des recommandations (“Le déniv est sympa mais la distance est trop importante.....a faire UNE fois ” [45189.xml]) peuvent être faites.

| Contenu  | Description   | Fréquence |
|--|---|-----------|
| Impressions personnelles                       | Observations subjectives ou commentaires sur la sortie dont l'information ne peut pas être mesurée                          | 10        |
| Références                                     | Mention d'une activité ou d'un fait trouvé dans le corps ou dans l'introduction   | 5         |
| Comparaison à une autre randonnée              | Comparaison entre le déroulement de la sortie avec une autre déjà réalisée dans le même endroit                             | 4         |
| Recommandation de la sortie à un interlocuteur | Avis sur la sortie réalisée dont le but est d'inviter d'autres skieurs à la répéter ou au contraire à l'éviter complètement | 3         |
| Synthèse                                       | Résumé de la sortie   | 4         |

TAB. 7.8: Contenu de la conclusion

Pour résumer, les structures des récits de randonnées peuvent être composées soit uniquement du corps, soit d'une introduction avec le corps, soit du corps et de la conclusion, soit de l'ensemble des composants (introduction, corps, conclusion). Le tableau 7.9 présente la fréquence dans le corpus textuel des différentes structures de récits d'activités. Il faut indiquer que, malgré l'identification des parties structurales, le contenu mis dans celles-ci varie beaucoup. Cela veut dire que, par exemple, un texte du corpus peut combiner les trois types d'introduction en même temps, l'auteur incluant dans la même section la description de la sortie, sa préparation et ses impressions personnelles, ce qui rend le récit plus riche en détails. De même, dans le cas du corps, une même structure peut être utilisée pour exprimer les deux types du corps (objectif et subjectif).

| Contenu                    | Nbre |
|----------------------------|------|
| Corps seul                 | 2    |
| Intro + Corps              | 3    |
| Corps + Conclusion         | 3    |
| Intro + Corps + Conclusion | 8    |

TAB. 7.9: Structures possibles de récits

### 7.1.3 Analyse du déroulement (Corps) transmis

Le déroulement de la sortie est décrit dans le corps du récit. C'est donc cette partie que nous allons chercher à générer. Pour cela, nous avons cherché à identifier dans le corps des récits d'activité des caractéristiques sur la manière dont le déroulement est transmis. Comme nous l'avons identifié dans l'analyse de fréquence, les mots employés dans les récits sont principalement des adjectifs et des adverbes, ceux-ci permettant de qualifier les informations. Il en ressort une appréciation générale sur la qualité de la sortie ressentie par le protagoniste. Par exemple, le déroulement peut avoir un ensemble d'informations projeté avec un sentiment d'accomplissement si tout ce qui a été planifié a été accompli ou de frustration dans le cas contraire.

Nous avons analysé à nouveau le corpus de modélisation afin de trouver quels étaient les ressentis les plus fréquents dans les récits de ski de randonnée. Quatre ressentis principaux ont été sélectionnés dont nous proposons les définitions suivantes :

- **Ressenti positif** L'expérience a été agréable avec éventuellement des difficultés mais qui n'ont pas créé de modifications de parcours notables. Les objectifs planifiés sont alors atteints. Ce type de ressenti regroupe :
  - **Positif sans contretemps** La montée et la descente ont été réalisées sans difficultés.
  - **Positif mais avec des défis** Il y a eu des difficultés surmontables rencontrées par les protagonistes (qui peuvent être jugées comme des défis attirants pour certains).
- **Ressenti nuancé** Bien que l'activité a été réalisée dans sa totalité, l'apparition d'incidents ou de difficultés ont empêché un déroulement totalement agréable. Dans ce cas, la personne exprime dans son récit

un ressenti avec un certain degré d'insatisfaction, mais pas assez pour déduire que l'activité a été un échec complet.

- **Ressenti négatif** D'après la personne qui raconte son vécu, l'activité a été un échec complet. L'apparition de nombreux incidents a rendu la progression tellement difficile qu'il a fallu faire des modifications par rapport à ce qui avait été prévu à la base. Ces modifications peuvent être l'abandon de l'objectif initial et par conséquent l'annulation de la sortie.

Dans la table 7.10 nous présentons le résultat de l'analyse qui montre le type de ressenti par texte. Dans le cas de ceux projetant un ressenti positif, nous incluons les deux types : *SC* signale les sorties sans contretemps, alors que *AD* indique les sorties portant des défis rencontrés.

| Texte du corpus | Ressenti positif | Ressenti nuancé | Ressenti négatif |
|-----------------|------------------|-----------------|------------------|
| 1. 35813.xml    | X(SC)            |                 |                  |
| 2. 37304.xml    | X(SC)            |                 |                  |
| 3. 38025.xml    | X(AD)            |                 |                  |
| 4. 38284.xml    | X(AD)            |                 |                  |
| 5. 38809.xml    | X(AD)            |                 |                  |
| 6. 44954.xml    |                  | X               |                  |
| 7. 44987.xml    |                  | X               |                  |
| 8. 44999.xml    |                  | X               |                  |
| 9. 45189.xml    |                  | X               |                  |
| 10. 45545.xml   | X(AD)            |                 |                  |
| 11. 45715.xml   |                  | X               |                  |
| 12. 46259.xml   |                  | X               |                  |
| 13. 48511.xml   |                  |                 | X                |
| 14. 53945.xml   |                  | X               |                  |
| 15. 54085.xml   |                  | X               |                  |
| 16. 54494.xml   |                  | X               |                  |
| <b>Total</b>    | 6                | 9               | 1                |

TAB. 7.10: Identification du ressenti par texte.

Nous présentons l'analyse des faits relatés par les types de ressenti dans la table 7.12. Nous avons réalisé cette analyse en appliquant la même méthode que celle présentée pour la modélisation de l'activité (Chapitre 6). Nous pouvons observer qu'il y a plus d'événements dans les récits avec un ressenti nuancé. La différence en quantité est particulièrement marquée pour la description des déplacements et des conditions de randonnée. Ce constat est appuyé par le nombre moyen de mots et de phrases présents pour chaque type de récits (Tableau 7.11). Nous pouvons observer la tendance d'utiliser plus de mots selon le ressenti. Lorsque l'expérience se passe sans problèmes le récit est assez court. Si l'expérience s'est déroulée avec difficultés (surmon-

tables ou pénibles), le nombre de mots est doublé en moyenne. Par contre, dans le cas d'une expérience négative, la quantité surpasse la moyenne des phrases<sup>2</sup>. Par rapport aux différences en termes de quantité de phrases, il est plus difficile de proposer une interprétation autour de la quantité identifiée (nous pourrions considérer des facteurs liées au style personnel).

|                            | Ressenti positif (sans contre-temps) [2] | Ressenti positif (avec défis)[4] | Ressenti nuancé[9] | Ressenti négatif[1] |
|----------------------------|--|----------------------------------|--------------------|---------------------|
| Nbre Moyen Mots            | 121                                      | 258                              | 258,7              | 406                 |
| Nbre Moyen Phrases         | 9  | 18,5                             | 14,5               | 33                  |
| Nbre Moyen mots par phrase | 13,4                                     | 13,94                            | 17,9               | 33                  |

TAB. 7.11: Nombre de mots et de phrases moyens par type de récits.

| Tags annotation    | Ressenti positif (sans contretemps) | Ressenti positif (avec défis) | Ressenti nuancé | Ressenti négatif |
|--------------------|-------------------------------------|-------------------------------|-----------------|------------------|
| Point_Depart       | 0                                   | 0,5                           | 1               | 0                |
| Déplacement        | 3,5                                 | 5,75                          | 8               | 13               |
| Pause              | 1                                   | 1,75                          | 2,89            | 4                |
| Incident           | 0                                   | 0,25                          | 0,11            | 0                |
| Terrain            | 1                                   | 3                             | 2               | 3                |
| Météo              | 1                                   | 1,25                          | 0,89            | 0                |
| Trajectoire        | 0                                   | 0                             | 0,11            | 0                |
| Condition_Physique | 0                                   | 0                             | 0,11            | 0                |
| Décision           | 0                                   | 1,5                           | 0,78            | 1                |
| Rencontre          | 0,5                                 | 1,25                          | 1,11            | 3                |
| Total              | 7                                   | 15,25                         | 17              | 24               |

TAB. 7.12: Identification du contenu par type de récit.

<sup>2</sup>Étant donné que nous avons qu'un texte avec un ressenti négatif, nous présentons cette conclusion comme une piste à suivre au travers de la collecte des autres textes avec le même ressenti

## 7.2 Élaboration du plan du récit

Comme exposé dans le Chapitre 2, les récits fictifs sont généralement composés d’existants (c’est à dire, les objets et contexte où l’histoire se déroule), d’événements (actions ou occurrences, enrichies ou déclenchées par des réactions) et d’interactions entre les personnages.

Dans le cas des récits d’activité que nous souhaitons générer, la source n’est pas créée mais collectée à partir de signaux objectifs réels. la génération dépend uniquement de la séquence d’activité produite à partir des données. Par conséquent, les récits produits manqueront de notes subjectives, exprimant le ressenti des personnages vis à vis du déroulement. Le plan du récit que nous définissons dans le cadre du récit d’activité est donc constitué uniquement à partir d’événements objectifs réalisés pendant l’activité (définition 4.3).

Nous avons défini précédemment (définition 4.7) que le plan du récit est défini en fonction d’un but communicationnel. Les événements du récit doivent donc être organisés non pas selon un ordre temporel mais pour satisfaire le but communicationnel. Pour rendre compte de ce but communicationnel, nous proposons d’utiliser le formalisme de Lehnert [Lehnert, 1981]. La section suivante présente comment ce formalisme (présenté dans la section 2.2.2.3) peut être utilisé dans ce contexte.

Le but communicationnel qui permet de choisir l’organisation est propre à l’auteur du récit. Bien que les analyses que nous avons réalisées ne permettent pas d’identifier explicitement le but communicationnel, elles nous permettent de mettre en avant le fait que les récits transmettent un ressenti sur la sortie narrée. Quatre types de ressenti ont été identifiés. La section 7.2.2 présente comment le ressenti sur la sortie est exprimé par le plan du récit.

### 7.2.1 Le modèle du plan du récit

Pour construire le plan du récit, nous avons utilisé le formalisme proposé par [Lehnert, 1981]. Il s’agit d’un modèle originellement construit pour l’analyse de récits textuels. Ce modèle exprime le plan du récit par un enchaînement du ressenti et des réactions des personnages, face aux événements racontés. Par exemple, si dans un récit un personnage a faim cela provoque de l’insatisfaction, il va alors vouloir chercher des aliments, lorsqu’il en trouve et qu’il mange, il ressent la satisfaction de ne plus avoir faim. Dans cet exemple, nous pouvons observer que les événements “avoir faim”, “chercher des aliments” et “manger une pomme” sont liés de manière causale par le ressenti du personnage. Le modèle de Lehnert permet donc d’exprimer les liens entre

les événements en fonction de leur effet sur le ressenti des personnages, ce qui constitue un moyen de définir des liens de cause à effet (liens causaux).

Le type de ressenti et les liens entre les événements permettent de définir des structures de base (*plot unit*, voir Figure 2.8) qui, combinées, expriment l'ensemble du plan. Celui-ci n'est donc pas prédéfini pour un type de récit (tel que le propose [Freytag, 1900] par exemple) mais résulte des événements eux-mêmes. Notre approche dépend uniquement des données collectées, l'adaptation aux événements proposés par ce modèle, est donc un atout important pour nos travaux.

### **7.2.1.1 Les unités d'information du plan du récit d'activité pour le ski de randonnée**

Afin de tester l'expressivité du modèle de Lehnert [Lehnert, 1981] vis à vis des récits réels, nous avons réalisé une analyse de textes sur certains récits de randonneurs. Pour ce faire, le corpus de modélisation a été pris à nouveau. La Figure 7.2 présente le récit 38284.xml. Il décrit la sortie faite vers le sommet du Charmant Som par la route des Cottaves. Ici, la montée au sommet a été réussie, malgré les conditions environnementales défavorables. L'introduction commente une décision de changement d'objectif extra-diététique qui se justifie par l'observation des conditions de risques en montagne ("Vu le risque de formation de plaque, on préfère s'arrêter aux cottaves et rester un maximum dans la forêt."). Le corps du texte insiste sur l'obstacle du grand froid qui a été surmonté, soit en adoptant le comportement adéquat ("un gros rythme d'entrée de jeu"), soit en utilisant l'équipement adéquat ("les gros gants, le casque ..."). Finalement, bien que la fin de la sortie ne soit pas relatée explicitement, il est possible d'inférer un retour sans complications grâce aux commentaires faits par rapport à la descente "La descente est franchement pas terrible jusqu'à l'attaque de la forêt").



*On n'a tout de même bien fait de ne pas partir pour le grand som comme prévu à la base. Vu le risque de formation de plaque, on préfère s'arrêter aux cottaves et rester un maximum dans la forêt. Bon le -19°C au parking nous fait hésiter mais en même temps on n'est pas venu jusqu'ici pour rien. On met un gros rythme d'entrée de jeu histoire de se réchauffer les doigts. Une fois les premiers onglés passés, la montée se passe sans encombre. On se cale derrière le dernier rocher avant l'attaque finale pour se camoufler avec les gros gants, le casque, la doudoune. Comme ça on n'a plus qu'à enlever les peaux au sommet. Pari réussi, malgré le vent et les -30 °C en ressenti au sommet, pas d'onglé!!! La descente est franchement pas terrible jusqu'à l'attaque de la forêt (neige croutée) mais franchement excellente sous les arbres où la neige n'a pas vu le soleil!!! Les trois chamois qu'on a croisés on l'air d'être du même avis ;-)*

FIG. 7.2: Récits provenant du site [skitour.fr](http://www.skitour.fr) décrivant la sortie faite pour atteindre le sommet du Charmant Som via la route des Cottaves (<http://www.skitour.fr/sorties/charmant-som,38284.html>, date de consultation : 14/03/18)

L'unité de décomposition du formalisme est l'*état affectif*. Selon Lehnert [Lehnert, 1981], les états affectifs signalent les réactions des personnages face à un événement narré dans un récit. Ces réactions sont identifiées lors de l'analyse textuel. Par exemple, si dans un récit apparaît une phrase telle que "L'homme a perdu son chien qu'il avait depuis qu'il était petit", nous pourrions déduire que cet événement a provoqué une réaction défavorable chez le personnage.

Cependant, dans le cas des récits d'activité que nous cherchons à générer, nous ne pouvons utiliser que ce qui a été observé ou vécu, sans pouvoir accéder aux souvenirs ou autres informations qui pourraient être données par un auteur (dans l'exemple précédent, savoir que le chien avait été le compagnon du personnage depuis longtemps). Nous pouvons observer que le texte de la figure 7.2 contient des phrases qui projettent des états affectifs à partir de ce qui a été fait à un certain moment de la sortie. La phrase "Pari réussi, malgré le vent et les -30 °C en ressenti au sommet" indique une réaction favorable qui a émergé lors de l'arrivée au sommet. Donc, afin d'adapter la notion d'états affectifs vis à vis les récits d'activité, nous proposons donc une adaptation des définitions qui sont données ci-après :

**Définition 7.1 État affectif d'un récit d'activité** Unité de base. Il indique le ressenti qu'un personnage a vis à vis d'un événement appartenant à la séquence de l'activité.

Ces états affectifs ne cherchent pas représenter des réactions complexes

mais à identifier si elles sont favorables ou non pour le personnage.<sup>3</sup> Ils sont classifiés en trois types : *positifs*<sup>4</sup>, *négatifs*<sup>5</sup> et *états mentaux*<sup>6</sup>. Dans le cadre du récit d'activité, ces états peuvent être précisés comme suit :

**Définition 7.2 État affectif positif « + » pour le récit d'activité** Ressenti positif d'un personnage par rapport à un événement appartenant à la séquence de l'activité.

Pour l'activité de ski de randonnée, nous pouvons considérer comme *ressenti positif* des événements qui indiquent :

- La réussite de l'atteinte d'un objectif.
- Les signes indiquant le bon état d'esprit ou état physique de celui qui fait l'activité.
- Des conditions environnementales qui n'empêchent pas le déroulement de l'activité.

Dans le récit de la figure 7.2, nous observons quelques phrases indiquant une réaction favorable, tel que l'arrivée au sommet (“Pari réussi, malgré le vent et les -30 °C en ressenti au somme”), une exécution sans problèmes grâce aux conditions environnementales (“la montée se passe sans encombre”), lesquels ont même offert des moments qui ont rendu plus agréable la sortie (“Les trois chamois qu'on a croisés on l'air d'être du même avis ;-”).

**Définition 7.3 État affectif négatif « - » pour le récit d'activité** Ressenti négatif d'un personnage par rapport à un événement appartenant à la séquence de l'activité.

Nous pouvons citer comme *ressenti négatif* des événements de la séquence qui indiquent :

- Des échecs pour accomplir les objectifs.
- Les signes indiquant le mauvais état d'esprit ou état physique de celui qui fait l'activité.
- Des conditions environnementales qui empêchent le déroulement de l'activité.

---

<sup>3</sup> *Affect states do not attempt to describe complex emotional reactions or states of desire in the detail that inference mechanisms would require* [Lehnert, 1981]

<sup>4</sup> *Events that please* [Lehnert, 1981]

<sup>5</sup> *Events that displease* [Lehnert, 1981]

<sup>6</sup> *Mental states (with neutral affect)* [Lehnert, 1981]

La phrase “Bon le -19°C au parking nous fait hésiter” (Figure 7.2) exprime le mécontentement. Cette réaction provient des conditions désagréables pour les personnes lors du début de la sortie.

**Définition 7.4 État mental « M » du récit d'activité** Il décrit une motivation ou intention de valence neutre. Cette information n'est pas présente dans les données mais est inférée lors de la création du plan du récit d'activité.

