ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2015-ENST-0073

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Informatique et Réseaux »

présentée et soutenue publiquement par

Antoine SAILLENFEST

le 25 novembre 2015

Modélisation Cognitive de la Pertinence Narrative en vue de l'Évaluation et de la Génération de Récits

Directeur de thèse : **Jean-Louis DESSALLES**

Jury

M. Nicolas SABOURET, Professeur, LIMSI, Université Paris Sud	Rapporteur
M. Nicolas SZILAS, Maître d'Enseignement et de Recherche, TECFA - FPSE, Université de Genève	Rapporteur
Mme Sylvie CATELLIN, Maître de Conférences HDR, CHCSC, Univ. de Versailles-Saint-Quentin	Examinatrice
M. Jean-Gabriel GANASCIA, Professeur, LIP6, Université Pierre et Marie Curie	Examinateur
Mme Catherine PELACHAUD, Directrice de Recherches, LTCI, Télécom ParisTech	Examinatrice
M. Jean-Daniel KANT, Maître de Conférences, LIP6, Université Pierre et Marie Curie	Invité
M. Jean-Louis DESSALLES, Maître de Conférences HDR, INFRES, Télécom ParisTech	Directeur de thèse

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

T H È S E

Résumé :

Une part importante de l'activité de communication humaine est dédiée au récit d'événements (fictifs ou non). Ces récits doivent être cohérents et intéressants pour être pertinents. Dans le domaine de la génération automatique de récits, la question de l'intérêt a souvent été négligée, ou traitée *via* l'utilisation de méthodes *ad hoc*, au profit de la cohérence des structures narratives produites. Nous proposons d'aborder le processus de création des récits sous l'angle de la modélisation quantitative de critères de pertinence narrative *via* l'application d'un modèle cognitif de l'intérêt événementiel. Nous montrerons que cet effort de modélisation peut servir de guide pour concevoir un modèle cognitivement plausible de génération de narrations.

La théorie de la Simplicité, qui est notre point d'entrée pour la modélisation des récits, est un modèle cognitif de l'intérêt événementiel. L'évaluation de l'intérêt repose principalement sur le calcul de l'inattendu d'un événement, défini comme écart de complexité cognitive, notion dérivée de la complexité de Kolmogorov. Les évaluations subjectives des événements, lorsqu'elles proviennent de plusieurs sources, peuvent conduire à des conflits d'ordre cognitif internes à un individu, quand un même état reçoit deux évaluations opposées. Nous montrons que les conflits cognitifs internes aux personnages sont sources de conflits narratifs, eux-mêmes sources d'intérêt pour l'audience et d'action pour les personnages.

Le cadre d'étude introduit par la théorie de la Simplicité et la notion de conflit cognitif nous amène à redéfinir formellement les modalités classiques de l'intérêt narratif (la surprise, le suspense et la curiosité). De plus, l'occurrence d'une situation inattendue fait apparaître un ou plusieurs conflits internes à un individu, ce qui l'amène à calculer une explication ou, lorsque ses désirs et préférences sont en jeu, à planifier une action pour favoriser l'occurrence d'un état désiré ou éloigner la perspective d'un état non-désiré.

Pour que le récit reste vraisemblable, les actions planifiées par les personnages doivent paraître intentionnelles pour l'audience. Nous montrons que les calculs de complexité cognitive peuvent servir à évaluer *a posteriori* le caractère intentionnel ou non des actions d'un personnage. Cela nous amène à formuler des critères de crédibilité des personnages dans un cadre narratif et à redéfinir formellement certaines notions classiques : les intentions, la responsabilité et le jugement moral.

Les modèles introduits dans cette thèse peuvent servir à analyser les récits, mais aussi à les générer. Nous montrons comment notre modèle de la surprise et des évaluations des intentions peut en principe servir de guide pour la conception d'un générateur de séquences narratives.

Abstract :

Humans devote a considerable amount of time to producing narratives. Whatever a story is used for (whether to entertain or to teach), it must be relevant. Relevant stories must be believable and interesting. The field of computational generation of narratives has explored many ways of generating narratives, especially well-formed and understandable ones. The question of what makes a story interesting has however been largely ignored or barely addressed. Only some specific aspects of narrative interest have been considered. No general theoretical framework that would serve as guidance for the generation of interesting and believable narratives has been provided. The aim of this thesis is to introduce a cognitive model of situational interest and use it to offer formal criteria to decide to what extent a story is relevant. Such criteria could guide the development of a cognitively plausible model of story generation.

Simplicity Theory, which is our starting point, is a complexity-based account for situational interest. Cognitive complexity is a notion derived from Kolmogorov complexity. Simplicity Theory defines unexpectedness (of a situation ; a key notion to evaluate the interest) as complexity drop. Individuals gather information from various external or internal sources to evaluate situations. Situations in which a same proposition is given two opposite evaluations generate cognitive conflicts. In stories, the introduction of narrative conflicts is a source of interestingness and drives characters to act. We show that narrative conflicts result from the characters' internal cognitive conflicts.

Many notions related to narrative interest can be formally redefined using the framework of Simplicity Theory (including surprise, suspense or curiosity). Unexpected situations create internal cognitive conflicts. While trying to solve these conflicts, individuals make attempts to explain what happened. They may also calculate plans of action to avoid the occurrence of an undesired state of affairs or favor the occurrence of a desired one.

Relevant stories must remain coherent. The characters' behaviour must be regarded as intentional by the audience. We assume that the a character's believability relies on computations based on cognitive complexity. This serves as a basis for defining notions related to characters' believability in narrative contexts (including intentions, responsibility and moral judgment).

The models introduced in this thesis can be used not only to evaluate but also generate stories. We show how our model of surprise and the evaluation of a character's intentions may serve as guidance for the development of a story generation system.

Remerciements

Je tiens à adresser mes remerciements les plus sincères à Jean-Louis Dessalles, mon directeur de thèse. Cela a été un réel plaisir d'échanger et de travailler avec lui pendant ces quatre années. Il m'a donné de nombreux conseils pertinents en matière de rédaction, de communication ou encore d'enseignement. De nos nombreux échanges ont émergé beaucoup d'idées ou de concepts ; on retrouvera une grande partie du fruit de ces échanges dans ce manuscrit.

J'adresse mes remerciements à celles et ceux qui ont accepté de participer à mon jury de thèse, à commencer par Catherine Pélachaud, la présidente dudit jury, ainsi que les rapporteurs Nicolas Sabouret et Nicolas Szilas. Je remercie également Sylvie Catellin et Jean-Gabriel Ganascia. Enfin je remercie Jean-Daniel Kant d'avoir accepté de faire partie de ce jury, lui qui fut à l'origine de ma rencontre avec Jean-Louis Dessalles.

Je tiens ensuite à remercier les équipes du laboratoire INFRES qui m'a accueilli pendant cette thèse ainsi que celles de Télécom ParisTech. En particulier, je tiens à remercier l'ensemble du groupe IC2 (Interaction Communication Cognition) et plus particulièrement l'équipe DBweb (*Database and web*). Je remercie Sylvain, Dong Bach, Quentin, Paul-Louis, Simon, Hoa, Cuauhtémoc et enfin Robin. Enfin, je remercie tout particulièrement Rémi ; tous les prétextes sont bons pour lui apporter du chocolat.

Je remercie aussi l'ensemble des membres de la Chaire MODIM, Modélisation des Imaginaires, Innovation et Création, dans le cadre de laquelle ces travaux ont été réalisés et la Fondation Télécom qui, au travers du programme Futur et Ruptures, a permis de les financer. J'adresse en particulier mes remerciements à Pierre Musso, directeur de la chaire MODIM, et à nouveau à Catherine Pélachaud, la présidente de mon jury de thèse.

Je voudrais ensuite remercier ma famille, et en particulier mes parents, qui m'ont permis, sans toujours l'avoir complètement compris, de suivre ce parcours d'étude et m'ont soutenu.

Je dois aussi remercier tous mes amis dont la présence a été nécessaire et importante pendant ces années de thèse. En particulier, je voudrais tout d'abord remercier d'anciens camarades d'École d'Ingénieurs : Olivier, Cécile, Julie, Jérémie, Célia, Nicolas (qui est devenu docteur le même jour que moi), Sébastien, Héléna

et Cyril (qui est devenu docteur bien avant moi). Ensuite, je voudrais remercier Damien et Mathieu. Je remercie également Mathilde et Alexandre. Enfin, j'adresse des remerciements particuliers à Quentin ; c'est une longue histoire.

Enfin, je remercie Nicolas qui m'a accompagné tout au long de cette thèse, qui a été présent matin midi et soir, dans les bons moments comme dans les moins bons. Il m'a donné de nombreux conseils pertinents. Ensemble, nous allons maintenant pouvoir nous consacrer à d'autres projets.

« Les miracles véritables, qu'ils font peu de bruit ! Les événements essentiels, qu'ils sont simples ! Sur l'instant que je veux raconter, il est si peu à dire qu'il me faut le revivre en rêve, et parler à cet ami. » Antoine de Saint-Exupery - Lettre à un otage

Table des matières

1 L'ordinateur qui rêvait d'écrire des histoires 13

2 Éléments de pertinence narrative 19

2.1 Pertinence structurelle dans les récits 20

2.1.1 Niveaux d'interprétation des narrations 22

2.1.2 Unité des narrations 23

2.1.3 Vraisemblance, cohérence et crédibilité dans le monde fictif . 25

2.1.4 Psychologie humaine, psychologie des personnages 26

2.1.5 Intentions et crédibilité des personnages 27

2.1.6 Responsabilité et jugement moral dans la fiction 28

2.2 Pertinence émotionnelle et tension narrative 30

2.2.1 *Narrative transportation* et émotions par procuration 31

2.2.2 La structure en 3 actes et le conflit narratif 32

2.2.3 L'inattendu comme source d'intérêt narratif 34

2.2.4 Rôle de l'inattendu dans l'expérience de la surprise 35

2.2.5 Le cas du suspense 37

2.2.6 Postdictibilité et baisse de la tension narrative 40

2.2.7 Lecture, relecture et attentes 41

2.3 Crédibilité contre Intérêt, le dilemme de l'auteur 42

2.4 Pertinence narrative, non-pertinence et critère de pertinence totale . 43

3 Différentes approches de la génération automatique des récits 49

3.1 Approches dans le processus de génération 51

3.1.1 Les systèmes de génération basés sur les personnages 52

3.1.2 Les systèmes centrés sur la fonction de l'auteur 55

3.1.3 Les systèmes hybrides 61
3.2 Les générateurs de tension narrative 63
3.3 Évaluation des modèles de génération de narrations 68
3.4 Approche cognitive de la génération de scénarios 71

4 Introduction au rôle de la complexité cognitive dans l'intérêt 73
4.1 Contraintes de la modélisation cognitive 74
4.1.1 Le modèle scientifique 74
4.1.2 La plausibilité cognitive 75
4.2 Introduction à la complexité au sens de Kolmogorov 76
4.3 Rôle de la complexité dans l'intérêt événementiel 81
4.3.1 Introduction à la théorie de la Simplicité 82
4.3.2 Calculabilité de la complexité cognitive 84
4.3.3 Complexité et génération des objets 84
4.3.4 Complexité et description des objets 87
4.4 Limites de la théorie des probabilités 90
4.5 Plagiat ou coïncidence ? La complexité comme facteur de décision . 93
4.5.1 Modélisation des coïncidences 94
4.5.2 Analyse en complexité d'un cas de plagiat 95
4.6 Complexité et pertinence émotionnelle dans la narration 101

5 Conflits cognitifs et conflits narratifs 109
5.1 Nécessité d'un état et conflits cognitifs internes à un individu 111
5.1.1 Nécessité et support d'un état 112
5.1.2 Le conflit d'ordre cognitif 113
5.1.3 Évaluation des nécessités 114
5.1.4 Quelques considérations au sujet de la résolution des conflits 115
5.2 Le conflit narratif : un ensemble de conflits cognitifs 116
5.3 Le conflit narratif, un élément essentiel de la pertinence narrative . 118

6 Complexité et modalités de l'intérêt narratif 123
6.1 Le cas de la surprise 124
6.1.1 Occurrence et évaluation d'une situation surprenante 125
6.1.2 L'événement surprenant source de conflits cognitifs 128

6.1.3 Un modèle de la surprise comme source de conflits 133
6.2 Curiosité et suspense 134
6.2.1 Le cas de la curiosité 135
6.2.2 Le cas du suspense 139
6.3 Pertinence émotionnelle et émotions par procuration 143
6.3.1 Identification aux personnages et empathie 144
6.3.2 Notations 145
6.4 Conclusion 146

7 Intérêt et crédibilité des personnages : un dilemme ? 149
7.1 Le dilemme intérêt-vraisemblance 150
7.2 Conflits cognitifs et nécessité des actions 151
7.3 Reconnaissance des intentions d'un acteur 153
7.3.1 Modélisation cognitive des intentions 153
7.3.2 Interprétation et prédictions du modèle 157
7.4 Responsabilité et jugement moral 163
7.4.1 Modélisation cognitive de l'attribution de responsabilité . . . 164
7.4.2 Modèle du jugement moral 166
7.4.3 Interprétation et prédictions du modèle 166
7.5 Analyse empirique et validation des modèles 168
7.6 Résoudre le dilemme intérêt vraisemblance 170

8 Complexité et analyse de scénario 175
8.1 Analyse en complexité 177
8.2 Bilan de l'analyse 186

9 Approche cognitive de la génération automatique des récits 191
9.1 Modèle de la surprise et génération de récits surprenants 192
9.1.1 Un modèle de génération récursif de séquences narratives surprenantes 192
9.1.2 Les structures événementielles enrichies CRRS (Conflit-Résolution-Réalisation-Surprise) 194
9.2 Conditions d'une approche cognitive de la génération de scénarios . 197
9.3 CAN : une procédure cognitivement plausible de résolution des conflits 199

9.3.1 L'abduction, un processus de raisonnement délibératif et de création littéraire . . . 199
9.3.2 La procédure Conflit-Abduction-Négation . . . 201
9.3.3 La procédure conflit-abduction : un modèle argumentatif pour la planification d'actions . . . 205
9.4 Mise en œuvre de la procédure conflit-abduction . . . 207
9.4.1 Représentation d'un monde fictif . . . 207
9.4.2 Détection des conflits cognitifs . . . 209
9.4.3 Le processus d'abduction . . . 211
9.4.4 La procédure conflit-abduction . . . 213
9.5 Génération récursive de séquences narratives . . . 214
9.5.1 Génération et intégration de conflits inattendus . . . 214
9.5.2 Génération d'une structure CRRS . . . 217
9.5.3 Génération récursive de la structure événementielle augmentée d'une séquence narrative localement surprenante . . . 220
9.5.4 Quelques considérations sur la génération du discours . . . 220
9.6 Évaluation du modèle et perspectives . . . 222

10 Conclusion et perspectives 227

A Annexe - Expériences 239
A.1 Rôle de la complexité de description dans la reconnaissance d'un possible cas de plagiat . . . 240
A.2 Intentions et Responsabilité - L'accident de la route . . . 243
A.3 Le dilemme moral - La mine inondée . . . 250

B Glossaire 265

Chapitre 1

L'ordinateur qui rêvait d'écrire des histoires

> « Concevons qu'on ait dressé un million de singes à frapper au hasard sur les touches d'une machine à écrire et que, sous la surveillance de contremaîtres illettrés, ces singes dactylographes travaillent avec ardeur dix heures par jour avec un million de machines à écrire de types variés. Les contremaîtres illettrés rassembleraient les feuilles noircies et les relieraient en volumes. Et au bout d'un an, ces volumes se trouveraient renfermer la copie exacte des livres de toute nature et de toutes langues conservés dans les plus riches bibliothèques du monde. » (Borel, 1913)

Le paradoxe des singes savants, dont on doit l'une des premières versions à Borel (1913) est généralement formulé ainsi : un singe immortel qui frapperait indéfiniment au hasard sur les touches d'un clavier produirait presque certainement n'importe quel texte de taille finie une infinité de fois. Ce singe produirait n'importe quel roman, ceux qui ont été écrits, ceux qui le seront un jour et même ceux qui ne le seront jamais.

Programmer un ordinateur à écrire des caractères aléatoires pendant un temps fini serait un effort vain pour générer des scénarios. D'une part parce que notre planète aura 'très certainement' disparu avant que l'ordinateur produise ne serait-ce qu'une phrase (de taille moyenne, pas trop courte) sensée dans une langue

quelconque. D'autre part parce qu'il faudrait 'très certainement' un temps bien supérieure à une vie humaine (voire même à plusieurs vies humaines) pour retrouver une telle phrase (si on suppose qu'elle a finalement été produite) au milieu de la suite de caractères aléatoires produits.

Une question se pose alors : comment 'programmer le singe savant' pour qu'il produise un bon récit dans un délai raisonnable ? Nous allons apporter des éléments de réponse dans cette thèse.

Enjeux scientifiques et économiques

Une grande part des productions humaines repose sur la production de bons récits, en particulier dans l'industrie du divertissement. La valeur d'un jeu vidéo, d'une pièce de théâtre, d'un film, d'une série ou encore d'un roman se mesure (entre autres) à la valeur narrative des récits autour desquels ils sont construits.

L'être humain est un narrateur. Une part importante de son activité de communication est dédiée à la production de récit (Norrick, 2000). Ces récits jouent un rôle important sur le plan social, ils permettent d'informer, facilitent la mémorisation d'événements et influencent fortement notre perception du monde et de son organisation. Être un bon narrateur est une qualité sur le plan des relations sociales, cela contribue à donner de la valeur à un individu. Étudier et comprendre les mécanismes cognitifs qui sont à l'œuvre dans la production des récits représente donc un enjeu scientifique majeur.

Produire des récits à forte valeur narrative représente également un enjeu sur le plan économique. Par exemple, dans le domaine du divertissement, à l'échelle mondiale l'industrie du jeu vidéo représentait 41.9 milliard de dollars de chiffres d'affaire en 2011 et devrait représenter 75 milliard de dollars de chiffre d'affaire en 2015. Pour la France, cette industrie représente environ 5000 emplois directs.

Par ailleurs, l'être humain est de plus en plus souvent amené à évoluer dans des environnements virtuels, au travers de la réalité augmentée par exemple, et est de plus en plus confronté à des agents virtuels intelligents avec lesquels il communique. La production de récits pertinents, éventuellement interactifs, représente un enjeu technique majeur en terme d'appréhension et de compréhension de ces environnements par les humains.

Enfin, le volume de données produites à l'échelle mondiale est en constante augmentation, ce qui rend de plus en plus difficile le traitement, l'analyse ou la compréhension de ces données. Déjà, la génération de séquences narratives à partir de données régulières et structurées permet de produire des articles de presse à partir de données boursières ou de résultats d'élections. Être capable de produire des narrations à partir de volumes importants de données généralement hétérogènes représente là encore un enjeu technique et économique majeur afin de rendre accessible automatiquement le contenu de ces données à une audience définie. Il ne s'agit pas d'extraire et de présenter toute l'information contenue dans les données, mais uniquement l'information pertinente qu'elles contiennent.

Notre travail de modélisation cognitive se situe au carrefour de la psychologie, des études littéraires, de l'analyse quantitative des récits et de l'intelligence artificielle (en particulier le sous-domaine de la génération automatique de récits). Dans nos travaux, nous faisons l'hypothèse qu'une meilleure compréhension des mécanismes cognitifs liés à la production de récits peut permettre, outre une analyse cognitive des récits, de développer de nouvelles approches pour la génération automatique de scénarios.

Une approche cognitive de la pertinence narrative

Le domaine de la génération automatique de récits est un champ de l'Intelligence Artificielle dont l'objet est la 'programmation du singe savant', c'est-à-dire le développement de programmes qui mettent en œuvre des méthodes permettant de produire des narrations (au sens large) *via* diverses approches motivées par des objectifs définis.

L'un de ces objectifs est généralement de produire des récits vraisemblables. En effet, un récit n'est pas qu'une suite d'événements ordonnés. Ces événements doivent se suivre de manière cohérente, se déroulent dans un monde qui doit apparaître comme étant vraisemblable et doivent se rapporter au sujet du récit. Ces contraintes doivent être respectées pour permettre, entre autres, une bonne compréhension et une bonne mémorisation des récits.

Or la valeur d'un récit vient aussi des effets qu'il produit sur son audience. La 'valeur émotionnelle' d'un récit vient de l'intérêt qu'il suscite, de la tension

narrative créée et maintenue tout au long du récit.

Créer des récits qui provoquent une réponse émotionnelle de l'audience est un sujet qui, jusqu'ici, a été relativement négligé dans le domaine de la génération automatique de récits, ou traité *via* l'utilisation de méthodes *ad hoc*.

Comme cela a été dit plus haut, comprendre les mécanismes cognitifs de la production des récits par un humain est source d'enjeux scientifiques et être capable de reproduire et automatiser ces mécanismes est source d'enjeux techniques et économiques majeurs. Dans nos travaux, nous abordons l'étude de la production des récits sous l'angle de l'analyse et la modélisation cognitive et vue du développement de modèles de génération automatique de récits. En particulier, nous nous intéressons à la modélisation cognitive de la pertinence narrative.

La pertinence narrative regroupe l'ensemble des critères qui contribuent à donner de la valeur, sur le plan narratif, à un récit. Cela regroupe à la fois des critères structurels, c'est-à-dire des critères qui contribuent à renforcer la qualité du récit du point de vue de la structure des événements qui le composent, ainsi que des critères de pertinence émotionnelle, c'est-à-dire des critères qui contribuent à rendre le récit plus intéressant.

La modélisation cognitive de l'intérêt narratif est notre point d'entrée dans l'étude de la pertinence narrative. Nous utiliserons un modèle mathématique de l'intérêt basé sur la notion de complexité, dérivée de la notion de complexité de Kolmogorov dans le domaine de la théorie de l'information. L'objectif de nos travaux est de prolonger ce modèle pour établir des critères d'évaluation de la pertinence d'éléments constitutifs des récits. Nous montrerons que ces modèles peuvent servir de guide en vue du développement de modèles de génération automatique de narrations pertinentes.

'Le singe savant' ne sait que taper sur les touches d'un clavier, il ne rêve pas d'écrire une histoire, il n'a pas d'autre objectif que celui de taper sans cesse sur les touches de son clavier. La machine artificiellement intelligente de demain rêvera peut-être d'écrire une histoire, nous allons essayer de l'aider à réaliser son rêve artificiel.

Plan de nos travaux

Dans le chapitre 2, nous détaillons la notion de pertinence narrative *via* deux approches de l'analyse des récits : l'approche structurelle et l'approche événementielle. L'approche structurelle nous amène à étudier les récits sous l'angle des événements qui le composent et qui doivent s'articuler de manière à rendre vraisemblable le monde (possiblement fictif) dans lequel ils se produisent. Cette étude structurelle nous conduit à l'étude des actions des personnages qui doivent donner l'illusion d'agir de manière rationnelle. Cette première partie de l'analyse permet de donner un aperçu de critères de ce que nous appelons la pertinence structurelle. L'approche événementielle nous amène à étudier les récits sous l'angle des effets qu'ils produisent sur une audience. En particulier, nous nous intéressons au rôle du conflit narratif dans la réponse émotionnelle à un récit et à l'inattendu comme facteur d'intérêt narratif. Cette seconde partie de l'analyse permet de donner un aperçu de critères de ce que nous appelons la pertinence émotionnelle.

Au travers de l'analyse de modèles de génération de scénarios existants dans le chapitre 3, nous montrons que la question de l'intérêt a souvent été négligée, ou traitée *via* l'utilisation de méthodes *ad hoc*, au profit de la cohérence des structures narratives produites. La pertinence structurelle des récits à donc été privilégiée par rapport à la pertinence émotionnelle. Notre approche consiste à modéliser cognitivement la pertinence narrative et à définir dans un même espace des critères de pertinence structurelle et de pertinence émotionnelle en vue de la création de modèles de génération de narrations cohérentes et intéressantes.

Dans le chapitre 4, nous introduisons la théorie de la Simplicité, un modèle cognitif de l'évaluation des probabilités subjectives d'occurrence de situations dans un cadre fictif ou non. Cette évaluation repose sur le calcul de l'inattendu d'une situation, défini comme écart de complexité cognitive, notion dérivée de la complexité de Kolmogorov.

Ces évaluations subjectives d'états du monde proviennent de sources diverses et peuvent conduire à des conflits d'ordre cognitif internes à un individu lorsqu'un même état reçoit deux évaluations opposées. Dans le cadre narratif, nous montrons dans le chapitre 5 que les conflits cognitifs internes aux personnages sont sources de conflits narratifs, eux-même sources d'intérêt et d'action des personnages.

Le cadre d'étude introduit par la théorie de la Simplicité et la notion de conflit cognitif nous amènent à proposer une définition et une évaluation de modalités classiques de l'intérêt narratif dans le chapitre 6. En particulier nous proposons un modèle cognitif de la surprise et de l'attitude d'un individu (fictif ou non) postérieure à l'observation de l'occurrence d'un événement surprenant.

Les conflits cognitifs placent les individus dans une situation d'inconfort sur le plan cognitif et peuvent donc les pousser à agir afin de résoudre lesdits conflits. Dans le cadre narratif, l'action des personnages doit, *a minima*, paraître intentionnelle pour l'audience. Nous montrons dans le chapitre 7 que les calculs de complexité cognitive peuvent servir à définir une évaluation de la reconnaissance des intentions d'un individu ainsi que des critères de crédibilité des personnages dans le cadre narratif.

Les modèles introduits dans nos travaux peuvent servir à analyser les récits (chapitre 8) ou à en générer. Nous utiliserons, dans le chapitre 9 le modèle de la surprise évoqué ci-dessus et les évaluations des intentions comme guides pour la conception d'un générateur de séquences narratives configurées par la surprise.

Chapitre 2

Éléments de pertinence narrative

Dans son dictionnaire de narratologie, Prince (1987) définit la narration comme le fait de raconter une séquence d'événements se rapportant à un sujet et constituant un tout ("*the recounting of a sequence of events that have a continuant subject and constitute a whole.*"). Cette définition nous renvoie directement à la définition aristotélicienne de la tragédie :

> [...] la tragédie est imitation d'une action menée jusqu'à sa fin et formant un tout, ayant une certaine étendue ; car il se trouve des choses qui forment un tout, mais n'ont aucune étendue. Forme un tout, ce qui a commencement, milieu et fin. (Aristote, b335, p.96 (1450b))

La définition de Prince établit que la narration est un objet qui provient d'un émetteur (celui qui la raconte) et s'adresse à une audience (celui à qui on la raconte). L'objet de cette thèse n'est pas de faire une théorie générale des narrations mais de présenter divers niveaux d'analyses utiles pour la modélisation puis la génération automatique de récits. Ainsi dans la suite de ce document, je considérerai l'auteur comme l'émetteur et l'audience comme le récepteur d'un récit. Dans certains cas particuliers, l'audience sera caractérisée par les termes 'lecteur' (dans le cas d'un récit écrit), 'spectateur' (dans le cas d'un récit présenté sous forme visuelle – théâtre, cinéma, etc.) ou 'auditeur' (dans le cas d'un récit oral).

Les définitions précédentes montrent la possibilité d'analyser les récits sous deux angles différents. Le premier angle est l'analyse événementielle du récit, celle qui considère l'événement comme la brique de base du récit. Le second sous lequel

peut être abordée l'analyse des récits est l'angle des effets qu'ils produisent sur l'audience, c'est-à-dire aux sources d'intérêt narratif. Nous allons aborder ces deux modes d'analyse successivement.

Ces deux approches constituent deux aspects de ce que nous appellerons la pertinence narrative des éléments du récits.

Le premier aspect de cette notion est la pertinence structurelle d'un élément du récit. Un élément structurellement pertinent contribue au respect des critères d'unité du récit, de cohérence et de vraisemblance (en particulier la vraisemblance des personnages dans le récit). Nous allons détailler ces notions dans la partie 2.1.

Dans un second temps, nous aborderons le second aspect de la pertinence narrative que nous nommons pertinence émotionnelle. Nous analyserons les récits sous l'angle des conflits narratifs qui les composent. Lesdits conflits sont sources d'intérêt narratif (on utilisera aussi le terme de tension narrative). Un élément du récit émotionnellement pertinent dans le récit contribue à produire un effet sur l'audience, à produire de l'intérêt ou à faire baisser la tension narrative en participant à la résolution des conflits auxquels les personnages sont confrontés. Ces sujets sont détaillés dans la partie 2.2

La pertinence narrative d'un élément dans le récit est une forme de caractérisation et d'évaluation de la valeur apportée par cet élément à la narration. Un auteur de récit doit sans cesse veiller à ce que l'ajout d'un élément au récit n'affecte pas négativement la crédibilité ou l'intérêt dudit récit, ce qui peut dans certains cas conduire à un dilemme (voir partie 2.3). Nous discuterons dans la partie 2.4 du statut des éléments non-pertinents et de la nécessité pour l'auteur de respecter un critère de pertinence totale (ou absolue).

2.1 Pertinence structurelle dans les récits

La définition aristotélicienne définit la narration comme l'imitation d'une action, celle de Prince comme le récit d'une séquence d'événements. On peut définir un événement comme un changement d'état d'état du monde fictif ("*a change from one state to another state*" (Toolan, 2012)). Prince (1987) propose une définition plus complète de notion d'événement :

> "*A change of state manifested in discourse by a process statement in the mode of Do or Happen. An event can be an action or act (when the change is brought about by an agent : "Mary opened the window") or a happening (when the change is not brought about by an agent : "the rain started to fall").*" (Prince, 1987, p.28)

Cette définition fait donc la distinction entre les événements qui sont réalisés par des entités agissantes (les acteurs) et ceux qui se produisent sans intervention d'un acteur (et qui ne dépendent que de la volonté de l'auteur qui contrôle le monde fictif).

L'acteur est une entité dotée de certains attributs (ou traits de personnalité) et capable de réaliser des actions. Les acteurs dans le cadre narratif sont les personnages des récits. Les événements dont la réalisation est provoquée par un acteur sont les actions. Les événements dont l'occurrence ne dépend pas d'un acteur sont des *happenings* ou *world events*.

Un événement est caractérisé par un changement d'état du monde. L'état du monde est un ensemble de propositions qui caractérisent le monde à un instant donné. Ces propositions peuvent être vérifiées ou niées (par exemple, "il fait beau" est une proposition dont la négation est "il ne fait pas beau").

La réalisation d'un événement peut être conditionnée à un ensemble de préconditions (ou circonstances). Ces circonstances sont des propositions (éventuellement niées) qui caractérisent l'état du monde fictif nécessaire à la réalisation d'un événement (par exemple, "La fenêtre n'est pas ouverte" et "Marie est près de la fenêtre" sont des préconditions de l'événement "Marie ouvre la fenêtre").

Les ensembles "préconditions – événement – changements dans l'état du monde", sont les composants de la causalité qui préside à l'évolution du monde fictif. Dans le monde fictif, tout n'est pas possible. Le monde fictif, comme le 'monde réel', est régi par un ensemble de lois causales, la causalité, qui déterminent ce qui peut se produire ou non en fonction de l'état du monde à un instant donné. Cette notion de causalité joue un rôle très important dans l'expérience narrative. Elle est ce autour de quoi est organisée notre compréhension et notre perception des événements dans le monde fictif (Trabasso and Sperry, 1985; van den Broek, 1988; Graesser et al., 1991). Nous verrons dans la suite que la causalité est une propriété inhérente

des récits, et qu'elle permet en particulier d'assurer la continuité et l'unité du sujet de la narration (Chatman, 1993).

2.1.1 Niveaux d'interprétation des narrations

Dans la définition de Prince, la narration est le récit d'une séquence d'événements qui se réalisent dans le monde fictif. Cette définition introduit une distinction entre les événements qui se produisent et la manière de les présenter à une audience. Aristote faisait une distinction fondamentale entre le *logos* – l'imitation d'une action réelle, c'est-à-dire l'histoire en elle-même – du *mythos* – la sélection et le possible réarrangement des éléments qui constituent le *logos* (Prince, 1987).

On trouve chez les formalistes russes une décomposition similaire des récits en trois niveaux d'interprétation : la *fabula*, le *sjuzet* et le niveau 'texte / medium' (Prince, 1987) (voir figure 2.1).

La *fabula* est la liste ordonnée chronologiquement de tous les événements de la narration qui se produisent dans le monde fictif entre le moment où l'histoire commence et le moment où elle se termine. Le récit présenté à une audience ne contient pas nécessairement tous les éléments de la *fabula* mais seulement un sous-ensemble d'événements pertinents au regard des intentions du narrateur. L'état dans lequel le monde se trouve au début de la *fabula* est l'état initial du monde. L'état dans lequel se trouve le monde à la fin de la *fabula* est l'état final du monde.

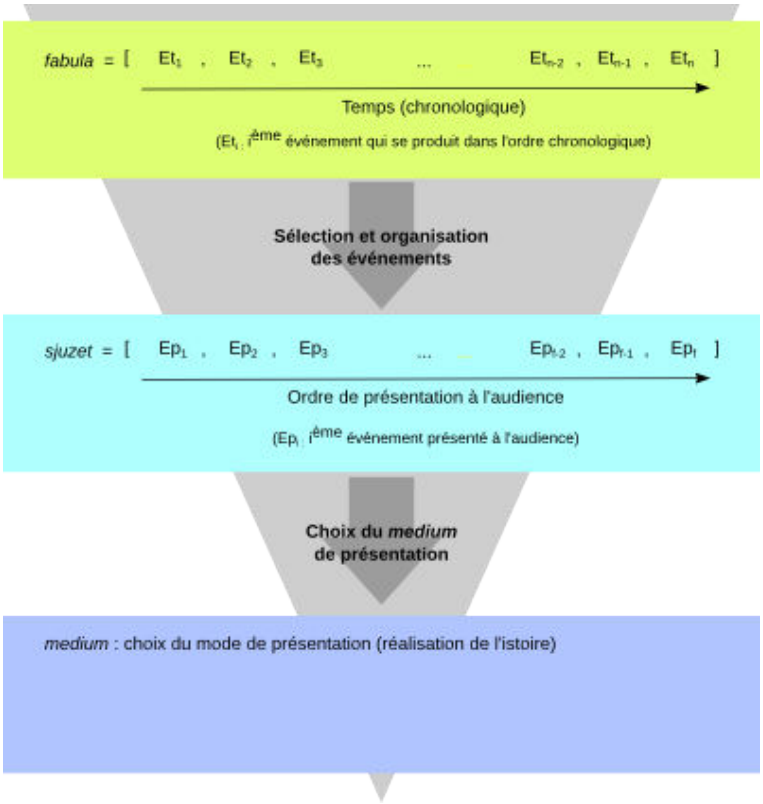
Le *sjuzet* est le sous-ensemble des événements présents dans la *fabula* rangés dans l'ordre dans lequel ils sont présentés à une audience.

Enfin, le dernier niveau d'interprétation 'texte / médium', aussi appelé 'discours', concerne tout ce qui se rapporte à la présentation en langage naturel des événements, au médium utilisé pour présenter le récit, à la mise en scène, etc.

Cette séparation en différents niveaux est artificielle du point de vue de l'auteur d'un récit. Cependant, elle est intéressante du point de vue de l'analyse et de la génération de scénarios car elle permet de traiter chacun de ces niveaux indépendamment. Nous allons principalement nous intéresser aux niveaux *fabula* et *sjuzet*.

Ces niveaux d'interprétations permettent d'analyser et de représenter les récits du point de vue événementiel. Cependant, une narration ne peut pas être résumée

FIGURE 2.1 – Niveaux d'interprétation des narrations



à un simple ensemble d'événements organisés et présentés à une audience. Un ensemble aléatoire d'événements ne constitue généralement pas une narration acceptable, les événements qui composent une narration doivent impérativement respecter le critère d'unité.

2.1.2 Unité des narrations

Le critère d'unité est destiné à assurer la cohérence du récit, sa continuité et sa focalisation autour d'un sujet unique (unicité de l'action).

> « Aussi, de même que dans les autres arts d'imitation, l'unité d'imitation résulte de l'unité de l'objet, de même l'histoire – puisqu'elle est imitation d'action – doit être imitation d'une action une et formant un tout ; et les parties constituées des actes accomplis doivent être agencées de façon que, si l'on déplace ou supprime

> l'une d'elles, le tout soit troublé et bouleversé. Car ce qui s'ajoute ou ne s'ajoute pas sans résultat visible, n'est en rien partie du tout. » (Aristote, b335, p.98 (1451b))

Le critère d'unité stipule que les éléments d'une narration doivent se rapporter à un sujet unique ou à un but communicatif unique (Chatman, 1993). Cela peut faire référence au récit d'un événement unique, à la focalisation autour d'un personnage unique ou bien encore à l'unicité du lieu dans lequel se passe le récit (Lavandier, 2004; Chion, 1985). Classiquement, l'unité est décomposée en unité de temps, de lieu et d'action.

Au sujet de l'unité de temps, Lavandier (2004) écrit que « l'histoire doit durer le temps nécessaire au protagoniste pour atteindre son objectif » (p.204), ce qui renvoie à la notion d'étendue d'Aristote (le temps pour passer de la mauvaise fortune à la bonne fortune) (Aristote, b335). L'unité de lieu renvoie à la situation du récit dans un lieu unique ou un ensemble restreint de lieux. Ce qui était valable pour le théâtre grec (ou il y avait un décor unique et donc un lieu unique) ne l'est pas nécessairement pour d'autres genres (cinéma, roman, etc.). Aussi l'unité de lieu et l'unité de temps ne sont pas des critères nécessaires pour former des narrations acceptables.

L'unité d'action, en revanche, « continue à être capitale » (Lavandier, 2004, p.204). C'est un critère de pertinence et de parcimonie. C'est un « principe selon lequel toutes les parties de l'œuvre (y compris les éventuelles sous-intrigues) doivent servir l'action » (Lavandier, 2004, p.566). Cela s'applique également aux détails introduits. On prête communément à Anton Tchekhov la suggestion de supprimer tout élément non pertinent du récit, ce qui est connu sous le nom de loi du fusil de Tchekhov : si l'auteur indique qu'il y a un fusil accroché au mur dans le premier chapitre, alors ce fusil doit absolument servir dans les chapitres suivants, sinon il n'a pas sa place dans le récit.

L'unité assure donc que les événements qui composent un récit sont pertinents et sont liés au sujet unique de la narration. L'un des éléments importants qui participent au respect du critère d'unité est la causalité, qui assure la cohérence entre les événements. L'ensemble des événements qui font évoluer l'état du monde, leurs causes et leurs effets, doivent respecter la causalité du monde, que leur occurrence

soit le fait d'un personnage ou 'du monde' (donc de l'auteur).

2.1.3 Vraisemblance, cohérence et crédibilité dans le monde fictif

Comme nous l'avons vu précédemment, la narration peut être définie comme une suite d'événements qui changent l'état du monde fictif. Ce monde fictif est le lieu dans lequel se déroule la narration. Il ne nous est pas accessible en totalité, seuls quelques éléments nécessaires à son établissement sont fournis à l'audience. Ces éléments peuvent être de natures différentes. Il peut s'agir de propositions qui caractérisent l'état du monde, personnages inclus, à un moment donné ("Yoda a un sabre laser" est une proposition qui décrit, à un moment donné, l'état du monde fictif dans lequel se déroulent les films de la série *Star Wars*). Il peut également s'agir de règles causales qui caractérisent la causalité du monde. Dans le monde fictif dans lequel se déroule *Star Wars*, il faut par exemple définir les modalités d'utilisation du sabre laser (objet purement fictif). Une règle causale associée au sabre laser est, par exemple, "un personnage utilise un sabre laser comme arme pour tuer un autre personnage" ; c'est un événement possible (ici une action) du monde fictif dans lequel se déroule l'action de Star Wars. Ces règles causales doivent être introduites par l'auteur et respectées par celui-ci ou les personnages-acteurs, sans quoi la vraisemblance du monde fictif peut être affectée.

On peut distinguer deux modalités de la vraisemblance : la cohérence du monde et la crédibilité du monde.

La cohérence renvoie directement au respect des règles causales introduites par l'auteur ou qui peuvent provenir du genre de la narration. En effet, certains genres sont associés à certaines attentes. Les films de la série Indiana Jones appartiennent au genre fantastique-aventure. Ainsi, dans le monde fictif dans lequel vit Indiana Jones, un état de fait tel que "être immortel" est un possible état d'un personnage et "devenir immortel" un possible événement.

Cependant, le monde dans lequel vit Indiana Jones ressemble beaucoup au monde "réel". Ainsi, des événements tels que l'occurrence d'un tremblement de terre ou la mort par crise cardiaque d'un protagoniste sont des événements 'rares' à la fois dans le 'monde réel' ainsi que dans le monde fictif dans lequel se situe

l'action des films de la série Indiana Jones. Un tel événement, s'il devait survenir sans avoir été 'préparé', et s'il devait modifier de manière significative la séquence des événements au moment approprié, aurait l'apparence d'une coïncidence et risquerait d'affecter très négativement la crédibilité du monde fictif (même si, du point de vue de la causalité, cela peut se produire, il n'y aurait pas d'incohérence). Dans Indiana Jones et la dernière croisade, le 'tremblement de terre' qui détruit le temple du Graal ne survient pas 'par hasard', il est la conséquence directe des actions des personnages (ils tentent de sortir le Graal du temple alors que c'est interdit, et cela active un mécanisme de destruction du temple). La survenue d'événements rares au moment approprié est souvent une faiblesse dans un scénario (sauf s'il s'agit de créer un effet comique), de tels événements donnent l'impression qu'à partir de maintenant, tout devient possible dans la suite de l'histoire et plus rien n'est vraiment inattendu.

La vraisemblance des personnages renvoie également à la notion de cohérence (ils doivent réaliser des actions qui respectent la causalité du monde) ainsi qu'à la notion de crédibilité. L'un des facteurs essentiel de la crédibilité des personnage est la capacité de l'audience à reconnaître les intentions desdits personnages. Les actions des personnages rationnels doivent être intentionnelles et reconnues comme telles par l'audience. Par ailleurs, les personnages crédibles peuvent être jugés, l'audience peut leur attribuer une responsabilité. Ces jugements d'intention et de responsabilité jouent un rôle important dans la formation d'une attitude de l'audience envers les personnages ainsi que l'attitude d'un personnage envers un autre personnage. La possibilité (pour l'audience ou d'autres personnages) de former des attitudes et des jugements à l'égard des personnages renforce la crédibilité des acteurs supposés rationnels (Ochs et al., 2009).

2.1.4 Psychologie humaine, psychologie des personnages

D'un point de vue logique, les personnages ne peuvent pas être considérés comme des personnes réelles. Cependant, quand on regarde de plus près la manière avec laquelle les personnes réelles perçoivent et analysent le comportement de celles et ceux qui les entourent dans le monde réel, la distinction qui est faite entre personnages réels et personnages fictifs est moins apparente. Personne ne

peut prétendre connaître une autre personne dans sa totalité (c'est-à-dire avoir accès à la totalité de l'état mental de cette autre personne). L'information sur ceux qui nous entourent est discontinue, schématique, incomplète. C'est valable y compris pour ceux dont nous sommes les plus proches. Ainsi, comme l'affirme Stanley Fish (1989), nous ne connaissons pas les gens réels beaucoup plus que les personnages d'un récit. Cette position est partagée par Hochman (1985). Bien que les personnages réels soient plus complets sur le plan ontologique, nous ne les connaissons pas parfaitement. Par contraste, les personnages de fictions sont plus schématiques, mais nous avons le sentiment que nous les connaissons mieux que les personnes réelles et qu'ils ont plus d'unité et de cohérence.

Il n'est donc pas déraisonnable de défendre l'idée que les personnages réels comme fictifs sont traités cognitivement de la même manière (Bortolussi and Dixon, 2003, p.140), et donc que nous pouvons traiter les personnages fictifs comme s'ils étaient réels (ce que nous faisons dans les travaux présentés ici).

2.1.5 Intentions et crédibilité des personnages

La crédibilité des personnages fictifs repose en grande partie sur la capacité d'une audience à reconnaître leurs intentions. L'audience attend d'eux qu'ils essayent de résoudre leurs problèmes et que chacune de leurs actions soit motivée. "*One important aspect of character believability is character intentionality [i.e.] the way in which the choice of actions and behaviors that a character makes appears natural (and possibly rational) to external observers.*" (Riedl and Young, 2005)

Un modèle classique en psychologie affirme qu'un personnage qui agit de manière intentionnelle doit nécessairement désirer la réalisation d'un certain état de fait (désir de résultat) et avoir un niveau de connaissance suffisant qui lui permet de déterminer comment son action peut mener à cet état de fait (Forguson, 1989). Ce modèle a été complété par Shaver (1985) qui a proposé une définition plus élaborée : une action est intentionnelle si l'agent (acteur) désire un résultat, est conscient des conséquences de son action ainsi que de sa capacité à pouvoir réaliser ladite action. Le modèle de Malle & Knobe (1997) détaille quant à lui cinq composants nécessaires qui permettent de déterminer si un résultat (ou un but) a été atteint de manière intentionnelle. L'agent doit désirer ce résultat, avoir une

connaissance qui lui permet de déterminer que son action lui permet d'atteindre ce résultat, réaliser l'action volontairement, être conscient du fait qu'il réalise l'action au moment où il la réalise, et être apte à réaliser l'action de manière fiable.

Ces modèles mettent l'accent sur la capacité à anticiper la réalisation d'un objectif, et donc à former des plans (un ensemble d'actions ordonnées) qui respectent la causalité du monde fictif. La causalité est intimement liée à la reconnaissance et l'attribution d'intentions aux personnages du récit (Graesser et al., 1991; Gerrig, 1993). Dans le 'monde réel' comme dans le monde fictif, notre compréhension du comportement des entités qui agissent dépend de notre capacité à comprendre les plans qu'ils forment pour atteindre leurs objectifs (Bortolussi and Dixon, 2003, p.116). Schank & Abelson (1977) ont d'ailleurs fait de la capacité d'une audience à inférer les plans des personnages un élément essentiel de leur modèle de compréhension des narrations.

Il faut donc fournir à l'audience suffisamment d'informations sur les désirs des personnages, l'état de leur connaissance du monde fictif et, en particulier, des règles causales qui font évoluer ce monde pour que l'audience soit capable de reconnaître les intentions desdits personnages.

Au delà de la reconnaissance des intentions, les actions des personnages affectent la manière dont nous les percevons ou dont ils sont perçus par les autres personnages du monde. En particulier, nous sommes capables de leur attribuer une forme de responsabilité.

2.1.6 Responsabilité et jugement moral dans la fiction

Au cours de la lecture d'un récit, le lecteur ne se contente pas de reconnaître la portée des actions des personnages mais est également capable de former des jugements (moraux) à leur encontre. Nous nous intéressons ici aux jugements de responsabilité, qui peuvent changer à la fois notre attitude à l'égard des personnages mais également celle des autres personnages de l'histoire à l'égard d'un personnage donné.

La causalité joue, là encore, un rôle très important. Les acteurs sont jugés, au moins en partie, responsables des conséquences de leurs actes (Shaver, 1985; Lagnado and Channon, 2008; Schlenker et al., 1994; Hart and Honoré, 1985; Mil-

ler and Gunasegaram, 1990; N'gbala and Branscombe, 1995). L'attribution de responsabilité et la reconnaissance des intentions d'un acteur semblent être deux concepts proches et intimement liés (Malle and Nelson, 2003). Laganado & Channon (2008) avancent l'idée que le niveau de responsabilité augmenterait en même temps qu'augmente l'intention de produire l'état de fait pour lequel l'acteur est jugé responsable. Cette automaticité est remise en cause par Pizarro et al. (2003), en particulier dans des cas dits 'déviants', ou le cours des événements n'est pas celui attendu par un acteur. Voici un exemple de cas 'déviant' :

John a l'intention de tuer Marc. Il prend une arme, puis monte dans sa voiture et se rend chez Marc pour le tuer. Il est un peu stressé et, alors qu'il est presque arrivé chez Marc, perd le contrôle de sa voiture et renverse un individu. Cet individu se trouve être Marc.

Dans ce cas, les personnes interrogées ont tendance à attribuer moins de responsabilité à John que s'il avait tué Marc en utilisant son arme. L'intention initiale de John n'a cependant pas varié. Or John est jugé responsable pour les actes qui se sont réellement produits, et l'intention *a posteriori* est également calculée en fonction de ce qui s'est produit. L'intention initiale n'a pas variée, mais John n'avait pas l'intention de tuer Marc en le renversant avec sa voiture.

L'un des facteurs de l'attribution de responsabilité est la prédictibilité de l'état de fait pour lequel un agent est jugé responsable. Lagnado & Channon (2008) ont montré que, plus un agent semble certain que le plan qu'il a décidé de réaliser mène à un certain état de fait, plus il est jugé responsable pour l'occurrence de cet état de fait. Là encore, il semble que les effets de la prédictibilité dans l'attribution de la responsabilité puissent être annulés (Pizarro et al., 2003). Si l'on reprend l'exemple précédent, le cas 'déviant', John prédit bien la mort de Marc avec un fort degré de certitude, il est certain que son plan va réussir, mais l'attribution d'une responsabilité atténuée est lié au caractère imprévu de l'accident. Le jugement de responsabilité semble donc dépendre de la prédictibilité des événements qui se sont produits, et non pas de la prédictibilité de ce qui avait été anticipé.

Si l'attribution de la responsabilité semble dépendre de notre capacité à inférer les plans d'un agent, la longueur de ces plans (ou chaînes causales) semble être

un facteur important pour le contrôle du niveau de responsabilité attribué. En particulier, il a été montré que que l'attribution de responsabilité diminue à mesure que la distance causale (en nombre d'événements présents dans la chaîne causale) augmente (Johnson and Drobny, 1985).

Un dernier facteur important dans le calcul du niveau de responsabilité attribuée à un agent semble être l'intensité de la réponse émotionnelle à un état de fait conséquence d'une action (Phares and Wilson, 1972).

Comme pour la reconnaissance des intentions, un observateur extérieur (par exemple un lecteur), doit avoir accès à l'état mental de l'acteur (par exemple un personnage de la narration) qui va être jugé par cet observateur. Le jugement de responsabilité présuppose la capacité à se projeter dans l'esprit d'un personnage, ce que l'on appelle classiquement la théorie de l'esprit (Baron-Cohen, 1999). Comparons les situations suivantes :

— Tom donne à manger à son fils Neil un plat. Neil mange ce plat pour la première fois. Personne ne sait que Neil est allergique à ce plat. Neil mange le plat et tombe gravement malade.

— Harry a toussé toute la journée. Son père, Louis, lui donne un médicament destiné aux adultes. Harry tombe rapidement gravement malade.

Classiquement, les personnes interrogées attribuent plus de responsabilité à Louis pour ce qui arrive à Harry qu'à Tom pour ce qui arrive à Neil. La raison vient de l'attribution d'un état d'esprit (ici le niveau de connaissance) différent à Louis et à Tom au moment de réaliser leur action (Notons ici que, dans les deux cas, peu de gens vont considérer que Louis et Tom ont 'blessé' leur enfant de manière intentionnelle).

2.2 Pertinence émotionnelle et tension narrative

Dans la partie précédente, nous avons analysé les récits en adoptant une approche événementielle. Nous avons détaillé divers niveaux d'interprétation des récits et détaillé des conditions d'acceptabilité des récits telles que la cohérence et la crédibilité du monde fictif et des personnages de la narration. Si cela peut suffire à rendre le récit acceptable, cela n'en fait pas forcément un bon récit.

Dans cette partie, nous allons aborder l'analyse des récits sous l'angle des effets

qu'ils produisent sur une audience.

2.2.1 *Narrative transportation* et émotions par procuration

Les humains adorent se laisser emporter dans un récit. Même lorsque nous sommes familiers avec certains genres de récits (récits d'aventure, d'horreur, de super-héros, etc.) et que nous nous attendons à ce que certains événements se produisent (par exemple la survie du héros et la mort du méchant), nous prenons du plaisir à vivre (et même à revivre) l'expérience de la narration. Plusieurs questions se posent à ce sujet : Pourquoi les événements racontés suscitent-ils de l'intérêt ? Pourquoi ressentons-nous les émotions des personnages ?

Une explication possible est le phénomène de *narrative transportation*. Ce phénomène se produit lorsque le lecteur ou le spectateur se perd dans l'histoire et se retrouve comme détaché du 'monde réel' à cause de l'attention porté au récit (Gerrig, 1993; Green et al., 2004). Ceci est lié au phénomène de suspension consentie de l'incrédulité (*willing suspension of disbelief*). Lavandier (2004) définit cette trêve de l'incrédulité comme l'≪ attitude d'un spectateur de fiction qui consiste à laisser son esprit cartésien au vestiaire et à accepter les conventions et les artifices de l'œuvre d'art ≫ (p.565) ≪ Quand la lumière s'éteint dans une salle de théâtre ou de cinéma, le spectateur sait que ce qu'il va voir n'est pas la réalité mais une représentation artificielle de celle-ci, une métaphore de la vie. Il est donc prêt à accepter un certain nombre d'entorses à la réalité qu'il connaît ≫ (p.249). Ainsi, nous ressentons les émotions comme si nous étions nous-mêmes partie intégrante des événements racontés.

Un autre phénomène est celui de l'identification aux personnages. Parce que nous ressentons une proximité avec certains personnages du récit, nous nous identifierions à eux et ressentirions les mêmes émotions qu'eux lorsqu'ils sont confrontés aux événements du monde fictif (Zillmann, 1995). Cohen (2001) décrit ce phénomène comme "*a process that consists of increasing loss of self-awareness and its temporary replacement with heightenend emotional and cognitive connections with a character*"(Cohen, 2001, p.251)

Cependant, la théorie de l'identification a ses limites. Imaginez un individu qui

se dirige vers un précipice mais n'en a pas conscience. Si l'audience a conscience de ce qui va possiblement se passer, et si elle est attachée à l'individu, alors l'audience va vivre un moment de suspense et ressentir une tension. Le personnage n'ayant pas conscience de ce qui peut lui arriver, cette tension n'est pas le résultat d'un partage d'émotion lié à l'identification au personnage.

Une théorie alternative concernant les émotions ressenties par l'audience d'un récit est la théorie de l'empathie, résumée ainsi par Brewer (1996), p.109 : "*the reader feels emotions for fictional characters that are like those the reader would feel for nonfictional individuals in similar circumstances*". Ainsi, d'après cette théorie, parce que l'audience éprouve de l'empathie pour le personnage qui s'apprête à chuter, celle-ci s'intéresse à ce qui peut arriver à cette personne et ressent des émotions qui dépendent de notre attitude envers ce personnage (si l'audience apprécie ou non ce personnage).

2.2.2 La structure en 3 actes et le conflit narratif

Dans sa définition de la narration (voir page 19), Aristote propose une structuration du récit en trois actes : le début – le milieu – la fin. Cette structuration définit l'évolution de la tension narrative au cours du récit.

Le début est une phase d'exposition et de mise en place de l'intrigue. Y sont introduits les personnages, les lieux, l'atmosphère ou encore le thème du récit. Cette phase suscite généralement la curiosité et l'intérêt de l'audience qui se pose des questions quant au déroulement de l'histoire et la direction que prend celle-ci. Généralement, le début du récit se termine par un événement catalyseur, un conflit (*inciting incident*) qui amorce la seconde phase du récit.

Cette seconde phase, le milieu, est une phase de développement de l'intrigue. Les conflits présentés dans la phase introductive sont ancrés et les obstacles rencontrés par les personnages qui tentent de résoudre ces conflits sont détaillés. Cette phase est souvent marquée par des temps forts, des éléments nouveaux et inattendus, qui entraînent généralement les protagonistes dans des directions nouvelles. Du point de vue rythmique, ces temps forts relancent l'histoire et ont pour fonction de maintenir le spectateur en éveil. Cette phase se termine généralement par un événement marquant, le *climax*, le point de tension maximale du récit. Il s'agit

par exemple, dans de nombreux récits, de l'ultime affrontement entre un héros et son *nemesis*, son ennemi principal.

La troisième et dernière phase du récit est une phase de résolution des conflits et d'achèvement de l'histoire. Le *climax* est suivi d'une phase de clôture du récit. Les quelques conflits restants sont résolus et le monde fictif retrouve une forme de 'stabilité' (mais son état n'est généralement pas le même qu'au début du récit, le monde a évolué d'un 'état stable' à un autre).

FIGURE 2.2 – L'arc dramatique d'Aristote

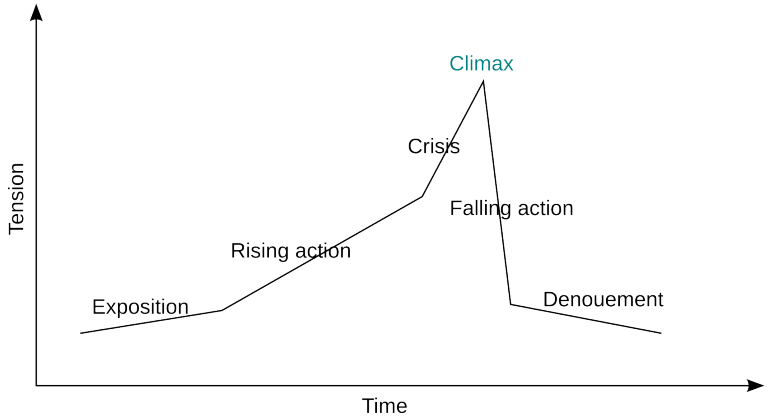
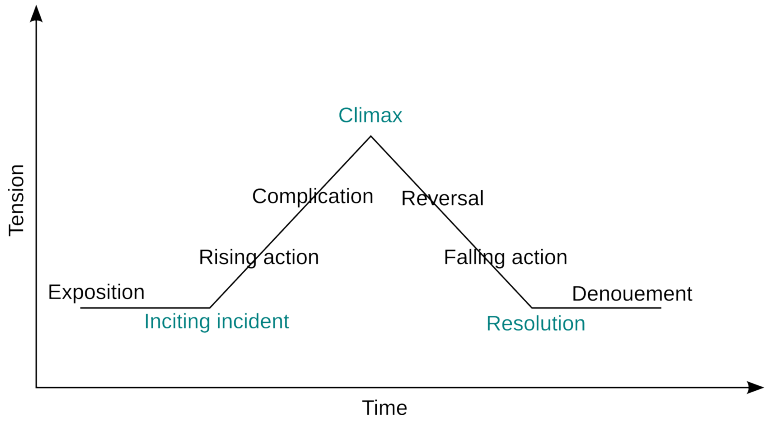


FIGURE 2.3 – L'arc dramatique de Freytag (pyramide de Freytag)



Cette structure en trois actes est une structure configurée par l'effet, la tension narrative, que le récit produit sur l'audience. La figure 2.2 représente la structu-

ration du récit (ou arc dramatique) proposée par Aristote. Une structure similaire bien que légèrement différente est proposée par Freytag (1894) (fig. 2.3).

Les récits sont généralement organisés autour d'un conflit narratif principal dont la présentation marque la fin de la première phase du récit et dont la résolution marque le passage à la troisième phase du récit (c'est une façon de redéfinir le critère d'unité, l'ensemble du récit est organisé autour d'un conflit narratif principal). Ainsi la structure en trois actes peut être résumée ainsi :

avant le conflit – du conflit à sa résolution – après la résolution du conflit

En littérature, le conflit narratif est classiquement défini comme l'incompatibilité entre les objectifs de deux personnages ou forces. Il y a généralement un conflit principal, qui est le fil conducteur du récit, et un ensemble de conflits secondaires qui participent à la résolution du conflit principal ou sont des 'obstacles' rencontrés par les personnages dans la réalisation de leurs plans. Ces conflits intermédiaires contribuent à maintenir l'attention de l'audience pendant le déroulement du récit.

2.2.3 L'inattendu comme source d'intérêt narratif

La notion de conflit dans le récit est très liée à la notion d'intérêt. La première vertu du conflit est en effet de faire monter la tension narrative (ou intérêt narratif) (Lavandier, 2004, p.40-41). Les travaux de Kintsch (1980) sur l'intérêt narratif ont montré qu'il y a deux sources majeures de l'intérêt narratif : une réponse émotionnelle à de grands 'thèmes' tels que la mort, le sexe ou encore la politique (ce qui est parfois appelé *emotional interest*) ainsi que la survenue d'événements surprenants ou inattendus.

Une distinction courante est faite entre intérêt individuel et intérêt situationel. La première approche s'intéresse aux différences individuelles qui peuvent exister dans la réponse émotionnelle face à une situation, la seconde s'intéresse aux facteurs objectifs de l'intérêt (cette distinction ne fait pas consensus (Hidi, 1990)). On peut toutefois noter que l'inattendu est une notion subjective. Imaginez un tirage du loto qui formerait un numéro de téléphone. Pour la plupart des individus, ce tirage n'a pas d'intérêt, car il n'apparaît pas comme inattendu. Pour celui à qui appartient ce numéro, ou quiconque qui reconnaîtrait ce numéro de téléphone, ce

tirage est inattendu (Dessalles, 2006). On peut également citer cet autre exemple de Teigen & Keren (2003) :

> "*an astrophysicist may be extremely 'surprised' to learn that the moon is one billion years older than previously thought, whereas a layman may declare himself not surprised at all ; not because the new estimate was expected, but because it did not conflict with any previously held beliefs.*" (Teigen and Keren, 2003)

Ainsi, le niveau d'intérêt d'une situation résulte rarement d'un calcul objectif. C'est le résultat d'un calcul subjectif qui dépend des différences individuelles entre les divers observateurs d'une situation donnée, de leurs états mentaux particuliers au moment de l'observation.

Dans le cadre narratif sont classiquement distinguées différentes modalités de l'intérêt narratif, en particulier la surprise, la curiosité et le suspense (Baroni, 2007). Nous allons analyser le sens commun donné à ces notions et les liens qui existent entre ces notions et la notion d'inattendu.

2.2.4 Rôle de l'inattendu dans l'expérience de la surprise

Surprise et inattendu sont deux notions proches et souvent confondues. Chacune des deux notions amène un même genre de questionnement sur les raisons de la survenue de tel ou tel événement. La notion de surprise est couramment associée à la notion de faible probabilité. Un événement dont l'occurrence est considérée comme rare serait donc surprenant. Cette relation inverse entre probabilité et surprise a été mise en évidence par Reisenzein (2000). Cependant, tous les événements dont la probabilité d'occurrence est rare ne sont pas nécessairement surprenants. Ou plus précisément, certains événements rares ne sont pas nécessairement perçus comme tels, et donc ne sont pas surprenants pour un individu non-conscient de leur rareté. Cela renvoie à la subjectivité du calcul de l'inattendu évoquée plus haut, c'est-à-dire à l'influence des connaissances particulières des individus dans le calcul de l'inattendu.

La notion de faible probabilité d'occurrence ne rend pas compte de manière satisfaisante de la surprise associée à certains événements. Par exemple, si vous ouvrez un pot pour prendre un gâteau et que vous réalisez en ouvrant ce pot

qu'il est vide, alors vous serez surpris parce que ce à quoi vous vous attendiez – y trouver un gâteau – ne correspond pas à ce que vous observez. Il n'y a pas de probabilité faible liée à l'absence d'un gâteau dans le pot, en revanche il existe une représentation du monde dans laquelle le gâteau se trouve dans le pot avec un certain degré de certitude, qui forme donc une attente. Si le pot est vide, cette attente est non-confirmée.

Une situation peut également être surprenante lorsqu'il n'y a pas d'attente. Par exemple, si quelqu'un entre dans votre bureau par erreur, vous allez être surpris alors que vous n'aviez pas formé dans votre esprit d'attente particulière. Il n'existe pas non plus de probabilité objective associée à la survenue de cet événement.

Dans un sens général, la notion de surprise rend compte de la réaction d'individus face à des situations perçues comme rares, des attentes non-confirmées et des événements non-anticipés.

Dans les cas où nous disposons de la probabilité objective d'occurrence d'un événement, alors cela nous fournit une mesure du niveau de surprise associé à la survenue dudit événement. Or nous ne disposons pas de la probabilité d'occurrence de la plupart des événements surprenants présents dans les récits. Aussi, Maguire et al (2011) font l'hypothèse que le niveau de surprise associé à l'occurrence d'un événement dépend plutôt du 'niveau de difficulté' pour intégrer cet événement dans une représentation existante du monde. Cette idée est illustrée par l'exemple suivant :

> *For example, if you found your car key was missing, and you had no way of explaining it, then you might experience a high level of surprise. However, if a plausible explanation subsequently emerged that allowed the anomaly to be resolved, such as realising that you left the key inside the car, then the experience of surprise should subside.* (Maguire et al., 2011)

Ainsi, le niveau d'inattendu dépendrait de la difficulté à rendre compte d'un événement dans un contexte existant. Ce niveau de difficulté dépend donc de la connaissance que l'on a du monde, ce qui inclut entre autres les 'probabilités d'occurrence'. De manière plus générale, un individu utilise sa connaissance de la causalité et de l'état du monde (réel ou fictif) pour intégrer les situations qu'il observe.

Dans le cadre narratif, il n'est pas rare qu'un auteur ne divulgue pas à l'audience immédiatement, de manière volontaire, un ou plusieurs éléments importants ou pertinents qui permettraient d'expliquer l'occurrence d'une situation présentée à cette audience. Cette rétention d'information est un procédé classique de création de surprise. Cela correspond à l'exemple classique de l'explosion (inattendue) d'une bombe, qui est surprenante pour l'audience, et dont l'explication (Qui a posé la bombe ? Pourquoi ? Dans quel but ?) est fournie ultérieurement. Outre la surprise initiale, une tension 'argumentative', ou explicative, est maintenue jusqu'à la révélation des éléments pertinents qui permettent d'expliquer, *a posteriori*, le fait surprenant. Ainsi, il s'agit d'un processus en deux temps, tout d'abord un élément surprenant qui ne s'intègre pas immédiatement dans la représentation qu'à l'audience du monde fictif, puis une tension narrative maintenue jusqu'à la révélation des éléments nécessaires à l'intégration de cette situation.

Certains procédés narratifs permettent de créer une surprise brève, c'est-à-dire sans volonté de la part de l'auteur de maintenir la tension suite à l'occurrence de l'événement surprenant ; un exemple de surprise brève dans un cadre narratif est le *deus ex machina*. Ce procédé classique consiste à résoudre un conflit par la survenue d'un événement non anticipé (et non anticipable). Un *deus ex machina* est une aide extérieure qui intervient pendant que l'un des protagonistes de l'histoire essaye de résoudre un conflit. C'est un événement inattendu ou improbable qui vient régler les problèmes du protagoniste à la dernière minute (Lavandier, 2004). L'expression qui signifie "Dieu issu de la machine" fait référence à la dramaturgie grecque dans laquelle un dieu faisait parfois son apparition par l'intermédiaire d'une grue à la fin des pièces. Le *deus ex machina* peut prendre bien des formes : une crise cardiaque opportune, la foudre qui vient abattre le méchant au bon moment ou encore la coïncidence salvatrice. Ici, la surprise est de courte durée, l'événement est assimilable à une loterie.

2.2.5 Le cas du suspense

> « Nous sommes en train de parler, il y a peut-être une bombe sous cette table et cette conversation est très ordinaire [...] et tout d'un coup, boum, explosion. Le public est surpris, mais, avant qu'il ne l'ait été, on lui a montré une scène

> absolument [...] dépourvue d'intérêt. Maintenant, examinons le suspense. La bombe est sous la table et le public le sait. [...] Le public sait que la bombe explosera à une heure et il sait qu'il est une heure moins le quart – il y a une horloge dans le décor ; la même conversation anodine devient tout à coup très intéressante parce que le public participe à la scène [...] Dans le premier cas, on a offert au public quinze seconde de surprise au moment de l'explosion. Dans le deuxième cas, nous lui offrons quinze minutes de suspense. La conclusion de cela est qu'il faut informer le public chaque fois qu'on le peut, sauf si la surprise est un twist, c'est-à-dire lorsque l'inattendu de la conclusion constitue le sel de l'anecdote. » Alfred Hitchcock dans (Truffaut and Hitchcock, 1975, p.59)

Le suspense est une modalité de la tension narrative (Baroni, 2007; Sternberg, 1992). Sternberg (1992) le définit ainsi :

> "*Suspense throws us forward to the opacity of the future [...] It works toward a temporal direction that is opaque (open, gapped, indeterminate) by nature : the future that darkens the resolution of conflict, character change, personnal dilemma, interpersonal entanglement, to-ing and fro-ing between ideologies, or any other developing entity in the world's arena.*" (Sternberg, 1992, p.531 et 537)

Pour Lavandier (2004), le suspense est une « attente angoissée, génératrice d'anxiété, vécue par le spectateur et créée par l'incertitude de la réponse dramatique. Appelé aussi tension. Il y a du suspense dans toute œuvre qui pose une question dramatique ou ironique et qui les exploite » (p.564).

Selon Zillmann (1991) (repris par Vorderer 1996) les facteurs qui provoquent la tension liée au suspense (de type 'détresse empathique') sont les suivants :

— il doit y avoir un personnage du récit pour lequel l'audience ressent de la sympathie (ou *a minima* ne le déteste pas).
— il doit y avoir une menace qui pèse sur ce personnage et qui est attendue (du point de vue de l'audience) mais néanmoins pas tout à fait certaine.

On peut décliner cette approche pour les personnages qui n'apparaissent pas sympathique, ce que Zillmann appelle *unempathic distress* :

— il doit y avoir un personnage du récit pour lequel l'audience ressent de

l'aversion (ou *a minima* ne l'aime pas).

— il doit y avoir une bonne fortune qui pèse sur ce personnage et qui est attendue (du point de vue de l'audience) mais néanmoins pas tout à fait certaine.

À la notion de sympathie pour tel ou tel personnage, Vorderer (1996) préfère parler de la relation au personnage. Cette notion couvre un nombre de cas plus large, par exemple les personnages que l'on préfère, ceux que l'on aime le plus, ceux auxquels nous nous sentons le plus liés ou encore ceux avec lesquels nous aimerions le plus nouer des contacts. Cette généralisation opérée par Vorderer vient du constat que, hormis dans la fiction populaire, les personnages ne sont pas aussi facilement séparables entre bons et mauvais, entre ceux que l'on aime et ceux que l'on n'aime pas.

D'après ce qui précède, nous pouvons constater que la notion de suspense est intimement liée à la notion d'inattendu de l'occurrence de situations (désirées ou non par l'audience selon la relation qu'elle a avec les personnages). Le suspense est une préparation à la survenue d'un événement généralement chargé émotionellement. La phase de suspense maintient la tension narrative en maintenant l'incertitude sur l'issue des événement, toute issue étant alors plus ou moins inattendue.

Truffaut (1975) proposait d'ailleurs une technique visant à produire du suspense en provoquant, *via* l'introduction d'une limite temporelle, une augmentation graduelle de l'inattendu de l'occurrence d'une situation désirée par un personnage :

> « Un exemple. Un personnage part de chez lui, monte dans un taxi et file vers la gare pour prendre le train. C'est une scène normale à l'intérieur d'un film moyen. Maintenant, si avant de monter dans un taxi, cet homme regarde sa montre et dit : "Mon Dieu, c'est épouvantable, je n'attraperai jamais mon train", son trajet devient une pure scène de suspense car chaque feu rouge, chaque croisement, chaque agent de la circulation, chaque panneau indicateur, chaque coup de frein, chaque manipulation du levier de vitesses vont intensifier la valeur émotionnelle de la scène. » (Truffaut and Hitchcock, 1975, p.16)

La phrase "Mon Dieu, c'est épouvantable, je n'attraperai jamais mon train" est une phrase d'accroche qui crée la tension en construisant une attente chez le spectateur/lecteur (cette technique de création de suspense par introduction d'une

phrase d'accroche a été utilisée dans le cadre de la génération automatique de récits par Miletitch et al. (2012))

2.2.6 Postdictibilité et baisse de la tension narrative

La tension narrative, quelle que soit la modalité qui en est la source, ne peut être maintenue indéfiniment, et doit retomber avant la fin du récit. Le critère de postdictibilité [1] consiste à restaurer la cohérence du monde en présentant à l'audience les éléments qui lui permettent d'intégrer toute situation du récit dans le monde fictif. Kintsch (1980) définit ainsi la postdictibilité :

> "*The text as a whole must hang together and make sense to the reader, so that he is able to construct a coherent macrostructure in which each text unit has its place and is meaningfully related to other sections of the text. Weird and unpredictable statements in a text are interesting only in so far as they are well motivated within the text as a whole, at least by hindsight.*" (Kintsch, 1980, p.89)

Ce critère est est un critère de restauration de la vraisemblance du monde fictif. C'est l'une des différences importantes qui peut exister entre le récit de faits réels et une fiction. Dans le monde réel, un événement non-expliqué nous met dans une situation d'inconfort mais est intéressant à raconter. Dans le cadre d'une fiction, le critère de postdictibilité oblige l'auteur à fournir une explication à toute situation présentée à l'audience avant la clôture du récit afin de ne pas laisser l'audience dans une situation d'inconfort à la fin du récit.