Un exemple d'état mental est montrée dans la phrase “On se cale derrière le dernier rocher avant l'attaque finale pour se camoufler avec les gros gants, le casque, la doudoune” (figure 7.2). Il montre comment le skieur, en raison des conditions environnementales et de l'éventuelle arrivée au sommet, décide de se préparer en avance pour faire face au froid.

L'ensemble des états affectifs identifiés lors de l'analyse du récit de la figure 7.2 produit la structure montrée dans la figure 7.3. Nous observons comment les états affectifs positifs ou négatifs sont associés aux événements réalisés dans l'activité (comme l'arrivée au parking) ainsi que les conditions de l'environnement (le froid ou la qualité de la neige). L'attribution du ressenti positif ou négatif dépend de l'impact des événements sur l'exécution de l'activité. En revanche, les états mentaux indiquent des intentions qui émergent à partir des états affectifs, soit qu'ils provoquent une décision pour les surmonter s'ils sont négatifs, soit qu'ils motivent la poursuite de l'activité telle qu'elle était prévue. Ainsi, ce récit est abstrait en réactions qui vont émerger à partir de comment l'agent de l'activité (ici, le skieur) fait face à situations positives ou négatives.

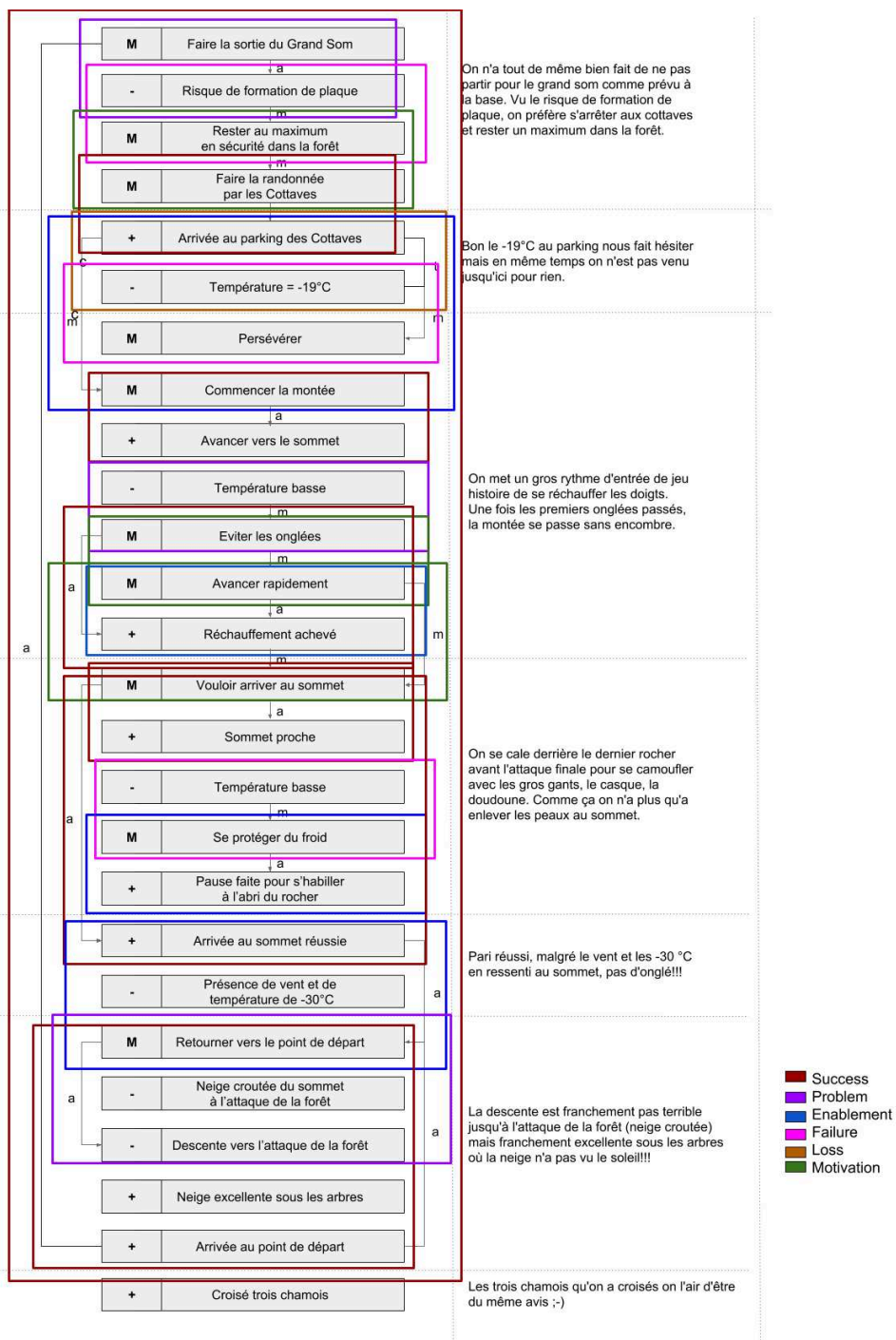


FIG. 7.3: Les *plots units* identifiés dans le texte de la figure 7.2

Donc, pour structurer un récit en suivant le modèle de Lehnert [Lehnert,

1981], nous n'avons pas besoin de connaître que le détail du ressenti des personnages. Cela est un grand avantage en comparaison avec les autres formalismes du récit présentés dans le Chapitre 2, dont il existe une construction plus complexe qui combine des informations plus précises (pensées, souvenirs, etc) et un schéma à suivre (par exemple, la pyramide de [Freytag, 1900] ou les fonctions du [Propp, 1970]).

Toutefois, le modèle de Lehnert a quelques limitations par rapport au contexte de la thèse. Celles-ci émergent du fait de travailler avec des récits dont le contenu est une activité humaine réelle. Lors de l'analyse, nous avons noté que les états mentaux n'étaient pas assez précis et que nous avons dû les raffiner. D'un côté, il y a des états indiquant des objectifs *obligatoires* à atteindre et qui sont forcément liés au parcours (comme l'arrivée au parking pour commencer la montée). D'un autre côté, il y a des états qui projettent des intentions qui *émergent* comme réponse aux conditions de randonnée (tel que vouloir surmonter des cailloux inattendus à cause d'un effondrement de neige).

L'analyse de fréquences des mots présentée dans la section 7.1.1 illustre cette distinction entre ce qui doit être fait et ce qui est survenu de manière inopinée. En tant qu'activité en extérieur, le ski de randonnée oblige les skieurs à réagir en permanence aux conditions environnementales, c'est pourquoi ces deux types d'intention correspondent à deux familles d'états mentaux réalisées en parallèle par les skieurs tout au long de la sortie. Nous posons les définitions suivantes pour différencier ces deux types d'états mentaux :

**Définition 7.5 État mental objectif « MOB »** Il décrit une motivation ou intention appartenant au déroulement attendu de l'activité.

Les *MOB* correspondent aux objectifs propres de l'activité. Dans le ski de randonnée, ceux correspondant aux étapes du parcours et incluant l'arrivée au sommet et le retour au point du départ. Ils peuvent être indiqués de manière explicite ou indirecte à partir du texte. Ainsi, nous pouvons inférer que la joie projetée dans la phrase "Pari réussi, malgré le vent et les -30 °C en ressenti au sommet" (figure 7.2) est le résultat d'avoir réussi à matérialiser cette intention implicite mais nécessaire.

**Définition 7.6 État mental ressenti « MRe »** Il décrit une motivation ou intention émergeant à partir des conditions environnementales imprévues.

Si nous observons la phrase "On met un gros rythme d'entrée de jeu histoire de se réchauffer les doigts." (Figure 7.2), l'intention derrière le surcroît d'effort est de diminuer la sensation de froid. Il s'agit d'une situation non

prévue dont il a fallu mettre une intention pour la surmonter. Grâce à ce *MRe* et à la manière de réagir pour les résoudre, l'activité peut continuer ou être modifiée, si besoin.

Nous avons déjà mentionné que la distinction entre positif et négatif est fortement liée à la réalisation des objectifs de base (les états mentaux appartenant à l'activité même) et à comment les événements inattendus permettent ou empêchent la progression de l'activité. Cependant, notre analyse permet d'observer qu'il y a des informations dans le récit d'une sortie qui n'affectent pas la réalisation de l'activité, mais qu'ils fonctionnent comme référence sur le contexte. Par exemple, la phrase "Les trois chamois qu'on a croisés on l'air d'être du même avis ;-)" (figure 7.2) apporte uniquement une référence amusante au récit, mais le fait de la mentionner n'a pas d'impact profond, positif ou négatif vis à vis du déroulement de l'activité.

Ainsi, pour préciser ce rôle du ressenti en fonction du rôle informatif de chaque état affectif, nous proposons une classification propre au domaine de l'activité de randonnée :

**Définition 7.7 *Navigation*** États affectifs exprimant la progression vis à vis des objectifs de l'acteur pendant l'activité

*Navigation* contient tous les états affectifs ayant un rapport avec le déroulement de l'activité même. Nous identifions dans cette catégorie les objectifs à accomplir (sous forme d'états mentaux du type *MOB*) et les réactions positives ou négatives qui émergent de ceux-ci. Les états affectifs du navigation n'arrivent pas par surprise, au contraire, ils sont connus ou attendus en avance. La phrase "la descente est franchement pas terrible jusqu'à l'attaque de la forêt" indique la réalisation réussie d'une étape obligatoire du parcours.

**Définition 7.8 *Environnement*** États affectifs positifs ou négatifs uniquement liés à des événements de l'environnement, en excluant les états mentaux

*Environnement* contient toutes les références concernant le contexte de réalisation de l'activité. Ici, nous pouvons inclure la température, les conditions du terrain ou d'autres aspects qui peuvent ou non affecter le déroulement de l'activité. Par exemple, l'indication de la température basse dans la phrase "Bon le -19°C au parking nous fait hésiter mais en même temps on n'est pas venu jusqu'ici pour rien."

**Définition 7.9 *Perception*** États affectifs qui traduisent un ressenti par rapport à l'environnement ou à la navigation

*Perception* présente des états affectifs qui émergent face aux conditions de l'environnement et qui sont trouvés lors de la réalisation de l'activité. Ces états affectifs positifs ou négatifs, ainsi que les états mentaux du type *MRes* répondent à ce qui n'a pas été prévu lors de la planification de l'activité (c'est à dire, les états affectifs *Navigation*).

Tous les concepts définis dans cette section servent à la construction d'un plan du récit d'activité. Ils prennent en compte les spécificités du récit d'activité et celle des connaissances basées sur le contenu objectif (données capteur). La section suivante présente comment ces états affectifs vont être unis pour composer le plan du récit.

### 7.2.1.2 Définition des *plots* pour le récit d'activité

Les états affectifs par eux mêmes ne construisent pas la structure d'un récit. Ceux-ci sont unis au travers de connexions que [Lehnert, 1981] identifie comme *liens causaux*.<sup>7</sup> Concrètement, ces liens servent pour signaler qu'un état mental a provoqué l'apparition d'un état affectif positif (par exemple, vouloir visiter un sommet cause une montée) ou négatif (vouloir se réchauffer provoque se couvrir du froid). De même, nous signalons le cas inversé : un état mental peut émerger d'un état affectif positif (tel que vouloir retourner au point de départ lors d'avoir arrivée au sommet) ou un état affectif négatif (un chemin bloqué cause vouloir modifier la route).

Il y a 4 types de liaison connectant une paire d'états affectifs : *Motivation* (m), *Actualisation* (a), *Équivalence* (e) et *Terminaison* (t). Sans donner une définition formelle, [Lehnert, 1981] donne des indications pour les distinguer :

- *Motivation* est utilisé pour lier un état affectif positif ou négatif vers un état mental et montre comment la réaction à un certain événement a causé la naissance d'un but ou d'un souhait pour obtenir quelque chose ou pour réaliser une action<sup>8</sup>.

**Exemple** Un personnage devient malade (état négatif) donc, il décide d'aller au médecin (état mental). La maladie *incite* au personnage à chercher attention médicale.

- *Actualisation* est utilisé pour lier un état mental vers un état affectif et indique comment un but est traduit en un événement portant un effet positif ou négatif<sup>9</sup>.

---

<sup>7</sup>Each link describes an oriented arc between two affect states [Lehnert, 1981]

<sup>8</sup>Motivation links describe causalities behind mental states [Lehnert, 1981]

<sup>9</sup>Actualisation links describe intentionalities behind events. [Lehnert, 1981]

**Exemple** Une couple veut fêter son anniversaire (état mental) donc, ils décident d'aller diner (état affectif positif). L'intention de célébrer *est traduite* ou *actualisé* en l'arrivée au restaurant.

- *Terminaison* relie deux états mentaux ou deux états affectifs positifs ou négatifs pour indiquer quand l'effet d'un état affectif est terminé ou obsolète<sup>10</sup>

**Exemple** Un chien perdu (état affectif négatif) a été retourné à son propriétaire (état affectif positif). La tristesse de ne pas avoir retrouvé l'animal *termine* ou *est substitué* par le fait de le voir à nouveau.

- *Équivalence* connecte deux états mentaux ou deux états affectifs positifs ou négatifs lorsqu'un même événement peut causer différentes réactions<sup>11</sup>.

**Exemple** Un personnage est promu à un nouveau poste (état affectif positif). Cet même événement *implique* ou *est égale* l'obtention d'un meilleur salaire (état affectif positif).

Il est possible d'identifier les liens causaux dans l'analyse des récits grâce à l'utilisation des marqueurs textuels, comme "donc" ou de les inférer d'après comment les événements du récit sont racontés. En revanche, étant donnée que nous allons travailler avec des données brutes, le modèle de [Lehnert, 1981] devient limité. Il n'y a pas de liens qui puissent être récupérables directement au travers de l'observation directe. Plus concrètement, l'ensemble d'information collectée ne présente pas aucune indication de causalité. Nous ne pouvons pas savoir avec certitude, par exemple, *pourquoi* une pause (comprise comme une arrête d'activité dans les données du déplacement) a eu lieu. Donc, vu la manque de ces liens sur les données, il faudra les créer liens à partir de ce qui a été enregistré. Ainsi, nous proposons la définition des liens causaux suivants :

**Définition 7.10 Lien causal du récit d'activité** Relation inférée entre une paire d'états affectifs attribués aux événements de la séquence de l'activité.

Nous adaptons également les définitions des 4 types de liens causaux :

---

<sup>10</sup> *The termination link is not used when an event per se is terminated, but when the affective impact of that event is surplanted or displaced.* [Lehnert, 1981]

<sup>11</sup> *Equivalent events and states are linked when multiple perspectives of a single affect state can be separated* [Lehnert, 1981]



**Définition 7.11 Motivation « m » pour le récit d'activité** Lien entre un ressenti (positif ou négatif) et un état mental, lequel montre comment un ressenti positif ou négatif fait émerger une motivation

**Exemple** Arriver au sommet (état affectif positif) déclenche ou motive vouloir retourner au point de départ (état mental MOb)

**Définition 7.12 Actualisation « a » pour le récit d'activité** Lien entre un état mental et un ressenti (positif ou négatif), lequel montre comment une intention peut amener à un événement de ressenti positif ou négatif

**Exemple** Vouloir manger (état mental MRes) pourrait amener ou être actualisé avec la prise d'une pause (état affectif positif)

**Définition 7.13 Terminaison « t » pour le récit d'activité** Lien entre deux états mentaux ou deux ressentis (positif ou négatif), qui modélise le fait que l'affect du deuxième événement prend le pas sur le premier.

**Exemple** La sensation de fatigue (état affectif négatif) *termine* ou *est substitué* avec la pris d'une pause (état affectif positif)

**Définition 7.14 Équivalence « e » pour le récit d'activité** Lien entre deux états mentaux ou deux ressentis (positifs ou négatifs) qui exprime des perspectives différentes à partir d'un même événement.

**Exemple** Le retour au point du départ (état affectif positif) implique ou *est égale* à la finalisation de la sortie (état affectif positif)

La figure 7.4 présente à nouveau l'analyse du récit trouvé dans la figure l'exemple du texte de la figure 7.2. Toutefois, celle-ci a été reformulée pour inclure le raffinement des états mentaux, ainsi que la classification en fonction du contenu (*Navigation*, *Ressenti* et *Environnement*). Les états affectifs sont maintenant distribués en différentes colonnes selon son rôle informative. Les trois contiennent des états affectifs positifs ou négatifs. Cependant, ceux du type *MOb* seront localisés dans *Navigation*, alors que ceux *MRes* dans la colonne *Perception*. Nous pouvons noter aussi que la plupart de tous eux se trouve sur la colonne *Navigation*. Cela illustre (et confirme) que le déroulement de la sortie est le plus important à transmettre dans les récits d'activité.

Cette figure inclut aussi les liens causaux pour mettre en relation tous états affectifs. Ils sont présents partout, afin de montrer *pourquoi* un certain événement a eu lieu. Grâce à eux, nous pouvons observer de manière plus

claire l'influence de l'environnement sur l'exécution de l'activité (*Navigation et Perception*). La phrase "On met un gros rythme d'entrée de jeu histoire de se réchauffer les doigts" projette deux états affectifs : un négatif (température basse) et un état mental Ressenti portant l'intention de mettre un effort supplémentaire pour éviter les onglées. Ainsi, ils sont connectés par le lien *motivation*. La phrase "On se cale derrière le dernier rocher avant l'attaque pour se camoufler avec les gros gants, le casque, la doudoune" indique d'un côté l'intention de se protéger du froid (état mental Ressenti) et d'un autre de couvrir le mieux possible avant d'arriver au sommet (état affectif positif). Pour cet raison, ils sont liés avec *actualisation*. Finalement, nous avons la phrase "Bon le -19° au parking nous fait hésiter mais en même temps on n'est pas venu jusqu'ici pour rien". Ici, la sensation de réussite causée par l'arrivée au parking (état affectif positif) est immédiatement remplacée ou *terminé* par la gêne occasionnée par le froid (état affectif négatif). Ceux-ci sont unis par *terminaison*.