Le non-respect du critère de postdictibilité peut, en particulier, affecter négativement la crédibilité d'un personnage. Si un personnage réalise une action inattendue, cela crée un effet de surprise immédiat, ce qui est intéressant localement. Mais cela affecte sa crédibilité si l'audience n'est pas capable d'expliquer cette action. Il y a un contrat tacite entre l'auteur et son audience, une sorte de promesse faite à l'audience que tout rentrera dans l'ordre, autrement dit que tout sera justifié d'ici la fin du récit. Si aucun élément ultérieur dans le récit ne permet d'expliquer l'action

1. En l'absence de traduction communément admise du terme anglais *postdictability* introduit dans Kintsch (1980), nous choisissons de traduire ce terme en français par 'postdictibilité' sur le modèle de *predictability*-prédictibilité (la traduction par 'postdictibilité' nous semble meilleure que la traduction par 'postdictabilité'). Nous utiliserons également l'adjectif dérivé 'postdictible'.

surprenante du personnage, ses intentions demeurent non-reconnues et, dans le pire des cas, le personnage risque d'apparaître comme étant non-rationnel (ce qui n'est pas souhaitable car à partir de ce moment tout devient possible de la part de ce personnage et il n'y a plus vraiment de surprise).

2.2.7 Lecture, relecture et attentes

L'expérience montre qu'un lecteur peut prendre du plaisir à relire un texte. Or la source principale d'intérêt narratif est l'inattendu. On pourrait penser qu'au moment de la relecture ce qui est présenté à l'audience est attendu et donc est inintéressant. Brewer (1996) fait un bilan de diverses théories qui peuvent expliquer ce phénomène.

Parmi ces théories, on trouve la 'suspension volontaire de mémoire' (*willing suspension of memory* ou *voluntary amnesia*). Il s'agit d'une forme particulière de la suspension volontaire d'incrédulité. Selon Gerrig (1989), nous aurions intégré cognitivement l'idée que les événements du 'monde réel' sont uniques. La relecture d'un roman ou le re-visionnage d'un film violerait cette idée. Ainsi l'audience ignorerait volontairement ce qu'elle connait du récit afin de maintenir "*an unfailing expectation of uniqueness*". Cela signifierait entre autres qu'une relecture provoquerait le même effet que la première lecture, ce qui est critiqué par Brewer (1996).

Une autre approche de la relecture concerne l'identification aux personnages. Dans la mesure où les états émotionnels de l'audience dépendent de l'état affectif des personnages auxquels elle s'identifie, alors une relecture produit le même effet que la lecture initiale car les états et émotions des personnages ne changent pas d'une lecture à l'autre. Lipsky (1956) avance l'idée, dans le cas du suspense, que le lecteur n'a plus de doute réel (direct) sur l'issue des événements, mais continue à partager à travers le processus d'identification le doute indirect (*vicarious doubt*) du personnage sur la suite des événements.

Enfin, d'autres théories font l'hypothèse que le problème viendrait des capacités mémorielles de l'audience. Il serait impossible pour quiconque de retenir tous les détails d'un récit et donc il resterait de la place pour l'incertitude au cours des diverses relectures ou divers re-visionnages d'un récit (Boulton, 1975; Beaugrande

and Colby, 1979; Brewer, 1996)

2.3 Crédibilité contre Intérêt, le dilemme de l'auteur

Une narration est une suite ordonnée d'événements qui forment un tout. Cette unité est assurée par le respect de la cohérence du monde (en particulier le respect de la causalité) ainsi que sa crédibilité. En particulier, les personnages-acteurs dans ce monde fictif doivent donner l'illusion de la rationalité, et donc leurs actions doivent être reconnues (éventuellement *a posteriori*) comme étant intentionnelles par l'audience.

Une narration est aussi constituée de conflits, auxquels les personnages sont confrontés et qu'ils tentent de résoudre. Ces conflits sont sources d'intérêt narratif. L'un des facteurs principaux de l'intérêt narratif est l'inattendu. Certains conflits sont sources d'évolution du monde fictif, ils poussent les personnages à agir pour résoudre ces conflits. Leur résolution doit intervenir avant la fin du récit, le critère de postdictibilité doit être respecté afin que le monde fictif retrouve une forme de stabilité et que l'audience ne soit pas dans une situation d'inconfort à la fin du récit.

Le respect de la vraisemblance du monde fictif, en particulier la cohérence de son évolution *via* des règles causales données et l'action de personnages crédibles qui le font évoluer, est nécessaire pour la compréhension de la narration par l'audience. La création de conflits narratifs, le maintien de l'intérêt créé par ces conflits et leur résolution donne une valeur narrative au récit.

Créer de l'intérêt en créant un conflit narratif peut parfois avoir un effet négatif (au moins localement) sur le maintien de la crédibilité du monde fictif. Prenons, l'exemple suivant :

Histoire de Mary et John : Mary et John sont très amoureux l'un de l'autre. Leur amour n'a cessé d'augmenter au fil du temps. Comme d'habitude, Mary déjeune avec John. Elle se lève, va dans la cuisine, saisit un couteau, revient à table et poignarde John dans le dos.

Dans cet exemple la dernière action de Mary est surprenante car elle ne s'intègre pas immédiatement dans notre représentation du monde fictif. L'audience pense, jusqu'au coup de couteau, que Mary aime John. Cette croyance est remise en cause lorsque Mary poignarde John. Soit Mary est folle, soit son action est bien intentionnelle mais l'audience n'est pas capable d'expliquer cet événement. Ainsi, à ce stade de l'histoire, la rationalité de Mary est mise en danger dans le but de créer de l'intérêt en présentant un événement surprenant.

Un auteur de fiction est sans cesse en équilibre entre, d'un côté la nécessité de respecter la vraisemblance du monde fictif et des personnages, et de l'autre la nécessité de produire des situations intéressantes. Nous appellerons cela le dilemme vraisemblance-intérêt :

Dilemme vraisemblance - intérêt : l'auteur est tiraillé entre deux objectifs parfois contradictoires que constituent le maintien de la vraisemblance du monde fictif (et des personnages qui évoluent dans ce monde et font évoluer son état) et la nécessité de produire des situations sources de tension narrative tout au long de la narration.

2.4 Pertinence narrative, non-pertinence et critère de pertinence totale

L'objet de cette thèse est la modélisation de la pertinence narrative en vue de la génération automatique de récits. Comme nous l'avons vu dans ce qui précède, les récits sont des objets complexes et la génération de récits doit respecter certaines règles qui ont été exposées plus haut. On prête à l'auteur russe Tchekhov le conseil d'ôter tout ce qui est inutile dans le récit, ce qui, dans une version positive, signifie qu'il ne faut conserver que ce qui est pertinent.

Qu'est-ce qu'un élément pertinent dans un récit ? Un élément pertinent est un élément qui est utile pour satisfaire un ou plusieurs critères qui donnent de la valeur au récit, soit en contribuant à maintenir la vraisemblance du monde dans lequel il se déroule (personnages inclus), soit en contribuant à produire de l'intérêt narratif (voir le tableau 2.1).

TABLE 2.1 – Quelques critères de pertinence narrative

Critère de pertinence	***Commentaire***
Unité du récit	Les événements qui composent un récit doivent être liés au sujet unique de la narration, à la résolution du conflit principal
Cohérence du monde fictif	La causalité doit être respectée, en particulier par les personnages dont l'action fait évoluer le monde.
Crédibilité du monde fictif (hors personnages)	Les règles causales doivent être manipulées avec précaution, l'occurrence d'événements rares peut affecter la crédibilité du monde. Cependant, un événement rare peut être la source du conflit principal.
Crédibilité des personnages	Les personnages doivent donner l'illusion de la rationalité, en particulier leurs actions doivent être reconnues comme étant intentionnelles *a posteriori*.
Respect d'un profil de tension	Le récit est organisé autour de l'apparition et la résolution d'un conflit principal. La phase de résolution du conflit est une phase pendant laquelle la tension est maintenue *via* l'introduction d'obstacles à la résolution du conflit principal. Le *climax* est le point de tension maximale, au moment duquel le conflit est en grande partie résolu.
Présence de séquences surprenantes	Des événements inattendus pour l'audience ou les personnages sont source de surprise. L'audience peut ressentir l'émotion ressentie par un personnage par procuration.
Présence de séquences de suspense	L'incertitude du développement événementiel est source de suspense. Par empathie envers un personnage, l'audience peut ressentir du suspense lorsque ce personnage fait face à la perspective d'un événement 'important' (que ce personnage en soit conscient ou non).
Postdictibilité	Outre l'événement qui est à l'origine de la formation du conflit narratif principal, toutes les situations du récit doivent pouvoir être expliquées rétrospectivement par l'audience.

Introduire un élément dans le récit a un coût, c'est un investissement réalisé dans le but d'obtenir un résultat. Tout élément qui ne serait pas utilisé pour donner de la valeur au récit serait une pure perte. Par ailleurs, l'audience intègre au fur et à mesure les éléments du récit qui lui sont présentés, qu'ils soient pertinents ou non. Un élément introduit n'est pas nécessairement immédiatement pertinent. Sa qualité d'élément pertinent peut apparaître ultérieurement.

Le contrat tacite entre l'auteur et son audience est un contrat de totale pertinence, c'est-à-dire que tout élément introduit (qui représente un coût cognitif payé par l'audience, au moins en termes de mémorisation) participera à l'obtention d'un gain ultérieurement, quelle que soit la nature de ce gain. Tout élément introduit dont l'utilité n'apparaîtrait pas à l'audience de manière rétrospective à la fin du récit est une faiblesse, voire même est une source d'interrogation et donc d'inconfort.

Dans la définition de la postdictibilité, Kintsch écrit : "*Weird and unpredictable statements in a text are interesting only in so far as they are well motivated within the text as a whole, at least by hindsight.*". Un événement du récit ne semble pas nécessairement devoir respecter ce critère, c'est l'événement qui est source du conflit narratif principal. L'apparition de ce conflit principal peut, par exemple, être consécutif à l'occurrence d'un événement assimilable à une loterie (un tremblement de terre, une crise cardiaque, une coïncidence, etc.). Ce genre d'événements est inattendu et le reste sans que la cohérence du monde soit négativement affectée. Le contrat entre l'auteur et l'audience n'est pas pour autant rompu, l'audience accepte cet événement car il est source de pertinence et surtout parce que cet événement est la raison pour laquelle l'histoire est racontée.

Toute autre manifestation trop évidente du hasard doit être, autant que possible, évitée. Le *deus ex machina* est un hasard providentiel, ce genre d'événements ne respecte pas le critère de postdictibilité, le *deus ex machina* ne s'explique pas. L'utilisation des *deus ex machina* n'est cependant pas interdite dans certains cas. Par exemple, l'utilisation d'un *deus ex machina* peut servir à la création d'un effet comique (la création d'un effet comique est un possible critère de pertinence narrative). Voici un exemple d'effet comique qui repose sur l'utilisation d'un *deus ex machina* dans cet extrait (du script) du film Monty Python and the Holy Grail (1975) :

[roar]

MAYNARD : *It's the legendary Black Beast of aaauuugh !*

ARTHUR : *Run away !*

ALL : *Run away ! Run away !*

[roar]

NARRATOR : *As the horrendous Black Beast lunged forward, escape for Arthur and his knights seemed hopeless. When, suddenly, the animator suffered a fatal heart attack. [ulk] The cartoon peril was no more. The Quest for the Holy Grail could continue.*

L'utilisation d'un *deus ex machina* peut aussi servir à respecter des règles liées au genre du récit, règles qui sont, dans ce cas, des critères de pertinence narrative (par exemple, l'utilisation d'un *deus ex machina* dans la tragédie grecque n'est pas nécessairement une faiblesse du récit).

D'une manière générale Lavandier (2004) met en garde contre l'utilisation des *deus ex machina*, les hasards providentiels. Tout d'abord parce que dans la vraie vie le hasard vient rarement régler nos problèmes. Dans le cadre fictif, le spectateur se laisse transporter dans le monde et les événements fictifs, il accepte de mettre son incrédulité au vestiaire. Or le *deus ex machina*, qui peut apparaître comme une volonté trop évidente (et souvent perçue comme une faiblesse) de l'auteur, tend à nous faire sortir du monde fictif (qui devient moins vraisemblable, dans lequel tout peut sembler permis et donc où plus rien n'est inattendu) et donc du récit.

D'une manière générale, le critère de postdictibilité doit donc être respecté pour l'ensemble des situations du récit à l'exception de l'événement perturbateur initial, celui qui est à la source du conflit principal autour duquel le récit est organisé, qui peut ne pas être postdictible. Ainsi, toute situation du récit doit rétrospectivement pouvoir être expliquée, ce qui, nous le verrons, revient à faire baisser rétrospectivement la complexité pour les générer. Parfois, cette baisse de complexité est très rapide, la situation devient brutalement postdictible, ce qu'illustre cet extrait du script du film Indiana Jones et la Dernière Croisade (Lucas and Meyjes, 1989) :

ELSA : *It would not be made out of gold.*

INDY picks up another cup—a simple earthenware jug.

INDY : *That's the cup of a carpenter.*

Le choix de la coupe la moins belle est surprenant (pour le spectateur) mais une explication (préparée par la phrase d'Elsa) est présentée immédiatement après et rend le choix postdictible. Nous appellerons ce phénomène une 'simplication', une baisse rapide de complexité qui est également source d'intérêt (une analyse plus complète de la scène de laquelle est extrait ce passage est réalisée dans le chapitre 8)

Nous pouvons maintenant définir le critère de pertinence narrative totale évoqué plus haut : tout élément qui ne contribue pas à la satisfaction d'un critère de pertinence narrative ne doit pas faire partie du récit. Ainsi, tout élément décrit ou toute anecdote doit définir des éléments de représentation du monde (y compris les règles causales et la représentation qu'ont les personnages du monde) qui seront utilisés, toute action doit servir à la réalisation de plans reconnus rétrospectivement comme étant intentionnels, tout événement doit participer à la résolution de conflits narratifs (afin, entre autre, de satisfaire le critère de postdictibilité) ou à l'apparition de nouveaux conflits narratifs (pour créer de l'intérêt narratif), l'ensemble des événements doit être postdictible sauf, éventuellement, l'événement perturbateur source du conflit principal autour duquel est construit le récit, etc.

L'objet de cette thèse est l'étude de critères de pertinence narrative. Dans un même espace nous allons définir des critères de pertinence liés à l'effet produit sur une audience (la tension créée) et des critères de pertinence liés au maintien de la vraisemblance du monde fictif (en particulier de la crédibilité du personnage *via* la reconnaissance de ses intentions). Nous montrerons alors comment la définition de ces critères peut servir de guide pour la génération de récits intéressants et vraisemblables dans lesquels chaque élément est pertinent. Auparavant, nous allons présenter diverses approches existantes de la génération de scénarios, ce qui nous permettra d'identifier les points forts et les faiblesses des approches actuelles en termes de pertinence des récits générés.

Chapitre 3

Différentes approches de la génération automatique des récits

Dans le chapitre précédent, nous avons mis en évidence deux angles possibles d'analyse des récits. Le premier d'entre eux consiste en une analyse que l'on peut qualifier de structurelle des éléments qui composent les récits. Le récit est considéré comme une suite d'événements cohérents, au niveau local (les événements se suivent en respectant la causalité dans le monde fictif), mais aussi au niveau global. Par ailleurs, les événements doivent être crédibles, en particulier les personnages doivent donner l'illusion de la rationalité, ce implique entre autres la possibilité pour une audience de reconnaître leurs intentions (éventuellement de manière rétrospective). Ces critères structurels sont des critères d'évaluation de ce que nous avons appelé la pertinence structurelle des éléments constitutifs du récit.

Le second niveau d'analyse porte sur les effets des récits sur une audience. La narration peut être décomposée en actes qui structurent la réponse émotionnelle de l'audience. Le récit est organisé autour d'un conflit narratif principal et de conflits intermédiaires qui sont sources d'intérêt et d'évolution du monde fictif, en particulier ces conflits poussent les personnage à agir. Les conflits introduits doivent tous être résolus à la fin du récit. Nous avons par ailleurs mis en évidence le rôle de l'inattendu dans la réponse émotionnelle à un récit. Ces critères émotionnels sont des éléments d'évaluation de ce que nous avons appelé pertinence émotionnelle des éléments constitutifs du récit.

Dans nos travaux, nous allons proposer une modélisation de critères de pertinence narrative en vue de la génération automatique de récits. Le domaine de la génération automatique de scénarios est une branche de l'intelligence artificielle dont le but est de développer, à partir d'un contexte initial donné, des algorithmes qui produisent des narrations. Le mot narration, dans ce contexte, a souvent été réduit aux aspects structurels définis plus haut. Ainsi, comme le souligne Pablo Gervàs (Gervàs, 2013) "*The underlying concept of "story" in SGAs (Story Generator Algorithms) is functional and does not imply any aesthetic notion. This is important because it sets the context for evaluation of generated stories, for which having a surface realization as a readable and appealing text is not necessarily a core issue*".

L'objectif principal de la génération de scénarios a, jusqu'ici, été la génération de *fabula* structurellement pertinentes, c'est-à-dire des suites d'événements cohérents (au niveau causal et au niveau global de l'intrigue) et crédibles. Pour y parvenir, diverses approches ont été explorées, qui peuvent être classées en trois catégories :

— les approches centrées sur la fonction de l'auteur
— les approches centrées sur la personnalité des personnages
— les approches hybrides.

Les aspects esthétiques de la narration, en particulier la génération d'histoires intéressantes, ont été souvent négligés. Il y a néanmoins eu des tentatives de contrôler la réponse émotionnelle. Nous verrons que plusieurs générateurs implémentent des méthodes *ad hoc* pour générer des récits intéressants.

Il ne s'agit pas dans cette partie de lister exhaustivement les générateurs de scénarios existants mais bien de donner un aperçu des travaux réalisés et d'identifier leurs limites. Nous pourrons alors nous situer dans ce paysage de la narration automatique et montrer en quoi notre approche diffère des précédentes.

La notion d'intérêt narratif sera notre point d'entrée pour la modélisation cognitive de la pertinence narrative en vue de l'analyse et la génération automatique de récits. Ce modèle cognitif et parcimonieux permettra de définir dans un même espace d'étude à la fois des critères de pertinence émotionnelle mais aussi de pertinence structurelle. Nous montrerons que cet effort de modélisation peut, en particulier, servir de guide pour la conception de modèles de générateurs automatiques de scénarios.

3.1 Approches dans le processus de génération

La comparaison entre les différents systèmes de génération de scénarios existants (ou ayant existé) ne peut pas vraiment se faire sur la base du contenu des narrations produites, chaque système utilisant un cadre qui lui est propre pour générer des narrations. Cependant, on peut comparer les différents systèmes sur la base des choix qui ont présidé à leur développement, la philosophie de leur conception. Bailey (1999) a proposé une classification des systèmes de génération de scénarios en trois catégories :

— Les modèles qui approchent le processus de narration sous l'angle d'un monde fictif dans lequel vivent des personnages et qui considèrent la narration comme la présentation acceptable des actions et des interactions des personnages.

— Les modèles qui approchent le processus de génération en prenant la perspective d'un auteur humain.

— Les modèles qui approchent le processus de génération sous l'angle de représentations abstraites des histoires (par exemple les approches basées sur des 'grammaires d'histoires' (Rumelhart, 1975)).

Cette séparation en trois catégories a été reprise par Mateas & Senghers (1999). Cependant, la distinction entre modèles centrés sur la fonction de l'auteur et l'approche sous forme de grammaires ou de scripts n'est pas très pertinente. Les structures de la grammaire peuvent être considérées comme des intentions particulières de l'auteur. Certaines grammaires narratives peuvent être utilisées comme modèles de compréhension des narrations mais également comme modèles pour guider le processus de génération. Ces modèles peuvent donc être considérés comme des modèles qui abordent le processus de génération sous l'angle de la modélisation d'intentions d'un auteur virtuel. Ainsi nous pouvons distinguer les deux types d'approches définies par Riedl (2004) :

1. Les approches qui abordent le processus de génération sous l'angle d'un monde fictif dans lequel évoluent des personnages. Ces systèmes modélisent les facteurs mentaux tels que les croyances, les désirs, les buts ainsi que les traits de personnalités qui affectent la manière dont les personnages agissent dans le monde virtuel et simulent les processus cognitifs qui amènent ces

personnages à agir (et réagir) dans l'environnement dans lequel ils sont situés. Les histoires émergent à partir des décisions prises par ces personnages.

2. Les approches qui abordent le processus de génération du point de vue de l'auteur dans lesquels l'auteur (simulé) a des intentions (explicites ou non) quant aux buts que doit atteindre l'histoire ou la structure de l'intrigue.

Cette distinction entre l'approche du point de vue de l'auteur et l'approche du point de vue des personnages amène des résultats différents au niveau des narrations produites. Comme le remarque Mark Riedl (2004), on observe en général que :

1. les approches centrées sur les personnages produisent des narrations dans lesquelles les personnages sont très crédibles mais dont la cohérence est moindre dans l'ensemble.
2. les approches centrées sur la fonction de l'auteur produisent des narrations très cohérentes mais dans lesquelles les personnages sont généralement moins crédibles.

Nous allons détailler chacune des approches définies ci-dessus et les illustrer en présentant quelques systèmes existants.

3.1.1 Les systèmes de génération basés sur les personnages

Dans ces systèmes, les personnages ont généralement des buts, des désirs et des traits de personnalité. La génération de narrations est le produit des actions des personnages dans le 'monde virtuel', c'est-à-dire des buts qu'ils décident d'atteindre, les décisions prises pour atteindre ces buts et leur comportement effectif dans le monde virtuel (Aylett, 1999). Cela donne généralement une impression de rationalité des personnages-acteurs du point de vue de l'audience. La crédibilité des personnages dans les narrations produites par ces systèmes est une conséquence logique de ce genre d'approche.

Cependant, comme le souligne Aylett (1999), le résultat de ce processus crée un paradoxe narratif (Aylett, 2000) : laisser la liberté à des personnages d'agir indépendamment des autres personnages rend incertain la génération d'une intrigue cohérente.

Ce problème peut être, au moins en partie, résolu par l'utilisation de contraintes sur l'intrigue afin de contraindre les personnages à suivre certaines voies ou à se comporter d'une certaine manière. Si cela peut servir à renforcer la cohérence de l'ensemble, cela peut aussi réduire la crédibilité des personnages (par exemple si on force les personnages à adopter subitement de nouveaux objectifs dans le but d'orienter leur comportement).

Dans les systèmes dont le résultat est le produit des actions des personnages, les personnages sont souvent traités comme des agents autonomes. Ce genre d'approche est donc préféré dans les systèmes interactifs, c'est-à-dire dans lequel un utilisateur extérieur humain peut influencer le cours des événements qui se déroulent dans le monde fictif.

Tale-Spin

Tale-Spin (Meehan, 1977) est un système qui produit des histoires de type 'fables d'Ésope'. Les narrations sont le produit des actions des personnages dans le monde fictif. Les personnages ont une connaissance partielle du monde fictif et ont des attitudes l'un envers l'autre. Avant de lancer l'algorithme, un utilisateur du système Tale-Spin décrit un ou plusieurs personnages et donne au moins un but à atteindre à l'un de ces personnages. Ensuite, le système Tale-Spin calcule des plans (des suites d'actions que les personnages réalisent) pour atteindre les objectifs donnés aux personnages.

Le modèle de raisonnement utilisé est basé sur le modèle des structures représentant la connaissance humaine de Schank & Abelson (1977) . Les structures (appelées *planboxes*) sont constituées de préconditions, d'une action et de postconditions. Pour chaque personnage, Tale-Spin utilise à la fois la planification et l'inférence pour déterminer les prochaines actions qu'il doit réaliser. L'utilisation de scripts ou de plans partiels est permise. À chaque étape de la planification pour un personnage donné, toutes les options possibles sont testées par l'algorithme jusqu'à que l'un des buts de ce personnage soit atteint ou qu'il échoue. Après la réalisation d'un morceau de plan ou un script, le mécanisme d'inférence met à jour le monde et l'attitude des personnages. Cette mise à jour peut éventuellement conduire à la génération de nouveaux buts pour les personnages.

Dans les narrations produites, les personnages donnent l'illusion d'être rationnels. Leurs actions respectent les règles causales du monde fictif et sont motivées par des buts en accord avec leurs désirs et leur attitude envers les autres personnages.

Cependant, la cohérence de la narration n'est pas assurée. Tout d'abord, le système permet aux personnages de générer de nouveaux buts, qui ne sont pas nécessairement liés aux buts initiaux (l'unité n'est pas assurée). Ensuite, le programme peut échouer car les personnages n'ont pas l'assurance de pouvoir calculer un plan qui leur permet d'atteindre leurs buts (la résolution des conflits n'est donc pas assurée). Voici par exemple un plan généré par Tale-Spin dans lequel le renard ne réussit pas à atteindre son but qui était de récupérer le fromage, ce qui devient impossible une fois que le corbeau l'a mangé :

> *Once upon a time there was a dishonest fox and a vain crow. One day the crow was sitting in his tree, holding a piece of cheese in his mouth. He noticed that he was holding the piece of cheese. He became hungry, and swallowed the cheese. The fox walked over to the crow. The end.* (Meehan, 1977)

Enfin, un reproche fréquemment adressé à Tale-Spin est qu'il ne cherche pas à produire des narrations qui produisent un effet sur les lecteurs. Ainsi, la plupart des récits produits sont peu intéressants.

Fabulist

Fabulist (Riedl and Young, 2004) est système de génération et de présentation de narrations. Dans ce système, la génération utilise l'algorithme IPOCL (*Intent-driven Partial Order Causal Link*). Cet algorithme prend en compte à la fois des éléments de causalité mais également les intentions (ou motivations) des personnages afin de produire des séquences d'événements cohérentes et d'assurer une forme de crédibilité des personnages.

Le système génère tout d'abord un plan qui permet d'atteindre des objectifs donnés, puis s'assure *a posteriori* que toutes les actions et les buts des personnages sont bien justifiés par les événements qui se produisent au cours de la narration.

3.1.2 Les systèmes centrés sur la fonction de l'auteur

Ces systèmes produisent généralement des narrations très cohérentes au niveau de l'intrigue. Dans ce genre de système, un 'auteur virtuel' fixe des buts auxquels la narration doit mener ou des éléments de structure. Les buts peuvent être donnés aux personnages ou consister en un état final du monde désiré par l'auteur (et déterminé indépendamment des buts éventuels des personnages). Dans ce genre d'approche, les actions réalisées par les personnages et les choix qu'ils font sont donc contraints par les objectifs fixés par l'auteur. La cohérence de l'intrigue est une conséquence générale de cette approche.

Cependant, dans les narrations produites, les buts des personnages peuvent entrer en contradiction avec les buts de l'auteur. L'auteur virtuel peut être amené à forcer un personnage à prendre une décision ou à donner un but nouveau à un ou plusieurs personnages pour satisfaire ses intentions (celles de l'auteur). Ces contraintes imposées aux personnages peuvent affecter leur crédibilité et donner l'impression qu'ils agissent de manière irrationnelle. Par ailleurs, ce genre d'approche contrainte par les désirs de l'auteur à tendance à limiter le nombre possible de narrations produites.

Author

Dans sa présentation du système Author (Dehn, 1981), Dehn s'oppose directement au système Tale-Spin, en lui reprochant de ne pas refléter le processus réel de création des narrations. Le modèle Author est organisé autour de quatre notions qui modélisent un auteur virtuel : ses intentions, sa capacité à reformuler les concepts, sa capacité 'à se souvenir' et son caractère opportuniste pour l'amélioration des narrations.

Les intentions de l'auteur consistent à créer une 'bonne histoire'. Aussi, il doit pouvoir intervenir sur le cours des événements et donc sur l'action des personnages. L'histoire qui en résulte est un agencement complexe des buts de l'auteur qui forment une sorte d'échafaudage pour l'histoire et ne doivent plus être visibles une fois la narration générée.

La reformulation des concepts est la capacité d'un auteur à reformuler ses objectifs pendant le processus de génération. Ainsi, l'objectif de rendre une histoire

plausible peut être reformulé en la nécessité de justifier la solution trouvée par un protagoniste de l'histoire pour résoudre un problème.

La capacité de l'auteur 'à se souvenir' est sa capacité, non seulement à connaître les éléments qui composent le monde fictif, mais aussi à garder en mémoire un ensemble d'épisodes ou de personnages qui peuvent être réutilisés. Ceci fournit à la fois du matériel pertinent à partir duquel l'auteur peut construire de nouveaux récits ainsi qu'une trace des objectifs qu'il poursuit (il peut donc, entre autres, les abandonner temporairement avant d'y revenir).

Enfin, l'auteur est capable de faire preuve de créativité. En particulier, il est capable d'avoir un comportement opportuniste et de modifier ses objectifs s'il identifie un meilleur objectif à suivre au cours du processus de génération.

Universe

Le système Universe (Lebowitz, 1985) utilise une approche de type 'planification' pour générer des histoires dont la fin est ouverte (comme dans les épisodes d'une série à la fin desquels tous les conflits ne sont pas résolus). Le système de planification assemble des fragments d'intrigues pour atteindre les buts de l'auteur - c'est-à-dire les objectifs que l'auteur veut voir atteints pendant l'épisode - ainsi que les objectifs secondaires qui peuvent être introduits au cours de la planification. Les fragments d'intrigues décrivent des situations qui se produisent dans le monde fictif à plusieurs niveaux d'abstraction. Ainsi, un fragment peut être décomposé en sous-objectifs jusqu'à ce qu'un niveau primitif soit atteint. Ces fragments sont contraints, des conditions doivent être remplies pour qu'ils puissent faire partie du plan global.

Ce processus de génération assure une cohérence de l'intrigue. Cependant, le système de planification ne cherche pas à évaluer si il est pertinent pour tel ou tel personnage de prendre part à tel ou tel morceau de l'intrigue. La crédibilité des personnages n'est donc pas assurée, parce que leur comportement n'est motivé par rien d'autre que les buts de l'auteur.

Tailor

Le système Taylor (Smith and Witten, 1991) est un système centré sur l'auteur et dont l'objectif est de planifier les actions d'un personnage principal. Le personnage principal se voit attribuer un objectif à atteindre, le système Taylor recherche alors une séquence d'actions que le personnage peut réaliser pour atteindre ce but. Le système autorise l'introduction d'un antagoniste pour créer un conflit. L'objectif de cet antagoniste est d'empêcher le personnage principal d'atteindre son but.

Ce système produit des narrations qui sont cohérentes sur le plan causal. Ceci est assuré par le système de planification. Cependant, le système ne contrôle pas la réussite du personnage principal. Ainsi, le personnage peut échouer à atteindre son objectif et donc l'intrigue dans son ensemble n'est pas cohérente. Par ailleurs, dans ce système la volonté de l'auteur est exprimée sous la forme d'un objectif donné au personnage principal. Ceci assure au personnage une forme de crédibilité. Cependant, l'antagoniste n'a pas d'objectif autre que d'empêcher le personnage principal d'atteindre son but et la crédibilité de l'antagoniste n'est pas assurée.

Mexica

Mexica (Pérez y Pérez and Sharples, 2001) est un système de génération de scénarios qui modélise un processus d'écriture créative (Sharples, 1999). Ce processus considère l'écriture comme un cycle d'engagement cognitif et de réflexion. Le processus d'engagement consiste à sonder notre mémoire à long-terme. Le processus de réflexion consiste à modifier et organiser les idées existantes.

Durant la phase d'engagement, Mexica cherche, dans une base d'histoires existantes, des éléments qui correspondent aux motifs désirés d'émotion et de tension de l'histoire qui est en cours de génération. Ces éléments correspondent à des actions qui sont choisies dans le but de compléter des séquences parce qu'elles avaient précédemment été utilisées de manière satisfaisante dans les histoires existantes. La phase d'engagement continue jusqu'à ce que le système se trouve face à une impasse, c'est-à-dire qu'il ne trouve plus d'élément dans la base d'histoires pour compléter celle qui est en cours de génération. La phase de réflexion commence alors.

Cette phase de réflexion consiste à lier les éléments de la phase d'engagement entre eux en utilisant un système de planification de type *partial-order planning*. Ce système de planification complète la séquence d'événements venant de la phase d'engagement par des événements intermédiaires qui permettent aux éléments trouvés dans la phase d'engagement de se produire. S'ensuit une nouvelle phase d'engagement.

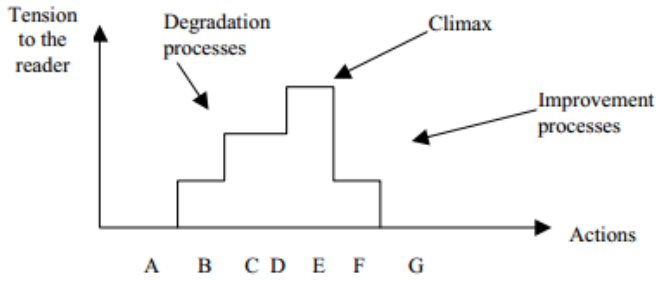
La phase de réflexion assure la cohérence de la narration produite par le système Mexica. En revanche, rien n'assure que l'intrigue soit cohérente ni que les personnages soient crédibles. La cohérence de l'intrigue peut éventuellement venir de la cohérence de l'intrigue d'une histoire de la base à partir de laquelle elle aurait été créée. Mais cela n'est pas garanti par la phase d'engagement. En ce qui concerne la crédibilité des personnages, Mexica contient un mécanisme d'ajout d'explications qui intervient une fois les phases d'engagement et de réflexion terminées et qui vise à rendre le comportement des personnages compréhensible.

Dans la phase d'engagement, Mexica cherche des événements qui doivent coïncider avec un motif de tension. Au cours du processus de génération, le système s'attache à respecter un profil de tension dramatique lorsqu'il ajoute un nouvel élément à l'histoire. Les séquences d'événements peuvent mener à des états du monde dégradés ou non (Bremond, 1966). Autrement dit certains événements tendent à renforcer les conflits existants et d'autres à les résoudre. Les phases de renforcement de conflit font monter la tension narrative tandis que les phases de résolution la font baisser. Dans le système Mexica, une histoire est considérée comme intéressante lorsqu'elle inclut un processus de dégradation (renforcement des conflits) suivi d'une amélioration (résolution des conflits). La représentation de la tension au cours de la narration fait partie de la connaissance nécessaire pour créer des séquences intéressantes.

Ainsi le système considère que la tension augmente lorsqu'un personnage est tué, lorsqu'un personnage est en danger ou encore quand les émotions de deux personnages entrent en conflit. L'évolution de la tension peut être la conséquence d'une action (par exemple si un personnage est attaqué alors sa vie est mise en danger) ou alors peut-être inférée (par exemple si deux personnages se retrouvent à aimer une même troisième personne alors il y a une compétition amoureuse). Les tensions associées à chaque "conflit" prennent des valeurs qui peuvent être

modifiées par l'utilisateur du système. Mexica mémorise l'évolution de la tension au cours de la génération et peut donc tracer des profils de tension (voir fig. 3.1 issue de (Pérez y Pérez and Sharples, 2001)).

FIGURE 3.1 – Représentation d'une histoire en termes de tension pour le lecteur pour le système Mexica (Pérez y Pérez and Sharples, 2001)



Façade

Façade (Mateas and Stern, 2003) est un système de génération de narrations basé sur un système de planification réactif. Ce système de planification sélectionne et ordonne des éléments d'intrigue (de faible 'granularité') appelé *beats*. Un *beat* est un petit morceau d'histoire (par exemple un dialogue ou une action) qui fait avancer l'histoire. Couramment, les *beats* sont de type 'action / réaction', c'est à dire qu'un personnage fait une action et un autre réagit à cette action. Cela permet une coordination entre les actions des personnages.

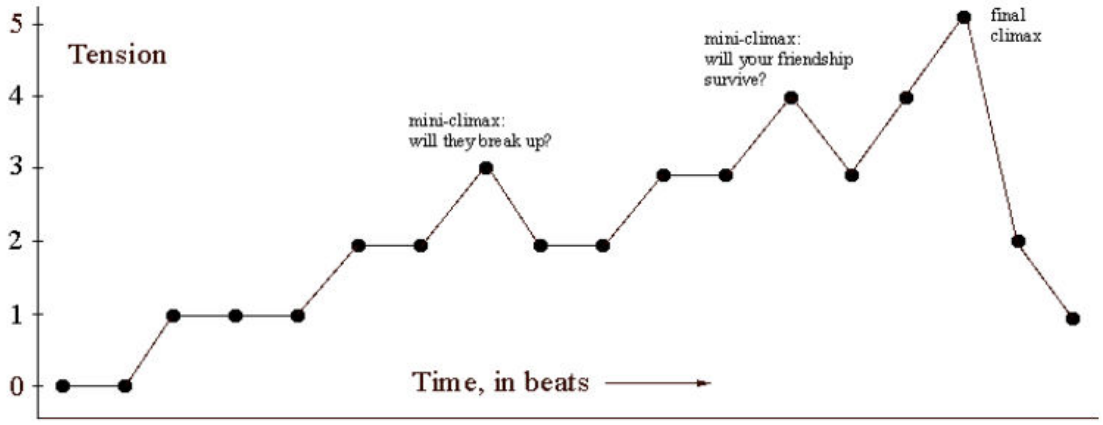
Dans ce système, un personnage ne gère pas la progression d'un autre. Les personnages tentent d'atteindre des objectifs liés à l'intrigue sans se soucier des besoins ou désirs des autres personnages.

Façade est un système interactif, un utilisateur humain peut intervenir au cours du processus de génération.

L'un des aspects importants du système Façade est le *beat manager*. L'objectif de ce *beat manager* est d'identifier les *beats* qui peuvent être appliqués en tenant compte de l'état courant du monde et ajoute celui qui correspond le mieux à la réalisation d'un arc dramatique donné. Un arc dramatique est une fonction qui représente la tension (ou l'intérêt) qu'une audience doit ressentir en fonction du

temps (mesuré en nombre de *beats* réalisés) (voir fig. 3.2 , reprise de Mateas & Stern (2003), inspirée de l'arc narratif aristotélicien). Ainsi, l'un des objectifs de Façade est de produire un certain effet sur le lecteur.

FIGURE 3.2 – Tension narrative idéale pour le système Façade (Mateas and Stern, 2003)



La manière dont les comportements sont scriptés dans les *beats* assurent une certaine crédibilité locale des personnages dans les narrations produites par ce système. Ce n'est pas nécessairement le cas au niveau global, les actions des personnages n'étant pas motivées par la réalisation d'objectifs individuels mais par des objectifs de l'auteur, liés à l'intrigue. Par ailleurs, la manière dont les *beats* sont ordonnés dans le but de suivre un arc dramatique assure une certaine forme de cohérence de l'intrigue. Cependant, l'éventail des comportements décrits dans les *beats* doit être limité si l'on ne veut pas que l'histoire dévie de son objectif. La recherche d'un effet sur le lecteur pourrait en effet amener le système à choisir des *beats* qui éloigneraient la narration du but qu'elle doit atteindre.

3.1.3 Les systèmes hybrides

Minstrel

Minstrel (Turner, 1993) est une approche hybride basée sur la notion de schémas. Ces schémas sont de deux types : les schémas de l'auteur, utilisés par Minstrel pour satisfaire des contraintes rhétoriques, et les schémas des personnages, employés par le système pour représenter les événements de l'histoire.

Minstrel possède un ensemble fini de buts de l'auteur qui correspondent à des contraintes de différents types : thématique de la narration, création d'effets dramatiques, maintien de la cohérence ou encore présentation de l'histoire. Ces objectifs de l'auteur sont associés à des ALP (*Author Level Plans*) qui indiquent explicitement comment créer des scènes de revanche, de déception ou encore comment créer une introduction et un dénouement dans le but d'avoir une histoire cohérente, etc.

Au niveau des personnages, le système utilise des objectifs (par exemple "satisfaire la faim d'un personnage", "aller dans un endroit donné", etc.) ainsi que des représentations des personnages, des monstres, des animaux, des objets, des croyances, des émotions, etc. pour construire des scènes.

Pour écrire des histoires Minstrel utilise 6 schémas prédéfinis, les PATs (*Planning Advice Themes*). La structure des histoires dans Minstrel est la suivante : Scènes d'introduction - Scènes liées à un thème (composées d'ensembles préconditions - décision - conséquences) - Scènes de dénouement.

Le processus de génération consiste tout d'abord à sélectionner un objectif au niveau de l'auteur, puis à le diviser en sous-objectifs. Ensuite, une phase de résolution de problèmes au niveau des personnages tente de résoudre ces sous-objectifs.

Ce processus hybride qui tente de satisfaire à la fois des buts au niveau de l'auteur mais également au niveau du personnage permet la génération d'histoires cohérentes, en particulier au niveau de l'intrigue. Cependant, ce système est contraint et le faible nombre de schémas possibles limite le nombre de possibilités pour produire des histoires.

Dans Minstrel, certains ALP sont dédiés à la génération de scènes qui provoquent un certain effet chez le lecteur. Par exemple, l'un des ALP est dédié à la

création de phases de suspense. Le choix des scènes dans lesquelles le programme décide de créer le suspense est déterminé par 1) l'importance des objectifs (des personnages) auxquelles ces scènes sont liées 2) l'importance de la scène dans l'histoire. Plus les objectifs des personnages sont importants et plus la scène est importante dans l'histoire, plus elle se prête à l'introduction d'une phase de suspense. Une fois la scène sélectionnée, deux techniques sont utilisées pour créer du suspense : la première est de montrer que le personnage éprouve des émotions fortes ("*Bebe was very scared*" dans l'exemple ci-dessous), la seconde consiste à ajouter une phase d'évitement qui échoue (*failed escape*) ("*Bebe tried to run away but failed*" dans l'exemple ci-dessous) :

Later, Bebe believed that he would die because he saw a dragon moving towards him and believed it would eat him. Bebe was very scared. Bebe tried to run away but failed ! ...

Virtual storyteller

Le système Virtual Storyteller (Theune et al., 2003) est une approche hybride. L'intrigue est créée par les actions des personnages. Cependant, les personnages posent des questions à un directeur virtuel qui guide leur action. Les personnages et le directeur virtuel sont capables de raisonner sur leur propre connaissance.

Les personnages sont capables de calculer des plans dans le but d'atteindre leurs propres objectifs en utilisant leur connaissance sur le monde (composée de faits et d'actions possibles). Le directeur possède des connaissances sur la structure des narrations. Il sait que les personnages et l'environnement doivent être introduits au début, que les personnages agissent pour atteindre leurs buts dans le milieu, et enfin que la fin doit être heureuse. Le directeur virtuel est capable de déterminer si les actions des personnages correspondent à cette structure.

Le directeur virtuel peut également empêcher certains personnages de réaliser certaines actions, en particulier pour éviter que l'histoire ne se termine trop rapidement ou encore pour éviter qu'un personnage principal ne soit tué. Ainsi, le directeur virtuel peut introduire de nouveaux personnages ou objets dans le monde, donner un but à un personnage ou encore interdire la réalisation d'une action prévue par un personnage.

La communication entre les personnages et le directeur virtuel se fait ainsi : les personnages demandent au directeur s'il peuvent se comporter d'une certaine manière dans le monde fictif et le directeur donne ou non son accord selon la conformité de ce comportement avec les règles qu'il connaît sur le déroulement des narrations.

Le système Virtual Storyteller est proche du système Tale-Spin. La différence majeure est la présence du directeur virtuel qui contrôle la structure de la narration.

3.2 Les générateurs de tension narrative

Il existe des systèmes qui sont entièrement dédiés à la création de tension narrative. La plupart de ces systèmes génère des récits à partir d'une séquence événementielle donnée en entrée du générateur.

Le modèle du conflit narratif

Le conflit narratif est généralement désigné comme étant source d'intérêt. Cependant, la notion de conflit narratif a rarement été explorée. Ware & Young (2010) proposent un modèle (computationel) du conflit ayant pour but la génération d'histoires intéressantes. Leur modèle du conflit est basé sur la notion de liens (causaux) conflictuels qui apparaissent lorsque deux plans valides dans l'espace de recherche considéré ne peuvent pas être exécutés tous les deux. L'exécution de l'un des plans empêche donc l'exécution de l'autre (et *vice versa*). Ainsi, dans le cadre de la génération d'un scénario, cela revient à déterminer un ensemble de liens conflictuels et à forcer les personnages de la narration à utiliser les opérateurs qui composent ces liens dans la réalisation de leurs intentions.

Un aspect intéressant du modèle de Ware et Young est la mise en évidence de deux notions que sont l'intensité du conflit et la résolution du conflit. Ces deux notions reposent une unique notion d'utilité, définie comme étant la satisfaction d'un personnage après la réalisation d'une étape (elle prend une valeur entre 0 et 1). L'intensité d'un conflit du point de vue d'un personnage A connaissant la dernière action s_B d'un personnage B est définie par $1 - utilite(A, s_B)$. Ceci

traduit l'idée suivante : plus l'action de B fait baisser la satisfaction de A, plus l'intensité du conflit augmente. La notion de résolution est quant à elle définie comme la variation d'utilité entre la première et la dernière étape réalisée par un personnage. Cette notion est donc une mesure de l'évolution de la satisfaction d'un personnage.

Ainsi, si l'on dispose de mesures d'utilité, ce modèle permet une évaluation du conflit, et donc de l'intérêt associé à ce conflit au cours du temps.

Suspenser

Suspenser (Cheong and Young, 2015) est un modèle de génération du suspense qui, contrairement à Minstrel où le suspense est généré au niveau de la *fabula*, génère le suspense au niveau du *sjuzet*. La *fabula* est l'un des éléments fourni en entrée de Suspenser. Ainsi, ce modèle peut être utilisé en complément de générateurs de *fabula* existants (dans Cheong & Young (2015) il s'agit d'IPOCL (Riedl and Young, 2004)).

Dans un premier temps, un générateur de squelette détermine quels éléments de la *fabula* sont importants. Le calcul du niveau d'importance des événements est basé sur la méthode de Trabasso et Sperry (Trabasso and Sperry, 1985) qui consiste à évaluer l'importance d'un événement en calculant le nombre de relations causales avec d'autres événements.

Une fois cette phase de sélection d'un squelette terminée, une phase de création du suspense est lancée. Dans cette phase, un organisateur de structure ajoute des étapes au squelette précédemment calculé. Ces étapes sont de deux types : celles qui menacent la réalisation d'un objectif d'un personnage, et celles qui favorisent la réalisation de l'objectif. Les étapes ajoutées sont donc principalement des étapes qui menacent la réalisation d'un objectif, la rendant donc plus incertaine. Pour déterminer si une action est menaçante ou non, le système mesure le suspense potentiel d'une action, qui correspond à la somme du suspense potentiel de ses conséquences. L'évaluation du suspense des conséquences est évalué par une heuristique prenant en compte l'inverse des distances causales de tous les liens dits menaçants (qui amènent donc à une non-réalisation de l'objectif considéré) et de tous les liens non-menaçants (qui amènent donc à la réalisation de l'objectif).

Ensuite, une évaluation du *sjuzet* évalue les incertitudes comme l'inverse du nombre de plans permettant la réalisation de l'objectif dans un espace de recherche donné.

Dramatis

Dramatis (O'Neill and Riedl, 2014) est un modèle computationel qui analyse des récits pour déterminer le niveau de suspense au cours de l'histoire et dont l'objectif est d'être utilisé dans des générateurs de scénarios existants. Dans un espace de recherche donné, la définition du suspense de O'Neill & Riedl est :

> *Given the belief that the protagonist can face a negative outcome, one can assume that outcome will occur and search for the single most likely plan in which the protagonist avoids that outcome.* (O'Neill and Riedl, 2014)

Ainsi, le modèle Dramatis est basé sur la recherche de l'échappatoire la plus vraisemblable. Dramatis est un modèle qui fonctionne en deux temps.

Premièrement, le modèle cherche à prédire les issues négatives auxquelles peuvent faire face les protagonistes des histoires analysées. Il se base pour cela sur des structures de type scripts qui représentent des structures typiques que l'on trouve dans les narrations (par exemple, quand il y a un espion dans un bar, une des issues négatives qui se produit typiquement est "l'espion est empoisonné"). Les scripts les plus pertinents sont choisis parmi ceux qui utilisent des événements qui se sont produits récemment dans l'histoire. Il peut également décider de piocher des informations provenant des ennemis du protagoniste (par exemple d'un discours prononcé par un ennemi du protagoniste).

La seconde phase du modèle Dramatis consiste à trouver et à évaluer les chances d'occurence d'une échappatoire qui permettrait au protagoniste d'éviter l'issue négative. Pour ce faire, le script précédemment trouvé est transformé en problème de planification et un algorithme trouve une solution presque optimale à ce problème de planification. L'évaluation des plans (ou échappatoires) calculés est la somme des coûts de chaque opérateur présent dans le plan. Pour calculer le coût des opérateurs, Dramatis utilise un modèle de la mémoire qui, au cours de la lecture de l'histoire analysée, mémorise des éléments de l'histoire (tels que

les lieux ou les personnages) et les événements (représentés comme des opérateurs STRIPS (Schank and Abelson, 1977)). Ce modèle de la mémoire sert à quantifier l'accessibilité des éléments de l'histoire dans la mémoire. Le coût d'un opérateur est donc une somme des coûts d'accessibilité des éléments qui le composent. Dramatis considère ensuite que les opérateurs dont le coût est le moins élevé sont ceux qui ont le plus de chance de se produire.

Prevoyant

Prevoyant (Bae and Young, 2014) est modèle de création de surprise qui utilise les *flash-backs* et les 'présages' (*foreshadowing*).

Les événements qui provoqueront la surprise sont sélectionnés parmi les événements importants du récit donné en entrée du générateur sous la forme d'un plan. Le calcul du niveau d'importance des événements est, là encore, basé sur la méthode de Trabasso et Sperry (1985) qui consiste à évaluer l'importance d'un événement en calculant le nombre de relations causales avec d'autres événements du récit.

Pour chaque événement important, le système détecte les événements qui sont liés (causalement) à cet événement et sélectionne ceux qui sont 'séparables', c'est-à-dire dont l'omission dans la chaîne d'événements qui composent le récit n'affecte pas le reste du récit (ces événements ne sont utiles que dans le cadre de la génération de l'événement important). Cet ensemble d'événements 'séparables' servira à expliquer *a posteriori* l'occurrence de l'événement surprenant, ils constituent le *flash-back*.

Le système évalue l'intensité de l'inattendu de l'événement surprenant en quantifiant l'impact de l'omission des éléments composant l'ensemble d'événements séparables sur la compréhension du récit par le lecteur (Prevoyant utilise un modèle du lecteur).

Le système utilise des 'présages' pour renforcer la postdictibilité de l'événement surprenant par le lecteur. Le système fait des allusions aux événements qui seront omis (par exemple en introduisant des préconditions de ces événements) et qui permettront, dans une certaine mesure, au lecteur d'anticiper ces événements omis, sans pour autant permettre au lecteur d'anticiper l'occurrence de l'événement surprenant. *A posteriori*, ces éléments de 'présage' renforceront la postdictibilité de

l'événement en permettant au lecteur de mieux intégrer les éléments qui composent le *flash-back*.

IDtension

IDtension (Szilas, 2007; Szilas and Richle, 2013) est un système de génération de récits dont le but est de générer des événements pertinents au regard d'un modèle de la tension dramatique. Un récit produit par IDtension a une structure qui contient de la 'circularité' (cette notion est au cœur de la création de tension dans ce système). Cette notion de 'circularité' renvoie à un paradoxe décrit ainsi :

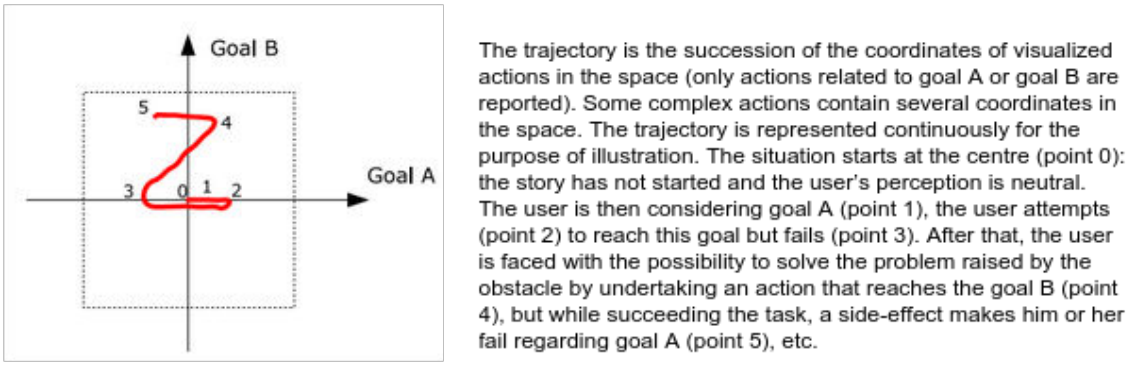
> "*The narrative paradox is the juxtaposition of two contradictory terms, that successive narrative actions attempt to resolve. Trying to "resolve the unresolvable" along the major part of a story inevitably creates tension. In some case, the story can ends by operating a major shift in the paradox, where the paradox's term are re-interpreted with a new point of view. In this case, the paradox is not solved but rather dissolved : it does not exist anymore. In other cases, the paradox remains until the end.*" (Szilas and Richle, 2013, p.258-259)

Et qui est illustré par cet exemple :

> "*John desires that princess Mary accepts him as a future spouse (goal A). For that purpose, he proposes to her (task 1), but she cannot accept, because he is not a prince (obstacle). So he decides to become a fake prince (goal B, triggered as a sub-goal), which he can perform via different tasks, grouped into a task module. Performing these tasks of appearing like a prince makes him a liar, which in turn makes Mary not want to marry him, via the side-effect (Adapted from the animated movie Aladdin, Disney, 1992).*" (Szilas and Richle, 2013, p.264)

IDtension génère donc des histoires en introduisant des événements qui tantôt vont favoriser la réalisation de l'un des objectifs du paradoxe, tantôt vont faire obstacle à la réalisation de cet objectif. L'histoire peut être représentée sous la forme d'une trajectoire dans un espace appelé *paradox space* qui illustre la 'circularité' des histoires (voir figure 3.3).

FIGURE 3.3 – *The paradox space* et une trajectoire possible dans cet espace (Szilas and Richle, 2013)



Pour favoriser ou faire obstacle à la réalisation d'un objectif, le système utilise des règles qui définissent le comportement des personnages. Les personnages peuvent, entre autres, encourager/décourager d'autres personnages de réaliser certaines tâches, peuvent donner des conseils aux autres personnages ou encore accepter/refuser d'aider d'autres personnages.

3.3 Évaluation des modèles de génération de narrations

Afin de pouvoir comparer, et éventuellement évaluer, les modèles de génération de scénarios, il nous faut définir une liste de critères de comparaison.

Un premier critère de comparaison concerne la philosophie qui a présidé au développement du modèle. Le modèle repose-t-il sur la création d'histoires comme produit des actions d'agents (dont le comportement a été modélisé) dans le monde fictif ou alors sur la génération de récits comme produit des intentions d'un auteur simulé ? Comme cela a été dit plus haut, chacune de ces approches produit des

narrations qui sont généralement de qualité variable sur le plan de la cohérence ou sur le plan de la crédibilité.

La cohérence et la crédibilité des narrations produites sont deux critères de comparaison (voire d'évaluation) possibles. Plus généralement, on peut définir un ensemble de critères d'évaluation liés à la pertinence structurelle des éléments des récits produits. Voici quelques critères de comparaison possibles :

— Est-ce que les narrations produites sont cohérentes ?
— Est-ce que les narrations produites sont crédibles ?
— Les personnages sont-il crédibles ? Donnent-il l'illusion d'être rationnels ?

De même on peut définir des critères d'évaluation liés à la à la pertinence émotionnelle des récits produits. Voici quelques critères de comparaison liés à la pertinence émotionnelle :

— Les narrations produites sont-elles intéressantes ?
— Les narrations produites contiennent-elles des séquences surprenantes ?
— Les narrations produites contiennent-elles des phases de suspense ?
— Le critère de postdictibilité est-il respecté dans les narrations produites ?
— Les narrations produites respectent-elles une montée graduelle de la tension narrative ? (respect de l'arc Aristotélicien (fig. 2.2) ou de la courbe de Freytag (fig. 2.3))

D'autres critères d'évaluations peuvent être pris en compte, par exemple la créativité du système (Pérez y Pérez and Sharples, 2004), l'originalité des narrations produites, ou encore la *replay value* (une évaluation de la diversité des narrations que le système peut produire).

La plupart des critères présentés ci-dessus peuvent difficilement donner lieu à une évaluation objective des narrations produites. L'objectif est de mettre en évidence les forces et les faiblesses des différentes approches ou de donner un aperçu de la philosophie adoptée par les créateurs des systèmes pour aborder la génération automatique de scénarios. Le tableau 3.1 donne un aperçu de l'évaluation de deux systèmes présentés plus haut, Mexica et Minstrel.

TABLE 3.1 – Comparaison des systèmes Mexica et Minstrel

Critère de comparaison	Mexica	Minstrel
Approche	*author based*	hybride : *author & character based*
Cohérence de l'intrigue	Non assurée, peut venir de la cohérence d'une histoire de la base donnée au système	Oui globalement, *via* la satisfaction d'objectifs de l'auteur ALP (*Author Level Plans*)
Cohérence causale	Narration cohérente grâce à une phase de réflexion	Oui, *via* une approche de planification et l'utilisation de scripts
Crédibilité des personnages	Non assurée, cependant un mécanisme d'ajout d'explications vise à rendre le comportement des personnages compréhensible	Crédibilité assurée *via* des schémas des personnages (qui définissent des objectifs, des croyances, des émotions, etc. des personnages)
Séquences surprenantes	Non-assuré, le système cherche à coïncider avec un profil de tension	Pas de volonté explicite de créer de la surprise
Séquences de suspense	Non-assuré, le système cherche à coïncider avec un profil de tension	Certains ALP sont dédiés à la création de phases de suspense
Narrations intéressantes	Oui, et respectent un profil donné de tension	Oui, des ALP sont dédiés à l'amélioration de la qualité littéraire des narrations produites
Montée graduelle de la tension	Oui, respect d'un arc narratif similaire à celui de Freytag (fig. 2.3)	Non
Postdictibilité	Respect du critère de postdictibilité non assuré, son respect peut être une conséquence de la coïncidence avec un profil donné de tension	Oui, *via* une approche semi-scriptée de la génération. Les récits produits ont une structure de type : Scènes d'introduction - Scènes liées à un thème - Scènes de dénouement

3.4 Approche cognitive de la génération de scénarios

Dans nos travaux, nous abordons la génération de récits sous l'angle des effets qu'elle produit sur une audience. Dans le chapitre 4 nous allons introduire un modèle cognitif de l'intérêt que nous utiliserons pour donner une définition cognitive de la notion de conflit narratif (voir chapitre 5) ainsi que de modalités de l'intérêt narratif (la surprise, la curiosité et le suspense, voir chapitre 6). Nous montrerons en particulier que les conflits narratifs sont source d'intérêt et peuvent être ramenés à des conflits cognitifs internes aux personnages du récit.

Notre approche de la génération de récits sera donc principalement basée sur la modélisation de personnages et de leurs actions dans le monde fictif. Dans le but de produire une tension narrative, nous envisageons le processus de génération de récits sous l'angle d'un processus de formation de conflits internes aux personnages, qui sont la traduction de conflits narratifs à l'échelle de la narration, et de la résolution de ces conflits internes *via* la réalisation d'actions perçues comme étant intentionnelles par l'audience (un modèle cognitif de la reconnaissance des intentions est présenté dans le chapitre 7). Ainsi, notre objectif premier est de produire un effet sur l'audience tout en maintenant la crédibilité des personnages (ce qui fait écho au dilemme intérêt-vraisemblance introduit p. 42).

Nous proposerons des éléments d'implémentation des modèles introduits précédemment dans le chapitre 9 dans le but de produire des séquences narratives intéressantes (configurées par la surprise), cohérentes et dans lesquelles les personnages donnent l'illusion d'agir de manière rationnelle (*via* la reconnaissance de leurs intentions par l'audience).

Ainsi, notre étude de la pertinence narrative peut être décomposée en trois temps :

— Nous proposerons des définitions cognitivement plausibles de critères de la pertinence narrative (tout d'abord des critères de pertinence émotionnelle (chapitres 4, 5 et 6) puis des critères de pertinence structurelle (chapitre 7)).
— Nous montrerons que les définitions des critères de pertinence narrative peuvent servir à analyser et évaluer un scénario fictif (chapitre 8)

— Nous montrerons comment les modèles cognitifs de critères de pertinence narrative peuvent servir de guide en vue de la création d'un générateur de séquences narratives pertinentes (partie 9)

Chapitre 4

Introduction au rôle de la complexité cognitive dans l'intérêt

L'objet principal de nos travaux est la modélisation de l'intérêt dans la fiction, nous abordons l'étude des récits sous l'angle de la pertinence émotionnelle, c'est-à-dire que nous nous intéressons aux éléments qui ont pour but de produire un effet sur l'audience, de susciter de l'intérêt dans le récit.

Le présent chapitre a pour but d'introduire un modèle cognitif de l'intérêt événementiel basé sur la notion de complexité cognitive. Après avoir présenté diverses conditions de la plausibilité cognitive, nous introduirons la notion de complexité cognitive, inspirée de la notion mathématique de complexité de Kolmogorov. Cette notion de complexité nous permettra de présenter la théorie de la Simplicité (Dessalles, 2006) et, en particulier, de donner une définition de l'inattendu événementiel comme saut de complexité cognitive.

Le choix de ce modèle n'est pas fortuit. Pour certains type de jugements, perçus comme étant des biais cognitifs, la théorie des probabilités échoue à rendre compte des phénomènes observés. Notre hypothèse est que ces jugements reposent sur des calculs effectifs de complexité, et que l'évaluation des probabilités est une conséquence de ces calculs. Au travers de l'exemple particulier d'un jugement portant sur la reconnaissance par un individu d'un cas de plagiat, nous montrerons

que les calculs de complexité permettent d'expliquer les jugements et de réaliser des prédictions (validées empiriquement).

Nous mettrons en évidence, au travers d'exemples issus de récits fictifs, l'utilisation de variations de complexité par des auteurs de fictions afin d'intéresser une audience.

Note préliminaire : L'analyse des biais cognitifs (partie 4.4) et les modèles liés à l'évaluation des cas de plagiat (partie 4.5) ainsi que l'expérience associée (annexe A.1) ont été publiés dans :
Saillenfest, A. & Dessalles, J.-L. (2015). Some probability judgments may rely on complexity assessments. Proceedings of the 37th Annual Conference of the Cognitive Science Society, 2069-2074. Austin, TX : Cognitive Science Society.

4.1 Contraintes de la modélisation cognitive

4.1.1 Le modèle scientifique

La notion de modèle est polysémique. Ici, nous nous intéressons au modèle scientifique, que l'on peut définir comme une représentation simplifiée d'un ensemble d'observations. L'un des modèles les plus élémentaires est la classification. Ce type de modèle possède le défaut de n'être souvent pas prédictif. Des modèles plus aboutis sont composés de lois (quantitatives ou non) qui relient des variables qui prennent leurs valeurs dans des ensembles définis.

Un modèle scientifique doit être cohérent, ainsi il ne doit pas s'auto-contredire. Par ailleurs, le modèle doit être cohérent avec des observations réalisables dans un espace défini. En effet, un modèle est un modèle de quelque-chose, autrement dit il doit rendre compte d'observations réalisées à l'aide de méthodes scientifiques liées aux aspects du monde modélisés. Ces observations doivent pouvoir être reproduites (ou ré-observées).

Le modèle doit également être explicatif. Cela signifie qu'il doit donner un sens, il doit apporter un gain de connaissance scientifique. Un modèle ne doit pas seulement décrire les données mais aussi définir une forme de langage destiné à en rendre compte.

Un bon modèle doit être prédictif, autrement dit il doit permettre de prédire le comportement de ce qu'il modélise et ce comportement doit pouvoir être testé et observé. Un bon modèle permet de couvrir un large spectre d'observations et doit permettre de réaliser un nombre important de prédictions.

Un bon modèle doit être réfutable, ce qui signifie qu'il permet d'établir des prédictions réfutables. Les prédictions réfutables sont des prédictions que l'on peut tester, et dont la validité remettrait en cause le modèle (le modèle, en l'état, serait donc réfuté, et pourrait être éventuellement modifié pour rendre compte de cette nouvelle observation). Un bon modèle ne doit pas rendre compte de tout, sans quoi il serait totalement inutile.

Le modèle doit être utile. La création d'un modèle repose sur une nécessité, le modèle est construit dans un but donné. Dans notre cas, il devra pouvoir être utilisé pour définir diverses notions classiques issues de l'étude des récits, être utilisé à des fins d'analyse des narrations et enfin guider le processus de réalisation d'un générateur automatique de scénarios.

Enfin, un bon modèle doit être parcimonieux. En particulier, le meilleur modèle réalise la meilleure compression des observations. Le principe de parcimonie renvoie au rasoir d'Occam, une heuristique fondamentale dans le domaine scientifique. Le rasoir d'Occam énonce que "les hypothèses suffisantes les plus simples sont les plus vraisemblables". La parcimonie d'un modèle est donc jugée au nombre d'hypothèses qu'il fait, à la qualité de la compression des données ainsi qu'à ses qualités 'esthétiques' (par exemple sa simplicité).

4.1.2 La plausibilité cognitive

La plausibilité cognitive d'un modèle est sa compatibilité avec le fonctionnement supposé du cerveau humain. C'est une notion floue qui rend difficile, sinon impossible, une classification des modèles sur la base d'un critère de 'cognitivité'. Cependant, certaines hypothèses, comme l'accès à un oracle ou à l'extension de mondes possibles, sont incompatibles avec la contrainte de plausibilité cognitive.

La plupart des modèles que nous allons développer sont inspirés de travaux réalisés dans des domaines variés (études littéraires, narratologie, intelligence artificielle ou encore psychologie expérimentale). Les prédictions de nos modèles seront

comparées aux observations ou aux prédictions de travaux existants dans ces divers domaines d'études.

Par ailleurs, nous avons nous-même testé empiriquement divers aspects des modèles mathématiques que nous présenterons (nous avons regroupé les expériences, leur *design*, les résultats et les discussions associées en annexe A ; nous ferons le lien, lorsque nous présenterons un modèle, avec l'expérience qui correspond au test de certaines prédictions de ce modèle).

Cela ne garantit pas que ces modèles rendent compte du fonctionnement réel du cerveau mais que, dans un certain domaine, rendent compte d'observations ou de résultats et donc, dans ce domaine, sont une représentation plausible du fonctionnement cognitif de l'être humain.

4.2 Introduction à la complexité au sens de Kolmogorov

La notion de complexité, au sens de Kolmogorov, trouve ses racines dans la calculabilité, l'apprentissage par induction et la notion mathématique de hasard (Li and Vitányi, 2008).

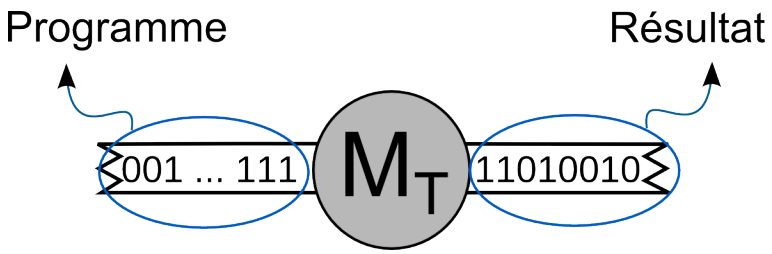
La complexité de Kolmogorov d'un objet peut être définie ainsi :

La complexité d'un objet est la taille du plus petit programme qui permet de le produire

Une machine de Turing universelle (qui est un modèle idéal de calculateur mécanique) reçoit des instructions (un programme) et produit un objet en fonction des instructions reçues (voir fig. 4.1). En théorie de la calculabilité, un programme peut être codé sous forme d'une chaîne de caractères binaires, c'est à dire une suite de 0 et de 1. La taille du programme correspond alors au nombre de 0 et de 1 qui composent ladite suite (on parle de nombre de *bits* d'information). On peut, de la même façon, calculer la taille (mesurée en *bits*) des objets (ou résultats) produits (voir fig. 4.1).

Si on considère un objet donné, il existe un ensemble (potentiellement infini) de programmes qui produisent cet objet. Parmi ces programmes (de longueur finie), il

FIGURE 4.1 – Machine de Turing



en existe au moins un dont la longueur est inférieure ou égale à celle de n'importe quel autre programme de cet ensemble. La longueur (ou taille) minimale d'un programme dans l'ensemble des programmes qui produisent un objet donné est la complexité de Kolmogorov de l'objet.

Formellement : soit X un objet observé en sortie d'une machine de Turing M_U. On note $l(p)$ la taille d'un programme p, $M_U(p)$ la réalisation du programme p au travers de la machine M_U et $K(X)$ la complexité de Kolmogorov de l'objet X.

$$K(X) = min_p(l(p)|M_U(p) = X) \tag{4.1}$$

Illustrons ce qui précède en analysant les objets suivants :

(X_1)0011001100110011001100110011001100110011001100110011

(X_2)0010010000111111011010101000100010000101101000110000

L'objet X_1 est particulièrement simple à décrire, il s'agit de 13 fois la chaîne 0011. Cette régularité peut être exploitée pour proposer un programme *court* qui permet de générer X_1 (par exemple le programme "écrire 13 fois 0101"). Ainsi, la complexité de l'objet X_1 est 'faible'.

X_2 ne présente pas *a priori* de régularité visible. Pour générer X_2, on n'identifie pas *a priori* de meilleur programme que celui qui consiste à épeler les *bits* qui composent X_2 un à un. Ce programme est de longueur au moins égale à la taille de l'objet X_2, c'est-à-dire 52 bits. Ainsi, on peut en conclure que la taille du programme minimal pour générer une chaîne binaire donnée est au plus égale à la taille de la chaîne (à une constante additive près). Les objets qu'aucun programme ne peut résumer sont dits incompressibles.

D'après ce qui précède on peut écrire que :

$$0 < K(X) < \text{ longueur de } X + O(1) \tag{4.2}$$

Dès lors que l'on parle de calcul de programme minimal, un problème se pose concernant le choix du langage utilisé pour représenter les données. On peut choisir d'utiliser un langage de programmation dit universel. Les langages de programmation classiques (python, Prolog ou encore LISP) sont des langages universels. Par définition, les langages universels sont tels que si un objet peut être reconstruit à partir d'une description dans n'importe quel langage, alors cet objet peut être reconstruit à partir d'une description dans un langage universel.

On peut remarquer que l'objet X_2 est la représentation binaire des premières décimales de π. Si la machine (choisir une machine et un langage est équivalent) a une liste de programme et que celui de π est l'un des premiers dans la liste, alors π est très simple pour cette machine. L'un des résultats fondamentaux relatif à la complexité de Kolmogorov est le théorème d'invariance qui dit que la taille de la description minimale d'un objet est invariante (à une constante près) entre les différents langages universels (Li and Vitányi, 2008, p.104-105).

On peut donc parler de la complexité de Kolmogorov d'un objet indépendamment du langage utilisé pour le décrire.

Pour résumer, la complexité de Kolmogorov d'un objet est la taille du plus court programme qui produit cet objet. On parle aussi de plus petite description de l'objet ou encore de son meilleur résumé. On peut aussi parler de compression maximale de l'objet (ce qui correspond à la compression maximale de l'information que contient l'objet en question).

Complexité et probabilité algorithmique

Dans la nature, les structures que l'on observe ne sont pas le fruit du hasard, elles sont le fruit de processus génératifs soumis à des règles (en particulier causales). Ces processus génératifs peuvent être considérés comme des programmes génératifs dans le monde (tel qu'on l'imagine). On peut définir la probabilité d'occurrence d'une situation comme la somme des probabilités cumulées de tous les programmes susceptibles de l'engendrer. Si on suppose que chaque bit d'un programme qui génère une situation donnée est fixée aléatoirement, alors la probabilité

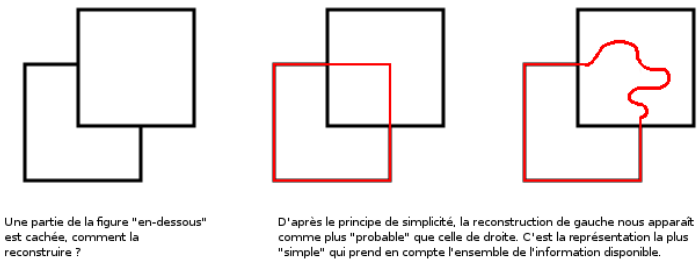
d'occurrence de s via ce programme est égale à $2^{-\text{longueur du programme}}$. Cette fonction décroit rapidement avec la taille des programme génératif, ainsi on peut considérer que la probabilité *a priori* d'une situation s et de l'ordre de celle de son programme génératif le plus court (c'est une version du rasoir d'Occam appliqué au monde tel qu'on se le représente) (Li and Vitányi, 2008, p.259-260).

Si on note $K(s)$ la taille de ce programme le plus court (par analogie avec la complexité de Kolmogorov), on peut donc approximer la probabilité d'occurrence $p(s)$ d'une situation s *via* la formule de la probabilité algorithmique suivante :

$$p(s) = 2^{-K(s)} \tag{4.3}$$

La complexité joue un rôle important en cognition, en particulier au niveau des processus perceptifs (Chater, 1999; Chater and Vitányi, 2003; Chater and Brown, 2008). L'exemple de la partie cachée (Dessalles, 2008b, p.28) ci-dessous illustre l'un des sens que l'on peut donner à la formule 4.3, c'est-à-dire notre propension à considérer les descriptions les plus simples comme les plus probables ($p(s)$ augmente à mesure que $K(s)$ diminue)

FIGURE 4.2 – L'exemple de la partie cachée



Dans ce qui précède, les notions de génération d'un objet et de description d'un objet ont été confondues. Un objet simple à observer est le produit d'un processus génératif simple. Or on peut trouver facilement des exemples pour lesquels l'équation définissant la probabilité algorithmique (équation 4.3) réalise de mauvaises prédictions.