L'union de paires d'états affectifs via des liens causaux permet de représenter des situations particulières décrites dans le récit. Par exemple, imaginons que dans un récit textuel fictif un personnage veut demander un prêt au banc et le banc l'approuve. Cette situation peut être décomposée en états affectifs ainsi : vouloir demander le prêt (état mental) est actualisée en l'obtention de l'argent. L'abstraction de ces situations en termes d'états affectifs et liens causaux est nommée *plot units* ou *plots unitaires* par Lehnert [Lehnert, 1981].

Dans le cadre de l'utilisation des plots unitaires pour le récit d'activité (qui sont composés des états affectifs vécus ou inférés à partir de données), ceux-ci peuvent être redéfinis comme suit :

**Définition 7.15 *Plot unitaire du récit d'activité*** Combinaison d'états mentaux et d'états affectifs du récit d'activité attribués à des événements vécus ou observés, lesquels sont reliés pour représenter des situations atomiques selon une typologie (Figure 2.8). Les plots unitaires ne comportent que deux états affectifs reliés par un seul lien.

**Exemple** Plot unitaire *Success* qui représente la prise d'une pause : Vouloir rester (état mental Ressenti) est actualisé avec l'arrête du déplacement (état affectif positif)

Les *plots units* peuvent être assemblés pour représenter des situations plus complexes qui sont nommées *plot complexe* (Figure 7.3). Si nous revenons à l'exemple du prêt, nous pourrions ajouter des autres états qui aident à comprendre la nature de la situation. Le même personnage veut demander un prêt au banc car il a été licencié. Le banc approuve la demande et grâce à

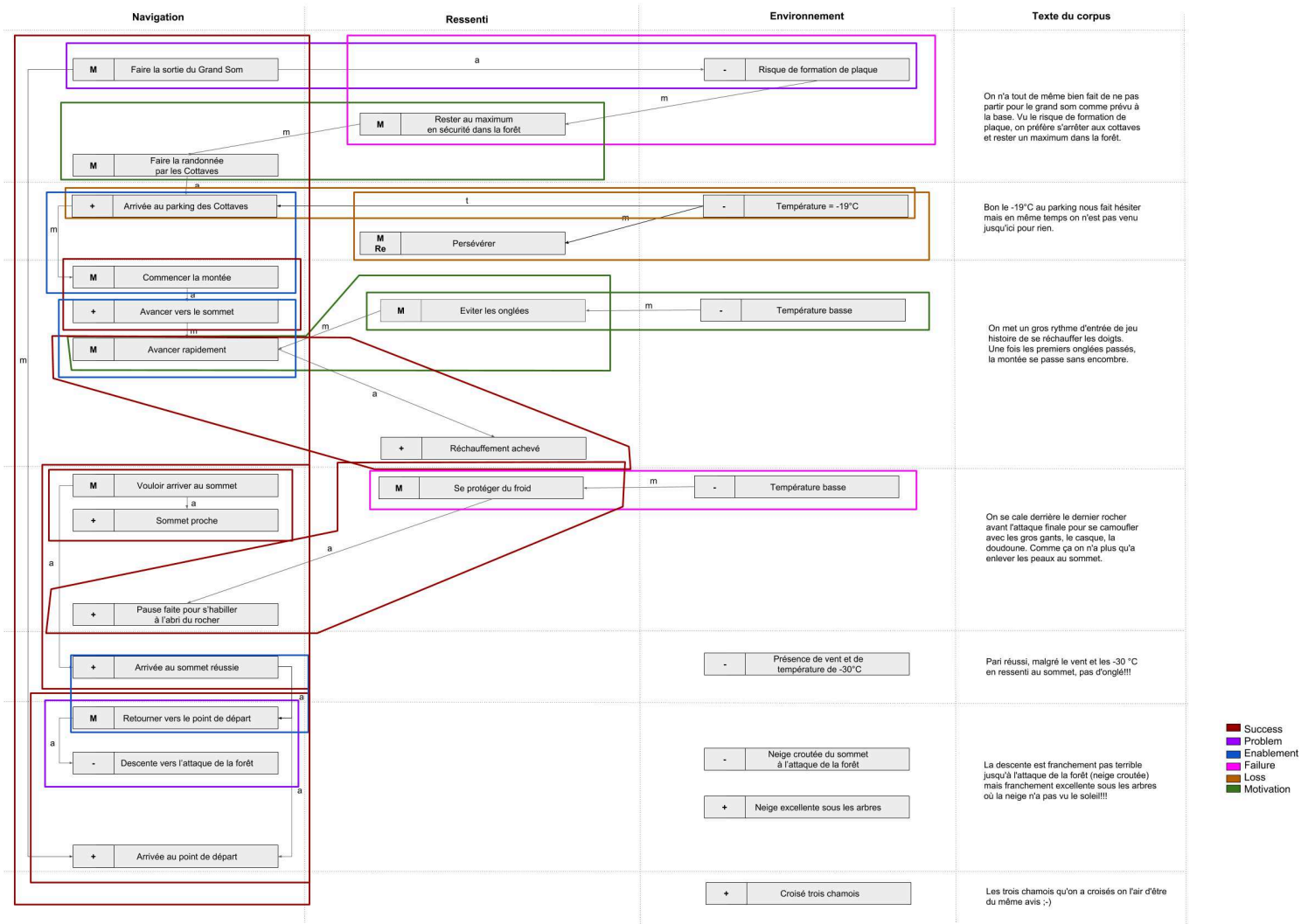


FIG. 7.4: Les plots units identifiés dans le texte de la figure 7.2 et présentés dans la figure 7.3, organisés selon les rôles informatifs associés aux états affectifs.

l'argent obtenu, il peut survivre jusqu'à trouver un autre poste. Cette situation est décomposée en états affectifs ainsi : perdre son travail (état affectif négative) motive vouloir demander le prêt (état mental) ; cette intention est actualisée en l'obtention de l'argent (état affectif positif), en terminant ainsi la situation négative de n'avoir plus de ressources pour survivre.

C'est ainsi que nous présentons la définition suivante pour les utiliser dans la modélisation de récits d'activité :

**Définition 7.16 *Plot complexe du récit d'activité*** Combinaison des plots units du récit d'activité reliés pour représenter des situations complexes selon une typologie.

**Exemple** Plot complexe *Intentional problem resolution* qui représente la prise d'une pause après un grand effort : Vouloir se reposer après l'activité exténuante (plot unitaire *Failure*). L'intention de rester motive arrêter le déplacement (*Success*) et grâce à cela, la fatigue disparaît (*Resolution*)

Si nous retournons à la figure 7.4, nous pouvons observer les plots unitaires identifiés pendant l'analyse. Donc, au lieu d'avoir un ensemble d'état affectifs connectés, nous pourrions faire une abstraction à un niveau plus haut. Un récit d'activité peut être compris comme un enchaînement de situations vécues ou faites par un agent (dans ce cas, le skieur). Dans la figure, celles-ci sont délimitées par lignes de couleur qui se superposent entre elles. Cela qui permet d'illustrer la complexité derrière d'un récit et comment tout le contenu est organisé et connecté de manière à faire progresser la raconté. Par exemple, arriver au parking (représenté comme *Success*) déclenche la montée au sommet (représenté comme *Enablement*). Dans le cas de notre contexte du travail, cette approche modulaire garantie l'obtention d'une structure causale à partir de seulement ce qu'il y a disponible, sans risquer l'apparition d'un problème d'incomplétude.

Pour résumer, l'adaptation du modèle d'analyse de [Lehnert, 1981] donne les bases idéales pour construire des récits dont le contenu dépend directement de ce qui a été enregistré dans les données. Il s'agit d'une approche flexible qui permet de donner une structure narrative aux événements portant seulement sur ce que les pratiquants de l'activité ont vécu. Toutefois, l'aspect le plus important est la capacité de lier les événements à partir des réactions qui émergent précisément du vécu et sans suivre une trame particulière. Bien que la liste des événements ne représente pas exhaustivement les point de vue des protagonistes sur la sortie, c'est à partir de ces réactions provenant du vécu que nous allons faire émerger le ressenti de l'expérience. La construction du

plan du récit à partir des données objectives et du ressenti de l'expérience et l'étape permettant de passer du monde *objectif*, dans lequel l'activité se passe dans la vie réelle et le monde *subjectif*, dans lequel l'expérience vécue de l'activité est retranscrite pour être transmise (cf. Figure 4.6).

## 7.2.2 Construction du modèle du plan du récit

Passer d'une séquence d'événements de l'activité en un plan du récit exprimé sous le formalisme des plots requiert de sélectionner les événements de l'activité à présenter dans le récit (c'est à dire de leur donner un rôle d'état affectif et de les lier) et de les organiser les uns par rapport aux autres pour construire une organisation logique et cohérente vis à vis de l'activité et du ressenti à transmettre.

L'analyse présente dans le tableau 7.12 montre que le contenu (les événements du récit) diffère d'un type de récit à l'autre. La section 7.2.2.1 s'intéresse à ces contenus et la section 7.2.2.2 présente comment structurer le contenu pour aboutir au plan du récit.

### 7.2.2.1 Contenu selon les types de récit

Nous avons observé pendant l'analyse des récits du corpus qu'il existait des différences dans les récits selon le ressenti à transmettre. L'inclusion de certains événements dépend donc du ressenti général à transmettre et de s'ils font partie ou non du déroulement prévu (comme des points de référence). Dans notre étude, les récits ont deux caractéristiques. La première est directement liée à la navigation (définition 7.7). Il y a les récits qui relatent une sortie dont les objectifs de parcours ont été atteints et ceux pour lesquels cela n'a pas été le cas. La seconde caractéristique est celle du ressenti global que le narrateur veut transmettre : positif ou négatif. Dans cette thèse ces deux caractéristiques sont celles qui traduisent le but communicationnel.

|         | Conforme aux objectifs de parcours | Modifié par rapport aux objectifs du parcours |
|---------|------------------------------------|---|
| Positif | récits positifs sans contretemps   | récits positifs avec défis                    |
| Négatif | récits nuancés                     | récits négatifs                               |

TAB. 7.13: Catégorisation des types de récit

Les types de récits identifiés dans l'analyse sont détaillés dans le Tableau 7.13. Ces caractéristiques permettent de définir le contenu. Un récit

d'activité peut contenir autant d'informations que le narrateur souhaite apporter. Par exemple, il peut détailler l'évolution des conditions météorologiques sur tout son parcours. Cependant, nous cherchons ici à définir le minimum des informations à apporter pour transmettre un ressenti sur une sortie identifiable. Deux types d'états affectifs sont à considérer. Les états affectifs qui sont présents afin d'identifier si la sortie s'est déroulée avec ou sans modifications des objectifs de parcours et les états affectifs qui sont pertinents uniquement pour transmettre le ressenti.

**Définition 7.17 *États affectifs indispensables*** États affectifs qui portent sur les éléments de Navigation (Définition 7.7)

Pour tous les types de récit il est nécessaire d'apporter des connaissances sur le parcours et dans le cas où le récit porte sur une modification, il faut que l'information sur la cause et la modification soient également présentes. Dans une sortie de randonnée à ski, chaque état affectif indispensable correspond alors à l'arrivée à un objectif (point de départ, arrivée, sommet) et à un problème qui a modifié le parcours (conditions environnementales ou physiologiques). En terme de contenu, des récits positifs avec défis et nuancés contiennent les mêmes événements, c'est à dire le déroulement prévu et les contrariétés rencontrées, la différence est dans la manière de les présenter.

**Définition 7.18 *États affectifs pertinents pour le ressenti*** États affectifs qui servent pour l'interprétation de la qualité de l'expérience vécue

Comme le montre l'analyse du corpus fréquentiel, en plus de la navigation, les éléments les plus fréquents sont les conditions environnementales. Ce contenu ajoute des connaissances sur l'environnement pour "projeter" le lecteur dans le ressenti du protagoniste. L'analyse du contenu par type de récits (section 7.1.3) montre également que les récits positifs avec défis et les récits nuancés apportent plus d'informations sur les déplacements et les conditions de randonnée.

Un seul récit négatif était présent dans le corpus analysé, il n'est donc pas pertinent de définir les états affectifs nécessaires et pertinents à partir de son analyse. Nous choisissons d'ajouter aux états affectifs indispensables la condition d'échec et la position (repère de navigation) où celle-ci a eu lieu.

La table 7.14 présente une synthèse du contenu pour chaque type de plan de récit.

|                                | <b>Positif sans<br/>contretemps</b>                                       | <b>Positif avec<br/>défis ou<br/>Nuancé</b>   | <b>Négatif</b>  |
|--------------------------------|---|---|---|
| États affectifs indispensables | Arrivée au point de départ, Arrivée au sommet, Arrivée au point d'arrivée | Arrivée au point de départ, Arrivée au sommet, Arrivée au point d'arrivée, Condition non prévue | Arrivée au point de départ, Arrivée au lieu d'échec (point de retour), Condition menant à l'échec, Arrivée au point d'arrivée |
| États affectifs pertinents     | -   | Arrivée aux étapes, Conditions de randonnée   | Arrivée aux étapes, Conditions de randonnée   |

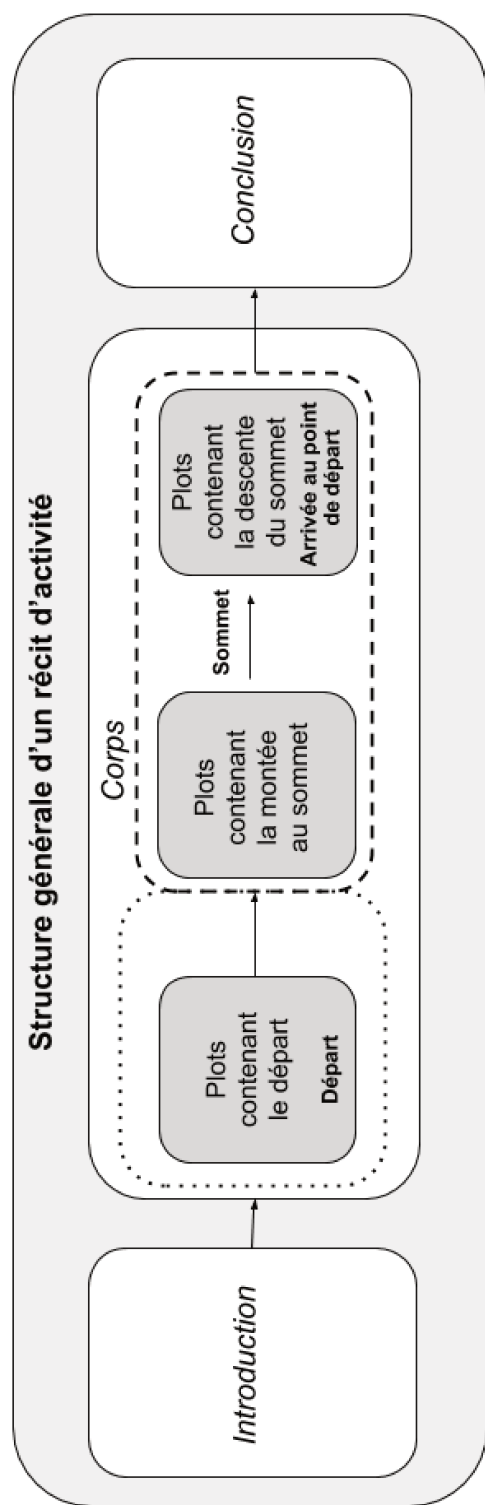
TAB. 7.14: Bilan de contenu selon le type de ressenti

### 7.2.2.2 Structure selon les types de récit

Le contenu du plan du récit que nous cherchons à construire constitue le corps du récit d'activité tel qu'identifié dans la section 7.1.2. Comme il a été mentionné précédemment, le plan est constitué d'au moins trois états affectifs indispensables : le départ, l'arrivée au point culminant (sommet ou lieu d'échec) et le retour au point de départ (Figure 7.5). Le départ est partagé par tous les types de plan, c'est dans la montée vers le sommet et la descente que va être faite la différence dans la structure.

Les états affectifs sont encapsulés dans des unités de plots qui sont ensuite organisés les uns par rapport aux autres jusqu'à fournir un plan complet. Nous avons indiqué brièvement qu'il y a 15 combinaisons de plots unitaires possibles qui sont fréquemment trouvés pendant l'analyse de textes écrits (figure 2.8). En raison de l'impossibilité de recueillir certains états mentaux, un nombre réduit d'unités de plot est utilisable pour le récit d'activité. Chaque plot a une sémantique qui lui est propre vis à vis du récit. Pour encapsuler les états affectifs identifiés précédemment, nous avons utilisé 5 d'entre eux dans nos plans du récit d'activité.

- *Success* Lien entre la survenue d'un événement observé ou fait par l'acteur conforme à un état mental MOB. Cet état mental est inféré comme étant l'objectif souhaité. L'événement est alors ressenti positivement "+" par l'acteur. Le plot unitaire *Success* est utilisé pour encapsuler



- ..... Partie de la structure du récit partagée par tous les types de plan
- Partie de la structure du récit où les plans vont se différencier selon le ressenti

FIG. 7.5: Structure générale d'un récit d'activité



tous les états affectifs que nous considérons comme conformes au déroulement prévu et pour lequel un ressenti positif veut être communiqué. Dans le cas des récits positifs avec un déroulement sans contretemps, tous les états affectifs (indispensables et pertinents) sont encapsulés dans trois unités de plot et deux plots complexes.