Supposons, par exemple, qu'à la loterie nationale sorte la combinaison 1-2-3-4-5-6. Contrairement à la prédiction de la formule 4.3, la plupart des individus

estiment que ce tirage est moins probable qu'un tirage ne présentant pas de forte régularité permettant de le décrire simplement. Autrement dit, la simplicité de ce tirage le fait apparaître comme étant moins probable.

Bien que ce tirage présente une régularité très forte, nous savons que le 'programme génératif' qui a permis d'engendrer cette situation n'est pas plus court que celui qui aurait pu permettre de générer n'importe quelle autre combinaison. Nous connaissons les règles causales du monde et nous savons que chaque chiffre a été tiré au hasard parmi un ensemble de chiffres possibles. Si on considère que ce tirage est le résultat de six tirages successifs indépendants, la complexité du tirage total est (à peu près) égale à six fois la complexité du tirage d'un numéro unique.

La simplicité pour décrire cette situation est fortuite, elle n'a rien à voir avec le mécanisme qui a permis de l'engendrer. Cette simplicité pour décrire le tirage vient de l'observation d'une régularité mathématique, chaque chiffre du tirage (sauf le premier) est égal au chiffre précédent incrémenté de 1, indépendante du mécanisme de tirage des chiffres.

La régularité observée de ce tirage permet de décrire plus simplement 1-2-3-4-5-6 que la plupart des combinaisons possibles. Décrire une combinaison qui ne présente pas de régularité revient à décrire chacun des chiffres qui la compose dans le pire des cas. Si la combinaison est ordonnée dans l'ordre croissant, on peut décrire le premier chiffre puis chacun des suivants comme étant la somme du précédent et d'un entier. Décrire un chiffre consiste, dans la plupart des cas, à décrire chaque *bit* de sa représentation binaire (de taille $log_2(n+1)$, la complexité C_n pour décrire un chiffre n est donc généralement de l'ordre de $C_n = log_2(n+1)$. La complexité de description C_{+i} de l'opération $+i$ (incrémentation de i) peut-être assimilée à la description de i, et donc $C_{+i} = log_2(i+1)$. Ainsi, la pour décrire le tirage 8-15-26-38-41-43, qui ne présente pas de régularité, consiste à décrire le chiffre 8, puis les incrémentations +7,+11,+12,+3,+2, ce qui fait une complexité totale $C_8 + C_{+7} + C_{+11} + C_{+12} + C_{+3} + C_{+2} \approx 16$ *bits*.

Dans le cas du tirage 1-2-3-4-5-6, seul le premier chiffre doit être décrit (cela coûte 1 *bit*), la description du mécanisme qui permet d'engendrer les autres chiffres est de taille constante, il faut $log_2(6)$ *bits* pour décrire le nombre d'utilisation de l'opérateur d'incrémentation et C_{+1} pour décrire l'opérateur. Nous assimilons la complexité pour décrire l'incrémentation à la complexité pour décrire la valeur de

cette incrémentation, ainsi $C_{+1} = 1$. Il faut donc un peu plus de 4 *bits* pour décrire le tirage 1-2-3-4-5-6.

Du point de vue de la génération, chaque tirage est de complexité équivalente. Mais certains sont plus simples à décrire que d'autres. Cette simplicité, liée à la détection d'une régularité indépendante des mécanismes de génération, rend l'occurrence d'un tirage simple plus inattendu pour un observateur. La plupart des individus estime d'ailleurs que l'occurrence d'un tirage simple à décrire, par exemple le tirage 1-2-3-4-5-6, est moins probable que celle d'un tirage qui ne présente pas de régularité (alors que les deux tirages ont une probabilité objective d'occurrence égale, ils sont également complexes à produire). Cela va à l'encontre des prédictions de l'équation 4.3 qui prédit que les structures simples à observer sont le résultat de processus génératifs simples.

L'exemple précédent illustre la nécessité, lorsque l'on traite de situations 'réelles', de ne pas confondre le processus de génération et le processus d'observation, à ne pas assimiler situation simple à décrire et processus de génération simple.

4.3 Rôle de la complexité dans l'intérêt événementiel

Comme nous l'avons dit précédemment, le problème de la théorie des probabilités algorithmiques est qu'elle mélange les notions d'observation et de génération de structures. Une structure simplement observable serait le produit d'un processus génératif simple. Or l'exemple des résultats de loterie présentant une forte régularité vient contredire cette idée : les loteries les plus simples à observer sont celles dont l'occurrence est perçue comme étant la moins probable (Dessalles, 2006). Autrement dit, les résultats de loterie les plus simples sont plus inattendus. La théorie des probabilités (y compris des probabilités algorithmiques) ne permet pas d'expliquer ce biais cognitif face à une simplicité jugée anormale.

Nous allons donc introduire la théorie de la Simplicité (Dessalles, 2006), qui définit l'inattendu d'une situation (ou d'un événement) comme étant le résultat d'un saut de complexité (cognitive, notion dérivée de la notion de complexité au sens de Kolmogorov).

4.3.1 Introduction à la théorie de la Simplicité

La théorie de la Simplicité définit l'inattendu U (pour *unexpectedness*) d'une situation en fonction de sa complexité de génération C_w et de sa complexité de description C_d :

$$U = C_w - C_d \tag{4.4}$$

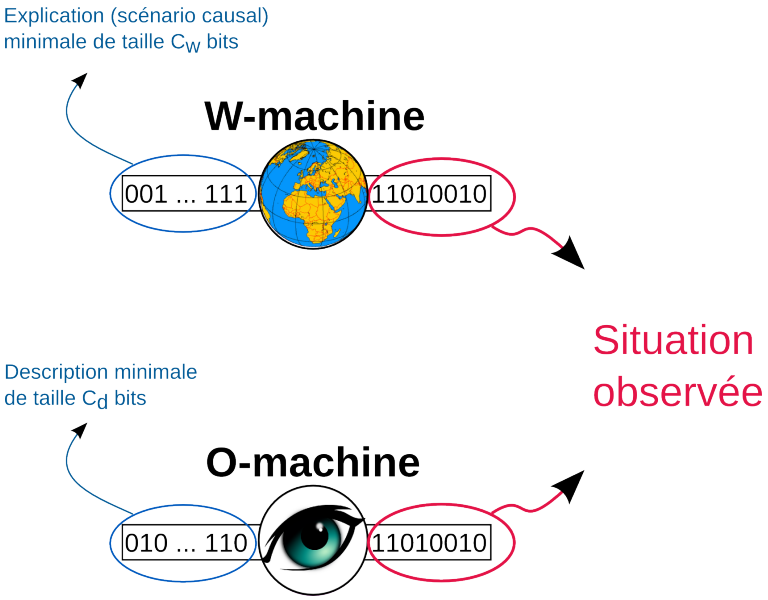
La complexité de génération C_w est contrainte par les connaissances qu'a un humain sur le monde qui l'entoure, les faits qu'ils connaît et les règles causales qui le régissent. La complexité de génération correspond à la complexité de la situation, considérée comme résultat de l'exécution d'un programme 'causal' à travers une machine-monde (*W-machine*, voir également la figure 4.3), c'est-à-dire à la taille de la plus courte hypothèse (causale) permettant de générer la situation.

La complexité de description C_d quant à elle est uniquement contrainte par les capacités cognitives de l'humain, il peut utiliser toutes les ressources à sa disposition pour calculer cette complexité. La complexité de description d'une situation correspond à la complexité de la situation, considérée comme étant le résultat de l'exécution d'un programme 'descriptif' à travers une machine-observation (*O-machine*, voir également la figure 4.3), c'est-à-dire à la taille de la plus courte description de la situation.

La machine-monde rend compte du fonctionnement du monde tel qu'un individu se le représente. Cela signifie qu'un programme dans la machine monde est une suite d'instructions de type 'événements' dont la réalisation fait évoluer l'état courant considéré du monde. La complexité de génération d'une situation générée *via* la machine-monde correspond à la taille minimale de la suite d'instructions à fournir au monde pour générer ladite situation (cette complexité est mesurée en *bits*).

La machine-observation rend compte des mécanismes de compression opérés par notre cerveau pour décrire les situations observées. Elle représente la connaissance d'un individu et ses capacités de calcul. Un programme exécuté par la machine-observation est une suite d'instructions visant à produire une description non ambiguë d'une situation. La complexité de description d'une situation observée en sortie de la machine-observation correspond à la taille minimale de la

FIGURE 4.3 – O-machine et W-machine



suite d'instructions à fournir à cette machine pour décrire ladite situation (cette complexité est également mesurée en *bits*).

Un événement inattendu est intéressant. Mais comme le note Kintsch (1980) les deux facteurs principaux de l'intérêt sont la réponse émotionnelle à de grands thèmes (mort, politique, etc.) et l'inattendu des situations. La théorie de la simplicité (Dessalles, 2006) définit l'intérêt d'une situation comme étant l'intensité de la réponse émotionnelle (notée E) à une situation. Cette intensité émotionnelle est la somme d'une réponse émotionnelle standard (notée E_h) et de l'inattendu de la situation U. E_h peut être vue comme un coût (évalué en *bits*), de type assurance, qu'un humain serait 'prêt à payer' pour qu'une situation qu'il désire se produise ou qu'une situation qu'il ne désire pas ne se produise pas. Formellement, pour une situation s, l'intérêt événementiel E (pour *Emotional response*) lié à l'observation de l'occurrence de s est défini ainsi :

$$E_h(s) = E(s) - U(s) \tag{4.5}$$

4.3.2 Calculabilité de la complexité cognitive

Un résultat classique concernant la complexité de Kolmogorov est sa non-calculabilité. La non-calculabilité de la complexité de Kolmogorov est souvent utilisée pour affirmer que l'être humain n'a pas accès à la complexité des objets auxquels il est confronté. Cette conclusion hâtive a poussé certains chercheurs à abandonner la notion de complexité au profit de l'inférence statistique (Griffiths and Tenenbaum, 2003) ou se sont tournés vers des mesures alternatives et calculables de complexité telles que la reconnaissance de motifs ou la complexité booléenne (Simon, 1972; Feldman, 2004).

Cependant, nous pouvons conserver la notion de complexité en considérant que l'humain a des ressources limitées. L'être humain possède une puissance de calcul qui lui permet, entre autres, d'identifier la répétition de motifs, ce qui lui permet ainsi de réaliser des opérations de compression face aux situations rencontrées. D'une manière plus générale, toute forme de reconnaissance de structures permet de réaliser une compression. Dès lors que nous disposons d'un modèle cognitif qui représente ces mécanismes, nous sommes en mesure de calculer la taille de la meilleure compression que peut réaliser un individu en un temps donné. On peut ainsi conserver la notion de complexité, en lui donnant ce sens de taille de la meilleure compression réalisée en un temps fini et en utilisant un modèle cognitif donné. En lui donnant ce sens, la complexité devient une notion calculable.

Les machines *W-machine* et *O-machine* ne sont pas des machines de Turing universelles, elles ont des ressources limitées (car le cerveau humain a des ressources limitées). Les complexités de génération et de description sont donc des notions calculables.

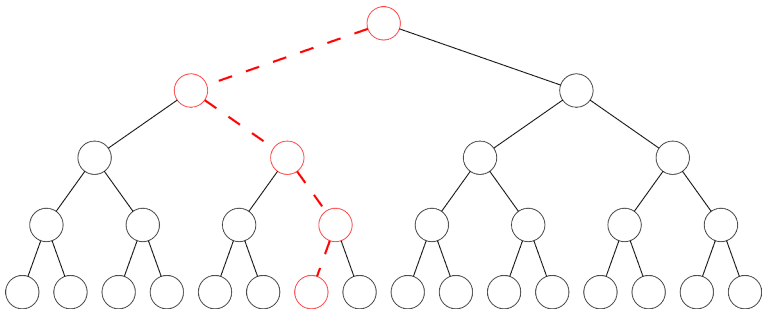
4.3.3 Complexité et génération des objets

Le calcul de la complexité de génération d'un objet (ou d'une situation) est subjectif. Il consiste à évaluer la taille minimale (exprimée en *bits* de génération) de la suite d'instructions à fournir à une certaine représentation interne à chaque individu pour produire ladite situation. C'est une mesure la plus petite chaîne causale (ou hypothèse causale) ayant permis l'occurrence de la situation.

Prenons, pour illustrer ceci, un modèle de génération inspiré du *pachinko* (un

jeu populaire au Japon) (voir figure 4.4). Dans ce jeu, une bille est lâchée à la racine de l'arbre binaire et, à chaque embranchement, tombe à gauche ou à droite jusqu'à atteindre une feuille de l'arbre. Soit la situation "la bille a atteint une feuille donnée de l'arbre". Pour calculer la complexité de génération de cette situation, il nous faut déterminer le nombre minimal de *bits* à fixer pour générer la situation. Dans notre exemple, à chaque embranchement, il nous faut dire si la bille a été à gauche ou à droite, ce qui revient à fixer un *bit* de génération à chaque embranchement. Comme la bille doit passer par 4 embranchements pour atteindre une feuille de l'arbre (qui en compte 16 soit 2^4), la complexité de génération de la situation $s =$"la bille a atteint une feuille donnée de l'arbre" vaut $C_w(s) = 4bits = log_2(16)bits$.

FIGURE 4.4 – Génération d'une situation *via* un arbre binaire



On peut étendre ce résultat à la génération d'une situation s, résultat d'une loterie susceptible de produire, de manière équivalente, n situations possibles :

$$C_w(s) = log_2(n) \tag{4.6}$$

Si on note $p = \frac{1}{n}$ la probabilité d'atteindre une feuille donnée, l'équation 4.6 peut être réécrite ainsi :

$$C_w(s) = -log_2(p) \tag{4.7}$$

Les cas pour lesquels on dispose de la probabilité d'occurrence d'une situation sont rares, et la formule 4.7 ne peut donc pas être utilisée.

Certaines situations s sont la conjonction de deux situations $s1$ et $s2$ ($s = s_1 * s_2$ en utilisant l'opérateur de composition $*$). La complexité de génération de la conjonction de deux situations s_1 et s_2 vaut :

$$C_w(s_1 * s_2) \leq C_w(s_1) + C_w(s_2||s_1) \tag{4.8}$$

où $C_w(s_2||s_1)$ est la complexité pour générer s_2 en supposant que l'état s_1 est (déjà) généré (dans le monde). La complexité de génération de $s_1 * s_2$ consiste tout d'abord à calculer la taille du scénario causal le plus court pour générer la situation s_1 puis, une fois celle-ci générée, à calculer la taille du scénario causal le plus court pour générer s_2. L'opérateur de composition n'est pas commutatif, $s_1 * s_2 \neq s_2 * s_1$, il définit un ordre de génération pour les situations qui composent la conjonction. L'égalité, dans l'équation 4.8, est atteinte si s_1 a été générée antérieurement à s_2.

Dans le cas où s_1 et s_2 sont le résultat de générations indépendantes alors $C_w(s_2||s_1) = C_w(s_2)$. L'indépendance de deux situations s_1 et s_2 est donc définie ainsi :

$$C_w(s_1 * s_2) = C_w(s_1) + C_w(s_2) \tag{4.9}$$

La formule 4.8 peut être généralisée à la composition de n événements $s = s_1 * s_2 * ... * s_n$ (tels que s_i est antérieur à s_{i+1} pour $i = 1, 2, ..., n-1$), ce qui donne :

$$C_w(s) = C_w(s_1) + C_w(s_2||s_1) + C_w(s_3||s_1 \& s_2) + ... + C_w(s_n||s_1 \& s_2 \& ... \& s_{n-1}) \tag{4.10}$$

Si chacune des n situations est indépendante des $n-1$ autres situations, l'équation 4.10 devient :

$$C_w(s) = C_w(s_1) + C_w(s_2) + C_w(s_3) + ... + C_w(s_n) \tag{4.11}$$

Supposons maintenant que l'on dispose d'une hypothèse causale H qui génère une situation s. $s = H * s$, ce qui signifie que s est le résultat du processus génératif H puis de la génération de s elle-même. Or H générant s, la complexité de génération de s une fois H exécuté est nulle $C_w(s||H) = 0$. Ainsi :

$$C_w(s) = C_w(H * s) = C_w(H) \tag{4.12}$$

Et donc si H est l'hypothèse causale la plus courte, c'est-à-dire la meilleure description des paramètres qui conduisent à la réalisation de H, on a $C_w(H) = l(H)$ où l est l'opérateur de mesure de la taille de H en *bits*. Ainsi, la complexité de s est la taille de l'hypothèse causale de taille minimale parmi toutes les hypothèses qui génèrent s, ce qui est une reformulation du rasoir d'Occam.

4.3.4 Complexité et description des objets

Le calcul de la complexité de description d'un objet (ou d'une situation) est subjectif, il dépend des connaissances propres à chaque individu et de ses capacités de calcul. Le calcul de la complexité de description d'un objet consiste à évaluer la taille minimale (exprimée en *bits* de description) de la suite d'instructions pour produire, en utilisant l'ensemble de la connaissance et des capacités de calcul de l'individu, une désignation non ambiguë de l'objet.

Reconsidérons l'exemple du *pachinko* (voir figure 4.4). On cherche à calculer la complexité de description de la situation "la bille a atteint une feuille donnée de l'arbre". Nous pouvons représenter chacune des n situations possiblement observables sous la forme d'une chaîne binaire de longueur $log_2(n)$ (il faut $log_2(n)bits$ pour représenter chacun des éléments d'un ensemble de n éléments non ordonnés). Comme cela a déjà été dit, si aucune 'régularité' ne permet de décrire simplement la chaîne, alors décrire la chaîne binaire consiste à épeler chacun des *bits* qui la composent un à un. Dans un ensemble de n éléments, un élément 'non-remarquable' a donc une complexité de description égale à $log_2(n)bits$. Ainsi, la complexité de description de la situation "la bille a atteint une feuille donnée de l'arbre" est égale à $4bits$ dans le cas de la figure 4.4 si la feuille en question n'est pas 'remarquable'.

La notion d'élément 'remarquable' est fondamentale. Comme nous l'avons dit, l'être humain est capable de comprimer l'information observée, en particulier il est capable d'exploiter des caractéristiques particulières des objets (ou situations) qu'il observe pour les distinguer d'autres objets (ou situations) similaires. Ainsi, si on ordonne les feuilles de l'arbre binaire, alors la feuille numéro 1 devient 'remarquable' car c'est celle qui a le plus petit numéro. De même, si toutes les feuilles sont colorées en bleu sauf une, qui est verte, la situation "la bille a atteint la feuille verte" est plus simple à décrire que n'importe quelle situation "la bille a

atteint une feuille bleue donnée" (il en reste $n-1$ équivalentes, ce qui va coûter $log_2(n-1)bits$ pour discerner une feuille bleue en particulier parmi l'ensemble des feuilles bleues). L'être humain exploite la caractéristique (la couleur de la feuille) qui permet de décrire de façon non-ambiguë la situation observée.

De manière plus formelle, soit une caractéristique (ou une classe) f (pour *feature*) et n objets (indiscernables) qui possèdent cette caractéristique. Ces n objets font partie d'un ensemble d'objets plus vaste comprenant N objets. Pour décrire complètement un objet de l'ensemble à N objets, si on n'exploite pas la caractéristique f, il nous faudrait $log_2(N)bits$ de description.

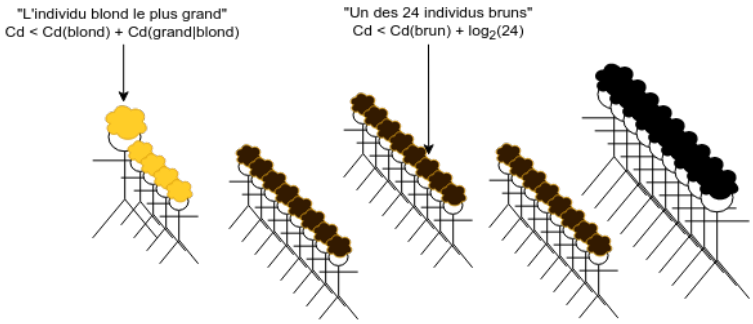
Si on cherche maintenant à décrire un objet qui possède la caractéristique f, on doit tout d'abord payer un coût (en *bits* de description) pour décrire f puis un coût pour déterminer l'objet parmi les n objets qui sont caractérisés par f. Ainsi, la complexité de l'objet est telle que :

$$C_d(objet) \leq C_d(f) + log_2(n) \tag{4.13}$$

Le coût de description de la caractéristique doit être pris en compte dans la description. S'il est trop élevé, nous pourrions nous retrouver dans une situation où $C_d(f) + log_2(n) > log_2(N)$. Dans ce cas, il n'est pas pertinent d'exploiter la caractéristique f pour décrire l'objet, car il existe une description de taille moins grande (celle qui consiste à décrire l'objet comme un élément de l'ensemble vaste). Ainsi, seules les caractéristiques suffisamment simples qui caractérisent des ensembles d'objets de taille suffisamment réduite peuvent servir à obtenir des descriptions de taille moindre. La couleur verte de la feuille ou le fait que ce soit celle qui 'a le plus petit numéro' sont des caractéristiques simples. Par ailleurs, ces caractéristiques caractérisent des ensembles contenant un unique élément et peuvent donc être exploitées.

L'utilisation de plusieurs caractéristiques particulières pour simplifier la description d'objets (ou de situations) est possible. Nous notons $C_d(f_2|f_1)$ la description de la caractéristique f_2 sachant que l'on a déjà décrit la caractéristique f_1. Dans la figure 4.5, on peut exploiter deux caractéristiques pour décrire simplement "l'individu blond le plus grand" (il est blond et il est grand), ce qui n'est pas le cas pour un personnage brun (les 24 personnages bruns sont non-discernables).

FIGURE 4.5 – Utilisation de caractéristiques pour décrire simplement des individus



Pour certaines caractéristiques, on peut ranger les individus ayant ces caractéristiques dans un ordre, leur donner un rang. Ainsi, si on considère la classe des "écrivains français", on peut ranger les éléments de cette classe en fonction de leur célébrité. Molière est plus célèbre que Pierre Boulle, on possède donc en utilisant la classe des écrivains français et en rangeant les éléments suivant une caractéristique de célébrité, une description plus courte de Molière que de Pierre Boulle car ce dernier se situe plus loin dans cette liste ordonnée.

Plus généralement, pour un objet d'une classe donnée *class* (*class* est une caractéristique, comme f définie ci-après), de rang *range via* un ordre dépendant d'une caractéristique f, la description de cet objet est telle que :

$$C_d(objet) \leq C_d(class) + C_d(f|class) + log_2(rang) \tag{4.14}$$

La formule précédente montre que les éléments les plus représentatifs (les plus simples) parmi les éléments d'une classe (simple) qui ont une caractéristique (simple) sont ceux de rang faible. Molière est l'un des écrivains français les plus célèbres (il est donc simple à décrire). Michael Jackson est l'un des chanteurs les plus célèbres (il est donc simple à décrire). Le doyen de l'humanité est l'homme le plus vieux (il est donc simple à décrire).

La complexité de description est subjective et dépend des connaissances de l'individu qui la calcule à un moment donné. On peut supposer que le rang de Modiano dans la liste des auteurs français les plus célèbres à évolué après l'obtention du prix Nobel de littérature. Par ailleurs, une possible meilleure description

de Modiano (par rapport à celles existantes avant l'obtention du prix Nobel) est (en l'an 2015) "c'est le dernier écrivain français à avoir obtenu le prix Nobel de littérature". De même un très vieil homme qui devient doyen de l'humanité devient plus simple à décrire à la mort du précédent doyen.

Lorsque l'on est face à la conjonction de deux situations s_1 et s_2, chacune des situations prise individuellement peut être décrite, et l'on obtient donc deux complexités $C_d(s_1)$ et $C_d(s_2)$. Or la description de la conjonction de ces situations $s_1 * s_2$ suit la formule suivant :

$$C_d(s_1 * s_2) \leq C_d(s_1) + C_d(s_2|s_1) \tag{4.15}$$

L'inégalité vient de possibles éléments de description communs aux deux situations. Ces éléments ne sont décrits qu'une fois et donc la description des deux situations est plus faible que la somme des descriptions individuelles. Nous verrons que cela permet, entre autres, d'expliquer l'intérêt porté par les humains aux coïncidences.

La formule précédente peut être étendue à la description de situations qui sont conjonctions de N situations s_1, s_2, . . . s_N. Ainsi, on a :

$$C_d(s_1 * s_2 * ... * s_N) \leq C_d(s_1) + C_d(s_2|s_1) + C_d(s_3|s_1 \& s_2) + ... + C_d(s_N|s_1 \& s_2 \& ... \& s_{N-1}) \tag{4.16}$$

4.4 Limites de la théorie des probabilités

Le cas des loteries remarquables illustre la limite de la théorie classique des probabilités. Deux événements de probabilité (objective) égale seront évalués comme ayant des probabilités (subjectives) différentes. Plusieurs travaux ont montré que la complexité de description influençait les évaluations de probabilité des événements.

L'un des principaux résultats de la théorie de la Simplicité est la définition d'une probabilité *ex post* p d'un événement via la formule :

$$p = 2^{-U} \tag{4.17}$$

où U est l'inattendu de l'événement.

Notre hypothèse est que le calcul des probabilités attribuées à l'occurrence d'un événement donné passe par le calcul d'un saut de complexité, et non l'inverse (Saillenfest and Dessalles, 2015). Autrement dit, l'être humain réalise d'abord un calcul de complexité et s'en sert pour évaluer la probabilité subjective d'occurrence d'un événement.

Comme nous l'avons dit plus haut, dans le cas des loteries remarquables, la théorie de la probabilité algorithmique réalise une prédiction inverse à celle constatée empiriquement, à savoir que les individus jugent très majoritairement l'occurrence de tirages remarquables comme étant moins probables. La formule 4.17, quant à elle, prédit que parmi un ensemble de situations d'égale complexité de génération (c'est le cas de l'ensemble des tirages de loteries) les plus simples à décrire seront perçues comme étant celles dont l'occurrence est la moins probable (autrement dit, la probabilité *ex post* (ou probabilité subjective) de ces tirages est faible (Dessalles, 2006; Maguire et al., 2013)). Ceci est conforme à l'observation, les tirages les plus inattendus sont ceux dont l'occurrence est perçue comme étant la moins probable. De la même manière, dans l'exemple du *pachinko*, si la bille tombe sur une feuille remarquable, cela sera inattendu et perçu comme étant 'un coup de chance'.

La formule 4.17 rend également compte du biais appelé 'erreur du parieur' (*gambler fallacy*) (Saillenfest and Dessalles, 2015). Ce biais cognitif peut être résumé ainsi : les individus jugent que la probabilité d'occurrence d'un événement de type 'tirage aléatoire' diminue après une occurrence de cet événement. Un joueur va avoir tendance à éviter de jouer un tirage qui est sorti récemment à la loterie (s'il est conscient que ce tirage est sorti récemment). Ce phénomène a été mis en évidence par Terrell (1994) qui a analysé les gains à la loterie du New Jersey et a montré que les tirages gagnants rapportent plus de gains s'ils sont à nouveau tirés, ce qui indique que les joueurs tiennent compte des tirages passés et entre en contradiction avec l'indépendance des tirages prise en compte par la théorie des probabilités.

Après l'occurrence d'un tirage à la loterie, ce tirage devient, pendant un certain temps, particulièrement simple à décrire (c'est l'un des derniers tirages dans la liste ordonnée des tirages passés). Or la complexité de génération du tirage n'est pas

affectée par l'occurrence de ce tirage. Une seconde occurrence de ce tirage serait donc très inattendue (c'est une application directe de la formule 4.4) et donc, d'après la formule 4.17, une seconde occurrence de l'événement est jugée comme étant peu probable pendant un certain temps (après ce 'certain temps', le rang du tirage dans la liste des derniers tirages ne permet plus d'économiser des *bits* de description).

Un autre biais classique qui met en échec la théorie des probabilités est le biais de représentativité (Simon, 1972). Les individus jugent les séquences de naissances au sein d'une même famille de type FFFFGGGG (tout d'abord quatre filles sont nées, puis ensuite quatre garçons) comme étant moins probables qu'une séquence plus complexe à décrire (FFGFGGFG par exemple). Là encore, la complexité de génération est la même pour chacune des séquences, mais la structure anormalement simple de la séquence FFFFGGGG la rend plus simple à décrire et donc la rend plus inattendue (voir équation 4.4), ce qui donne ce sentiment d'improbabilité (voir équation 4.17) (Saillenfest and Dessalles, 2015).

L'erreur de conjonction est un autre biais cognitif mis en évidence par Tversky & Kahneman (1983). Pour les individus interrogés, il est moins probable que Paul, un ex-activiste écologiste, conduise un 4x4 plutôt qu'il conduise un 4x4 fonctionnant au GPL (les 4x4 sont réputés consommer plus d'énergie qu'un véhicule standard, le GPL est réputé être moins polluant que le gasoil). Du point de vue de la théorie des probabilités, c'est absurde, l'ensemble des 4x4 inclut les 4x4 fonctionnant au GPL, le jugement de probabilité devrait donc être inverse. Nous allons montrer que la théorie de la Simplicité permet de réaliser des prédictions conformes à l'observation et rend donc compte de ce biais (Saillenfest and Dessalles, 2015).

Considérons deux situations : "Paul conduit un 4x4" (s_1) et "Paul conduit un 4x4 qui fonctionne au GPL" (s_2). Si on suppose que Paul est déjà dans le contexte, la situation s_1 peut être décrite en utilisant le concept d'activiste-écologiste G (pour *Green*) et le concept de 4x4 f_1.

$$C_d(s_1) = C_d(G) + C_d(f_1|G)$$

Pour décrire s_2 il nous faut introduire le concept de GPL f_2.

$$C_d(s_2) = C_d(G) + C_d(f_1|G) + C_d(f_2|G\&f_1)$$

(nous ignorons volontairement les autres caractéristiques communes aux deux situations, telles que la notion de 'conduire').

Analysons maintenant les situations s_1 et s_2 du point de vue de leur génération. Un 4x4 est généralement alimenté avec du gasoil. La situation s_1 évoque une situation 'typique' (ou représentative d'un cas général), c'est-à-dire une situation dans laquelle le 4x4 fonctionne au gasoil. Le GPL étant plus "écolo-*friendly*", la contradiction entre le passé d'activiste-écologiste de Paul est moins flagrante dans le cas de la situation s_2 que dans le cas de la situation s_1. Cela signifie donc que le scénario causal de taille minimale pour expliquer s_2 est moins complexe que celui qui permet d'expliquer s_1. Autrement dit : $C_w(s_2) < C_w(s_1)$. Ainsi, si on compare l'inattendu des situations s_1 et s_2, on obtient :

$$U(s_1) - U(s_2) = C_w(s_1) - C_w(s_2) + C_d(f_2|G\&f_1) > 0$$

Ainsi, la théorie de la Simplicité prédit que la situation s_1 sera perçue comme étant plus inattendue, et donc moins probable, que la situation s_2. Cette prédiction est conforme avec ce qui est appelé l'erreur de conjonction (*conjunction fallacy*). Notre explication n'est pas sans lien avec celle de Maguire et al. (2013). Les différents modèles introduits par ces auteurs correspondent aux scénarios causaux sur lesquels reposent les évaluations de $C_w(s_1)$ et $C_w(s_2)$. Bien que leur approche soit basée sur la notion de complexité, leur description est exprimée en termes de distributions de probabilités. Nous avons montré ci-dessus que le détour par les probabilités n'est pas nécessaire et que l'ensemble du phénomène peut être analysé exclusivement en utilisant la notion de complexité.

4.5 Plagiat ou coïncidence ? La complexité comme facteur de décision

Le plagiat est défini comme le fait d'utiliser le travail ou les mots d'une personne sans la créditer. Face à un cas de possible de plagiat, un individu doit décider si

les ressemblances perçues entre deux textes donnés relèvent de la coïncidence ou d'un acte délibéré de plagiat. La théorie des probabilités ne parvient pas à rendre compte de notre attitude, en particulier notre intérêt, pour les coïncidences. Elle ne peut pas non plus être utilisée pour déterminer si un possible plagiat est un cas avéré de plagiat ou une simple coïncidence.

Dans un premier temps nous allons procéder à l'analyse en complexité des coïncidences, nous montrerons que la théorie de la simplicité permet d'évaluer l'inattendu d'une coïncidence, et donc ouvre la voie à une quantification de l'intérêt des individus pour les coïncidences.

Nous allons étendre notre étude à l'analyse en complexité des cas possibles de plagiat. En particulier, notre objectif est de proposer une modèle du jugement face à un possible cas de plagiat. Nous allons montrer que l'application directe de la théorie de la simplicité aux cas de plagiats permet de mettre en évidence un certain nombre de facteurs qui influencent le jugement des individus lorsqu'il s'agit de décider si un possible cas de plagiat est un cas flagrant de plagiat ou une coïncidence (certaines des prédictions du modèle ont été testées empiriquement).

4.5.1 Modélisation des coïncidences

D'une manière générale, on parle de coïncidence lorsqu'une situation qui se produit peut être décrite en faisant référence à une situation similaire. On peut distinguer deux cas. Le premier est le cas de situations qui, réunies en une seule et unique situation (*via* l'opérateur de conjonction), sont anormalement simples à décrire. C'est le cas du tirage de loterie qui forme le numéro de téléphone de votre meilleur ami. Chaque numéro tiré est une situation non inattendue. La conjonction de ces situations forme le tirage, dont la description est anormalement simple, et qui est perçu comme étant inattendu.

Nous allons nous intéresser au cas d'une situation qui se produit, et que l'on peut décrire par rapport à une autre situation. Soit s_2 la situation qui se produit et s_1 la situation qui sert de référence. Produire la situation s_2 nécessite $C_w(s_2)bits$. Pour décrire s_2, il faut tout d'abord décrire s_1, puis préciser comment s_1 est transformée en s_2 (on notera A l'algorithme qui transforme s_1 en une partie de s_2) puis décrire ce qui reste à décrire dans s_2. On fait l'hypothèse que la description de

l'algorithme A ne dépend pas de la description de s_1, ainsi on a $C_d(A|s_1) = C_d(A)$.

La description de s_2 (en supposant que s_1 a été déjà décrite) est donc telle que :

$$C_d(s_2|s_1) \leq C_d(A) + C_d(s_2|s_1\&A) \tag{4.18}$$

Ainsi, la description de la coïncidence $s_1\&s_2$ est telle que :

$$C_d(s_1\&s_2) \leq C_d(s_1) + C_d(A) + C_d(s_2|s_1\&A) \tag{4.19}$$

La co-occurrence de s_1 et s_2 est une coïncidence si les occurrences respectives de ces situations sont supposées indépendantes :

$$C_w(s_1\&s_2) = C_w(s_1) + C_w(s_2) \tag{4.20}$$

En utilisant l'inéquation 4.19 ainsi que les équations 4.20 et 4.4, l'inattendu de la coïncidence est tel que :

$$U(s) \geq C_w(s_1) + C_w(s_2) - C_d(s_1) - C_d(A) - C_d(s_2|s_1\&A) \tag{4.21}$$

Autrement dit, une coïncidence est d'autant plus inattendue que :

1. les situations observées sont difficiles à produire ($C_w(s_1)$ et $C_w(s_2)$ élevés)
2. la situation de référence est simple à décrire. ($C_d(s_1)$ faible)
3. l'algorithme qui permet d'associer la situation de référence à la situation observée est simple à décrire ($C_d(A)$ faible)
4. la situation observée est simple à décrire une fois l'analogie avec la situation de référence a été présentée ($C_d(s_2|s_1\&A)$ faible)

4.5.2 Analyse en complexité d'un cas de plagiat

Supposons qu'un individu se trouve face à un texte t_2 et trouve que ce texte ressemble fortement à un autre texte t_1 qu'il connaît. L'individu se pose alors la question de savoir si le texte t_2 est le résultat d'un acte de plagiat ou s'il s'agit d'une coïncidence. Le plagiat sera d'autant plus flagrant que la coïncidence est inattendue (Saillenfest and Dessalles, 2015).

Soit la situation T_1 qui consiste à produire le texte t_1, et T_2 celle qui consiste à produire t_2. En utilisant l'équation 4.21, l'inattendu de la coïncidence $T_1\&T_2$ est tel que :

$$U(T_1\&T_2) \geq C_w(T_1) + C_w(T_2) - C_d(T_1) - C_d(A) - C_d(T_2|T_1\&A) \tag{4.22}$$

Dans le cas du plagiat, l'algorithme A correspond à l'ensemble des transformations appliquées à t_1 pour le transformer en t_2.

La description de l'événement T_1 correspond à la description du texte t_1. Ainsi, $C_d(T_1) = C_d(t_1)$. De même $C_d(T_2|T_1\&A) = C_d(t_2|T_1\&A)$.

Ainsi, l'inéquation 4.22 devient :

$$U(T_1\&T_2) \geq C_w(T_1) + C_w(T_2) - C_d(t_1) - C_d(A) - C_d(t_2|T_1\&A) \tag{4.23}$$

L'hypothèse de plagiat est d'autant plus forte en tant qu'hypothèse explicative que la coïncidence est inattendue. Autrement dit, plus la coïncidence est inattendue, plus le plagiat est flagrant. L'analyse des différents termes de l'inéquation 4.23 permet de réaliser un ensemble de prédictions qui influencent notre perception du plagiat.

Prédictions du modèle du plagiat et validation expérimentale

Dans le cas du plagiat, le problème auquel fait face un individu est de décider si la ressemblance entre deux textes est une coïncidence ou s'il y a plagiat de la part de l'auteur du second texte (dans l'ordre chronologique). Dans ce qui précède, nous avons montré que l'application directe de la théorie de la Simplicité permet de quantifier l'inattendu de la coïncidence entre le contenu de deux textes.

Le calcul a permis de mettre en lumière quelques termes qui semblent influencer le jugement des individus face à un possible cas de plagiat. Dans cette partie, nous allons passer en revue un certain nombre de prédictions réalisées à partir des équations de la partie 4.5.2. L'objectif est de montrer que, face à un cas de plagiat possible, les individus calculent effectivement les termes de complexité (Saillenfest and Dessalles, 2015).

La complexité de génération d'un texte : L'hypothèse de la coïncidence repose en grande partie sur ce terme. En effet, une coïncidence est une situation complexe à produire mais anormalement simple à décrire *via* l'utilisation d'éléments d'une situation déjà décrite. Le calcul de la complexité de génération d'un texte nécessite de se placer dans un état épistémique dans lequel le texte n'existe pas encore (sinon il fait partie du monde, il a déjà été généré et sa complexité de génération est donc nulle). Générer un texte consiste généralement à produire N mots qui forment le contenu du texte, nous allons faire une approximation en considérant que ces mots sont choisis parmi les n mots les plus courants de la langue dans laquelle le texte est écrit. Si on considère un modèle de génération du texte dans lequel chacun des mots est choisi aléatoirement parmi les n mots les plus courants, alors la génération de chacun des mots nécessite $log_2(n)$ *bits*. La production du texte complet nécessiterait donc $N * log_2(n)$ *bits*. Ce modèle de génération d'un texte est très peu réaliste. Un meilleur modèle consisterait à supposer que chaque mot est choisi parmi les plus pertinents en fonction du contenu du texte. On pourrait par exemple utiliser un critère de type TF-IDF, c'est-à-dire en regardant la fréquence et le nombre d'occurrences des mots dans le texte. Quoi qu'il en soit, la complexité de génération d'un texte T avec un tel modèle serait toujours linéaire en fonction de la taille du texte $taille(T)$ (exprimée en nombre de mots, de caractères, de lignes, etc.). Ainsi, nous proposons une approximation de la complexité de génération d'un texte (hors hypothèse de plagiat) qui serait de la forme :

$$C_w(T) = taille(T) * K \tag{4.24}$$

où K est une constante. Cela nous permet donc de prédire, d'après l'équation 4.23, que plus le texte soupçonné d'être le produit d'un plagiat est long, plus la coïncidence est perçue comme inattendue, et donc plus l'hypothèse de plagiat est privilégiée. Cela correspond à l'intuition générale. Par ailleurs, un test expérimental montre un effet clair de la taille du texte sur la perception des cas de plagiat (voir annexe A.1). De meilleures approximations sont toutefois possibles pour le calcul de la complexité de génération d'un texte. On peut par exemple prendre en compte la densité des mots pertinents dans le texte que l'on soupçonne être le

produit d'un plagiat. La complexité de génération du texte serait une fonction croissante en fonction de cette densité, ce qui aurait donc pour effet de rendre une coïncidence plus intéressante, car plus inattendue, mais tout en faisant pencher la balance en faveur de l'hypothèse de plagiat.

La complexité de description d'un texte : D'après l'équation 4.23, on peut voir que la taille minimale de l'algorithme qui permet de décrire le texte supposé être un plagiat joue un rôle dans la reconnaissance du plagiat. Nous allons détailler ici quelques éléments de description du texte supposé résulter d'un acte de plagiat qui interviennent dans le calcul de cette complexité de description.

Tout d'abord, il faut noter que l'algorithme 'par défaut' pour décrire un texte qui n'existe pas encore, et qui doit donc être discriminé des autres textes potentiels, requiert la description de chaque mot (la description doit être non ambiguë). La complexité de cet algorithme par défaut est à peu près égale à la complexité de génération du texte, il faut la même quantité d'information pour décrire un mot et pour le générer à partir d'une loterie. Ainsi, la plupart des textes n'apparaissent pas comme étant 'inattendus', en tout cas au moment de leur écriture. *A posteriori*, certains textes deviennent anormalement simples, parce qu'ils sont célèbres, écrits par des auteurs célèbres ou encore parce qu'ils présentent une singularité qui rend leur description simple. L'écriture de ces textes peut apparaître comme inattendue *a posteriori*, lorsque le texte est devenu simple à décrire, mais elle n'est généralement pas inattendue au moment où le texte est écrit.

Dans le cas du plagiat, l'algorithme pour décrire un texte supposé être le résultat d'un plagiat est plus court si on connaît le texte plagié car on peut exploiter leurs ressemblances. Une description plus courte que celle qui consiste à décrire mot par mot le texte peut donc se décomposer en trois étapes (ce dont rend compte l'équation 4.23) :

1. décrire un texte de référence t_1 (le texte supposé plagié), ce qui coûte $C_d(t_1)$ $bits$.
2. décrire l'algorithme A qui transforme le contenu du texte supposé plagié en le contenu du second texte t_2, ce qui coûte $C_d(A)$ $bits$.

3. décrire tout ce qui dans t_2 n'a pas été décrit jusqu'ici (finir la description de t_2), ce qui coûte $C_d(t_2|T_1\&A)$ *bits* (en particulier, la description de t_1 et l'algorithme A ne nous donnent pas d'informations sur la localisation de t_2, qui est donc un élément qui reste à décrire à ce stade).

Décrire un texte consiste à donner tous les éléments nécessaires pour le rendre unique. Par exemple, pour décrire un texte on peut indiquer où il se trouve dans un livre puis donner son contenu. Mais, si l'on s'intéresse au premier chapitre du premier livre de La Recherche du Temps Perdu de Marcel Proust, alors cette description qui vient d'être donnée (c'est-à-dire "le premier chapitre du premier livre de La Recherche du Temps Perdu de Marcel Proust") est plus simple que celle consistant à donner la liste de tous les mots qui composent ce texte.

Un facteur qui peut influencer la taille minimale de la description d'un texte est sa localisation. Localiser un passage dans un livre consiste à déterminer une page parmi l'ensemble des pages du livre. Ainsi, pour un livre B, de taille $taille(B)$, la complexité de la localisation L est donc de l'ordre de :

$$C_d(L) = log_2(taille(B)) \tag{4.25}$$

Ainsi, la formule 4.23 prédit que la coïncidence est d'autant plus inattendue que les livres dans lesquels les textes se trouvent sont courts. Et donc l'hypothèse de plagiat devient une explication plus satisfaisante.

Parfois, lorsque le texte se trouve en début de livre par exemple, décrire la page à laquelle se trouve le texte permet d'économiser des *bits* de description. Si le texte se trouve à la page $page(T)$, alors dans ce cas, la complexité pour décrire la localisation dudit texte est de l'ordre de :

$$C_d(L) = log_2(page(T)) \tag{4.26}$$

Ainsi, l'équation 4.23 prédit qu'un plagiat sera plus évident si les textes impliqués se trouvent dans des endroit remarquables (au début par exemple) des livres dans lesquels ils se trouvent.

Jusqu'ici, les évaluations de la complexité de description ont été données en supposant les livres dans lesquels les textes se trouvent déjà connus. Or pour

décrire l'endroit où se trouve un texte, il faut également décrire le livre dans lequel il se trouve. Ainsi, une évaluation de la complexité de description d'un texte peut donc consister à décrire successivement (et donc à additionner les complexités de descriptions de ces éléments) :

1. l'auteur, ce qui coûte $C_d(auteur)$ *bits*
2. le livre (sachant l'auteur), ce qui coûte $C_d(livre|auteur)$ *bits*
3. l'endroit où se trouve le texte dans le livre, ce qui coûte $C_d(localisation|livre\&auteur)$ *bits*
4. le contenu du texte, ce qui coûte $C_d(contenu)$ *bits*

La machine observation cherche la meilleure description possible du texte. Une œuvre comme la Recherche du Temps Perdu ne nécessite pas forcément de préciser l'auteur, ni le contenu du texte.

Ce que prédit le modèle de l'équation 4.23, c'est que tout ce qui contribue à simplifier la description des textes va contribuer à renforcer l'hypothèse de plagiat. Ainsi, le plagiat sera jugé comme plus évident si le texte plagié est le premier chapitre du premier livre de La Recherche du Temps Perdu que si c'est le chapitre 5 du 4è livre d'un auteur peu connu.

Décrire l'algorithme qui transforme un texte plagié en un autre texte consiste à décrire les modification apportées au contenu du texte plagié pour le transformer en une copie modifiée. Ce que prédit le modèle de l'équation 4.23, c'est que le plagiat est d'autant plus flagrant que l'algorithme qui transforme le texte plagié en une copie modifiée est simple à décrire. Ainsi, une copie pure et simple du texte plagié rend le plagiat plus évident que les cas où le plagiat consiste à découper le texte plagié en plusieurs parties et à insérer ces parties dans divers endroits d'un livre.

Validation expérimentale : En annexe A.1 se trouvent les résultats d'une expérience qui valide des prédictions présentées ci-dessus. En particulier, les prédictions suivantes sont testées et validées :

— Plus le texte soupçonné d'être le produit d'un plagiat est long, plus la coïncidence est perçue comme inattendue, et donc plus l'hypothèse de plagiat est

privilégiée.

— La coïncidence est d'autant plus inattendue que les livres dans lesquels les textes se trouvent sont courts. Et donc l'hypothèse de plagiat devient une explication plus satisfaisante.

— Un plagiat sera plus évident si les textes impliqués se trouvent dans des endroits remarquables (au début par exemple) des livres dans lesquels ils se trouvent.

— Le plagiat est d'autant plus flagrant que l'algorithme qui transforme le texte plagié en une copie modifiée est simple à décrire.

4.6 Complexité et pertinence émotionnelle dans la narration

Le modèle de la théorie de la simplicité donne une définition de l'inattendu en utilisant la notion de complexité. Sont inattendues les situations qui sont anormalement simples à décrire. Les deux facteurs de l'intérêt événementiel sont la complexité de génération (qui est une complexité attendue) et la complexité de description (qui est une complexité constatée). Comme on l'a vu, de nombreux facteurs peuvent affecter la complexité de description d'une situation. La théorie de la simplicité rend donc compte de notre attrait général pour des événements qui concernent des personnalités qui sont simples à décrire (les célébrités, nos proches, nos amis ou toute personnalité simple pour une raison simple), qui se déroulent dans des lieux simples à décrire (lieux connus, grandes villes, lieux où nous avons vécu, lieu où nous travaillons, etc.) ou encore les événements dont le thème est simple (une passion, un livre qui a été lu, un film qui a été vu, etc.). La plupart des situations du quotidien ne sont pas inattendues, car elles sont aussi complexes à discriminer qu'à produire.

Comme nous l'avons montré en introduction, l'inattendu est un facteur majeur de l'intérêt narratif. Il n'y a pas véritablement lieu de penser que l'être humain réalise un calcul de l'intérêt différent dans le cadre d'une fiction de celui qu'il réalise lorsqu'il est confronté à un événement dans le monde réel. Par ailleurs, notre représentation du monde réel est très parcellaire, nous ne connaissons qu'une petite

partie des éléments qui le composent tout comme nous connaissons peu d'éléments qui constituent le monde fictif (ceci explique peut-être en partie pourquoi nous nous projetons aussi facilement dans ce monde fictif). De plus, une grande partie des informations que nous avons sur le monde réel ne provient pas d'une expérience sensible de ce monde mais au travers de médias divers, une sélection des informations ayant été opérée avant leur transmission. Ainsi, le calcul de l'intérêt n'est pas *a priori* moins facile à réaliser dans le cadre narratif que dans les cas où nous sommes confrontés à des événements qui se produisent dans le monde réel.

Certains effets de simplicité dans le récit sont évidents. Notre intérêt est accru lorsque les situations qui nous sont présentées mettent en scène les personnages principaux du récit ou se déroulent dans des lieux simples (par exemple l'Étoile de la Mort dans la série de films de l'univers de Star Wars).

Par ailleurs, il n'est pas rare que des personnages aient un signe distinctif très particulier, ce qui rend leur description simple et permet de les discriminer facilement des autres personnages. On peut citer par exemple la cicatrice en forme d'éclair de Harry Potter ou les attributs d'Indiana Jones, son fouet et son chapeau Fedora.

Voici un autre exemple de l'utilisation de la simplicité de description pour produire un effet sur l'audience, extrait de "La Bibliothèque Universelle" de Kurd Lasswitz qui a pour cadre une bibliothèque qui contiennent un très grand nombre de livres dont le contenu est une suite aléatoire de caractères d'un ensemble donné de caractères possibles (cette œuvre a inspiré l'écriture de "La Bibliothèque de Babel" de Borges) :

> « « Comment ? » s'écria la maîtresse de maison. « Tout est dans la bibliothèque ? Tout Goethe y est aussi ? La Bible ? Les œuvres complètes de tous les philosophes ayant existé ? »
> « Sans oublier toutes les interprétations auxquelles personne n'a encore pensé. Tu y trouveras également toutes les œuvres perdues de Platon ou de Tacite, ainsi que leurs traductions. Mais également l'ensemble de nos œuvres futures à tous les deux, tous les discours parlementaires, aussi bien ceux qui ont été oubliés que ceux qui n'ont pas encore été prononcés, le traité universel de Paix mondiale et l'histoire des guerres du futur qui s'ensuivent. »

> « Ainsi que le registre impérial des horaires de trains, mon oncle ! » s'exclama Suzanne. « C'est ton livre préféré. » » in "La Bibliothèque Universelle" de Kurd Lasswitz (1904)

Dans cet exemple, l'auteur prend comme exemples d'œuvres qui se trouvent dans la bibliothèque des livres simples à décrire, soit parce que les auteurs desdits livres sont très célèbres (Goethe, Platon, Tacite), soit parce que le livre en lui-même est très célèbre (La Bible), mais également des œuvres qui sont simples pour les personnages principaux du récit qui sont eux-mêmes simples ("nos œuvres futures à tous les deux", "« C'est ton livre préféré. »").

Le thème du plagiat est celui du court récit "Une carrière" de Jean Giraudoux. Voici deux extraits qui montrent comment l'auteur manipule la simplicité de description des œuvres plagiées pour créer de l'intérêt narratif :

> « W. O'Duffin, suivant en cela l'exemple de nombreux universitaires américains, venait de publier, coup sur coup, un premier roman : Les Trois Mousquetaires, par O'Duffin et Dumas ; un second roman : Monte-Cristo, des mêmes auteurs, et, toujours par les mêmes, une tragédie : La Dame aux Camélias. » in "Une Carrière" de Jean Giraudoux

> « Elle lut, d'abord hésitante, puis s'abandonnant à une joie qui révéla enfin ses préférences :
> – L'Odyssée, par MM. O'Duffin et Homère... Je comprends, s'écria-t-elle, ce sont vos mémoires. » in "Une Carrière" de Jean Giraudoux

Un autre aspect de la manipulation de la complexité de description dans le cadre narratif est la référence à un objet précédemment introduit. En effet, au cours de la lecture d'un récit fictif, de nombreuses situations sont décrites et stockées en mémoire. Les séquences de descriptions peuvent être vite ennuyeuses car la complexité de description fait baisser l'intérêt du récit. Mais les situations, une fois décrites, sont alors simples d'accès en mémoire (pendant un temps au moins). Ainsi, une situation, ou un objet, va être décrit (parfois de manière complexe, ce qui ne rendra pas cette séquence du récit intéressante), et, ultérieurement dans le récit, être réutilisé. Cette fois-ci, il sera alors plus simple à décrire car l'algorithme

pour le décrire consistera à faire référence non pas à l'objet mais à une précédente occurrence dudit objet. Voici un exemple de manipulation de la complexité de description *via* la référence à une situation antérieure dans le Horla :

> « 19 août. – Je sais... je sais... je sais tout ! Je viens de lire ceci dans la Revue du Monde scientifique : « Une nouvelle assez curieuse nous arrive de Rio de Janeiro. Une folie, une épidémie de folie, comparable aux démences contagieuses qui atteignirent les peuples d'Europe au moyen âge, sévit en ce moment dans la province de San-Paulo. Les habitants éperdus quittent leurs maisons, désertent leurs villages, abandonnent leurs cultures, se disant poursuivis, possédés, gouvernés comme un bétail humain par des êtres invisibles bien que tangibles, des sortes de vampires qui se nourrissent de leur vie, pendant leur sommeil, et qui boivent en outre de l'eau et du lait sans paraître toucher à aucun autre aliment.
> « M. le professeur Don Pedro Henriquez, accompagné de plusieurs savants médecins, est parti pour la province de San-Paulo afin d'étudier sur place les origines et les manifestations de cette surprenante folie, et de proposer à l'Empereur les mesures qui lui paraîtront le plus propre à rappeler à la raison ces populations en délire. »
> Ah ! Ah ! je me rappelle, je me rappelle le beau trois-mâts brésilien qui passa sous mes fenêtres en remontant la Seine, le 8 mai dernier ! Je le trouvais si joli, si blanc, si gai ! L'Être était dessus, venant de là-bas, où sa race est née ! Et il m'a vu ! Il a vu ma demeure blanche aussi ; et il a sauté du navire sur la rive. Oh ! mon Dieu ! » (in "Le Horla" de Guy de Maupassant)

Dans cet extrait, le personnage dit qu'il a compris ce qui se passait, et se met à décrire une situation proche de la sienne mais qui se produit au Brésil. Ainsi, la complexité pour produire la situation "Le personnage principal est hanté par un être invisible (le Horla)" est peu ou prou égale à celle de faire venir ledit être invisible du Brésil. La complexité de description de la situation brésilienne est faible car similaire à celle du personnage décrite précédemment dans l'œuvre. Ainsi l'eau et le lait sont évoqués précédemment.

> « 10 juillet. [...] Le 6 juillet, avant de me coucher, j'ai placé sur ma table du vin, du lait, de l'eau, du pain et des fraises.
> On a bu - j'ai bu - toute l'eau, et un peu de lait. On n'a touché ni au vin, ni au

pain, ni aux fraises.
Le 7 juillet, j'ai renouvelé la même épreuve, qui a donné le même résultat.
Le 8 juillet, j'ai supprimé l'eau et le lait. On n'a touché à rien. » in "Le Horla" de Guy de Maupassant

Ou encore la fuite du personnage de chez lui :

« 16 août. - J'ai pu m'échapper aujourd'hui pendant deux heures, comme un prisonnier qui trouve ouverte, par hasard, la porte de son cachot. J'ai senti que j'étais libre tout à coup et qu'il était loin. J'ai ordonné d'atteler bien vite et j'ai gagné Rouen. Oh ! quelle joie de pouvoir dire à un homme qui obéit : « Allez à Rouen ! » » in "Le Horla" de Guy de Maupassant

Enfin, la résolution du problème vient d'une situation précédemment introduite par l'auteur, qui explique comment l'être a pu venir du Brésil. Cette situation est à la fois de complexité de génération nulle (car elle s'est produite et qu'il ne reste pas d'incertitude liée à sa réalisation) mais aussi de complexité de description faible car déjà décrite précédemment. Voici l'extrait de la nouvelle à laquelle le personnage fait référence :

« 8 mai. [...] Vers onze heure, un long convoi de navires, traînés par un remorqueur, gros comme une mouche, et qui râlait de peine en vomissant une fumée épaisse, défila devant ma grille.
Après deux goélettes anglaises, dont le pavillon rouge ondoyait sur le ciel, venait un superbe trois-mâts brésilien, tout blanc, admirablement propre et luisant. Je le saluai, je ne sais pourquoi, tant ce navire me fit plaisir à voir. » in "Le Horla" de Guy de Maupassant

Ainsi, cela fait baisser subitement la complexité de génération de la situation à laquelle est confrontée le personnage principal sans nécessiter un grand nombre de *bits* de description.

Comme nous l'évoquions en introduction, la satisfaction du critère de postdictibilité passe par une baisse de la complexité de génération des situations présentées. Nous proposons d'appeler 'simplication' cette baisse de complexité de génération. Si un élément s' est présenté qui fait baisser la complexité de génération d'une

situation s, la 'simplication' de s se mesure ainsi :

$$SIMPLICATION(s) = C_w(s) - C_w(s||s') \tag{4.27}$$

Comme l'illustre l'exemple précédent extrait du Horla, les situations de *climax* sont caractérisées par une simplication élevée d'un état du monde, ce qui fait retomber de manière brutale la tension narrative (ce qu'illustre la courbe de tension de Freytag (voir figure 2.3)). Dans l'exemple de simplication présenté page 46 extrait du script d'Indiana Jones et la Dernière Croisade, l'élément présenté au spectateur ("*That's the cup of a carpenter.*") provoque une baisse brutale de la complexité de génération du choix de la coupe la moins belle (notons également que la phrase précédente d'Elsa avait préparé cette explication).

La 'simplication' d'une situation est source d'intérêt, le spectateur est satisfait de pouvoir 'intégrer' (c'est-à-dire calculer une explication) rétrospectivement les éléments du récit dans sa représentation du monde fictif. Une 'simplication' forte, c'est-à-dire une baisse rapide de la complexité de génération est d'autant plus satisfaisante.

Les auteurs préparent généralement ce saut de complexité. Cela peut être vu comme une sorte de réciproque au conseil de Tchékhov (si un auteur veut produire un effet *via* une baisse de complexité dans une scène, le ou les éléments nécessaires à cette baisse de complexité doivent avoir été introduits dans une scène antérieure et être simplement accessibles). Ceci est lié à la technique classique connue sous le nom de "*set up / payoff*" ou "*planting / payoff*". Cette technique consiste à introduire, dans la structure narrative des éléments – le *set up* - qui n'ont pas nécessairement une utilité au moment où il sont introduits (et peuvent même être répétés) dans le but de produire ultérieurement un effet sur l'audience dans le récit (le *payoff*) (autrement dit, pour produire un effet au moment du *payoff*, il faut avoir introduit les éléments nécessaires à la production de cet effet (le *set up*)).

Voici deux exemples d'application de cette technique de création d'intérêt dans un récit :

— un élément est introduit dans le récit afin de provoquer ultérieurement une baisse de complexité de génération. Par exemple, dans le film "James Bond – Casino Royale", James Bond fait un arrêt cardiaque suite à un empoi-

sonnement. À ce moment, le scénariste veut créer un effet - un *payoff* - en permettant à James Bond de se sortir rapidement de cette situation désespérée, autrement dit vaut faire baisser la complexité de la situation "James Bond survit" qui augmente rapidement (il est très inattendu que James Bond s'en sorte vivant à ce moment). Pour cela, un défibrillateur a été introduit antérieurement dans le récit. Un défibrillateur apparaît furtivement dans la voiture de James Bond après que James Bond a appuyé par erreur sur un bouton dans une scène antérieure. C'est un *setup*, l'objet n'a pas d''utilité' au moment où il est introduit dans l'histoire. Le défibrillateur est utilisé pour sauver James Bond lorsque, ultérieurement, il fait un arrêt cardiaque. Le défibrillateur devient alors un élément de baisse de complexité qui est simple à produire et à décrire car introduit peu de temps auparavant (c'est un exemple de simplication).

— un élément est introduit dans le récit de manière a produire une attente qui, au moment du *payoff*, ne se réalisera pas, de faire anticiper à l'audience le développement actionnel d'une séquence narrative similaire qui apparaît ultérieurement. Le *payoff* sera donc caractérisé par un saut de complexité de génération. Un exemple typique est le suivant : dans le film "Indiana Jones - Raiders of the Lost Ark", Indiana Jones, qui tient son fouet à la main, est confronté à un ennemi armé d'un sabre. Dans cette situation de duel, le spectateur s'attend à ce qu'Indiana Jones utilise son fouet ; il y a un *set up* au début du film, Indiana Jones désarme l'un de ses ennemis à l'aide de son fouet ($U = 0$). Cela ne se produit pas, Indiana Jones utilise son revolver (C_w augmente très rapidement) puis, par simplication baisse à nouveau rapidement (le revolver a été introduit précédemment - un autre exemple de *set up*). Dans cet exemple, la technique *set up / payoff* est donc utilisée deux fois, une première fois pour provoquer une hausse de complexité - créer un effet de surprise - et une seconde fois pour provoquer une baisse rapide de complexité - une simplication.

On commence à voir apparaître les mécanismes utilisés par les auteurs de fictions pour faire varier les complexités de génération et de description dans le but de produire des effets sur le lecteur. Dans la suite, nous allons tenter de formaliser ces mécanismes, en redéfinissant des notions classiques de la narration comme le

conflit narratif, la surprise, le suspense ou encore la curiosité. Puis, nous verrons comment définir dans un même espace les contraintes de cohérence du monde et des personnages, en particulier comment la complexité affecte notre perception des intentions des personnages.

Chapitre 5

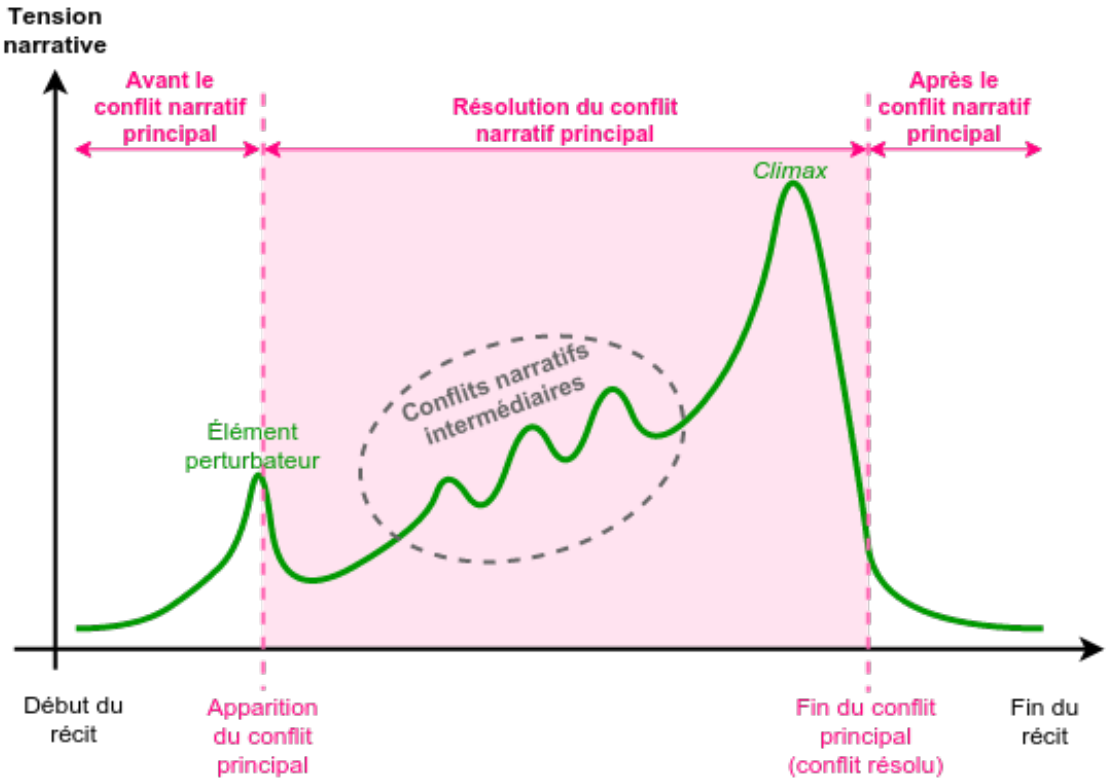
Conflits cognitifs et conflits narratifs

Un récit peut généralement être décomposé en trois parties (ou actes du récits). Le début du récit est une phase au cours de laquelle sont généralement introduits le monde fictif et les principaux personnages. Dans cette phase le monde fictif et les personnages se trouvent généralement dans un état que l'on peut qualifier de stable, ce qui ne signifie pas que l'état du monde est figé mais que le cours des événements qui s'y produisent est 'ordinaire'. La fin de cette phase est caractérisée par l'occurrence d'un événement perturbateur, dans le sens où il vient perturber de manière notable le cours ordinaire des événements. La seconde phase du récit débute alors, les personnages vont agir dans le but de rétablir la stabilité du monde, de mettre fin aux conséquences de l'élément perturbateur, et ils rencontrent généralement un ensemble d'obstacles qui maintiennent l'intérêt de l'audience. Cette phase s'achève généralement après un *climax*, une séquence pendant laquelle la majeure partie des problèmes qui maintiennent le monde dans un climat d'instabilité vont être résolus et qui provoque un pic de tension chez l'audience. La dernière phase est une phase de stabilité retrouvée, ce qui ne signifie pas que le monde est dans le même état de stabilité qu'au début du récit, il peut avoir évolué.

Dans le cadre narratif, la notion de conflit narratif désigne une incompatibilité entre les divers objectifs de personnages ou de 'forces'. L'élément perturbateur est la source du conflit principal du récit, celui dont la résolution marquera la

fin de la seconde phase du récit. Ainsi, nous définissons schématiquement le récit comme une opération de résolution d'un conflit narratif principal, global qui est généralement perturbée par des obstacles qui sont autant de sources de conflits narratifs intermédiaires (d'envergure généralement locale) (voir figure 5.1).

FIGURE 5.1 – Arc narratif et structure du récit configurée par les conflits narratifs qui le composent



L'objectif de ce chapitre est de définir la notion de conflit narratif. Pour cela, nous allons introduire la notion de nécessité, une méthode d'évaluation des états du monde non binaire. Nous montrerons que cette notion de nécessité des états du monde peut être utilisée pour définir une notion de conflit cognitif interne à un individu. Après avoir montré le rôle de la complexité dans l'évaluation des nécessités, nous montrerons que les conflits narratifs peuvent être ramenés à des conflits d'ordre cognitif internes aux personnages, ce qui nous permettra de donner

une définition des conflits narratifs basée sur la notion de complexité. Enfin, nous montrerons que la notion de conflit est source de tension narrative, et donc un élément essentiel de la pertinence émotionnelle dans les récits. Nous montrerons par ailleurs que le conflit peut être source d'action des personnages du récit, il sera donc nécessaire d'étudier la résolution de conflits par des personnages rationnels afin d'établir des critères de pertinence structurelle dans les récits.

5.1 Nécessité d'un état et conflits cognitifs internes à un individu

Lorsqu'il s'agit de représenter la connaissance d'un individu concernant son environnement, il est usuel d'adopter une représentation binaire des états des propositions qui caractérisent cet environnement. Ainsi, une proposition telle que "il fait beau" est désignée comme étant 'vraie' (*True* au sens de la logique booléenne) ou 'fausse' (*False*). Cet état n'est pas permanent, il dépend du contexte dans lequel se trouve l'individu. Cette représentation binaire des états des propositions qui composent la connaissance des individus ne rend pas compte des incertitudes qu'un individu, qui vit dans un monde complexe et a une connaissance subjective très partielle de ce monde, peut avoir.

Par ailleurs, certains états du monde peuvent recevoir des évaluations distinctes venant de sources diverses. Un individu peut par exemple observer un état de fait, et donc considérer que la proposition désignant cet état de fait est 'vraie', et dans le même temps lire dans le journal que cette proposition est 'fausse'. Deux sources, l'observation directe d'un fait et la transmission d'une observation *via* un *medium*, fournissent deux évaluations différentes d'une même proposition. Les sources des évaluations sont diverses : perception directe, déclaration d'autrui, préférence, mémoire d'une évaluation passée, une déduction, etc. Par ailleurs, certaines sources peuvent être perçues comme plus ou moins 'crédibles' par un individu. Une observation directe d'un agent peut être perçue comme une source plus fiable qu'une expérience indirecte par exemple (Thiriot and Kant, 2008).

Enfin, il est important de noter que certaines propositions peuvent être des objets de désir. Ainsi, "il fait beau" peut être 'faux' mais également désiré.

Ces remarques nous amènent donc à introduire le modèle des nécessités (Dessalles, 2008a) qui est un modèle d'évaluation des états représentants la connaissance d'un individu. Cette notion de nécessité permettra de définir la notion de conflit cognitif interne à un individu.

5.1.1 Nécessité et support d'un état

Afin de rendre compte de l'incertitude dans la représentation d'une connaissance et la possibilité de recevoir diverses évaluations distinctes pour un même élément de cette connaissance, nous prendrons comme modèle d'évaluation des états un modèle basé sur les attitudes que nous avons envers ces états, c'est-à-dire l'ensemble des croyances et des désirs associés à ces états (et leurs correspondants négatifs). Ces attitudes seront considérées comme graduelles, ainsi un état de fait T pourra recevoir une évaluation $\nu(T)$ qui est une valeur positive ou négative. $|\nu(T)|$ désigne l'intensité de cette évaluation. Nous appellerons nécessités d'un état ces évaluations (Dessalles, 2008a). Chaque nécessité a une intensité et une valence.

On considérera les nécessités comme cohérentes, c'est-à-dire que la négation $\neg T$ d'un état de fait T reçoit une évaluation opposée à celle de l'état de fait.

$$\nu(T) = -\nu(\neg T) \tag{5.1}$$

Nous proposons de fixer un seuil I_0 et de ne nous intéresser qu'aux nécessités d'un état dont l'intensité est supérieure à ce seuil. Nous appelons support d'un état de fait l'ensemble des évaluations dépassant ce seuil d'intérêt. Nous faisons, en introduisant la notion de support, l'hypothèse que l'être humain ne considère à un moment donné, pour un état de fait particulier, qu'un ensemble restreint d'évaluations (celles qui appartiennent au support).

On dit de ce support qu'il est cohérent si, pour un état de fait donné, l'ensemble des évaluations qui appartiennent à ce support, et donc dont l'intensité dépasse le seuil d'intérêt I_0, sont toutes de même signe. Formellement, le support $Support(T)$ d'un état de fait T est cohérent si :

$$\forall \nu_1(T), \nu_2(T) \in Support(T); \nu_1(T) * \nu_2(T) > 0 \tag{5.2}$$

5.1.2 Le conflit d'ordre cognitif

Lorsque le support d'une évaluation est incohérent, nous considérons que l'individu fait face à un conflit d'ordre cognitif. Formellement, un conflit cognitif autour d'un état de fait T est caractérisé par :

$$\exists \nu_1(T), \nu_2(T) \in Support(T); \nu_1(T) * \nu_2(T) < 0$$

Ce que nous appelons conflit cognitif est donc un conflit entre (au moins) deux nécessités d'un même état de fait (Saillenfest and Dessalles, 2014b). Il ne s'agit pas ici de lister l'ensemble des possibles conflits auxquels peut être confronté un individu, il nous faudrait pour cela dresser une liste des sources possibles d'évaluation des états de faits. On peut néanmoins citer les cas de conflits courants (ou classiques) suivants :

1. un état désiré mais dont la non-occurrence est crue (ou inférée)
2. un état non-désiré mais dont l'occurrence est crue (ou inférée)
3. un état de fait dont l'occurrence ainsi que la non-occurrence sont crues (ou inférées).

Du point de vue cognitif, la notion de support d'un état ne suppose pas qu'à un instant donné un ensemble important d'évaluations est pris en compte par les individus. Par ailleurs, il n'est pas nécessaire qu'un état de fait soit complètement instancié pour pouvoir être évalué (la croyance en la réalisation ou la non-réalisation de cet état peut être plus ou moins forte).

Il est important de noter que les conflits cognitifs sont des conflits internes à un individu, c'est donc une notion subjective (il ne s'agit pas de conflits entre individus). Les évaluations d'un état de fait particulier dépendent des connaissances, de l'expérience ou encore des désirs de l'individu qui évalue cet état. Ces évaluations peuvent être exprimées et confrontées à celles d'autres personnes, ce qui peut donner lieux à des conflits (ici on utilise le sens commun du mot) entre personnes ou des batailles argumentatives.