- *Failure* Lien entre la survenue d'un événement observé ou fait par l'acteur par rapport à un état mental MOB. Cet état mental est inféré comme étant l'objectif souhaité. L'événement est alors ressenti négativement "-" par l'acteur car il ne correspond pas à l'état mental MOB (il doit y renoncer). Le plot unitaire *Failure* est utilisé pour encapsuler les états affectifs liés aux échecs par rapport aux objectifs fixés. Dans le cas des récits de sortie de ski de randonnée, ce plot unitaire n'est utilisé que pour les récits négatifs (qui sont les seuls à avoir un échec), ils encapsulent l'arrivée au lieu d'échec (état affectif négatif).
- *Complex negative event* Lien entre la survenue de deux événements ressentis négativement par le protagoniste. L'unité de plot *Complex negative event* est utilisée pour encapsuler les deux événements négatifs liés à la définition de l'échec de la randonnée dans le cas des récits négatifs (qui sont les seuls à avoir un échec), le lieu et les conditions.
- *Problem* Lien entre un événement et un état mental qui a pour but de réagir vis à vis de cet événement. L'événement est considéré comme porteur d'un ressenti négatif car il entraîne une modification de "correction" par rapport à ce qui était prévu. L'état mental est inféré comme un MOB (voir définition 7.5) lorsque la correction entraîne un changement d'objectif de navigation et un MRe (voir définition 7.6) lorsqu'il entraîne un événement de réaction aux conditions sans changement d'objectif de navigation. Ce plot unitaire est utilisé pour encapsuler les états affectifs portant sur les conditions non prévues pour les récits avec défis ou Nuancés. La différence entre un récit nuancé et un récit positif avec défis est formalisée ici par le type d'état mental inféré. Si c'est un MOB, c'est un récit positif avec défis sinon, c'est un récit nuancé.
- *Intentional Problem Resolution et Fortuitous Problem Resolution* Lien entre une unité de plot qui exprime un problème (*Problem*) et sa résolution. La résolution d'un problème rencontré est ressentie positivement et est donc représentée sous la forme d'un état affectif positif. Celui-ci est soit le résultat d'une action du protagoniste, soit une modification dans les conditions de randonnée indépendante de sa volonté. Dans le cas où le protagoniste est à l'origine de la résolution (par exemple il s'est couvert pour faire face à une chute de la température), le plot complexe est dit *Intentional Problem Resolution*, sinon c'est un *Fortuitous*

*Problem Resolution.* Ces plots sont utilisés pour les récits positifs avec défis ou nuancés et pour les récits négatifs. Ils lient les conditions de randonnée lorsque celles-ci ont été modifiées par rapport à un problème.

|                               | <b>Positif sans<br/>contretemps</b>                                       | <b>Positif avec<br/>défis ou<br/>Nuancé</b>   | <b>Négatif</b>   |
|-------------------------------|---|---|--|
| <i>Success</i>                | Arrivée au point de départ, Arrivée au sommet, Arrivée au point d'arrivée | Arrivée au point de départ, Arrivée au sommet, Arrivée au point d'arrivée, Arrivée aux étapes | Arrivée au point de départ, Arrivée au point d'arrivée                                   |
| <i>Problem</i>                | -   | Condition de randonnée non prévue   | Arrivée aux étapes, Conditions de randonnée  |
| <i>Problem Resolution</i>     | -   | Correction sur les conditions de randonnée  | Conditions de randonnée  |
| <i>Failure</i>                | -   | -   | Arrivée au lieu d'échec (point de retour)  |
| <i>Complex negative event</i> | -   | -   | Arrivée au lieu d'échec (point de retour) et Condition de randonnée ayant mené à l'échec |

TAB. 7.15: Plots pour l'encapsulation des états affectifs liés au contenu

Une fois tous les états affectifs liés au contenu des événements encapsulés dans des plots, ceux-ci doivent être liés pour pouvoir constituer un graphe (le plan du récit complet). Nous avons cherché à définir ces liens par les résultats de l'analyse de textes de corpus (réalisée comme pour l'exemple de la section 7.2.1.1). Les liens causaux utilisés pour exprimer les relations entre les plots d'encapsulation du contenu ont pour conséquence la création de nouveaux

plots complexes. Il existe plusieurs solutions pour lier les événements entre eux. Lehnert et al. indiquent d'ailleurs qu'il est possible que d'autres plots complexes existent en plus de ceux qui sont présentés dans [Lehnert, 1981]. C'est d'autant plus vrai pour l'analyse de récit d'activité pour lesquels les états mentaux ne sont que rarement décrits. Ainsi, lors de l'analyse des récits du corpus nous avons choisi de ne pas chercher toutes les structures possibles mais d'identifier celles qui étaient caractéristiques des liens entre les faits marquants de la sortie (encapsulés dans les unités de plot).

Après cette analyse seuls trois types de plots (en plus de ceux exposés précédemment) sont présents dans les récits de randonnées :

- *Enablement* Lien qui unit un événement à un état mental. L'état mental n'est devenu possible que par l'arrivée de l'état affectif (lié à l'événement) qui est alors perçu comme quelque chose de positif.
- *Perseverance* Lien qui unit deux états mentaux pour exprimer l'équivalence entre les deux. Le second état mental est alors une réaffirmation du premier.
- *Nested subgoals* Lien de buts/sous-buts entre des états affectifs. Un état mental (objectif) est ici atteignable via une succession d'états mentaux intermédiaires. Les états affectifs qui permettent d'actualiser les états mentaux des sous-buts (*Success*) actualisent également l'état mental parent

Dans le cas d'un récit positif sans contretemps, le contenu est représenté par un ensemble de plots *Success*. Les liens entre ces plots traduisent alors une succession d'étapes (étape 2 après le succès de l'étape 1) ou une structuration hiérarchique entre les étapes (l'étape 3 considérée réussie signifie qu'il faut réussir l'étape 4 et l'étape 5). Dans le cas où le contenu est limité aux états affectifs indispensables, seul le plot *enablement* est nécessaire pour permettre les liaisons entre les *success* terminés par l'arrivée au point de départ et initiés par l'objectif de monter jusqu'au sommet et entre les *success* terminés par l'arrivée au sommet et initiés par l'objectif de redescendre au point de départ. Le plan du récit obtenu est alors comme celui présenté sur la figure 7.6.

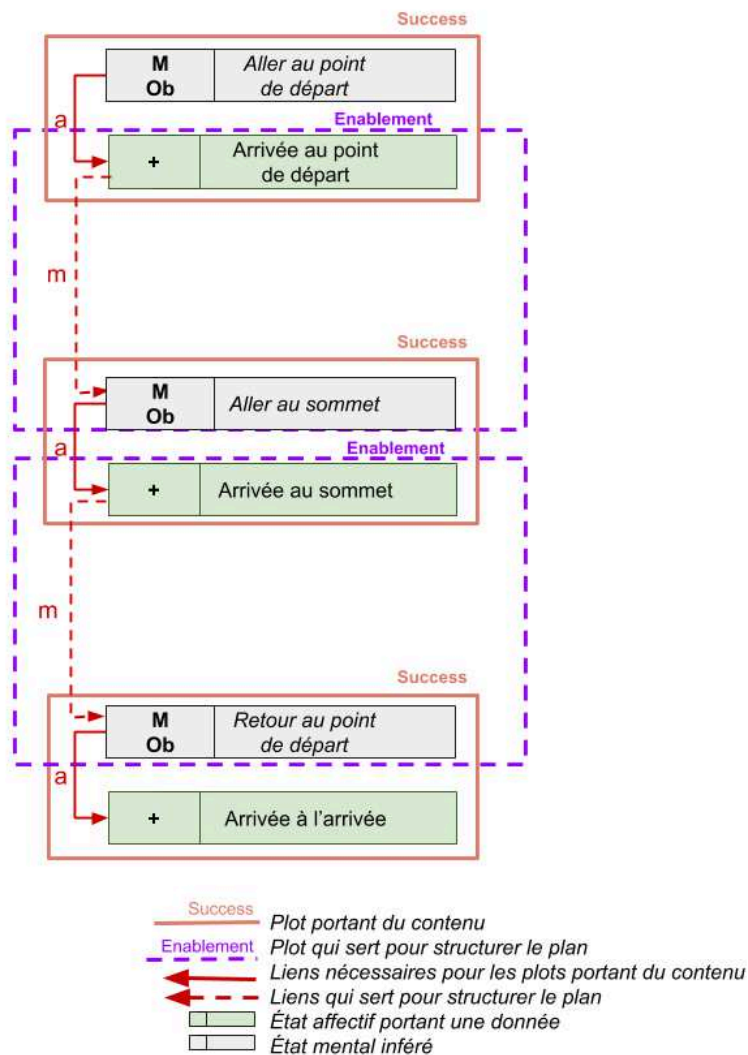


FIG. 7.6: Le plan de récit pour un récit positif sans contretemps

Dans le cas d'un récit positif avec défis ou nuancé, en plus des plots *Success* pour le point de départ, le sommet et le retour au point de départ, il y a des plots *Success*, des plots *Problem* et des plots *Problem résolution* pour les différentes étapes à intégrer. Ces différents plots sont à intégrer soit pendant la montée, soit pendant la descente, ce qui signifie que la progression de l'une ou l'autre (ou les deux) est détaillée. Une manière d'exprimer le lien entre les étapes est d'ajouter des MOB intermédiaires pour chaque étape, qui sont liés avec des liens de motivation (m) à l'état mental parent (*Aller au sommet* ou *Retourner au point de départ*). La figure 7.7 montre comment peuvent être intégrées des étapes pendant la montée. Dans le cas d'un récit dont le

déroulement suit ce qui est prévu, les différentes étapes sont des *Success* au sein desquelles peuvent s'intégrer les plots correspondants à la rencontre d'un problème :

- Si les conditions de randonnée présentent un problème pour lequel le protagoniste choisit de ne pas faire de correction (par exemple, averse de neige désagréable mais pas dangereuse) alors le plot *Problem* est à ajouter.
- Si les conditions de randonnée présentent un problème pour lequel le protagoniste choisit de ne pas faire une correction mais qu'il perçoit positivement l'arrêt de la condition négative, alors c'est un plot *Fortuitous problem resolution* qu'il faut lier au *Success*.
- Si les conditions de randonnée présentent un problème pour lequel le protagoniste choisit de faire une correction alors le plot *Intentional problem resolution* qui en découle est un sous but du *Success* de l'étape.

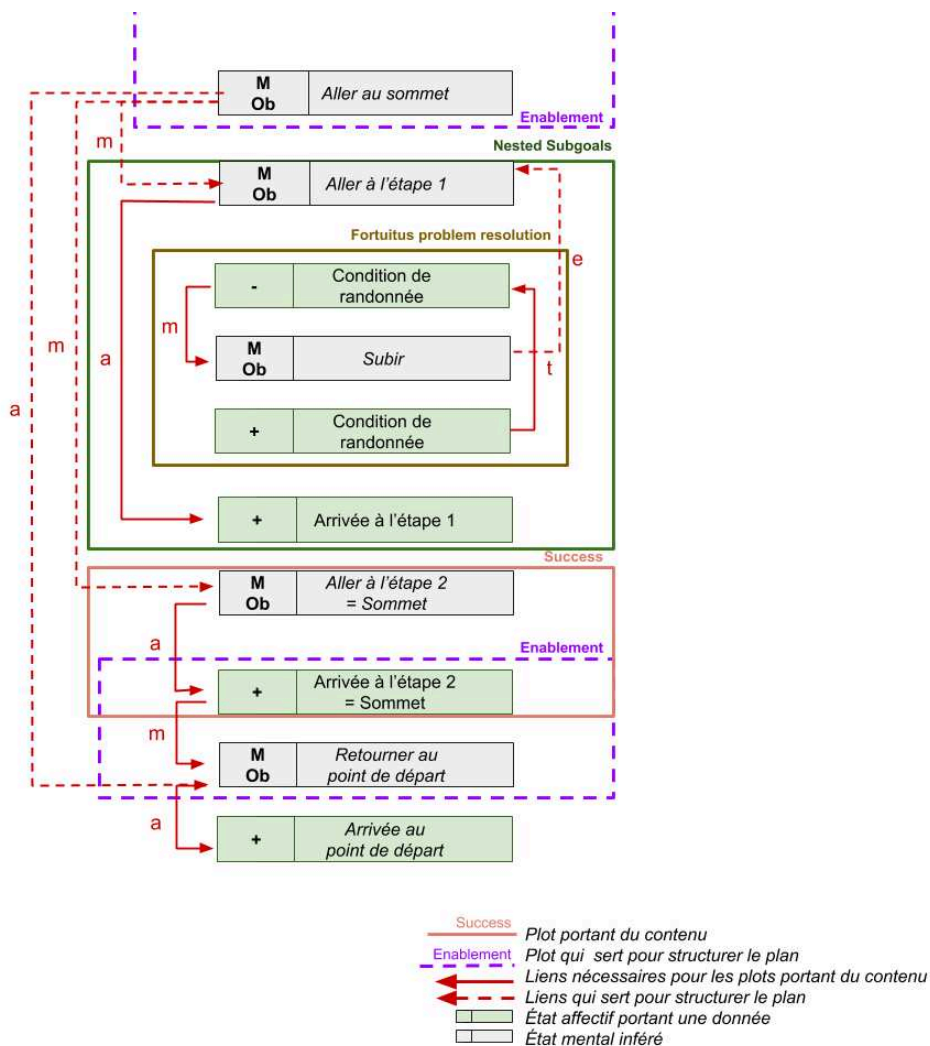


FIG. 7.7: Intégration d'étapes pendant la montée

La relation entre ces plots et le plot *Success* dans lequel ils s'intègrent est alors réalisée à l'aide de liens entre les états affectifs en fonction de leur sémantique (avec des liens causaux). Ainsi lorsque le protagoniste choisit de ne pas faire de correction, implicitement il réaffirme que son état mental reste inchangé (*Aller à l'étape n*). Dans ce cas, le lien avec le plot unit *Success* est réalisé en ajoutant un lien d'équivalence entre l'état mental du *Problem* ou *Fortuitous problem resolution* (qui exprime que le protagoniste ne fait pas de correction) et l'état mental du *Success* (qui exprime l'objectif de l'étape). Le plot ainsi créé est le plot *Perseverance*. La figure 7.8 présente graphiquement cette relation entre un plot *Fortuitous problem resolution* et le plot *Success* dont il dépend.

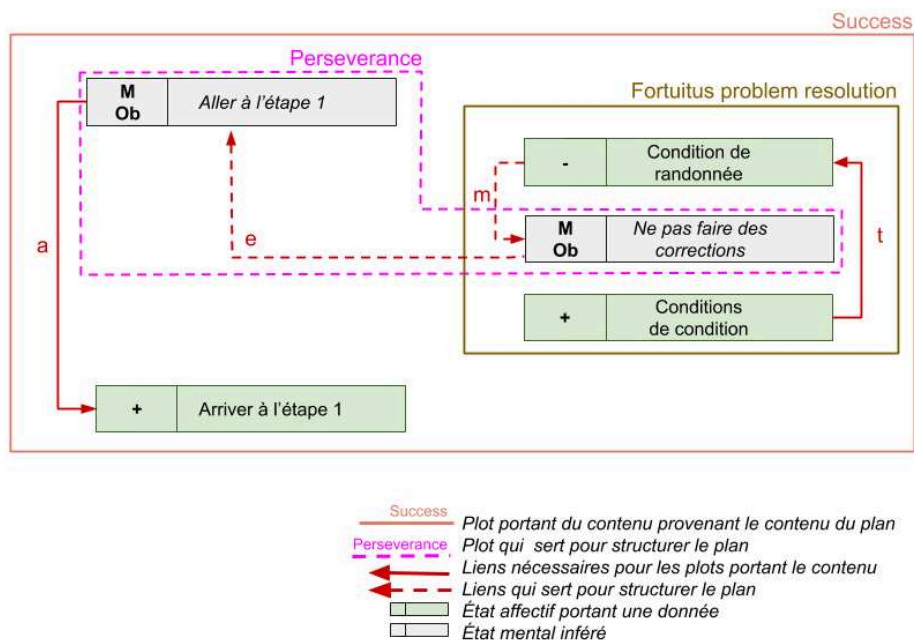


FIG. 7.8: Intégration de plots avec difficultés et sans modification des conditions par le protagoniste

Dans le cas où le protagoniste choisit de réagir, alors la résolution (*Intentional problem resolution*) devient un sous but de l'étape. Un lien de motivation (m) entre les états mentaux permet alors de définir de nouveau un *Nested subgoal*.

Enfin, dans le cas où le récit est un récit négatif, c'est à dire pour lequel il y a eu un abandon, la structure est totalement différente car elle repose sur des plots avec des ressentis négatifs et s'articule autour de l'arrivée au sommet mais à un point d'échec. En raison du faible nombre de textes disponibles pour identifier la structure de ce type de récit, nous n'avons pas fait d'analyse de textes pour définir la structure, celle-ci est issue uniquement d'une réflexion sur la sémantique des plots et leur corrélation avec la mise en avant des caractéristiques de l'échec.

Le début est le même que pour les autres types de récit car le protagoniste ne commence pas sa sortie en sachant que celle-ci va échouer. L'arrivée au point de départ est donc encapsulée dans un plot *Success*, suivi par un état mental qui exprime l'objectif d'aller au sommet. Cet état mental est alors lié au *Success* de l'arrivée au point de départ par un *Enablement*. Le même schéma est respecté pour le retour au point de départ. Par rapport à la structure du récit positif sans contretemps (représenté graphiquement sur la figure 7.6), seul le plot unit central (celui de l'arrivée au sommet) est changé.

Les états affectifs considérés à la place de l'arrivée au sommet correspondent à l'arrivée sur le lieu d'échec et sur les conditions de randonnée. Ces deux états affectifs sont encapsulés dans un même plot qui est un *Complex negative event*. Ce plot constitue l'origine de l'état mental *Retourner au point de départ* permettant ainsi de définir un plot *Problem*. Il met fin à l'état mental initial (*Aller au sommet*) et le remplace par *Retourner au point de départ*. Les deux états mentaux sont alors liés par un lien de terminaison (t) définissant alors un plot *Change of mind*. La figure 7.9 présente graphiquement le plan du récit que nous proposons pour les récits de ski de randonnée négatifs.