5.1.3 Évaluation des nécessités

Communément, un état de fait est dit nécessaire lorsqu'il ne peut pas ne pas être. Dans un sens proche, une situation est d'autant plus nécessaire que sa non-occurrence est inattendue. Cette dernière proposition nous fournit une possible évaluation de la nécessité d'une situation T :

$$\nu(T) = U(\neg T) \tag{5.3}$$

En utilisant le sens commun du mot 'nécessité', nous aurions tendance à dire que plus un état de fait est inattendu, moins cet état de fait est nécessaire. L'équation 5.3 nous permet de dire que $\nu(\neg T) = U(T)$ et, comme les évaluations sont cohérentes (voir équation 5.1), $\nu(\neg T) = -\nu(T)$. Ainsi, une autre possibilité pour évaluer une nécessité d'un état de fait T est :

$$\nu(T) = -U(T) \tag{5.4}$$

Comme nous l'avons dit plus haut, une autre source d'évaluation d'un état peut provenir d'un désir lié à la perspective d'occurrence de cet état. Ainsi, une autre possible évaluation de nécessité pour l'état T est :

$$\nu(T) = \epsilon(T)(E(T) - U(T)) \tag{5.5}$$

si l'occurrence de T est désirée, $\epsilon(T) = 1$ et la nécessité est positive, sinon $\epsilon(T) = -1$ et la nécessité est négative (rappelons ici que E désigne une intensité émotionnelle (voir l'équation 4.5), on a $E \geq 0$).

On appellera 'intensité émotionnelle perspective' d'une situation T la valeur $|E(T) - U(T)|$. $U(T)$ correspond ici essentiellement à un inattendu causal, la complexité de description de T rend compte d'une éventuelle ambiguïté qui pourrait subsister dans la description de T. Si la situation est complètement définie, $U(T) = C_w(T)$. L'inattendu vient donc des 'incertitudes' liées à la production de la situation T (de la complexité pour produire T). D'après la formule 4.5, plus T est simple à décrire, plus l'intensité de E augmente. L'intensité émotionnelle perspective (et donc la nécessité) d'une situation augmente à mesure que la situation est simple à décrire.

Cette évaluation traduit, en particulier, l'idée qu'une situation désirée est d'autant moins nécessaire qu'elle est inattendue.

Prenons un exemple pour illustrer le calcul des nécessités. Un individu désire avoir un million d'euros sur son compte en banque (proposition que nous notons T_{1MEur}), l'émotion que cette éventualité peut susciter chez lui est généralement assez faible, car cette situation est généralement très inattendue : $\nu_1(T_{1MEur}) = E(T_{1MEur}) - U(T_{1MEur})$ est faible. Par ailleurs, l'individu est quasiment certain qu'il ne possède pas ce million d'euros, donc $\nu_2(T_{1MEur}) = U(\neg T_{1MEur}) \approx 0$. On peut aussi dire que, pour cet individu, il serait très inattendu d'avoir ce million d'euros donc on a une troisième évaluation $\nu_3(T_{1MEur}) = -U(T_{1MEur}) << 0$. ν_1 et ν_3 sont les deux seules nécessités qui ont la possibilité de mener à un conflit cognitif. Dans la plupart des contextes dans lequel l'individu se trouve, il n'y a pas de conflit, car la situation T_{1MEur} est trop inattendue et donc ν_1 est faible. Si le million d'euros est 'à portée de main' de l'individu, ν_1 augmente et ν_3 diminue, quand ν_1 et ν_3 sont simultanément dans le support de T_{1MEur}, un conflit est créé autour de cet état.

5.1.4 Quelques considérations au sujet de la résolution des conflits

Une fois un conflit détecté vient une phase de diagnostic. Cette phase vise à fournir une explication de l'état conflictuel détecté ou à calculer un plan pour résoudre le conflit si celui-ci est anticipé. Dans les deux cas, il s'agit de calculer une hypothèse qui explique l'incohérence du support des états sources du conflit ou va correspondre au plan que l'on va réaliser dans le but de résoudre le conflit.

Soit un conflit autour d'un état T désiré mais dont l'occurrence est inattendue (suffisamment pour qu'il y ait conflit) et H une hypothèse causale. C'est un conflit que l'on cherche à résoudre en calculant un plan (représenté par une hypothèse) qui permet de produire l'état désiré.

Nous pouvons évaluer l'occurrence de l'état T sous l'hypothèse H : $\nu(H * T) = -U(H * T)$. Nous souhaitons que l'hypothèse résolve le conflit. Une bonne candidate doit donc faire baisser l'intensité de cette évaluation (c'est-à-dire rendre l'occurrence de T moins inattendue). L'hypothèse H doit donc permettre une

évaluation pour T qui fait sortir l'évaluation de T sous l'hypothèse H du support de T.

En utilisant la loi de composition de la complexité de génération (voir équation 4.8), et sachant que la complexité de génération d'une situation est égale à la complexité d'une hypothèse causale qui la génère (voir équation 4.12), la meilleure hypothèse causale pour résoudre le conflit doit vérifier :

$$H = argmin_k[C_w(H_k) + C_w(T||H_k)] \tag{5.6}$$

C'est là encore une version du rasoir d'Occam. Parmi toutes les hypothèses causales, il faut privilégier 'la plus simple qui nous rapproche le plus du but à atteindre' (il y a une double optimisation).

5.2 Le conflit narratif : un ensemble de conflits cognitifs

L'objectif de ce chapitre est de donner une définition cognitive, en termes de complexité, du conflit narratif. Nous avons introduit précédemment les notions de nécessité et de conflits cognitifs, ce qui nous a permis de faire le lien entre calculs de complexité et détection de conflits internes aux individus. Il nous faut maintenant montrer les liens entre conflits internes à un individu et la notion de conflit narratif.

Comme nous le rappelions en introduction de ce chapitre, le conflit narratif est défini comme l'incompatibilité entre les objectifs de deux personnages ou forces. Au cœur de la notion de conflit narratif se trouve l'idée qu'un ensemble d'actions intentionnelles de personnages (planifiées dans le but d'être réalisées) pour atteindre un ou plusieurs objectifs se trouvent contrecarrées par des événements inattendus (des obstacles). Ces événements peuvent être des actions d'autres personnages ou non. Il est classique de distinguer plusieurs sortes de conflits en fonction des forces en présence. Classiquement, on distingue trois formes de conflits narratifs :

1. le conflit entre personnages, c'est-à-dire un conflit qui oppose au moins deux protagonistes dont les objectifs sont contradictoires. Un exemple typique est

le conflit "héros-*nemesis*" (Batman contre le Joker ou encore l'affrontement final entre les trois personnages principaux dans "Le Bon, la Brute et le Truand").

2. le conflit entre un personnage et la nature (c'est-à-dire entre un personnage et des éléments non contrôlés par une entité rationnelle appartenant au monde fictif). Dans ce cas, ce sont des *world events* qui viennent perturber la réalisation des plans d'un ou plusieurs protagonistes du récit (un exemple typique est l'arrivée d'une tornade dans le film Twister).

3. le conflit interne à un personnage. Dans ce type de conflit, le personnage est opposé à lui-même, il doit faire face à des incohérences internes. Un exemple classique de conflit est le dilemme moral. On peut citer comme exemple ce dilemme moral classique : un protagoniste doit décider s'il se met en danger pour sauver un autre protagoniste.

Les conflits entre personnages ou entre un personnage et le monde fictif sont souvent considérés comme externes. Or ces conflits peuvent être ramenés à un ensemble de conflits internes aux protagonistes impliqués dans ces conflits. Prenons l'exemple du duel au pistolet entre deux personnages. Aucun des protagonistes impliqués ne souhaite mourir, mais les chances de mourir sont suffisamment élevées pour au moins l'un d'entre eux, ce qui, chez ce protagoniste au moins, provoque l'apparition d'un conflit cognitif interne. Dans la plupart des duels, chacun des personnages se met suffisamment en danger pour que la situation provoque l'apparition d'un conflit cognitif chez chacun d'eux. D'une manière générale, lorsqu'il y a conflit (dans le sens 'affrontement') entre deux protagonistes, on peut décomposer ce conflit narratif en deux conflits d'ordre cognitif dans l'esprit de chacun des protagonistes (qui sont ici traités comme seraient traités des être humains réels (voir la remarque faite p.27 à ce sujet)).

De la même manière, un conflit entre un personnage rationnel et une force quelconque non rationnelle du monde fictif est équivalent à un conflit cognitif interne à l'échelle du personnage.

Le conflit narratif interne à un personnage est également décomposable en un ensemble de conflits cognitifs, au sens défini plus haut, c'est-à-dire de conflits entre nécessités d'états du monde. L'exemple classique de conflit interne est le dilemme

moral. Par exemple, une héroïne doit mettre sa vie en danger pour sauver un protagoniste auquel elle est attachée. Ce dilemme est une conjonction de deux conflits : le premier dont la source est le danger qui pèse sur la vie de l'héroïne (c'est nécessaire pour sauver un autre protagoniste mais très peu nécessaire car non-désiré par l'héroïne), le second dont la source est le danger qui pèse sur la vie du protagoniste à sauver (ce danger est nécessaire car sa non-occurrence est inattendue - le protagoniste est en danger - et non-nécessaire car non-désiré par l'héroïne). Ainsi le dilemme moral est bien une conjonction de deux conflits. Il y a dilemme parce que la résolution de l'un empêche la résolution de l'autre.

Ainsi, d'une manière générale, nous proposons :

Un conflit narratif, c'est-à-dire l'incompatibilité entre les objectifs de deux personnages, **peut être assimilé à un ensemble de conflits d'ordre cognitifs** à l'échelle des protagonistes rationnels qui sont impliqués dans le conflit.

La figure 5.2 résume ce qui a été dit précédemment.

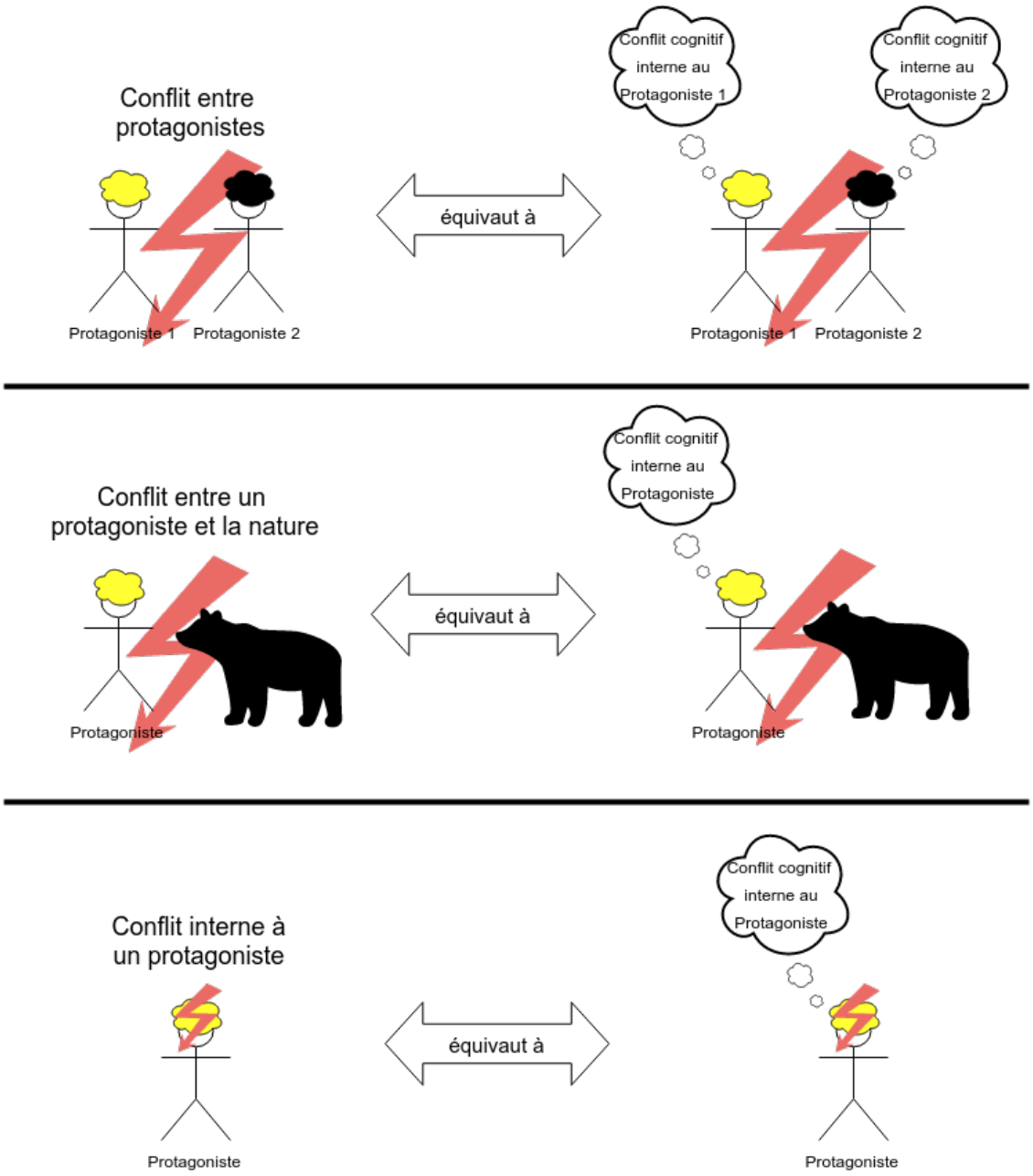
5.3 Le conflit narratif, un élément essentiel de la pertinence narrative

Dans ce court chapitre, nous avons mis l'accent sur la notion de conflit. La notion de conflit narratif est fondamentale dans l'acte de création d'un récit, c'est à la fois un élément qui crée et structure la réponse émotionnelle du récit, mais également un élément qui structure la construction événementielle du récit.

Comme le montre la figure 5.1 en introduction de ce chapitre, le récit est structuré par l'apparition de conflits narratifs et leurs résolutions. L'apparition et la résolution des conflits est source de tension narrative.

Yves Lavandier affirme que les conflits narratifs sont sources d'intérêt (Lavandier, 2004). La décomposition des conflits narratifs en conflits cognitifs permet de faire le lien entre le conflit narratif et les calculs de complexité introduits dans le chapitre 4. En particulier, les sources des conflits narratifs sont les croyances (plus ou moins fortes) et les désirs des personnages sur l'état du monde fictif dans lequel ils évoluent (car les nécessités sont évaluées *via* l'inattendu des états et l'émotion

FIGURE 5.2 – Décomposition des conflits narratifs en conflits cognitifs internes aux protagonistes impliqués



attachée à ces états).

L'occurrence d'une situation inattendue est surprenante (car inattendue), mais elle est aussi source de conflits cognitifs car elle provoque *a minima* un conflit interne à la connaissance. Les conséquences de cet événement inattendu peuvent provoquer d'autres conflits internes, par exemple parce que cela empêche un individu d'atteindre un objectif qu'il cherchait à atteindre ou encore parce que l'occurrence de la situation surprenante fait apparaître un danger qui n'existait pas auparavant.

Ainsi, un personnage confronté à l'occurrence d'une situation inattendue va être surpris (et, par procuration, l'audience aussi - voir page 31), et cela va provoquer l'apparition de conflits cognitifs internes à ce personnage qui vont former des conflits narratifs.

Le conflit narratif structure aussi la construction événementielle du récit. En effet, les personnages impliqués dans des conflits peuvent être amenés à résoudre les conflits en calculant tantôt une hypothèse explicative tantôt un plan (c'est-à-dire un ensemble ordonné d'actions) à réaliser. Cette seconde possibilité montre que le conflit, parce qu'il met les protagonistes du récit dans des situations d'inconforts va souvent les pousser à agir, ce qui va faire évoluer l'état du monde et donc faire progresser le récit. La résolution des conflits devant intervenir avant la fin du récit, cela signifie que la plupart des actions notables des personnages seront réalisées dans la seconde phase du récit. En cela, le conflit structure la répartition des événements (en particulier des actions des personnages) dans le récit.

Le conflit (narratif ou cognitif) est à la fois un élément qui contribue à la pertinence émotionnelle mais également à la pertinence structurelle des récits. C'est donc un élément fondamental de la pertinence narrative. Nous allons exploiter la position centrale du conflit pour prolonger notre étude du rôle joué par la complexité dans la pertinence émotionnelle ainsi que pour étudier le rôle de la complexité dans la pertinence structurelle des récits, et en particulier dans la reconnaissance des intentions des personnages.

Dans le chapitre 6, nous allons prolonger notre étude du rôle de la complexité dans la réponse émotionnelle à un récit en étudiant des modalités particulières de l'intérêt narratif : la surprise, la curiosité et le suspense. Nous proposerons par ailleurs un modèle de l'attitude des individus suite à l'occurrence d'une situation

surprenante, basé sur la notion de conflit cognitif.

Dans le chapitre 7, nous nous intéresserons à la pertinence structurelle des récits et en particulier à la résolution des conflits par les individus *via* la réalisation d'actions jugées intentionnelles.

Chapitre 6

Complexité et modalités de l'intérêt narratif

Comme nous l'avons vu, les conflits qui structurent la narration peuvent être ramenés à des conflits cognitifs chez l'audience. La source d'intérêt de ces conflits est liée à la notion d'inattendu comme modèle pour l'évaluation des états de faits.

Dans cette partie, nous allons nous intéresser à des modalités particulières de l'intérêt narratif : la surprise, la curiosité et le suspense. Nous allons proposer une étude de la surprise et des conséquences de l'occurrence d'un événement surprenant. Cela nous amènera à proposer un modèle de la surprise configuré par les conflits engendrés par l'occurrence d'un événement surprenant. Nous verrons que l'incertitude liée à la résolution de ces conflits est également créatrice de tension narrative, de curiosité et de suspense. Nous étudierons alors ces deux notions et nous proposerons une évaluation de l'intensité de la tension associée à la curiosité et au suspense.

Dans un dernier temps, nous monterons que les émotions ressenties par l'audience d'une séquence narrative le sont parfois par l'intermédiaire des personnages impliqués dans la séquence. Après avoir présenté quelques éléments théoriques au sujet des émotions ressenties par procuration, nous introduirons une notation pour rendre compte de l'aptitude d'un observateur à prendre le point de vue de divers personnages observés afin d'évaluer leur attitude. Cette notation sera abondamment utilisée dans la suite de la présentation de nos travaux.

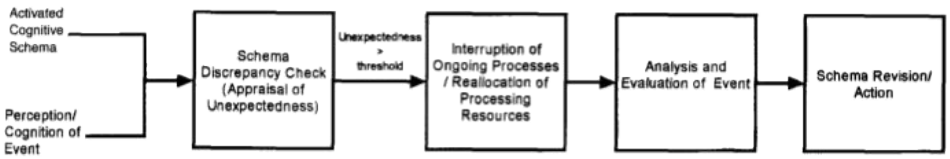
Note préliminaire : La définition de la surprise comme inattendu a été publiée dans :
Saillenfest, Antoine & Dessalles, Jean-Louis (2014). Can Believable Characters Act Unexpectedly ?, Literary and Linguistic Computing (2014) 29(4) : 606-620.

6.1 Le cas de la surprise

La surprise est une émotion de courte durée. La surprise est, par ailleurs, une notion subjective, ce qui surprend un individu ne surprendra pas nécessairement un autre individu. Enfin, la surprise est très souvent associée à la notion d'inattendu. *Common-sense psychology conceptualizes surprise as a peculiar state of mind, usually of brief duration, caused by unexpected events of all kinds* (Macedo et al., 2009).

Meyer *et al.* (1997) proposent un modèle de la surprise en quatre grandes étapes (voir le schéma reproduit figure 6.1). Le modèle de Meyer *et al.* (1997) ne se limite pas uniquement à la tension brève créée par la surprise mais également aux opérations mentales postérieures à la détection de l'événement surprenant. Les quatre étapes sont les suivantes : 1) un événement inattendu est identifié 2) l'émotion provoquée par la surprise est ressentie et le sujet focalise son attention sur l'événement surprenant 3) L'événement est analysé (*via* une analyse des causes et des conséquences) 4) le sujet met à jour son schéma mental (sa représentation du monde).

FIGURE 6.1 – *The cognitive-psychoevolutionary model of surprise* (Meyer et al., 1997) (version simplifiée extraite de (Reisenzein, 2000b))



Le modèle précédent rend compte de la volonté des individus d'intégrer les éléments surprenants dans leur représentation du monde, d'expliquer l'occurrence

de ces événements. Dans le cadre narratif, le critère de postdictibilité devant être respecté, les événements surprenants doivent être 'intégrés' à notre représentation du monde afin qu'ils ne soient plus surprenant rétrospectivement. Dans le monde réel, un individu peut échouer à trouver une explication à l'occurrence d'un événement surprenant.

Dans cette partie, nous allons analyser tout d'abord la détection d'un événement surprenant, et nous donnerons une définition de l'intensité émotionnelle associée. Nous montrerons que l'inattendu, défini comme un saut de complexité, permet de couvrir tous les cas possibles de surprise. Ensuite, nous montrerons que l'occurrence d'un événement surprenant est source de conflits cognitifs.

Ensuite, nous nous intéresserons à la résolution des conflits créés par l'occurrence d'un événement surprenant. La résolution de certains conflits consiste à expliquer l'occurrence de cet événement. D'autre conflits poussent le personnage à agir pour résoudre le conflit.

L'étude de l'attitude d'un individu suite à l'observation d'un événement surprenant nous amènera à présenter un modèle de cette attitude basé sur les conflits cognitifs qui apparaissent suite à cette observation.

6.1.1 Occurrence et évaluation d'une situation surprenante

D'après Ortony & Partridge (1987), il existe trois types d'événements surprenants : les événements rares, les erreurs d'anticipation ou encore les événements non anticipés. Qu'ils soient rares, mal-anticipés ou non-anticipés, nous allons montrer que l'inattendu, défini comme saut de complexité (voir équation 4.4), rend compte de la réponse émotionnelle d'un individu face à ces différents types d'événements surprenants.

La notion d'inattendu a souvent été mal définie. Dans le cas du modèle *cognitive-psychoevolutionary model of surprise* (Meyer et al., 1997), la notion d'inattendu est définie comme équivalente à la notion de *schema-discrepancy*. Selon la théorie des schémas (Schank and Abelson, 1977; Rumelhart, 1980), la perception, les pensées et les actions humaines sont contrôlées par un vaste ensemble complexe et organisé de structures de connaissance (ou croyances) appelées schémas. Ces schémas peuvent être considérés comme un ensemble de croyances sur les objets,

les événements (ou les séquences d'événements) et les situations. Les schémas serviraient à pouvoir interpréter les situations présentes, passées ainsi qu'à prédire l'occurrence d'événements futurs (et permettraient donc de prendre des décisions concernant nos actions futures).

La notion de divergence par rapport à un schéma attendu est généralement réduite à un saut de probabilités subjectives entre la probabilité de l'événement le plus probable et celle de l'événement qui se produit (Meyer et al., 1997). L'inattendu, comme *schema-discrepancy*, couvre les erreurs d'anticipation et, dans une moindre mesure, les événements rares. Les événements non anticipés, par définition, ne divergent pas d'un schéma attendu.

Nous proposons dans notre étude de définir la réponse émotionnelle associée à l'occurrence d'un événement surprenant comme étant le résultat du calcul de l'inattendu tel que nous l'avons défini dans le chapitre 4 (c'est-à-dire comme un saut de complexité cognitive (voir l'équation 4.4)) (Saillenfest and Dessalles, 2014a). Formellement :

$$SURPRISE(s) = U(s) \tag{6.1}$$

Nous allons tout d'abord montrer que l'occurrence d'une situation surprenante est associée à une augmentation, un saut, de complexité de génération d'une situation.

Tout d'abord, nous avons montré que, si nous disposons de la probabilité p d'occurrence d'une situation s, alors nous pouvons évaluer la complexité pour générer ladite situation *via* la formule $C_w(s) = -\log_2(p)$ (voir la partie 4.3.3). Ainsi, plus la valeur de p est faible, plus la situation est complexe à produire et plus elle est inattendue (d'après l'équation 4.4 qui définit U comme une fonction croissante de C_w).

On dispose cependant rarement de la probabilité d'occurrence d'une situation. Comment, par exemple, définir la probabilité qu'un pot que nous avons rempli de gâteaux hier soit vide aujourd'hui ? Dans l'histoire de Mary et John (voir page 42), quelle est la probabilité que Mary attaque John avec un couteau au moment où cette situation se produit ?

Le calcul de la complexité de génération d'une situation ne nécessite pas de

connaître la probabilité objective d'occurrence de la situation. On peut définir des modes de calcul de la complexité de génération, par exemple en utilisant la notion de schémas ou toute autre représentation de la connaissance (et de la causalité).

La théorie des nécessités définit la nécessité d'une proposition par l'inattendu de sa négation. Plus une proposition est inattendue, plus la croyance sa négation est forte ($\nu(s) = U(\neg s)$). Ainsi, lorsqu'une situation est anticipée, elle nous apparaît comme étant très nécessaire (et donc $U(\neg s) >> 0$ [1]). Cette forte nécessité nous fournit une mesure de la surprise associée à la non-réalisation de la situation. On a donc $SURPRISE(\neg s) = U(\neg s)$ dans le cas de la non-réalisation de situations dont l'occurrence est anticipée mais qui ne se produisent pas (ce qui correspond donc à des erreurs d'anticipation).

Lorsque la situation est non-anticipée, cela signifie qu'elle n'est ni attendue ni inattendue avant sa réalisation. Sa réalisation provoque donc un saut de complexité de génération immédiat. On peut supposer que ce saut est maximal, la complexité de génération atteint un seuil (nous nous plaçons au moment de l'occurrence de l'événement surprenant, la complexité n'a pas éventuellement été réévaluée *via* un raisonnement ultérieur).

Ainsi, pour résumer :

— l'occurrence d'une situation dont la probabilité p est faible provoque un saut de complexité de génération $C_w = -\log_2(p) >> 0$.

— la non-occurrence $\neg s$ d'une situation s anticipée provoque une surprise de l'ordre de $U(\neg s)$, qui mesure la nécessité de s antérieure à la réalisation de $\neg s$. On a donc $SURPRISE(\neg s) = U(\neg s)$.

— l'occurrence d'une situation non-anticipée provoque une augmentation immédiate (et supposément maximale) de complexité de génération pour cette situation.

La seule prise en compte des sauts de complexité de génération ne peut pas expliquer le phénomène de la surprise. En effet, la surprise est une émotion subjective, elle dépend donc de notre expérience. Le calcul de la complexité de génération est subjectif, l'évaluation de la difficulté de l'occurrence de certaines situations varie d'un individu à l'autre, ce qu'illustre l'exemple de Teigen & Keren (2003) reproduit ci-dessous :

1. si $x >> 1$ alors $log_2(x) >> 0$

"*an astrophysicist may be extremely 'surprised' to learn that the moon is one billion years older than previously thought, whereas a layman may declare himself not surprised at all ; not because the new estimate was expected, but because it did not conflict with any previously held beliefs.*" (Teigen and Keren, 2003)

Ainsi, dans un cadre événementiel, une situation est plus surprenante pour un individu donné si les personnes impliquées sont des proches de cet individu, ou si les personnes impliquées sont simples à décrire pour une raison simple (par exemple si cela concerne le président des États-Unis). Dans le cadre narratif, les situations rares sont plus surprenantes, et donc intéressantes, si elles concernent les personnages principaux du récit, ou si elles se produisent dans des lieux simples (par exemple, dans le monde fictif de Star Wars, l'Étoile de la Mort est un lieu simple). D'une manière générale, cette contrainte de simplicité dans la fiction populaire va pousser de nombreux auteurs à situer leur action dans des lieux très simples (palais, capitales de pays, lieux extra-ordinaires (au sens littéral du mot), etc.), et vont faire aussi intervenir des personnages simples (des rois/reines, des personnages importants sur un plan politique, artistique, etc.).

6.1.2 L'événement surprenant source de conflits cognitifs

Supposons qu'un individu ouvre une boîte pour prendre un gâteau (il pense donc qu'elle contient des gâteaux), puis, une fois la boîte ouverte, il réalise qu'elle est vide. L'individu est donc surpris. Mais au-delà de la surprise, l'absence de gâteaux dans la boîte est source d'un conflit entre une croyance préalable et une observation directe. Voici comment traduire cela en termes de nécessités :

- **Avant l'événement surprenant** la situation "il y a des gâteaux dans la boîte" (notée g) est très nécessaire pour l'individu. L'absence de gâteaux $\neg g$ est très inattendue et donc $\neg g$ est très peu nécessaire. Une nécessité $\nu(\neg g) = -U(\neg g) << 0$ est donc affectée à $\neg g$.
- **La situation surprenante est observée.** Lorsque l'individu découvre que la boîte est vide, la surprise est de l'ordre de $U(\neg g)$. L'observation est source de nécessité, la situation $\neg g$ devient très nécessaire. Une nécessité $\nu'(\neg g) = U(g) >> 0$ est donc affectée à $\neg g$.
- **La situation surprenante est source de conflit** $\nu * \nu' < 0$. La situation

$\neg g$ est donc source de conflit cognitif. L'observation de $\neg g$ vient contredire une croyance antérieure à son occurrence, ce qui provoque une incohérence dans la représentation qu'à l'individu du monde et donc un conflit d'ordre cognitif.

D'une manière générale, la surprise est source de conflit car une situation auparavant très peu nécessaire le devient lorsque l'événement surprenant est observé. Cela crée un conflit entre une croyance préalable et une observation (c'est donc un conflit entre deux croyances). Ou en d'autres termes, ce qui semblait auparavant inattendu se produit. Ou ce qui était attendu ne se produit pas. Le cas des situations non-anticipées est un peu différent. Rétrospectivement, après l'occurrence de la situation surprenante non-anticipée, on peut supposer qu'un individu alloue une nécessité négative antérieure à la réalisation de l'événement. Autrement dit, l'individu ne s'attendait pas à l'occurrence de l'événement et se dit rétrospectivement que l'événement n'aurait pas dû se produire.

Le mécanisme général de formation d'un conflit au sein de la connaissance d'un individu suite à l'occurrence d'une situation s est donc le suivant (voir également la figure 6.2) :

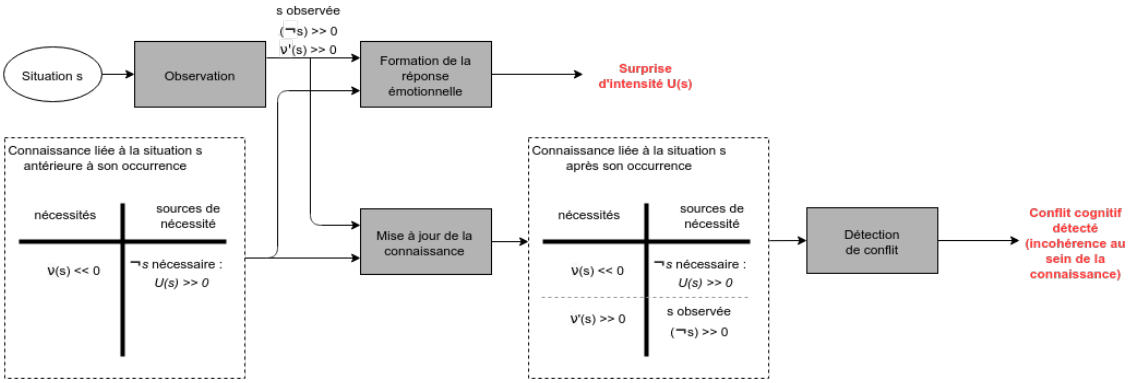
— Avant l'occurrence de s, la situation s est très peu nécessaire car inattendue et donc est affectée d'une nécessité ν négative.

— L'observation de l'occurrence de s rend s nécessaire. s est donc affectée d'une nécessité ν' positive.

— Les nécessités ν et ν' sont sources de conflit. Un conflit entre deux croyances apparaît. Ce conflit traduit une incohérence au sein de la représentation qu'a l'individu du monde.

Le mécanisme détaillé ci-dessus concerne une incohérence au sein de la connaissance, une situation que l'on pensait inattendue au préalable est observée. Une autre source de nécessité vient des désirs d'un individu. Or une situation inattendue qui se produit, et devient donc nécessaire, peut être non-désirée, et donc est affectée d'une nécessité négative. Si la situation inattendue en elle-même n'est pas non-désirée, elle peut néanmoins contribuer à rendre une situation non-désirée plus nécessaire qu'auparavant. Autrement dit, la situation surprenante peut être source d'une perspective de conflit portant sur un désir. Par exemple, si la foudre vient s'abattre sur la cabane dans laquelle se trouve un individu et que celle-ci

FIGURE 6.2 – Formation d'un conflit lié à une incohérence au sein de la connaissance d'un individu suite à l'occurrence d'une situation surprenante s



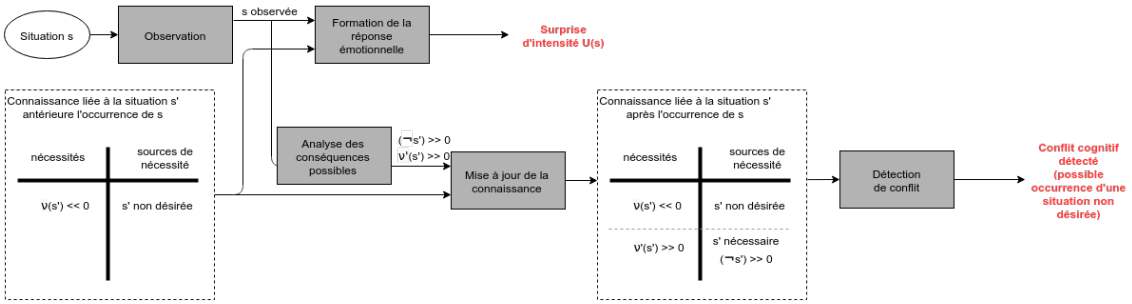
s'enflamme, alors cela va mettre la vie de l'individu en danger, or celui-ci ne veut pas mourir (il y a donc un conflit anticipé suite à l'occurrence de l'événement surprenant).

Formellement, l'occurrence de la situation s rend une situation s' non-désirée nécessaire. La formation du conflit est donc la suivante (voir également la figure 6.3) :

— s' est non-nécessaire car non-désirée, s' est donc affectée d'une nécessité ν négative.
— L'observation de l'occurrence de s rend s' nécessaire. s' est donc affectée d'une nécessité ν' positive.
— Les nécessités ν et ν' sont sources de conflit. Un conflit entre une croyance et un désir apparaît. Ce conflit correspond à la possible occurrence d'une situation non-désirée suite à l'occurrence de l'événement surprenant.

De même, si l'occurrence de s éloigne la perspective d'occurrence d'un état désiré, alors cet état devient non-nécessaire car inattendu, mais est nécessaire car désiré, ce qui est (là encore) source de conflit cognitif entre une croyance et un désir.

FIGURE 6.3 – Formation d'un conflit lié à la possible occurrence d'une situation s' non-désirée par un individu suite à l'occurrence d'une situation surprenante s



Résolution des conflits

Les conflits consécutifs à l'occurrence d'un événement surprenant viennent d'une incohérence entre deux croyances ou une croyance et un désir. Ainsi, afin de retrouver un état stable, l'apparition d'un conflit est suivie d'une phase de raisonnement qui a pour but de résoudre le conflit.

Dans le cas d'un conflit entre deux croyances, ce raisonnement délibératif est de type diagnostic. La résolution du conflit consiste à fournir une explication, c'est-à-dire à trouver un ensemble de causes et d'effets qui permettent d'intégrer l'événement surprenant dans notre représentation du monde. Reproduisons l'exemple suivant qui illustre ce processus :

> "*For example, if you found your car key was missing, and you had no way of explaining it, then you might experience a high level of surprise. However, if a plausible explanation subsequently emerged that allowed the anomaly to be resolved, such as realising that you left the key inside the car, then the experience of surprise should subside.*" (Maguire et al., 2011)

La phase de résolution consiste à former une ou plusieurs hypothèses, plus ou moins valables (la meilleure étant celle à conserver). Dans l'exemple de Mary qui poignarde John (voir page 42), une explication possible est : Mary est une espionne et n'a jamais aimé John. La taille de l'hypothèse explicative mesure ce que nous définirons comme étant une forme de curiosité (voir partie 6.2.1). Le mot curiosité est utilisé dans un sens proche de son sens commun. Nous sommes

curieux de savoir ce qui s'est passé, ce qui a produit ce que nous avons observé. Plus l'explication semble fiable (semble être 'la bonne explication'), moins sa complexité de génération est élevée et moins l'événement est inattendu (donc surprenant) rétrospectivement. Dans ce cas, la curiosité peut donc être vue comme le reste de surprise associée à un événement après son occurrence (notons ici que ce n'est pas le sens donné à cette notion dans le cadre narratif, ce qui sera discuté dans la partie 6.2.1).

Dans le cadre narratif, l'audience est généralement maintenue pendant un temps variable dans cet état d'incertitude, mais le critère de postdictibilité doit être respecté et tous les événements doivent être intégrés à la fin du récit. L'auteur doit donc fournir tous les éléments nécessaires pour que, rétrospectivement, aucun événement du récit ne soit surprenant, l'audience doit être en mesure d'expliquer tout ce qui lui a été présenté rétrospectivement.