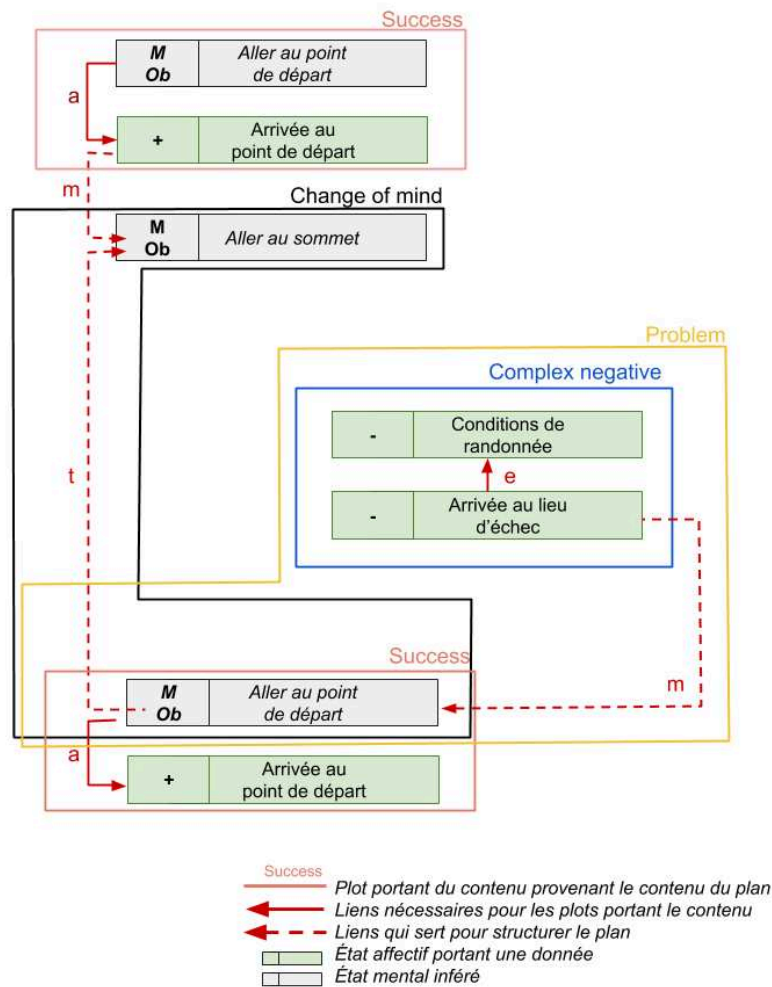


FIG. 7.9: Plan du récit pour les récits négatifs



### 7.2.2.3 Résumé du passage de la séquence d'événements valide au plan du récit

À partir de l'étude de corpus nous avons proposé un contenu et une structure pour générer des récits d'activité du ski de randonnée à partir des seules données contenues dans des événements qu'il est possible de capter. Les structures que nous proposons constituent des plans du récits avec la connaissance minimale à apporter pour satisfaire un but communicationnel dans le récit. Cependant, nous avons utilisé un modèle de récit modulaire qui permet d'inclure des *plots* en fonction des connaissances captées. Ainsi, les structures minimales peuvent être complétées par l'ajout de *plots* correspondant à l'ajout d'information. Nous proposons un algorithme de production de plan du récit à partir d'une séquence valide d'événements qui prévoit de tels ajouts (1).

---

**Algorithme 1** : Création du plan de récit

---

- 1 : Sélection des événements de la séquence par rapport aux informations indispensables et pertinentes pour le type de récit
  - 2 : Création des états affectifs correspondants
  - 3 : Pour chaque état affectif indispensable créer un plot l'encapsulant
  - 4 : Remplacer les plots des templates (tels que présentés dans la section 7.2.2.2 par les plots obtenus à l'étape 3)
  - 5 : Pour chaque état affectif pertinent créer un plot encapsulant
  - 6 : Lier les plots obtenus à l'étape précédente aux plots de la structure (avec des liens d'équivalence (e) ou de motivation (m))
- 

Pour créer les plans du récit, nous n'avons considéré que les caractéristiques liées au suivi du parcours (*Navigation*) et celle du ressenti sur la sortie. Ces deux caractéristiques ne définissent pas à elles seules un but communicationnel mais elles permettent de mettre en évidence comment des aspects objectifs sur l'activité (liés à la navigation) et des aspects subjectifs (le ressenti) peuvent être combinés pour produire un récit d'activité.

Nous nous appuyons sur les plots composants les plans du récit pour générer les phrases qui constituent le récit textuel.

## 7.3 Du plan du récit aux textes

La génération du récit textuel écrit que nous proposons repose uniquement sur la traduction en langue naturelle des connaissances exprimées dans un

plan du récit tel que défini dans la section 7.2.1. pour cela, nous utilisons les principes de génération automatique de texte (GAT). La génération d'un texte à partir d'une représentation sémantique se fait généralement en trois grandes étapes [Reiter et Dale, 2000] : 1) La sélection des éléments du discours et leur organisation ; 2) la planification des phrases — syntaxe, lexicalisation (choix du vocabulaire) et l'agrégation de phrases – et 3) la réalisation de surface et le formatage. Nous avons appliqué ces processus pour générer une présentation textuelle d'un plan d'un récit.

Dans notre cas, l'ensemble des éléments du discours et l'organisation sont entièrement définis au sein du plan du récit. Nous avons défini les règles nécessaires à ce processus pour des plans de récit constitués des éléments indispensables (résultant des 4 premières étapes de l'algorithme 1. Nous nous appuyons sur l'ensemble des éléments du modèle de plan de récit (section 7.1), c'est à dire :

- les états affectifs
- les liens
- les plots (unités et complexes)

Nous avons défini des règles de génération (étape 2 de processus de GAT) pour chacun de ces éléments :

- Les plots qui donnent lieu à des phrases (une ou plusieurs) qui sont construites autour des valeurs portées par les états affectifs positifs et négatifs
- Les liens causaux entre les plots qui permettent de générer des connecteurs entre les phrases.

Les figures 7.10, 7.11 et 7.12 présentent les liens entre le contenu des plans du récit et la structure du texte qui est générée à partir de ces plans.

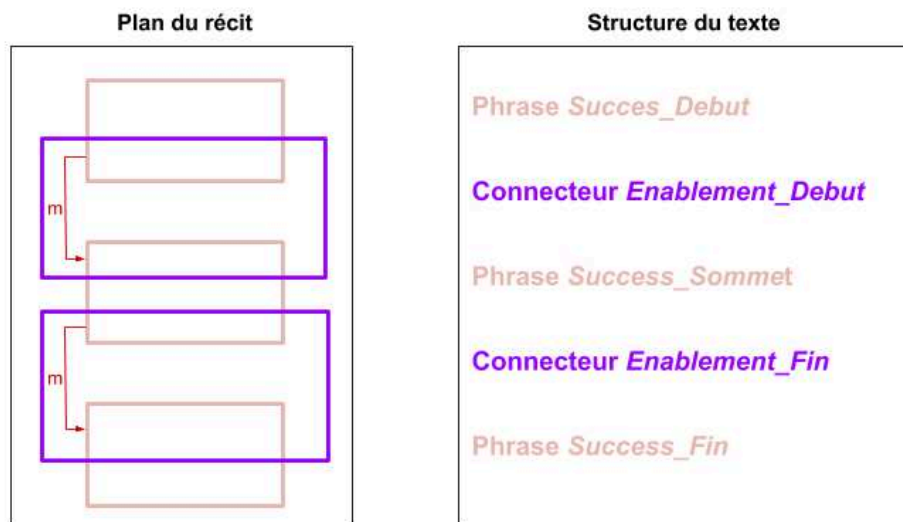


FIG. 7.10: Correspondance entre la structure du texte généré et le plan d'un récit positif sans contretemps

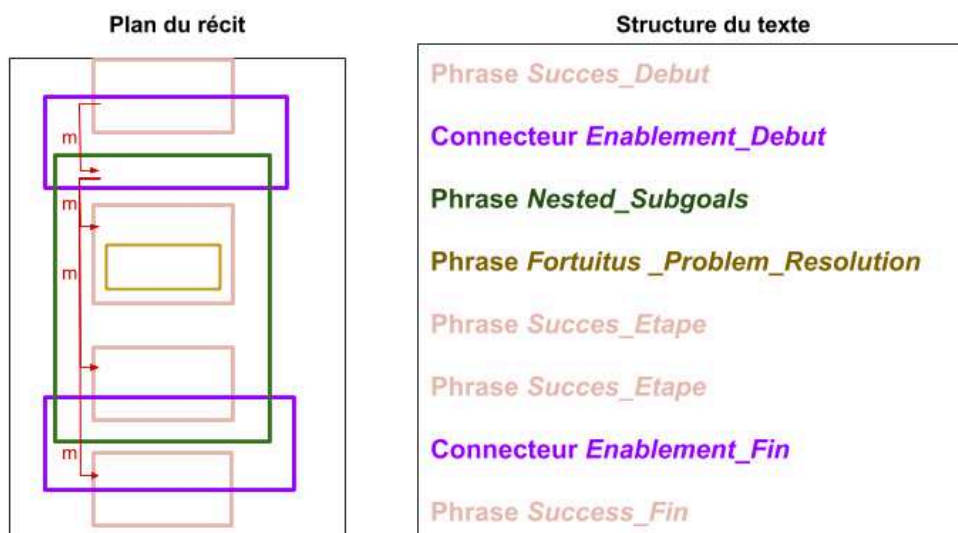


FIG. 7.11: Correspondance entre la structure du texte généré et le plan d'un récit nuancé sans correction des conditions par le protagoniste

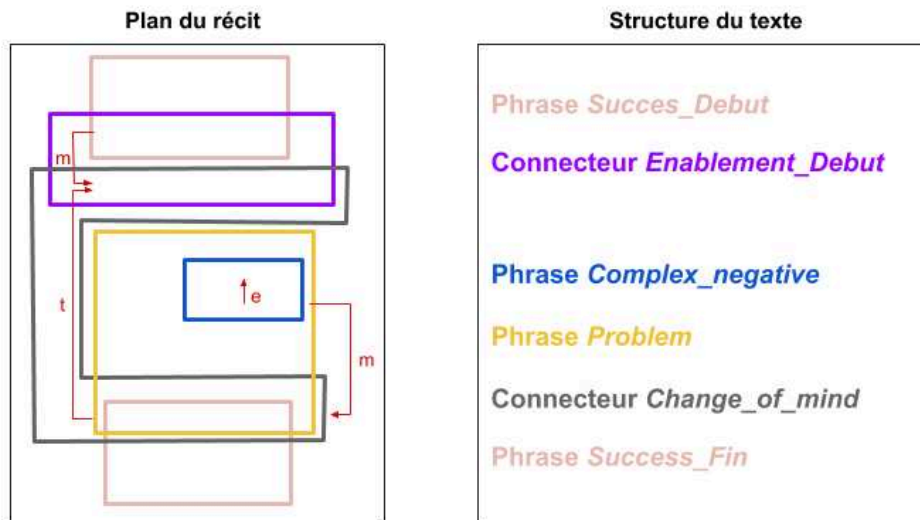


FIG. 7.12: Correspondance entre la structure du texte généré et le plan d'un récit négatif

Les connecteurs (*Enablement\_Debut*, *Enablement\_Fin*, *Problem* et *Change\_of\_mind*) ne dépendent pas des valeurs des données de randonnée. Ils correspondent à des plots construits sur des liens causaux liant des plots qui encapsulent les données elles-mêmes. Donc, pour le processus de GAT les règles de génération ne sont pas paramétrées.

Deux façons de connecter des éléments textuels sont possibles : utiliser des adverbes coordonnant (*enfin*, *plus...*) ou des phrases. Nous avons sélectionné cette dernière solution car les phrases permettent d'indiquer que l'activité est continue. Le tableau 7.16 présente les phrases choisies pour chaque connecteur.

| Connecteur              | Texte   |
|-------------------------|---|
| <b>Enablement_Debut</b> | De là, PRONOM PERSONNEL 3EME PERSONNE SINGULIER monter                                  |
| <b>Enablement_Fin</b>   | D'où PRONOM PERSONNEL 3EME PERSONNE SINGULIER descendre                                 |
| <b>Problem</b>          | Ce qui PRONOM PERSONNEL 3EME PERSONNE SINGULIER amener à vouloir modifier les objectifs |
| <b>Change_of_mind</b>   | PRONOM PERSONNEL 3EME PERSONNE SINGULIER abandonner le parcours                         |

TAB. 7.16: Phrases correspondant aux connecteurs

Pour générer les phrases nous nous appuyons sur les données captées

(selon l'ontologie définie dans la section 6.2) (Table 7.17)

| Phrase                              | Éléments de l'ontologie  |
|-------------------------------------|--|
| <b>Success_Debut</b>                | <i>Departure.take_place_at</i><br><i>Departure.Start_Time</i><br><i>Departure.has_actor</i>  |
| <b>Success_Sommet</b>               | <i>Activity_Event.has_end_natural_element</i><br><i>Activity_Event.Start_Time</i><br><i>Activity_Event.has_actor</i>   |
| <b>Success_Fin</b>                  | <i>Arrival.take_place_at</i><br><i>Arrival.Start_Time</i><br><i>Arrival.has_actor</i>  |
| <b>Success_Etape</b>                | <i>Activity_Event.has_end_natural_element</i><br><i>Activity_Event.Start_Time</i><br><i>Activity_Event.has_actor</i>   |
| <b>Nested_Subgoals</b>              | <i>Activity_Event.Sommet.has_end_natural_element</i><br><i>Activity_Event.Sommet.has_actor</i><br><i>Activity_Event</i>  |
| <b>Fortuitus_Problem_Resolution</b> | <i>Condition_Event1</i><br><i>Condition_Event2</i><br><i>Condition_Event1.Start_Time</i><br><i>Condition_Event2.Start_Time</i><br><i>Condition_Event1.takes_place_at</i><br><i>Activity_Event.Sommet.has_actor</i> |
| <b>Complex_Negative</b>             | <i>Condition_Event</i><br><i>Condition_Event.Start_Time</i><br><i>Condition_Event1.takes_place_at</i>  |

TAB. 7.17: Phrases correspondant aux connecteurs

Les phrases sont générées par des *templates* (Table 7.18)

| Phrase                              | Phrase Template  |
|-------------------------------------|--|
| <b>Success_Debut</b>                | time( <i>Departure.Start_Time</i> ), sujet( <i>Departure.has_actor</i> ), partir comme prévu de lieu( <i>Departure.take_place_at</i> ).  |
| <b>Success_Sommet</b>               | time( <i>Activity_Event.Start_Time</i> ), sujet( <i>Activity_Event.has_actor</i> ) arriver à l'objectif : lieu( <i>Activity_Event.has_end_natural_element</i> ).   |
| <b>Success_Fin</b>                  | time( <i>Arrival.Start_Time</i> ), sujet( <i>Arrival.has_actor</i> ) être de retour à lieu( <i>Arrival.take_place_at</i> ).  |
| <b>Success_Etape</b>                | time( <i>Activity_Event.Start_Time</i> ), sujet( <i>Activity_Event.has_actor</i> ) arriver à l'objectif de l'étape : le lieu( <i>Activity_Event.has_end_natural_element</i> ).   |
| <b>Nested_Subgoals</b>              | Pour atteindre l'objectif de lieu( <i>Activity_Event.Sommet.has_end_natural_element</i> ), sujet( <i>Activity_Event.Sommet.has_actor</i> ) réaliser calculNombre([ <i>Activity_Event</i> ]) étapes.  |
| <b>Fortuitus_Problem_Resolution</b> | time( <i>Condition_Event1.Start_Time</i> ), description( <i>Condition_Event1</i> ) être difficile. A lieu( <i>Condition_Event1.takes_place_at</i> ), sujet( <i>Activity_Event.Sommet.has_actor</i> ) continuer. time( <i>Condition_Event2.Start_Time</i> ), description( <i>Condition_Event2</i> ), s'améliorer. |
| <b>Complex_Negative</b>             | time( <i>Condition_Event.Start_Time</i> ), lieu( <i>Condition_Event1.takes_place_at</i> ), description( <i>Condition_Event</i> ) devenir insurmontables.   |

TAB. 7.18: Templates utilisés pour générer les phrases

Ces *templates* sont une illustration des phrases qui pourraient constituer le texte d'un récit. Aucune étude n'a pas été menée pour les construire. Dans les figures 7.13, 7.14, 7.15, 7.16 et 7.17 nous présentons des exemples des phrases générées.

**À 08:16, Catherine part comme prévu du parking.**

FIG. 7.13: Exemple de phrase *Success\_Debut*

**De là, elle monte.**

FIG. 7.14: Exemple de phrase *Enablement\_Debut*

**À 11:57, Catherine arrive à l'objectif : le Grand Rocher.**

FIG. 7.15: Exemple de phrase *Success\_Sommet*

**D'où elle descend.**

FIG. 7.16: Exemple de phrase *Enablement\_Fin*

**À 13:47, Catherine est de retour au parking.**

FIG. 7.17: Exemple de phrase *Success\_Fin*

Finalement, nous présentons dans les figures 7.18, 7.19, 7.20 et 7.21 des exemples des récits positifs sans contretemps obtenus automatiquement. Ceux-ci sont entièrement générés à partir de données collectées dans le Corpus Parallèle.

À 16:46, Frank part comme prévu du parking. De là, il monte.  
À 17:00, Frank arrive à l'objectif : le Col de la Maurienne,  
d'où il descend. À 18:14, Frank est de retour au parking.

FIG. 7.18: Récit provenant de l'enregistrement N° 5 (Date : 11/02/17, Sommet : Col de la Maurienne)

À 09:53, Frank part comme prévu du parking. De là, il monte.  
À 13:40, Frank arrive à l'objectif : le Col de l'Aigleton,  
d'où il descend. À 14:31, Frank est de retour au parking.

FIG. 7.19: Récit provenant de l'enregistrement N° 7 (Date : 18/03/16, Sommet : Col de l'Aigleton)

À 16:46, Frank part comme prévu du parking. De là, il monte.  
À 18:00, Frank arrive à l'objectif : le Chamechaude, d'où il  
descend. À 18:14, Frank est de retour au parking.

FIG. 7.20: Récit provenant de l'enregistrement N° 10 (Date : 30/03/16, Sommet : Chamechaude)

À 08:16, Catherine part comme prévu du parking. De là, elle  
monte. À 11:57, Catherine arrive à l'objectif : le Grand  
Rocher, d'où elle descend. À 13:47, Catherine est de retour  
au parking.

FIG. 7.21: Récit provenant de l'enregistrement N° 11 (Date : 22/01/17, Sommet : Grand Rocher)



## 7.4 Conclusion de la modélisation du point de vue subjectif de l'activité : le récit

Nous avons présenté dans ce chapitre une méthode à base de modèles qui permet de passer du monde objectif de l'activité vers le monde subjectif de l'expérience. Ce passage est assuré par la transformation de données issues du monde de l'activité en des données propres au monde du récit. Le premier modèle est celui du plan du récit. Nous avons proposé une adaptation du modèle de Lehnert pour tenir compte des spécificités du récit d'activité et plus particulièrement du ski de randonnée : à la dimension navigation – soumise aux aléas du contexte de l'activité – est associée la dimension de ressenti – d'où émerge un récit sensible à l'expérience vécue. Une fois ce plan établi nous avons montré comment ce plan peut être utilisé pour générer le récit de l'activité à l'aide de techniques de génération automatique de textes. La modélisation du plan du récit que nous proposons permet de transmettre une perspective sur ce qui a été vécu lors de la réalisation de l'activité. Cette approche est novatrice vis à vis de l'état de l'art de la génération automatique de récits car elle s'appuie sur des données objectives qui sont produites sans suivre un modèle narratologique préétabli.

Nous avons proposé également une méthode pour définir les plans du récit à partir d'une analyse de corpus. Cette méthode permet d'identifier le contenu et la structure des récits. Elle a été appliquée dans le cadre du ski de randonnée pour identifier trois types de récits et les trois plans du récit qui leur correspondent. D'abord, grâce à une analyse de textes, il a été possible d'identifier quels sont les aspects les plus importants à inclure dans le récit (la description du parcours et les conditions environnementales). Une fois ces éléments caractérisés, la classification de ces éléments en fonction de la fréquence de leur présence au sein d'un récit permet d'identifier les types de structure.

Avec ces connaissances, nous avons construit le modèle du récit. Pour passer de la séquence d'événements au *plan du récit*, il faut une sélection du contenu et une organisation qui transmette un ressenti. Nous proposons une approche qui structure le récit en s'adaptant au contenu (*états affectifs positifs et négatifs* directement issus des données captées) et en prenant en compte le ressenti à transmettre tout au long du processus de transformation de la séquence d'événement au plan du récit (choix des événements à présenter et des *plots* à appliquer). Ce processus nous permet de proposer une approche flexible.

Enfin, nous avons aussi abordé le passage de ces structures vers le texte en montrant comment cela pourrait être appliqué. Les choix au niveau du

lexique ne sont faits qu'à titre d'exemple mais permettent de faire passer la sémantique des *plots* utilisés dans les plans du récit. L'étude des corpus a montré que des phrases dans les récits ne portaient pas toutes sur le déroulement de la sortie, elles permettent par exemple de transmettre des expériences passées, des souvenirs ou des avis. Les récits textuels que nous générons ne contiennent aucun de ces éléments. Les choix de génération que nous avons faits doivent être mis en œuvre pour pouvoir conclure sur l'intérêt ou non de chercher à introduire des connaissances qui ne sont pas directement liées au déroulement de l'activité.

# Chapitre 8

## Conclusion

La génération automatique de récits à partir de données réelles est un domaine de recherche qui est encore peu développé. Il s'agit pourtant d'une approche novatrice pouvant apporter une perspective nouvelle à la transmission d'informations et de connaissances. Par ailleurs, le développement de ce domaine apporterait également un nouvel éclairage sur l'activité humaine pour les sciences cognitives. Dans ce manuscrit, nous présentons une approche de génération de récits permettant de relater – à partir de données captées – une expérience de randonnée à ski qui respecte les canons de la narratologie. Les contributions de cette approche sont résumées dans la section 8.1. Le travail effectué lors du développement de cette thèse a conduit à adopter certaines contraintes et a permis de mesurer le chemin à parcourir. Ces limitations de l'étude sont brièvement résumées section 8.2. Toutefois, ces limites ont permis d'ouvrir un ensemble de perspectives passionnantes pour la recherche en narratologie informatique et en informatique ambiante qui sont esquissées section 8.3.

### 8.1 Contributions de la thèse

Au cours de ce mémoire de thèse, nous avons présenté une nouvelle approche pour la génération automatique de récits. Cette approche est influencée par différents travaux en Informatique, en Narratologie, et en Sciences Cognitives. La nouveauté de ce domaine a conduit cette thèse à identifier de nombreux verrous scientifiques et méthodologiques qui ont orienté la démarche. La méthode adoptée comprend notamment la nécessité de redéfinir la notion de récit, d'étudier des corpus de récits humains pour en extraire la connaissance, de collecter des corpus réels, de modéliser le processus de génération, de le mettre en œuvre et de l'évaluer.

La première partie de la thèse a permis de définir la notion de récit d'activité ainsi que ses composantes (Chapitre 4). Cette définition se base sur l'état de l'art en narratologie (Chapitre 2). Cet état de l'art a permis de parcourir les éléments fondamentaux d'un récit et de constater la nécessité de nouvelles définitions adaptées au type de récits abordé dans cette thèse. Ainsi, nous proposons comme première contribution la définition de la notion de *récits d'activité*, ainsi que ses éléments de base, provenant de la théorie narrative, mais adaptés au contexte de la thèse. Nous y intégrons la notion du *But communicationnel*, facteur qui va guider la construction du récit d'activité.

La deuxième partie de la thèse décrit l'approche de génération automatique, une contribution qui se démarque des recherches en génération de récits qui sont principalement focalisées sur la génération de récits fictifs (Chapitre 3). En effet, dans les récits fictifs, le monde dans lequel évolue les personnages est un monde fermé et omniscient dans le sens où les actions et motivations/émotions des personnages sont connues et malléables. Dans notre cas, nous travaillons dans un monde ouvert dans lequel beaucoup d'informations n'étaient pas observables, qu'elles soient objectives (p.ex., état de la neige, rencontre avec des animaux sauvages) ou subjectives (p.ex., satisfaction d'une descente, motivation). Nous avons alors adopté une approche par modèles afin d'obtenir le maximum de connaissances sur le monde et sur l'activité. Pour acquérir ces modèles nous avons collecté des corpus afin de recueillir des connaissances sur trois axes : les données disponibles concernant l'activité, les éléments objectifs et subjectifs mentionnés dans les récits, la structure et les paramètres structurant des récits. Au total, trois corpus ont été construits (Chapitre 5) : deux corpus textuels provenant d'un site web du domaine et un troisième, dit parallèle, qui est composé de données de capteurs acquises pendant des sorties de ski de randonnées et des récits écrits par les acteurs de ces sorties. Ces corpus constituent une ressource intéressante pour le domaine notamment le corpus parallèle qui est, à notre connaissance, le seul à avoir été constitué.

Les analyses de corpus et le soutien d'un expert du domaine ont permis de modéliser l'activité (Chapitre 6). Deux modèles ont été construits : un modèle du domaine sous forme d'ontologie et un modèle de l'activité sous forme de modèle de tâches. Le premier modèle permet de représenter les données provenant du monde physique en des instances de concepts formels. Par ailleurs, l'ontologie permet de représenter tous les concepts manipulés durant le processus de génération. Ainsi, les concepts de l'activité et du plan de récit y sont également représentés. Le modèle de tâches permet, quant à lui, de vérifier la validité de la séquence d'activités extraite des données. Pour prendre en compte la notion de monde ouvert, le modèle peut également être utilisé pour identifier les tâches manquantes au sein de l'activité pour rendre

une séquence valide. Ainsi, une séquence incomplète pourrait quand même être utilisée pour générer un récit, soit en mentionnant de manière explicite les tâches manquantes, soit en utilisant des ellipses.

Le passage du monde objectif de l'activité vers le monde subjectif de l'expérience vécue constitue une grande difficulté pour le domaine de la génération de récits réels car c'est à ce moment que le ressenti subjectif est communiqué et qu'une perspective narrative est prise. Le chapitre 7 présente la méthode que nous avons employée pour aborder cette difficulté. Nous avons observé que les auteurs des récits ont tendance à structurer la rédaction de l'expérience en fonction de leur ressenti au moment des faits, notamment en fonction de la conformité avec leur attentes. Nous avons donc proposé un modèle de génération de plan de récit capable de fonctionner à partir des impressions des acteurs tout en étant flexible en termes structurels. Pour cela, nous avons adapté le modèle de Lehnert [Lehnert, 1981], initialement proposé pour l'analyse de récits, à la génération de récits. Ce modèle très flexible permet de construire des intrigues de plus en plus complexes à partir d'intrigues unitaires. En utilisant cette adaptation nous avons pu construire les bases de trois stéréotypes de plans de récit correspondant à trois ressentis communiqués : expérience positive sans contretemps, expérience positive avec défis et expérience négative. Ces bases sont ensuite instanciées et complétées en fonction des données. Le plan de récit final peut ensuite être rendu en langage naturel.

La thèse apporte ainsi une démarche complète de génération de récits qui apporte une solution au problème de génération de récits à partir de données brutes réelles dans la cas du ski de randonnée. Une implémentation possible de cette démarche de génération a été mise en place comme preuve de concept.

## 8.2 Limitations

Les contributions de la thèse ont permis de produire des récits adoptant une perspective subjective à partir de traces de l'activité humaine exécutée. Cependant, cette approche reste restreinte par trois types de limitations :

### Ressources

Les ressources disponibles pour effectuer ce type de recherche restent limitées. S'il existe un grand nombre de corpus textuels disponibles, peu d'entre eux sont annotés et lorsqu'ils le sont, le schéma d'annotation repose sur une théorie narrative bien précise [Finlayson, 2013]. La collecte de données pour constituer le corpus parallèle est une tâche très lourde qui a pris un temps important

durant le doctorat. En conséquence, malgré les efforts pour réaliser cette collecte le corpus parallèle reste restreint et insuffisant pour effectuer une validation définitive de l'approche complète.

### **Modélisation**

Les récits générés sont incomplets dans le sens où il n'a pas été possible d'inclure d'autres sources d'information essentielles à l'écriture d'un récit, tels que les opinions, motivations, souvenirs ou références vers des autres événements. Cependant, des travaux récents entrepris au sein de l'équipe de recherche ont montré qu'il était possible d'extraire des informations physiologiques pouvant servir de support à l'inférence de ressenti (fatigue, récupération, attente...).

Les modèles obtenus sont très dépendants du cas d'usage. Il conviendrait de vérifier la généralité de ceux-ci en les testant sur de nouvelles sorties en dehors des massifs considérés et en les portant dans des domaines connexes (comme la randonnée en été par exemple).

### **Mise en œuvre**

La validation de la séquence d'activités par modèle de tâches permet de vérifier si les tâches ont été réalisées dans un ordre conforme à l'activité. Cependant, nous avons dû compléter cette vérification avec un modèle d'édition afin de prendre en compte les tâches manquantes ou non observables. Pour être complet, la démarche aurait dû exploiter les résultats de cette recherche de tâches manquantes dans les traitements suivants, ce qui n'a malheureusement pas pu être le cas par manque de temps.

Le texte produit reste très simple. Le récit attendu est composé de phrases pouvant refléter les ressentis à partir des données de l'acteur de la sortie. Le récit est toujours rapporté à la troisième personne, afin de respecter une démarche éthique par rapport à celui ou celle qui a vécu l'expérience. De nombreux perfectionnements au niveau syntaxique et lexical auraient dû être implantés dans l'étape de génération de textes afin de rendre les récits générés plus vivants et attrayants.

## **8.3 Perspectives**

Toutes les limitations citées plus haut ouvrent des perspectives de recherche pour des travaux ultérieurs. Plusieurs pistes pour les aborder peuvent être envisagées.

Nous avons montré que le manque d'information subjective dans les données peut être abordé par une bonne connaissance du domaine. Suite à notre étude de corpus, nous avons proposé qu'un plan de récit ait une structure qui obéisse à un but communicationnel particulier. Ces buts résultent des ressentis projetés par les skieurs. Nous avons ensuite considéré que ces ressentis font partie d'un ensemble de stratégies mis en place pour justifier les motivations des skieurs et, de là, leur actions. Cependant, le modèle que nous proposons n'est pas assez fin pour intégrer ces paramètres et leur influence sur la construction du plan du récit. Par exemple, pour transmettre un but qui réveille un sens de prévention chez le lecteur, un récit d'activité devrait comporter des défis, des conseils et du vocabulaire technique qui spécifie pourquoi il est important de faire attention à certains endroits. Une perspective intéressante de recherche serait donc d'analyser plus finement les corpus afin de faire émerger ces paramètres.

Une autre perspective intéressante serait de prendre en compte l'audience du récit. Nous avons développé des récits en suivant un but communicationnel particulier influençant les ressentis d'intérêt à transmettre dans le récit mais sans prendre en compte le lecteur. Par exemple, une personne qui cherche à partager son expérience avec une communauté de skieurs réguliers adoptera une forme particulière en considérant ce que les lecteurs connaissent déjà (sous-activité usuelle, route pour aller à un certain point, jargon du domaine). Si la même personne souhaite partager son récit avec des débutants de l'activité, alors le langage et les éléments inclus seront bien différents. La prise en compte de l'audience est essentielle pour que le but communicationnel puisse être atteint.

Étant donnée la complexité d'obtenir un corpus d'évaluation complet, la démarche proposée dans cette thèse n'a été testée que sur un seul cas d'usage. Une perspective évidente est de l'appliquer à d'autres cas d'usage dans lequel la compréhension de l'activité humaine est essentiel. Un exemple dans lequel l'analyse de l'activité humaine est utilisé est la conception de dispositifs interactifs. Les méthodologies de conception font nécessairement l'étude de l'activité humaine dans laquelle le dispositif devra s'inscrire. L'analyse fine des protagonistes (humains et/ou machine) ainsi que la procédure permet de guider les choix de conception et de mesurer les erreurs de conception afin d'améliorer l'utilisabilité du dispositif. Par exemple l'usage de scénarios est très répandu au début de la conception de systèmes interactifs pour faciliter la communication entre les acteurs de l'activité et ceux de l'équipe de conception [Carroll, 2000] lors de la définition du contexte d'usage. Ces scénarios représentent un cas typique de la réalisation de l'activité par des utilisateurs cibles selon les experts et prennent la forme de récits. Mais la communication avec les experts de l'activité ne se limite pas au début de la conception, elle

est nécessaire tout au long du processus de conception pour pouvoir se reposer sur la “réalité terrain”, c’est à dire sur la réalisation effective de l’activité. Ainsi, des traces de simulations, des observations d’interaction réelle peuvent être acquises afin d’affiner, de corriger ou de valider les choix de conception. Ces traces sont, comme nous l’avons mentionné plus haut, un ensemble de données brutes difficiles à interpréter pour le non-expert. La présentation de ces traces (les observations d’une activité humaine) sous forme de récits permettraient d’être analysés par les experts de l’activité. De récents travaux dans notre équipe [Caffiau et Portet, 2017] ont montré l’intérêt de la génération de textes à partir de modèles de tâches pour la conception de SI. Une étape suivante est d’utiliser la génération de récit.

Une autre perspective qui émerge des progrès récents en intelligence artificielle est d’utiliser l’apprentissage automatique pour concevoir un système de génération qui soit entièrement dépendant des données d’entrées. Bien que de nombreux systèmes basés sur l’apprentissage profond se réclament de la génération automatique d’histoire [Roemmele, 2016b, Martin *et al.*, 2017], ils se rapprochent plus de la restructuration de texte que de la génération de récits à partir de données brutes. Il n’en reste pas moins qu’avec le développement de ces méthodes pour la fiction, étudier la possibilité d’appliquer l’apprentissage automatique à tout ou partie de la démarche présentée représente un intérêt scientifique indéniable (tant que les corpus sont suffisants).

Finalement, un dernier défi à résoudre est l’évaluation des récits d’activité. Nous avons vu dans le chapitre 2 que l’évaluation des récits automatiques est encore peu développée. Etant donné l’approche que nous proposons, il est nécessaire de mettre en place des protocoles permettant d’observer les effets de la génération de récits sur l’audience. Par ailleurs, il conviendrait, comme dans les autres tâches du TALN, de définir des mesures automatiques de performances permettant de vérifier l’évolution de la modélisation de la solution en cours de développement. Dans une étude qui n’est pas rapportée dans cette thèse, nous avons développé et testé une démarche d’évaluation permettant de mesurer la pertinence des buts communicationnels. Cette démarche devrait être poursuivie.

Les contributions de la thèse permettent de mettre en lumière la complexité qui se cache derrière une notion aussi familière que le récit. Ce manuscrit rapporte la thèse développée pour permettre la génération de récit à partir de données réelles. Elle s’inspire autant que possible du processus de construction des récits observé en science cognitive. Bien que cette étude contienne un certain nombre de limites, l’approche proposée reste novatrice et originale. Les travaux développés sont encourageants pour continuer l’exploration de ce sujet qui, s’il respecte ses promesses, permettra de rendre la génération automatique de récits, cette faculté uniquement détenue par l’humain, possible



en interaction homme machine et plus largement en informatique.

# Appendices

## Analyses par types de récits

| Mot      | Fréquence | Normalisé |
|----------|-----------|-----------|
| ski      | 4         | 0,004     |
| neige    | 4         | 0,004     |
| jusqu    | 3         | 0,003     |
| attaque  | 3         | 0,003     |
| fond     | 3         | 0,003     |
| tout     | 3         | 0,003     |
| forêt    | 2         | 0,002     |
| Pierre   | 2         | 0,002     |
| vue      | 2         | 0,002     |
| descente | 2         | 0,002     |
| bon      | 2         | 0,002     |
| bien     | 2         | 0,002     |
| poudre   | 2         | 0,002     |
| rando    | 2         | 0,002     |
| eau      | 1         | 0,001     |
| soleil   | 1         | 0,001     |
| sortie   | 1         | 0,001     |
| douzaine | 1         | 0,001     |
| pistes   | 1         | 0,001     |
| devenu   | 1         | 0,001     |

TAB. 1: Les 20 mots les plus fréquents dans les récits avec le ressenti positif (*sans contretemps*)

| Mot        | Fréquence | Normalisé |
|------------|-----------|-----------|
| neige      | 11        | 0,011     |
| bien       | 9         | 0,009     |
| montée     | 8         | 0,008     |
| plus       | 8         | 0,008     |
| som        | 6         | 0,006     |
| descente   | 5         | 0,005     |
| bon        | 5         | 0,005     |
| forêt      | 5         | 0,005     |
| charmant   | 4         | 0,004     |
| ski        | 4         | 0,004     |
| trace      | 4         | 0,004     |
| donc       | 4         | 0,004     |
| sommet     | 4         | 0,004     |
| peu        | 4         | 0,004     |
| chartreuse | 4         | 0,004     |
| couloir    | 4         | 0,004     |
| collet     | 4         | 0,004     |
| belle      | 4         | 0,004     |
| très       | 4         | 0,004     |
| tour       | 4         | 0,004     |

TAB. 2: Les 20 mots les plus fréquents dans les récits avec le ressenti positif (*avec défis*)

| Mot      | Fréquence | Normalisé |
|----------|-----------|-----------|
| bien     | 17        | 0,017     |
| trace    | 16        | 0,016     |
| neige    | 13        | 0,013     |
| descente | 12        | 0,012     |
| sommet   | 12        | 0,012     |
| départ   | 11        | 0,011     |
| couloir  | 11        | 0,011     |
| jusqu    | 11        | 0,011     |
| peu      | 10        | 0,01      |
| faire    | 9         | 0,009     |
| poudre   | 9         | 0,009     |
| plus     | 9         | 0,009     |
| montée   | 9         | 0,009     |
| niveau   | 8         | 0,008     |
| bon      | 8         | 0,008     |
| puis     | 8         | 0,008     |
| fait     | 8         | 0,008     |
| ski      | 8         | 0,008     |
| piste    | 7         | 0,007     |
| fois     | 7         | 0,007     |

TAB. 3: Les 20 mots les plus fréquents dans les récits avec le ressenti nuancé

Nous présentons ensuite l'analyse par bigrammes.

| Mot    | Fréquence | Normalisé |
|--------|-----------|-----------|
| ça     | 5         | 0,005     |
| peu    | 4         | 0,004     |
| sale   | 4         | 0,004     |
| puis   | 4         | 0,004     |
| gosse  | 4         | 0,004     |
| plus   | 3         | 0,003     |
| casse  | 3         | 0,003     |
| bon    | 3         | 0,003     |
| bas    | 3         | 0,003     |
| alors  | 3         | 0,003     |
| bien   | 3         | 0,003     |
| neige  | 2         | 0,002     |
| très   | 2         | 0,002     |
| montée | 2         | 0,002     |
| foret  | 2         | 0,002     |
| truc   | 2         | 0,002     |
| si     | 2         | 0,002     |
| retour | 2         | 0,002     |
| quand  | 2         | 0,002     |
| pire   | 2         | 0,002     |

TAB. 4: Les 20 mots les plus fréquents dans les récits avec le ressenti négatif

| <b>Bigramme</b>    | <b>Fréquence</b> |
|--------------------|------------------|
| ski fond           | 3                |
| fond bon           | 2                |
| bonne jusqu        | 1                |
| descente fond      | 1                |
| neige geléecroutée | 1                |
| fait bien          | 1                |
| jusqu forêt        | 1                |
| rando seb          | 1                |
| loin attaque       | 1                |
| lasse 3            | 1                |
| saison fond        | 1                |
| pistes dos         | 1                |
| jusqu haut         | 1                |
| devenu neige       | 1                |
| loin dernier       | 1                |
| fouetter visage    | 1                |
| forêt vraiment     | 1                |
| tout remonté       | 1                |
| 3 départ           | 1                |
| observent loin     | 1                |

TAB. 5: Les 20 bigrammes les plus fréquents des récits avec un ressenti positif (sans contretemps)

| <b>Bigramme</b>   | <b>Fréquence</b> |
|-------------------|------------------|
| charmant som      | 4                |
| beau tour         | 3                |
| sentier dété      | 2                |
| soleil brume      | 2                |
| niveau neige      | 2                |
| doit être         | 2                |
| faire tour        | 2                |
| dernière collet   | 2                |
| pré batard        | 2                |
| aucune trace      | 2                |
| combe lif         | 2                |
| plus lourde       | 2                |
| neige plus        | 2                |
| dernière montée   | 2                |
| montée collet     | 2                |
| plus grâce        | 1                |
| tranquille neige  | 1                |
| diriger jugé      | 1                |
| posée randonneurs | 1                |
| ancienne mettre   | 1                |

TAB. 6: Les 20 bigrammes les plus fréquents des récits avec un ressenti positif (avec défis)

| <b>Bigramme</b>    | <b>Fréquence</b> |
|--------------------|------------------|
| ancienne piste     | 4                |
| trace faite        | 4                |
| charmant som       | 3                |
| trace montée       | 3                |
| couloir nw         | 3                |
| bon ski            | 3                |
| déjà bien          | 2                |
| peut être          | 2                |
| ski bien           | 2                |
| poudre lourde      | 2                |
| jusqu crête        | 2                |
| heureusement trace | 2                |
| neige dure         | 2                |
| va falloir         | 2                |
| fallait aime       | 2                |
| heureusement faite | 2                |
| descendant miam    | 2                |
| agréable skier     | 2                |
| col porte          | 2                |
| sans trop          | 2                |

TAB. 7: Les 20 bigrammes les plus fréquents des récits avec un ressenti nuancé

Et finalement, les trigrammes

| <b>Bigramme</b> | <b>Fréquence</b> |
|-----------------|------------------|
| sale gosse      | 4                |
| voire bien      | 2                |
| casse dalle     | 2                |
| gosse si        | 2                |
| col porte       | 2                |
| puis plus       | 2                |
| journée çà      | 2                |
| plus lourd      | 2                |
| si venu         | 2                |
| bien loooooourd | 2                |
| hui petit       | 1                |
| revenir épingle | 1                |
| 17 dont         | 1                |
| bien collective | 1                |
| place truc      | 1                |
| crevés sale     | 1                |
| peu heure       | 1                |
| couenne dit     | 1                |
| puis place      | 1                |
| montée avalée   | 1                |

TAB. 8: Les 20 bigrammes les plus fréquents des récits avec un ressenti négatif

| <b>Trigramme</b>         | <b>Fréquence</b> |
|--------------------------|------------------|
| ciel tout bleu           | 1                |
| neige forêt vraiment     | 1                |
| qualité neige forêt      | 1                |
| rando ski fond           | 1                |
| première sortie saison   | 1                |
| paysages cela fait       | 1                |
| encore classique lasse   | 1                |
| bleu remonté jusqu       | 1                |
| classique lasse faire    | 1                |
| observent loin attaque   | 1                |
| splendide sommet aucun   | 1                |
| devenu belle neige       | 1                |
| bergerie neige cartonnée | 1                |
| usante bon va            | 1                |
| croisé douzaine personne | 1                |
| soleil superbe trace     | 1                |
| plaindre croisé douzaine | 1                |
| neige cartonnée descente | 1                |
| bonne jusqu attaque      | 1                |
| terrible passage vue     | 1                |

TAB. 9: Les 20 trigrammes les plus fréquents des récits avec un ressenti positif (sans contretemps)

| <b>Trigramme</b>            | <b>Fréquence</b> |
|-----------------------------|------------------|
| dernière montée collet      | 2                |
| détend profite couche       | 1                |
| taillées besoin crampons    | 1                |
| tourne descend rappel       | 1                |
| semble bonne direction      | 1                |
| vierge feront demitour      | 1                |
| nécessitent bon remplissage | 1                |
| quand sagit décider         | 1                |
| remontée express pré        | 1                |
| ca devrait être             | 1                |
| ensuite longue descente     | 1                |
| route monte col             | 1                |
| casque doudoune comme       | 1                |
| direction charmette lors    | 1                |
| voie normale ami            | 1                |
| dété lassés décide          | 1                |
| va si mal                   | 1                |
| 30 °c ressenti              | 1                |
| passage skis sac            | 1                |
| réchauffer doigts fois      | 1                |

TAB. 10: Les 20 trigrammes les plus fréquents des récits avec un ressenti positif (avec défis)

| <b>Trigramme</b>            | <b>Fréquence</b> |
|-----------------------------|------------------|
| heureusement trace faite    | 2                |
| breche pourettes filles     | 1                |
| jamais rattrapé redescend   | 1                |
| ancienne piste seules       | 1                |
| grand espace refaire        | 1                |
| prolongé descente neige     | 1                |
| neige dure skis             | 1                |
| bergerie fait cueillir      | 1                |
| reste neige lever           | 1                |
| montfromage puis montée     | 1                |
| miam miam descente          | 1                |
| prenons trace montée        | 1                |
| enneigée chevauchée blanche | 1                |
| bonne rigolade descente     | 1                |
| celui là ailleurs           | 1                |
| encore vierge reviendra     | 1                |
| deux puis rejoins           | 1                |
| hésiter tonio voulait       | 1                |
| envie réitérer sortie       | 1                |
| mort va falloir             | 1                |

TAB. 11: Les 20 trigrammes les plus fréquents des récits avec un ressenti nuancé

| <b>Trigramme</b>         | <b>Fréquence</b> |
|--------------------------|------------------|
| jusqu chamechine casse   | 1                |
| porte montée avalée      | 1                |
| alors petit tricot       | 1                |
| mi herbe mi              | 1                |
| dommage autant raclette  | 1                |
| enquille nw belle        | 1                |
| quoi retourne col        | 1                |
| choisissant bien croise  | 1                |
| lourd voire bien         | 1                |
| nargue sms réponse       | 1                |
| dont premier casse       | 1                |
| être terrible fois       | 1                |
| copains allés buter      | 1                |
| trop forme aujourd       | 1                |
| mecs crevés sale         | 1                |
| transfo parfois peu      | 1                |
| bonne poudre choisissant | 1                |
| collective gucem bas     | 1                |
| mi rocher 5              | 1                |
| jusqu bas ça             | 1                |

TAB. 12: Les 20 trigrammes les plus fréquents des récits avec un ressenti négatif



# Bibliographie

- [Adam, 1999] ADAM, J. (1999). *Le récit*. Numéro 2149 de Que sais-je ? Presses Universitaires de France, 6ème édition.
- [Ahson et Ilyas, 2010] AHSON, S. et ILYAS, M. (2010). *Mobile Web 2.0 : Developing and Delivering Services to Mobile Devices*. CRC Press.
- [Amaro et al., 2014] AMARO, K. R., BEETZ, M. et CHENG, G. (2014). Understanding human activities from observation via semantic reasoning for humanoid robots. *In IROS Workshop on AI and Robotics*.
- [Amos-Binks et al., 2016] AMOS-BINKS, A., ROBERTS, D. L. et YOUNG, R. M. (2016). Summarizing and comparing story plans. *In Proceedings of 7th Workshop on Computational Models of Narrative*, Krakow, Poland.
- [Aristote, 1980] ARISTOTE (1980). *Poétique, texte, traduction, notes par Roselyne Dupont-Roc et Jean Lallot*. Éditions du Seuil.
- [Aylett, 1999] AYLETT, R. (1999). Narrative in virtual environments-towards emergent narrative. *In Proceedings of the AAAI fall symposium on narrative intelligence*, Massachusetts, United States.
- [Baader et Werner, 2002] BAADER, F. et WERNER, N. (2002). *Description Logic Handbook*, chapter Basic Description Logics., pages 47–100. Cambridge University Press.
- [Baez et al., 2014] BAEZ, B., CAFFIAU, S., GARBAY, C. et PORTET, F. (2014). Task based model for récit generation from sensor data : an early experiment. *In Proceedings of 5th Workshop on Computational Models of Narrative*, Quebec city, Canada.
- [Baez et al., 2015] BAEZ, B., CAFFIAU, S., GARBAY, C. et PORTET, F. (2015). Generating recit from sensor data : evaluation of a task model for story planning and preliminary experiments with gps data. *In Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*, Brighton, United Kingdom.

- [Bal, 2009] BAL, M. (2009). *Narratology : Introduction to the Theory of Narrative*. University of Toronto Press.
- [Baron et al., 2006] BARON, M., LUCQUIAUD, V., AUTARD, D. et SCAPIN, D. L. (2006). K-made : un environnement pour le noyau du modèle de description de l'activité. *In Proceedings of the 18th Conference on l'Interaction Homme-Machine*, Montreal, Canada.
- [Barthes, 1966] BARTHES, R. (1966). Introduction à l'analyse structurale des récits. *Communications*, 8(1) :1–27.
- [Blachon, 2016] BLACHON, D. (2016). *Reconnaissance de scènes multimodale embarquée*. Thèse de doctorat, Université Grenoble Alpes.
- [Blachon et al., 2014] BLACHON, D., PORTET, F., BESACIER, L. et TASSART, S. (2014). Recordme : A smartphone application for experimental collections of large amount of data respecting volunteer's privacy. *In International Conference on Ubiquitous Computing and Ambient Intelligence*, Belfast, United Kingdom.
- [Black et Wilensky, 1979] BLACK, J. B. et WILENSKY, R. (1979). An evaluation of story grammars. *Cognitive Science*, 3(3) :213 – 229.
- [Bremond, 1973] BREMOND, C. (1973). *Logique du récit*. Éditions du Seuil.
- [Bringsjord et Ferrucci, 1999] BRINGSJORD, S. et FERRUCCI, D. (1999). *Artificial Intelligence and Literary Creativity : Inside the Mind of Brutus, a Storytelling Machine*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.
- [Bringsjord et Ferrucci, 2000] BRINGSJORD, S. et FERRUCCI, D. A. (2000). Artificial intelligence and literary creativity : Inside the mind of brutus, a storytelling machine. *Computational Linguistics*, 26(4) :642–647.
- [Caffiau, 2009] CAFFIAU, S. (2009). *Approche dirigée par les modèles pour la conception et la validation des applications interactives : une démarche basée sur la modélisation des tâches*. Thèse de doctorat, ISAE-ENSMA Ecole Nationale Supérieure de Mécanique et d'Aérotechnique-Poitiers.
- [Caffiau et Portet, 2017] CAFFIAU, S. et PORTET, F. (2017). Natural Language Generation to support the understanding of task models : a preliminary study. *In AFIHM*, éditeur : *29ème conférence francophone sur l'Interaction Homme-Machine*, page 12 p., Poitiers, France. AFIHM, ACM.

- [Caffiau *et al.*, 2010] CAFFIAU, S., SCAPIN, D., GIRARD, P., BARON, M. et JAMBON, F. (2010). Increasing the expressive power of task analysis : Systematic comparison and empirical assessment of tool-supported task models. *Interacting with Computers*, 22(6) :569–593.
- [Callaway et Lester, 2002] CALLAWAY, C. B. et LESTER, J. C. (2002). Narrative prose generation. *Artificial Intelligence*, 139(2) :213 – 252.
- [Campion, 2015] CAMPION, B. (2015). Évaluer le récit comme acte cognitif. quel cadre pour les approches expérimentales ? <http://journals.openedition.org/narratologie/7212>. consulté le 10 septembre 2018.
- [Carroll, 2000] CARROLL, J. (2000). *Making Use : Scenario-Based Design of Human-Computer Interactions*. MIT Press.
- [Cavazza *et al.*, 2007] CAVAZZA, M., LUGRIN, J.-L., PIZZI, D. et CHARLES, F. (2007). Madame bovary on the holodeck : immersive interactive storytelling. *In Proceedings of the 15th ACM international conference on Multimedia*, Augsburg, Germany.
- [Chanquoy et Alamargot, 2002] CHANQUOY, L. et ALAMARGOT, D. (2002). Mémoire de travail et rédaction de textes : évolution des modèles et bilan des premiers travaux. *L'Année psychologique*, 102(2) :363–398.
- [Chatman, 1977] CHATMAN, S. (1977). *Story and discourse : Narrative structure in fiction and film*. Cornell University Press.
- [Cook et Thomas, 2005] COOK, K. A. et THOMAS, J. J. (2005). *Illuminating the path : The research and development agenda for visual analytics*. IEEE Computer Society : Chicago.
- [Day *et al.*, 2004] DAY, D. S., MCHENRY, C., KOZIEROK, R. et RIEK, L. (2004). Callisto : A configurable annotation workbench. *In Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal.
- [Elhadad et Robin, 1996] ELHADAD, M. et ROBIN, J. (1996). An overview of SURGE : a reusable comprehensive syntactic realization component. *In Eighth International Natural Language Generation Workshop*, Brighton, United Kingdom.
- [Elson, 2012] ELSON, D. K. (2012). *Modeling Narrative Discourse*. Thèse de doctorat, Columbia University, New York City.

- [Erlich, 1973] ERLICH, V. (1973). Russian formalism. *Journal of the History of Ideas*, 34(4) :627–638.
- [Fetzer, 2002] FETZER, J. H. (2002). *Consciousness Evolving*. Numéro 34 de Advances in Consciousness Research. John Benjamins Publishing Company.
- [Finlayson, 2008] FINLAYSON, M. A. (2008). Collecting semantics in the wild : The story workbench. In *Proceedings of the AAAI Fall Symposium on Naturally-Inspired Artificial Intelligence*, Menlo Park, USA. AAAI Press.
- [Finlayson, 2013] FINLAYSON, M. A. (2013). A survey of corpora in computational and cognitive narrative science. *Sprache und Datenverarbeitung (International Journal for Language Data Processing)*, 37(1–2) :113–141.
- [Fisher, 1985] FISHER, W. (1985). The narrative paradigm : An elaboration. *Communications Monographs*.
- [Flower et Hayes, 1981] FLOWER, L. et HAYES, J. R. (1981). A cognitive process theory of writing. *College composition and communication*, 32(4) : 365–387.
- [Fludernik, 2009] FLUDERNIK, M. (2009). *An introduction to narratology*. Routledge.
- [Forster, 2016] FORSTER, E. (2016). *Aspects of the Novel*. Hodder & Stoughton.
- [Fournier-Viger, 2005] FOURNIER-VIGER, P. (2005). *Un modèle de représentation des connaissances à trois niveaux de sémantique pour les systèmes tutoriels intelligents*. Université de Sherbrooke.
- [Freytag, 1900] FREYTAG, G. (1900). *Freytag's Technique of the Drama : An Exposition of Dramatic Composition and Art*. Scott, Foresman.
- [Friedenberg, 2008] FRIEDENBERG, J. (2008). *Artificial Psychology : The Quest for What It Means to Be Human*. Taylor & Francis.
- [Galofaro et Kubas, 2016] GALOFARO, F. et KUBAS, M. M. (2016). Dei genitrix : A generative grammar for traditional litanies. In *Proceedings of the 2016 Workshop on Computational Models of Narrative*, Krakow, Poland.
- [Gatt et Krahmer, 2018] GATT, A. et KRAHMER, E. (2018). Survey of the state of the art in natural language generation : Core tasks, applications and evaluation. *Journal of AI Research*, 61 :65–170.

- [Genette, 2007] GENETTE, G. (2007). *Discours du récit*. Points. Série Essais. Éditions du Seuil.
- [Gennari *et al.*, 2003] GENNARI, J. H., MUSEN, M. A., FERGERSON, R. W., GROSSO, W. E., CRUBÉZY, M., ERIKSSON, H., NOY, N. F. et TU, S. W. (2003). The evolution of protégé: an environment for knowledge-based systems development. *International Journal of Human-computer studies*, 58(1) :89–123.
- [Gervás, 2013] GERVÁS, P. (2013). Propp’s morphology of the folk tale as a grammar for generation. In *Proceedings of the 4th Workshop on Computational Models of Narrative*, Berlin, Germany.
- [Gervás, 2016] GERVÁS, P. (2016). Computational drafting of plot structures for russian folk tales. *Cognitive Computation*, 8(2) :187–203.
- [Gervás *et al.*, 2005] GERVÁS, P., DÍAZ-AGUDO, B., PEINADO, F. et HERVÁS, R. (2005). Story plot generation based on cbr. *Knowledge-Based Systems*, 18(4) :235–242.
- [Gervás *et al.*, 2015] GERVÁS, P., LEÓN, C. et MÉNDEZ, G. (2015). Schemas for narrative generation mined from existing descriptions of plot. In *Proceedings of the 6th Workshop on Computational Models of Narrative*, Atlanta, Georgia.
- [Gil, 2005] GIL, Y. (2005). *The Semantic Web – ISWC 2005 : 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings*. Information Systems and Applications, incl. Internet/Web, and HCI. Springer.
- [Göhring et Volk, 2011] GÖHRING, A. et VOLK, M. (2011). The text+ berg corpus an alpine french-german parallel resource. In *Actes de la 18e Conférence Annuelle sur le Traitement Automatique des Langues Naturelles (TALN-2011)*, Montpellier, France.
- [Grégoire et Karsenti, 2013] GRÉGOIRE, P. et KARSENTI, T. (2013). Le processus de révision et l’écriture informatisée—description des utilisations du traitement de texte par des élèves du secondaire au québec. *Alsic. Apprentissage des Langues et Systèmes d’Information et de Communication*, 16.
- [Greimas, 1966] GREIMAS, A. (1966). *Semantique structurale, recherche de methode*. Langue et Langage. Larousse.

- [Gruber, 2008] GRUBER, T. (2008). Ontology (computer science) - definition. In LIU, L. et ÖZSU, M. T., éditeurs : *Encyclopedia of Database Systems*, Heidelberg. Springer-Verlag.
- [Harrell, 2007] HARRELL, D. A. (2007). *Theory and technology for computational narrative : an approach to generative and interactive narrative with bases in algebraic semiotics and cognitive linguistics*. Thèse de doctorat, University of California, San Diego.
- [Harrison et al., 2017] HARRISON, B., PURDY, C. et RIEDL, M. O. (2017). Toward automated story generation with markov chain monte carlo methods and deep neural networks. In *Workshop : Intelligent Narrative Technologies (AAAI 2017)*, Snowbird, USA.
- [Herman, 1999] HERMAN, D. (1999). *Narratologies : New Perspectives on Narrative Analysis*. Theory and interpretation of narrative series. Ohio State University Press.
- [Herman, 2009a] HERMAN, D. (2009a). *Basic Elements of Narrative*. Wiley-Blackwell.
- [Herman, 2009b] HERMAN, D. (2009b). Narrative ways of worldmaking. *Narratology in the age of cross-disciplinary narrative research*, 20 :71.
- [Hovy, 1988] HOVY, E. (1988). *Generating Natural Language Under Pragmatic Constraints*. Lawrence, Erlbaum Associates, Hillsdale, NJ.
- [Huntington et al., 2017] HUNTINGTON, J. L., HEGEWISCH, K. C., DAUDERT, B., MORTON, C. G., ABATZOGLOU, J. T., MCEVOY, D. J. et ERICKSON, T. (2017). Climate engine : cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding. *Bulletin of the American Meteorological Society*, 98(11) :2397–2410.
- [Jain et al., 2017] JAIN, P., AGRAWAL, P., MISHRA, A., SUKHWANI, M., LAHA, A. et SANKARANARAYANAN, K. (2017). Story generation from sequence of independent short descriptions. In *Proceedings of Workshop on Machine Learning for Creativity (SIGKDD'17)*, Halifax, Canada.
- [Jasinskaja et Roßdeutscher, 2009] JASINSKAJA, K. et ROSSDEUTSCHER, A. (2009). Through narrative planning towards the preverbal message : A drt-based approach. *Ms. IMS, University of Stuttgart. Under review*.

- [Kim, 2004] KIM, J. H. (2004). *Beyond Role and Play – Tools, Toys and Theory for Harnessing the Imagination. The book for Solmukohta 2004*, chapter Immersive Story : A View of Role-Played Drama, pages 31–39. Ropecon ry, Helsinki.
- [Labov, 1972] LABOV, W. (1972). *Language in the Inner City. Studies in the Black English Vernacular*, chapter The Transformation of Experience in Narrative Syntax, pages 354–396.
- [Lacy, 2005] LACY, L. (2005). *Owl : Representing Information Using the Web Ontology Language*. Trafford Publishing.
- [Landa et Onega, 2014] LANDA, J. A. G. et ONEGA, S. (2014). *Narratology : an introduction*. Routledge.
- [Lebowitz, 1984] LEBOWITZ, M. (1984). Creating characters in a storytelling universe. *Poetics*, 13(3) :171–194.
- [Lehnert, 1981] LEHNERT, W. G. (1981). Plot units and narrative summarization. *Cognitive Science*, 5(4) :293–331.
- [Lehnert et Ringle, 2014] LEHNERT, W. G. et RINGLE, M. H., éditeurs (2014). *Strategies for Natural Language Processing*. Taylor & Francis.
- [Leskens *et al.*, 2017] LESKENS, J. G., KEHL, C., TUTENEL, T., KOL, T., HAAN, G. d., STELLING, G. et EISEMANN, E. (2017). An interactive simulation and visualization tool for flood analysis usable for practitioners. *Mitigation and Adaptation Strategies for Global Change*, 22(2) :307–324.
- [Li et Riedl, 2015] LI, B. et RIEDL, M. O. (2015). Scheherazade : Crowd-powered interactive narrative generation. *In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15*, pages 4305–4306, Austin, Texas. AAAI Press.
- [Limbourg et Vanderdonckt, 2004] LIMBOURG, Q. et VANDERDONCKT, J. (2004). Comparing task models for user interface design. *The handbook of task analysis for human-computer interaction*, 6 :135–154.
- [Lucquiaud, 2005] LUCQUIAUD, V. (2005). Proposition d'un noyau et d'une structure pour les modèles de tâches orientés utilisateurs. *In Proceedings of the 17th Conférence on l'Interaction Homme-Machine*, pages 83–90, Toulouse, France. ACM.

- [Mandler et Johnson, 1977] MANDLER, J. M. et JOHNSON, N. S. (1977). Remembrance of things parsed : Story structure and recall. *Cognitive psychology*, 9(1) :111–151.
- [Mani, 2012] MANI, I. (2012). Computational modeling of narrative. *Synthesis Lectures on Human Language Technologies*, 5(3) :1–142.
- [Mani, 2013] MANI, I. (2013). Computational narratology, in : Hühn, peter et al. (eds.) : The living handbook of narratology. [http://wikis.sub.uni-hamburg.de/lhn/index.php/Computational\\_Narratology](http://wikis.sub.uni-hamburg.de/lhn/index.php/Computational_Narratology). Consulté le 17 février 2017.
- [Mann et Thompson, 1988] MANN, W. C. et THOMPSON, S. A. (1988). Rhetorical structure theory : Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3) :243–281.
- [Martin et al., 2017] MARTIN, L. J., AMMANABROLU, P., WANG, X., HANCOCK, W., SINGH, S., HARRISON, B. et RIEDL, M. O. (2017). Event representations for automated story generation with deep neural nets. *In KDD 2017 Workshop on Machine Learning for Creativity*.
- [Mateas et Stern, 2005] MATEAS, M. et STERN, A. (2005). Structuring content in the façade interactive drama architecture. *In YOUNG, M. et LAIRD, J., éditeurs : Proceedings of the First Artificial Intelligence and Interactive Digital Entertainment International Conference (AIIDE)*, pages 93–98, Marina del Rey, États-Unis.
- [McGuinness et al., 2004] MCGUINNESS, D., VAN HARMELEN, F. et al. (2004). Owl web ontology language overview. *W3C Recommendation*, 10(10) :2004.
- [McIntyre et Lapata, 2009] MCINTYRE, N. et LAPATA, M. (2009). Learning to tell tales : A data-driven approach to story generation. *In SU, K., SU, J. et WIEBE, J., éditeurs : Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 217–225, Suntec, Singapour. ACM.
- [Meehan, 1977] MEEHAN, J. R. (1977). Tale-spin, an interactive program that writes stories. *In Fifth International Joint Conference on Artificial Intelligence*, volume 77, pages 91–98, San Mateo, Californie. Morgan Kaufmann.



- [Meister, 2009] MEISTER, J. (2009). *Handbook of Narratology*, chapter Narratology, pages 329–350. Walter De Gruyter, Berlin.
- [Nonnon, 2000] NONNON, É. (2000). Ce que le récit oral peut nous dire sur le récit. *Repères, recherches en didactique du français langue maternelle*, 21(1) :23–52.
- [Ogata, 1998] OGATA, T. (1998). Proposal of computational narratology. *In PSJ SIGNotes Computer and Humanities*. Abstract No.041 - 008 (écrit en japonais).
- [Ogata et Akimoto, 2016] OGATA, T. et AKIMOTO, T. (2016). *Computational and Cognitive Approaches to Narratology*. IGI Global, Hershey, USA, 1ère édition.
- [O’Neill, 2013] O’NEILL, B. (2013). *A computational model of suspense for the augmentation of intelligent story generation*. Thèse de doctorat, Georgia Institute of Technology.
- [Peinado et Gervás, 2006] PEINADO, F. et GERVÁS, P. (2006). Evaluation of automatic generation of basic stories. *New Generation Computing*, 24(3) : 289–302.
- [Pérez y Pérez, 1999] PÉREZ y PÉREZ, R. (1999). *MEXICA : A computer model of creativity in writing*. Thèse de doctorat, University of Sussex.
- [Pérez y Pérez, 2007] PÉREZ y PÉREZ, R. (2007). Employing emotions to drive plot generation in a computer-based storyteller. *Cognitive Systems Research*, 8(2) :89–109.
- [Perron, 2003] PERRON, P. (2003). *Narratology and text : subjectivity and identity in New France and Québécois literature*. University of Toronto Press.
- [Portet et al., 2009] PORTET, F., GATT, A., HUNTER, J., REITER, E. et SRIPADA, Y. (2009). Le projet BabyTalk : génération de texte à partir de données hétérogènes pour la prise de décision en unité néonatale. *In TALN 2009*, Senlis, France.
- [Propp, 1970] PROPP, V. (1970). Morphologie du conte. *Éditions du Seuil*.
- [Reagan et al., 2016] REAGAN, A. J., MITCHELL, L., KILEY, D., DANFORTH, C. M. et DODDS, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1) :31.

- [Reiter et Dale, 2000] REITER, E. et DALE, R. (2000). *Building natural language generation systems*. Cambridge university press.
- [Reiter et al., 2008] REITER, E., GATT, A., PORTET, F. et VAN DER MEULEN, M. (2008). The importance of narrative and other lessons from an evaluation of an NLG system that summarises clinical data. *In Proceedings of the 5th International Conference on Natural Language Generation (INLG-08)*, pages 147–153, Salt Fork, Ohio, United States.
- [Riedl et Young, 2010] RIEDL, M. O. et YOUNG, R. M. (2010). Narrative planning : balancing plot and character. *Journal of Artificial Intelligence Research*, 39(1) :217–268.
- [Roemmele, 2016a] ROEMMELE, M. (2016a). Writing stories with help from recurrent neural networks. *In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI’16.
- [Roemmele, 2016b] ROEMMELE, M. (2016b). Writing stories with help from recurrent neural networks.
- [Rowe et al., 2009] ROWE, J. P., MCQUIGGAN, S. W., ROBISON, J. L., MARCEY, D. R. et LESTER, J. C. (2009). Storyeval : An empirical evaluation framework for narrative generation. *In Intelligent Narrative Technologies II, Papers from the 2009 AAAI Spring Symposium, Technical Report SS-09-06*, pages 103–110, Stanford, États-Unis. AAAI.
- [Rudolph, 2011] RUDOLPH, S. (2011). Foundations of description logics. *In Reasoning Web. Semantic Technologies for the Web of Data*, pages 76–136. Springer.
- [Russell et Norvig, 2010] RUSSELL, S. et NORVIG, P. (2010). *Intelligence artificielle : Avec plus de 500 exercices*. Pearson Education France.
- [Sacco et Tzitzikas, 2009] SACCO, G. et TZITZIKAS, Y. (2009). *Dynamic Taxonomies and Faceted Search : Theory, Practice, and Experience*. The Information Retrieval Series. Springer Berlin Heidelberg.
- [Sanou et al., 2008] SANOU, L., GIRARD, P., GUITTET, L. et CAFFIAU, S. (2008). Tester la conformité d’une ihm à son modèle de tâches. *In Proceedings of the 20th Conference on l’Interaction Homme-Machine*, pages 159–162. ACM.
- [Scardimalia et Bereiter, 1987] SCARDIMALIA, M. et BEREITER, C. (1987). *The psychology of written composition*. Lawrence Erlbaum.

- [Schank et Abelson, 1977] SCHANK, R. et ABELSON, R. (1977). *Scripts, Plans, Goals, and Understanding : An Inquiry Into Human Knowledge Structures*. Artificial Intelligence Series. Taylor & Francis.
- [Schank, 1982] SCHANK, R. C. (1982). *Dynamic Memory : A Theory of Reminding and Learning in Computers and People*. Cambridge University Press.
- [Schank et Tesler, 1969] SCHANK, R. C. et TESLER, L. (1969). A conceptual dependency parser for natural language. *In Proceedings of the 1969 Conference on Computational Linguistics*, pages 1–3, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [Sebillotte, 1991] SEBILLOTTE, S. (1991). Décrire des tâches selon les objectifs des opérateurs, de l’interview à la formalisation. *Le travail humain*, pages 193–223.
- [Sharples, 1999] SHARPLES, M. (1999). *How we write : Writing as creative design*. Psychology Press.
- [Staab et Studer, 2004] STAAB, S. et STUDER, R. (2004). *Handbook on Ontologies (International Handbooks on Information Systems)*. SpringerVerlag.
- [Theune et al., 2004] THEUNE, M., RENSEN, S., op den AKKER, R., HEYLEN, D. et NIJHOLT, A. (2004). Emotional characters for automatic plot creation. *In International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, pages 95–100, Darmstadt, Allemagne. Springer.
- [Thorvaldsdóttir et al., 2013] THORVALDSDÓTTIR, H., ROBINSON, J. T. et MESIROV, J. P. (2013). Integrative genomics viewer (igv) : high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2) :178–192.
- [Todorov, 1969] TODOROV, T. (1969). *Grammaire du Decameron*. La Haye.
- [Todorov, 2013] TODOROV, T. (2013). *Introduction à la littérature fantastique*. Poétique. Éditions du Seuil.
- [Turner, 1993] TURNER, S. R. (1993). *Minstrel : A Computer Model of Creativity and Storytelling*. Thèse de doctorat, University of California Los Angeles, Los Angeles, États-Unis.

[Ware *et al.*, 2014] WARE, S. G., YOUNG, R. M., HARRISON, B. et ROBERTS, D. L. (2014). A computational model of plan-based narrative conflict at the fabula level. *IEEE Transactions on Computational Intelligence and AI in Games*, 6(3) :271–288.