Dans le cas d'un conflit anticipé lié à la possible occurrence d'une situation non-désirée (ou, ce qui est équivalent, à la possible non-occurrence d'une situation désirée), la résolution du conflit consiste à imaginer un plan, c'est-à-dire une suite d'événements (généralement des actions à réaliser) qui permettront d'éviter (ou de s'éloigner de) l'occurrence de l'état non-désiré (ou de permettre l'occurrence de l'état désiré). Nous définirons la notion de suspense comme une incertitude liée à la résolution du conflit (voir partie 6.2.2).

Révision des connaissances et planification des actions

La première révision des connaissances intervient directement après l'occurrence de l'événement surprenant. Ensuite, l'analyse des possibles conséquences, de la possible occurrence de situations suite à l'occurrence de l'événement surprenant provoque également des changements dans notre représentation du monde.

Le choix d'une hypothèse satisfaisante (après son calcul) concernant une incohérence au sein de la connaissance d'un individu a pour conséquence une révision de ladite connaissance. Dans l'histoire de Mary et John (voir partie 42), supposons que notre meilleure hypothèse est "Mary est une espionne et elle n'a jamais aimé John". Cette hypothèse modifie plusieurs éléments de notre représentation du monde fictif, par exemple :

— "Mary aime John" devient "Mary n'aime pas John" (avec une certaine nécessité)
— "Mary est une espionne" est ajouté à la connaissance (avec une certaine nécessité)
— "John est blessé" (voire "John est mort") est ajouté à la connaissance (avec une certaine nécessité)

Attention, bien que l'hypothèse "Mary est une espionne et elle n'a jamais aimé John" soit considérée comme étant notre meilleure candidate, elle résout le conflit lié à la tentative de meurtre de John en créant d'autres conflits, par exemple un conflit autour de la situation "Mary est une espionne", qui est supposé mais inattendu *a priori*.

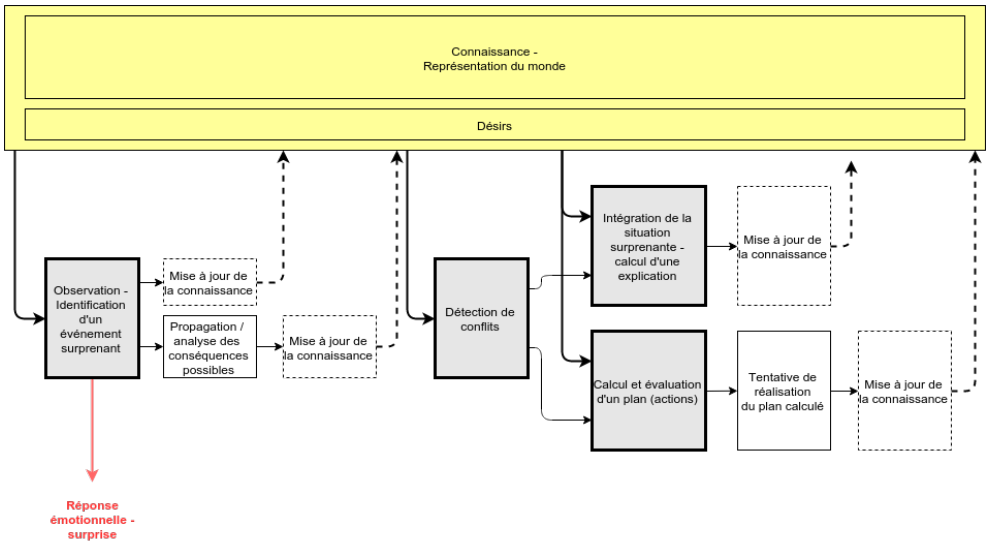
Dans le cas où l'événement surprenant fait apparaître la perspective d'un conflit anticipé, l'individu est amené à calculer un plan d'action pour éloigner la perspective de conflit. Il faut distinguer le calcul du plan et sa réalisation effective. Le calcul du plan permet d'anticiper les éventuelles conséquences de la réalisation du plan, l'humain est capable de simuler mentalement la réalisation du plan pour voir s'il résout effectivement le conflit sans créer une situation encore plus conflictuelle. S'il décide finalement de réaliser le plan, alors cela va modifier l'état du monde. Si le plan n'est pas réalisé dans sa totalité ou si la simulation de la réalisation du plan est erronée (si l'individu a oublié de prendre en compte des conséquences de ses actions futures par exemple), alors les changements dans la connaissance peuvent être différents de ceux qui ont été simulés.

6.1.3 Un modèle de la surprise comme source de conflits

L'analyse ci-dessus nous a permis de montrer que l'occurrence d'un événement surprenant est source de conflits. Il peut s'agir d'incohérences internes liées à une proposition de la base de connaissances de l'individu (*ie.* des conflits entre croyances), de conflits liés à l'occurrence (ou l'occurrence attendue) d'états non-désirés ou encore de conflits liés à la non-occurrence (ou la non-occurrence attendue) d'états désirés (*ie.* des conflits entre croyances et désirs).

L'étude de l'occurrence d'un événement surprenant nous amène à présenter un modèle de la surprise basé sur la notion de conflits (voir figure 6.4).

FIGURE 6.4 – Modèle de la surprise configuré par les conflits cognitifs



6.2 Curiosité et suspense

L'étude de la surprise nous a amené à évoquer les notions de curiosité et de suspense. La surprise, la curiosité et le suspense sont trois modalités de la tension émotionnelle (Sternberg, 1992; Brewer, 1996; Baroni, 2007). Dans un cadre narratif, Brewer (1996) définit ainsi la curiosité :

> "*Curiosity. We hypothesize that curiosity is produced by including a crucial event early in the event structure. In a text with curiosity structure (unklike a surprise discourse) the discourse contains enough information about the earlier events to let the reader know that the information has been omitted from the discourse. This type of discourse organization causes the reader to become curious about the omitted information, and the curiosity is resolved by providing enough information later in the text to allow the reader to reconstruct the missing event*" (Brewer, 1996, p.112)

Voici un exemple de séquence narrative configurée par la curiosité (repris de (Baroni, 2007, p.120) et adapté) :

— *On est dans le cadre d'un duel entre un héros et un méchant.*

— Le spectateur entend les tirs. Le résultat du duel est dissimulé.

— Des indices trompeurs sont fournis au spectateur.
— Le méchant s'écroule.

Par ailleurs, Brewer (1996) définit ainsi le suspense :

"*Suspense. We hypothesize that suspense is produced by including an initiating event or situation in the underlying event structure. An initiating event is an event that has the potential to lead to a significant outcome (good or bad) for one of the main characters in the narrative. In addition, we hypothesize that the event structure must contain the outcome of the initiating event. [...] The initiating event causes the reader to become concerned about the potential consequences for the character, the intervening material prolongs the suspense, and the eventual occurrence of the outcome resolves the suspense.*" (Brewer, 1996, p.113)

Voici un exemple de séquence narrative configurée par le suspense (repris de (Baroni, 2007, p.120) et adapté) :

— Le héros et le méchant se provoquent en duel.
— Le duel commence, le spectateur attend que les tirs aient lieu.
— Les tirs se produisent.

6.2.1 Le cas de la curiosité

Nous souhaitons ici rendre compte du sens commun du mot curiosité. Nous allons nous intéresser à la curiosité créée par l'occurrence d'une situation, passée ou futur. Nous allons étudier l'hypothèse suivante, qui porte sur les sources de curiosité et les facteurs qui influencent le niveau de curiosité :

La curiosité d'une situation s apparaît lorsqu'une source de nécessité fait apparaître une nécessité $\nu(s)$ qui crée un conflit cognitif autour de cette situation, l'intensité de la curiosité $CURIOSITE(s)$ liée à l'occurrence de cette situation est égale à l'intensité de cette nécessité.

D'après Baroni (Baroni, 2007, p.108), la curiosité (comme modalité de l'intérêt narratif) est fondée sur une incertitude concernant "ce qui s'est passé", c'est-à-dire qu'elle est basée sur un raisonnement de type diagnostic. Baroni (2007) définit ainsi

le diagnostic :

> « *diagnostic :* anticipation incertaine, à partir d'indices, de la compréhension d'une situation narrative décrite provisoirement de manière incomplète » (Baroni, 2007, p.110)

Dans son sens commun, par opposition au sens 'narratif', le mot curiosité ne renvoie pas nécessairement à l'occurrence d'une situation passée. Un individu peut être curieux de savoir si un événement va se produire. Dans son sens général, la notion de curiosité renvoie à une volonté de connaître (ou de savoir).

Soit s une situation, dont l'occurrence passée a été observée par un individu (s est complètement définie par cet individu). Si s n'est pas source de conflit cognitif pour cet individu, c'est que l'occurrence de s n'a pas créé d'incohérence dans la représentation du monde de cet individu, ce qui s'est produit était 'attendu' et l'occurrence de s n'est pas source de curiosité.

Supposons que s est une situation future, dont l'occurrence éventuellement incertaine est anticipée par un individu, et qui n'est pas source de conflit cognitif pour cet individu. La curiosité traduit un désir (voire un besoin) de savoir, ce qui suppose un bénéfice lié à ce savoir. Si l'occurrence incertaine de s n'est pas conflictuelle, c'est qu'elle ne représente pas d'enjeu émotionnel ni ne risque d'affecter la cohérence de notre représentation du monde. Il n'y a donc ni désir ni besoin de savoir si s va se produire, s n'est pas source de curiosité.

Nous défendons donc l'idée que la curiosité est liée à un conflit (éventuellement une perspective de conflit) autour d'une situation.

Soit s une situation dont l'occurrence, observée par un individu, est source de conflit cognitif pour cet individu. L'occurrence a été observée, donc l'occurrence est passée. L'occurrence de s est source de conflit parce que s était inattendue *a priori* (ce qui est source de nécessité $\nu(s) = -U(s)$, l'occurrence de s est *a priori* non-nécessaire) mais a été observée (s est nécessaire *a posteriori*). L'individu, surpris par l'occurrence de s, est curieux de savoir comment s a pu se produire, et ce tant qu'*a posteriori* l'inattendu de l'occurrence de s reste positif. Si un individu entre chez lui et voit un trou dans son plafond, il est tout d'abord surpris puis ensuite curieux de savoir ce qui s'est passé, ce qui a provoqué l'occurrence de cette

situation surprenante. Tant qu'il n'a pas trouvé une explication qui fait disparaître le conflit autour de l'occurrence de cette situation, l'individu est curieux de savoir ce qui s'est passé. À mesure qu'il trouve des indices qui lui permette de trouver une meilleure hypothèse, sa curiosité diminue jusqu'à disparaître lorsqu'il a trouvé une hypothèse satisfaisante (une hypothèse qui fait disparaître le conflit autour de l'occurrence de la situation surprenante). Ainsi, dans ce cas, le niveau de curiosité associé à l'occurrence de la situation surprenante s correspond à $U(s)$, c'est-à-dire à $|\nu(s)|$.

Notre hypothèse concernant les sources et l'intensité de la curiosité nous permet de prédire : plus l'occurrence passée d'une situation, source de conflit, est inattendue pour un individu, plus la curiosité de cet individu à propos de l'occurrence de cette situation est grande. En particulier, cela prédit que l'intensité de la curiosité est accrue lorsque les situations observées (par un individu et qui sont sources de curiosité pour cet individu) se produisent dans des lieux simples ou impliquent des personnes simples (pour cet individu).

Considérons maintenant le cas d'une situation s dont la perspective d'occurrence est source de conflit cognitif. s se produira peut-être, dans le futur. Une source de nécessité $\nu(s)$ de s est son intensité émotionnelle perspective (voir 5.1.3). Cette intensité émotionnelle perspective représente un enjeu, un gain ou une perte possible. Cet enjeu pousse les individus à chercher à savoir ce qui va se passer. Nous pensons que plus l'intensité émotionnelle perspective est forte, plus un individu est curieux de savoir ce qui va se passer. Ainsi, un individu qui peut gagner une somme d'argent suite à la possible signature, dans quelques heures, d'un contrat sera plus curieux de savoir si le contrat va être signé si la somme en jeu vaut un million d'euros que si la somme vaut cent euros. De même, si la signature du contrat peut faire perdre une somme d'argent, la curiosité est plus grande si la somme vaut un million d'euros que si elle vaut cent euros. La curiosité dépend aussi de la simplicité de la situation dont la perspective d'occurrence crée un conflit. Un individu sera plus curieux si sa mère risque de perdre mille euros que si l'un de ses cousins (moins simple à décrire généralement) risque de perdre mille euros. Ainsi, dans le cas de la perspective d'occurrence d'une situation s, source de conflit cognitif, l'intensité de la curiosité liée à s correspond à l'intensité émotionnelle perspective

de s, c'est-à-dire à $|\nu(s)|$.

Notre hypothèse concernant les sources et l'intensité de la curiosité nous permet de prédire : plus l'occurrence future d'une situation source de conflit cognitif est nécessaire, (*i.e.* plus l'intensité émotionnelle perspective de cette situation est élevée) pour un individu, plus la curiosité de cet individu à propos de l'occurrence future de cette situation est grande. En particulier, cela prédit que l'intensité de la curiosité est accrue lorsque les situations anticipées (par un individu et sources de curiosité pour cet individu) se produisent dans des lieux simples ou impliquent des personnes simples (pour cet individu).

Traitons maintenant le cas d'un événement (observé) qui a produit une situation partiellement connue (non observée). L'individu sait que cette situation s'est produite mais ne l'a pas observée. Ceci correspond à la situation dans laquelle se trouve le spectateur dans l'exemple de séquence configurée par la curiosité proposé par (Baroni, 2007, p.120) ci-dessus : il y a un duel, les tirs ont eu lieu, le spectateur les a entendus, le spectateur sait que cet événement a provoqué l'occurrence d'une situation 'indéterminée'. La situation qui s'est produite est 'indéterminée', mais on suppose que l'audience est capable de déterminer un ensemble fini de situations possibles qui peuvent correspondre à la situation dont l'occurrence a eu lieu. L'incertitude liée à l'occurrence de chacune de ces situations crée un conflit cognitif dès lors qu'il s'agit d'une situation affectée d'une intensité émotionnelle perspective (source de nécessité). Après avoir entendu les tirs, il y a une perspective de conflit autour de la mort du héros (non-désirée mais son occurrence est 'possible' ('suffisamment' nécessaire)), et autour de la mort de l'ennemi (désirée mais non-nécessaire). Chacune des situations de l'ensemble des possibles situations qui est affectée d'une intensité émotionnelle perspective est à l'origine d'un 'conflit anticipé' (ou pour être plus précis le conflit porte sur la future observation de la situation, qui hérite de l'émotion de la situation elle-même). L'occurrence de la situation est passée mais l'observation de la situation est future. Il y a donc bien un conflit anticipé autour de l'observation de la situation. D'après ce qui a été dit précédemment, l'intensité de la curiosité associée à chacune des possibles situations produites qui sont affectées d'une nécessité héritée de leur intensité émotionnelle perspective est égale à la valeur absolue de cette nécessité.

Notre hypothèse concernant les sources et l'intensité de la curiosité nous permet de prédire : plus l'observation future de l'occurrence passée d'une situation source de conflit cognitif est nécessaire (*i.e.* plus l'intensité émotionnelle perspective liée à l'observation de cette situation est élevée) pour un individu, plus la curiosité de cet individu (à propos de l'observation future de l'occurrence passée de cette situation) est grande. En particulier, cela prédit que l'intensité de la curiosité d'un individu est accrue lorsque les situations passées possibles (sources de curiosité pour cet individu) se produisent dans des lieux simples ou impliquent des personnes simples (pour cet individu).

Si on reprend l'exemple du duel entre un héros et un méchant, et si on suppose que les tirs ont été entendus, alors tout indice tendant à nous laisser penser que le héros ou le méchant a été tué fait monter la curiosité respective liée à l'occurrence des situations possibles "Le héros est mort" ou "Le méchant est mort". Ces indices éventuels rendent ces situations plus nécessaires (leur intensité émotionnelle perspective augmente car elles deviennent moins inattendues).

Pour résumer ce qui précède, nous avons émis l'hypothèse et donné des arguments en faveur d'une définition de la curiosité basée sur la notion de nécessité d'une situation. Nous avons proposé de définir ainsi la curiosité (le sens commun du mot) : la curiosité d'une situation s apparaît lorsqu'une source de nécessité fait apparaître une nécessité $\nu(s)$ qui crée un conflit cognitif autour de cette situation, l'intensité de la curiosité $CURIOSITE(s)$ liée à l'occurrence de cette situation est égale à l'intensité de cette nécessité :

$$CURIOSITE(s) = |\nu(s)| \tag{6.2}$$

Cette définition permet de réaliser des prédictions qui peuvent être testées (ce que nous n'avons pas eu l'occasion de faire mais peut être envisagée dans le cadre d'une poursuite de nos travaux).

6.2.2 Le cas du suspense

Le suspense est fondé sur une incertitude concernant une issue potentielle, conséquence de la situation actuelle, c'est-à-dire qu'il est basé sur un raisonnement

de type pronostic visant à déterminer l'issue des événements.

Baroni (2007) définit le pronostic de la manière suivante :

> « *pronostic :* anticipation incertaine d'un développement actionnel dont on connaît seulement les prémisses. » (Baroni, 2007, p.110)

Lorsque nous sommes confrontés à un cas de suspense, l'occurrence d'une situation s anticipée est incertaine ($U(s) > 0$). Par ailleurs, la non-occurrence de s est également possible et incertaine ($U(\neg s) > 0$). Le suspense s'accompagne généralement d'attentes quant à la résolution de la phase de suspense. Autrement dit, l'audience désire la réalisation (ou la non-réalisation) de l'une des conséquences possibles de l'action en cours.

Supposons que la situation s est désirée, alors cette situation reçoit deux évaluations de signes opposées. Le suspense nous place, chronologiquement, avant l'occurrence de la situation qui terminera la phase de suspense. La possibilité de réalisation de s est donc quantifiée par $U(s)$. À un instant donné, on peut évaluer la réalisation possible de s en calculant l'hypothèse H la plus simple qui expliquerait le mieux l'occurrence future de s ($H = argmin_{Hyp}[C_w(Hyp) + C_w(s||Hyp)]$). L'inattendu de s, ainsi que le suspense lié à l'occurrence possible de s, augmente donc avec la taille de la meilleure hypothèse (a un moment donné) pour générer s : $U(s) = C_w(H) + C_w(s||H) - C_d(s)$.

L'intensité du suspense dépend aussi de la valeur que l'on accorde à la situation, quantifiée par $E_h(s)$. Ainsi, plus s est désirée, plus la tension créée par le suspense est grande.

Nous proposons de définir la tension liée au suspense portant sur l'occurrence d'une situation anticipée s désirée comme l'intensité émotionnelle anticipée associée à la possible occurrence de s. Or d'après la formule 4.5, l'intensité émotionnelle associée à l'occurrence d'une situation est égale à $E_h(s) + U(s)$. Ainsi, par application directe de la théorie de la Simplicité, nous définissons le suspense associé à la possible occurrence d'une situation s désirée par la formule :

$$SUSPENSE(s) = E_h(s) + (C_w(H) + C_w(s||H) - C_d(s)) \tag{6.3}$$

Dans les phases de suspense, la taille de la meilleurs hypothèse causale aug-

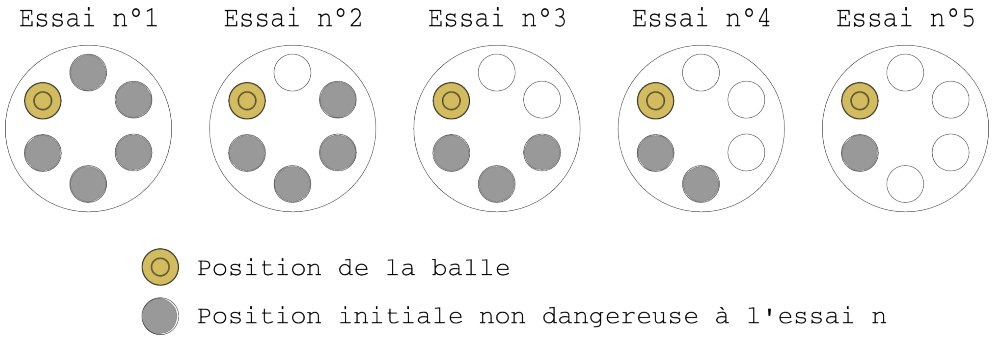
mente avec le temps à cause d'une limite de la phase de suspense qui a été présentée. Plus le temps passe, plus la limite est proche. Plus la limite est proche, plus la situation s semble difficile à produire et donc plus la taille de la meilleure hypothèse H (et donc $U(s)$ ainsi que $SUSPENSE(s)$) augmente.

Analysons l'exemple suivant (inspiré d'un épisode du roman "Pétronille" d'Amélie Nothomb) :

Histoire de roulette russe
L'héroïne de l'histoire est désespérée. Dans un moment de folie, elle décide de mettre sa vie en danger en jouant à la roulette russe. Elle place une balle dans le barillet puis le fait tourner.
L'héroïne appuie une première fois sur la gâchette. Rien ne se passe.
L'héroïne appuie une seconde fois sur la gâchette. Rien ne se passe.
L'héroïne appuie une troisième fois sur la gâchette. Rien ne se passe.
L'héroïne appuie une quatrième fois sur la gâchette. Rien ne se passe.
L'héroïne appuie une cinquième fois sur la gâchette ...

Cet exemple est configuré de sorte à ce que la tension due au suspense augmente au fur et à mesure des essais. À chaque essai, deux issues sont possibles, soit l'héroïne meurt (situation m) soit elle ne meurt pas (situation $\neg m$). Tant que le jeu n'est pas terminé, ces deux issues sont possibles, chacune étant incertaine. Cependant, l'issue $\neg m$ est désirée (le lecteur ne veut pas voir l'héroïne mourir). Ainsi, il nous faut calculer l'évolution de $U(\neg m)$ au cours de cet épisode pour voir l'évolution de la tension narrative.

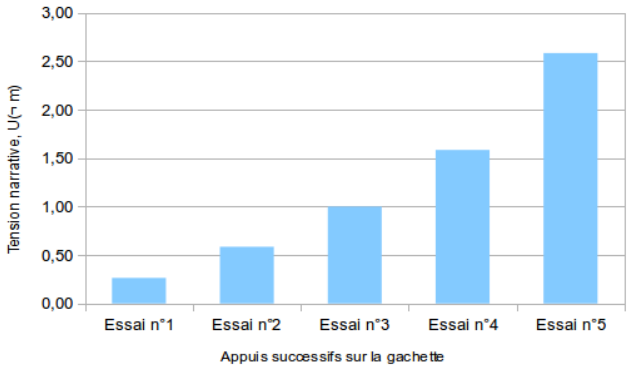
FIGURE 6.5 – Illustration de la roulette russe



À chaque essai, l'inattendu de la situation $\neg m$ dépend de la taille de l'hypothèse minimale qui permet de produire cette situation. Dans le cas présent, à l'essai numéro n, la situation $\neg m$ reste vraie si et seulement si le barillet se trouve, à l'état initial, dans une position telle qu'aucun des appuis successifs ne provoque le tir de la balle. Ainsi, comme on le voit sur la figure 6.5, lors du premier appui sur la gâchette, 5 positions du barillet sur les 6 possibles ont pour conséquence $\neg m$. Lors du second appui sur la gâchette, 4 positions initiales sur 6 ont pour conséquence $\neg m$. Ainsi, la position initiale du barillet étant déterminée aléatoirement, cela coûte $\log_2(6)$ bits de génération. Pour déterminer une position initiale non dangereuse à l'essai n, cela coûte $\log_2(n)$ bits de description. Ainsi, lors du n^{eme} appui sur la gâchette , $U(\neg m) = \log_2(6) - \log_2(6-n) - C_d(\neg m)$ bits (et dans le même temps $U(m) = \log_2(6) - \log_2(n) - C_d(m)$ bits).

On peut donc tracer l'évolution de la tension associée à cet épisode (voir 6.6)

FIGURE 6.6 – Évolution de la tension narrative dans l'histoire de la roulette russe



L'inattendu de la situation anticipée évolue au cours temps. Si un film montre une bombe cachée près d'un groupe de personnes et sur laquelle se trouve une minuterie qui affiche le décompte des secondes restantes avant l'explosion de la bombe. Plus le temps passe, moins les individus qui se trouvent près de la bombe ont la possibilité de fuir le danger. Ainsi, l'état du monde dans lequel les individus proches de la bombe sont sains et saufs après l'explosion apparaît comme étant de plus en plus non-nécessaire (l'inattendu de cet état augmente) pour le spectateur et donc la tension liée au suspense augmente à mesure que les secondes passent.

Analysons cet autre exemple : un personnage important de l'histoire est menacé de mort par un second personnage. Ce second personnage décide de jouer à pile ou face la survie du personnage principal. Il sort une pièce et la lance. Si elle retombe sur pile, le personnage important aura la vie sauve, sinon il se fait tuer sur le champ.

Imaginons que la scène est filmée et que l'on nous montre la pièce en train de tourner dans les airs au ralenti. Repousser le dévoilement de l'issue du pile ou face fait monter la tension du suspense. Cependant l'incertitude (objective) de l'occurrence de chacune des issues possibles ne varie pas au cours du temps, chacune des issues possibles à une probabilité égale à 50% de se produire. Cela montre que les individus ne tiennent pas compte des probabilités objectives d'occurrence.

Notre hypothèse est que le ralenti introduit une forme de granularité dans la causalité. L'action pile-ou-face a deux conséquences possibles, chacune des deux pouvant se produire avec une probabilité égale à 50%. L'action du pile-ou-face montrée au ralenti est une succession d'événements (des tours de pièces en l'air) qui sont autant d'étapes qui augmentent, de manière subjective, la complexité causale pour générer chacune des issues (et donc leur incertitude). Le spectateur passe donc d'un raisonnement sur une loterie (50% de chances de faire pile, 50% de chances de faire face) à un raisonnement sur la causalité. Plus le temps passe, plus l'issue du pile ou face approche et moins la pièce a la possibilité de faire des tours complets (ce genre de scène n'est jamais montrée en plan large, le spectateur sait que la pièce va retomber sur une surface mais ne voit pas la distance à la surface restante, il ne peut donc pas anticiper le résultat en utilisant cette distance). À chaque apparition de la face désirée (le côté pile) pendant que la pièce est en l'air, l'hypothèse causale la plus simple augmente et le suspense aussi.

6.3 Pertinence émotionnelle et émotions par procuration

Les agents rationnels sont capables de se 'projeter dans l'esprit d'autres agents', autrement dit de réaliser des opérations mentales en prenant les valeurs supposées de ces agents, l'évaluation supposée de l'état du monde par ces agents. On dira

également qu'un agent prend le point de vue d'un autre agent. C'est une version très primaire de ce que l'on appelle communément la théorie de l'esprit (Baron-Cohen, 1999).

6.3.1 Identification aux personnages et empathie

Jusqu'ici, nous avons régulièrement dit que les calculs de complexité étaient subjectifs. La notion d'inattendu, en particulier, est une notion subjective.

Analysons l'histoire suivante : une personne marche dans la rue en regardant vers le ciel. Il y a, à quelques mètres devant elle, une plaque d'égout ouverte. La personne continue son chemin, ne voit pas la plaque d'égout ouverte et tombe dans le trou.

La personne qui tombe dans le trou est surprise, sa chute est inattendue de son point de vue. Pour un observateur de la scène, la chute dans le trou est peu inattendue. L'observateur sait que la plaque d'égout est ouverte, 'voit' la personne marcher vers le trou, 'voit' que cette personne n'a pas vu le trou et anticipe la chute de la personne. En revanche, l'observateur sait que la personne qui est tombée dans le trou a été surprise par sa chute. Du point de vue de l'intérêt, la chute, bien qu'attendue par l'observateur, est intéressante.

Une audience confrontée à un récit, fictionnel ou non, accepte généralement de 'suspendre de manière consentie son incrédulité' (*willing suspension of disbelief*) et se retrouve absorbée dans le récit. Cette absorption dans le récit a des conséquences importantes, en particulier l'audience semble partager les émotions d'un certain nombre de personnages auxquels elle est capable de s'identifier. Brewer (1996) définit ainsi la théorie de l'identification : "*the reader of fiction identifies with the character and thereby comes to feel the emotions that are being felt by the fictional character*". Cette théorie explique pourquoi, dans l'histoire qui précède, un observateur est affecté par la chute du personnage dans le trou. En se projetant dans l'esprit de la victime qui tombe dans le trou, l'observateur est capable d'utiliser l'évaluation supposée de l'état du monde par la personne qui tombe pour évaluer les émotions ressenties par cette personne.

Cependant, la théorie de l'identification ne permet pas d'expliquer pourquoi l'observateur ressent une tension de type suspense avant que la personne tombe

dans le trou. En effet, la personne qui tombe dans le trou ne sait pas *a priori* que la plaque d'égout est ouverte, cette personne ne fait donc pas l'expérience du suspense. Le suspense ressenti par l'observateur ne vient donc pas d'une possible identification au personnage dans ce cas.

Une théorie alternative concernant les émotions ressenties par le lecteur d'un récit est la théorie de l'empathie, résumée ainsi par Brewer : "*the reader feels emotions for fictional characters that are like those the reader would feel for non-fictional individuals in similar circumstances*". D'après cette théorie, l'observateur éprouve de l'empathie pour la personne qui s'apprête à tomber dans le trou, s'intéresse à ce qui peut lui arriver et ressent donc des émotions qui dépendent de l'attitude de cet observateur à l'égard de la personne qui tombe (si l'observateur apprécie ou non la personne).

Le fait que, dans le cadre narratif, l'audience ait une connaissance possiblement différente de celle des personnages n'empêche pas pour autant l'identification. Il semble en effet que l'audience soit capable de maintenir à jour divers états de la connaissance, le sien et celui des personnages qui évoluent dans le monde fictif. La représentation mentale de la connaissance qu'ont les personnages du monde fictif fait partie de la représentation mentale que l'audience a du monde fictif, et l'audience est capable d'adopter le point de vue d'un personnage en fonction des données à propos de ce personnage dont elle dispose (nous appelons représentation mentale un ensemble qui contient entre autres des connaissances sur l'état du monde, des désirs ou encore des connaissances sur la causalité dans le monde fictif).

6.3.2 Notations

Pour rendre compte des différents calculs de complexité selon le point de vue adopté par un observateur $\mathcal{O}$ d'une même situation, d'un même état de fait, nous allons adopter dans la suite de ce manuscrit la notation suivante :

— $F^{\mathcal{O}\to\mathcal{B}}$ désignera le calcul d'une fonction F par l'observateur $\mathcal{O}$ qui prend le point de vue d'un individu $\mathcal{B}$ (ce que l'observateur pense être l'état mental de l'individu $\mathcal{B}$ au moment du calcul).

— $F^{\mathcal{O}}$ désignera le calcul d'une fonction F par l'observateur $\mathcal{O}$ en utilisant son

propre point de vue (tout ce qu'il sait sur ce qu'il observe). On notera que $F^{\mathcal{O}} = F^{\mathcal{O}\to\mathcal{O}}$

— Par extension, $F^{\mathcal{O}\to\mathcal{B}\to\mathcal{C}}$ désignera le calcul d'une fonction F par l'observateur $\mathcal{O}$ qui prend le point de vue d'un individu $\mathcal{B}$ qui lui-même prend le point de vue d'un individu $\mathcal{C}$.

Dans la suite, on notera :

— $\mathcal{O}$ un observateur (en particulier nous considérerons que dans le cadre narratif une audience est un observateur du monde fictif).

— $\mathcal{A}$ un acteur (en particulier nous considérerons que dans un cadre narratif un personnage est un acteur).

En particulier, l'équation 4.4 de l'inattendu calculé par un observateur $\mathcal{O}$ qui utilise sa propre évaluation des états du monde devient :

$$U^{\mathcal{O}} = C_w^{\mathcal{O}} - C_d^{\mathcal{O}}$$

L'équation 4.4 de l'inattendu calculé par un observateur $\mathcal{O}$ qui prend le point de vue d'un personnage-acteur $\mathcal{A}$, devient :

$$U^{\mathcal{O}\to\mathcal{A}} = C_w^{\mathcal{O}\to\mathcal{A}} - C_d^{\mathcal{O}\to\mathcal{A}}$$

Se projeter dans l'esprit d'un autre individu représente certainement un coût cognitif qui affecte l'intensité de la réponse émotionnelle calculée en prenant le point de vue de cette individu. Nous pensons que nous sommes moins surpris si nous faisons l'expérience de la surprise en se mettant à la place d'un individu au lieu d'en faire l'expérience directement. Dans l'histoire de l'homme qui tombe dans le trou, l'individu est certainement plus surpris que nous. Cette 'taxe cognitive' dépend certainement de la complexité de l'individu dans l'esprit duquel se fait la projection (dans la suite, nous ne tiendrons pas compte de cette 'taxe cognitive' à payer pour prendre le point de vue d'un personnage).

6.4 Conclusion

Dans ce chapitre nous avons montré les liens qui peuvent exister entre calculs de complexité et les modalités classiques de l'intérêt narratif ainsi que la nécessité

de fixer une notation pour rendre compte de la subjectivité de ces calculs.

Face à un récit, l'audience semble très opportuniste. Tout ce qui provoque une émotion chez un personnage se traduit par une tension chez l'audience (en tout cas dès qu'il s'agit d'un personnage avec laquelle elle peut procéder à une forme d'identification). Ainsi, un auteur qui désire intégrer des éléments pertinents sur le plan émotionnel peut, au choix, intégrer des éléments qui créent une tension directe chez l'audience, ou indirecte (c'est-à-dire les éléments qui sont sources de tension ressentie par procuration).

Si on cherche à déterminer si un événement, ou l'occurrence d'une situation s, est pertinent sur le plan émotionnel, comme créateur de tension, il faut se demander si s peut être source d'intérêt, en particulier de surprise, de curiosité ou de suspense (ressenti directement ou par procuration).

Notons Per l'ensemble des personnages d'un récit et Π_e l'ensemble des éléments du récit qui composent la pertinence émotionnelle. Nous proposons alors que, pour une situation s :

$$s \in \Pi_e \Leftarrow \exists \mathcal{B} \in Per \cup \mathcal{O}, SURPRISE^{\mathcal{O}\to\mathcal{B}}(s) > 0 \tag{6.4}$$

$$s \in \Pi_e \Leftarrow \exists \mathcal{B} \in Per \cup \mathcal{O}, CURIOSITE^{\mathcal{O}\to\mathcal{B}}(s) > 0 \tag{6.5}$$

$$s \in \Pi_e \Leftarrow \exists \mathcal{B} \in Per \cup \mathcal{O}, SUSPENSE^{\mathcal{O}\to\mathcal{B}}(s) > 0 \tag{6.6}$$

Ces équations signifient que toute situation dont l'occurrence crée une tension de type surprise, curiosité ou suspense est un élément de la pertinence émotionnelle d'un récit.

L'analyse de la surprise et le modèle configuré par les conflits que nous avons proposé montre qu'une séquence configurée par la surprise peut être source d'action pour les personnages. Ces derniers peuvent en effet être amenés à résoudre des conflits anticipés suite à l'occurrence d'une situation surprenante et donc à former des plans et tenter de les réaliser. Comme nous l'avons dit dans le chapitre 5, le conflit est un élément central de la pertinence narrative, il participe à la fois à la création de tension mais pousse également les personnage à agir, ce qui en fait un

élément de pertinence structurelle.

La résolution des conflits anticipés consiste à calculer un plan puis à le réaliser. Les actions que réalisent les personnages doivent, du point de vue de l'audience, être perçues comme étant réalisées de manière intentionnelle afin que le personnage donne l'illusion de la rationalité. Dans la partie suivante, nous allons donc définir des critères de pertinence structurelle, en particulier des critères liés à la reconnaissance des intentions des personnages du récit par l'audience.

Chapitre 7

Intérêt et crédibilité des personnages : un dilemme ?

Un bon récit doit aussi apparaître vraisemblable pour celui qui le lit. En introduction, nous avons montré que les événements du récits doivent respecter des règles dans le monde fictif. En particulier, les actions des personnages doivent respecter ce que les personnages savent du monde mais également être perçues comme étant réalisées de manière intentionnelle. Comme nous l'avons déjà indiqué, un personnage de roman doit donner l'illusion de la rationalité, en particulier l'audience doit être capable de reconnaître ses intentions et éventuellement le juger moralement.

Dans certains cas, susciter de l'intérêt tout en maintenant l'illusion de la rationalité des personnages semblent être des objectifs difficiles à concilier, ce que nous avons nommé dilemme intérêt-vraisemblance (voir partie 2.3). Après avoir présenté un modèle cognitif du sens commun de la notion d'intentionnalité en utilisant les notions de nécessité et de complexité introduites précédemment puis détaillé les prédictions de ce modèle, nous verrons comment ce modèle permet de définir des critères de pertinence narrative à respecter dans la génération d'un récit.

Note préliminaire : La majorité des définitions introduites dans ce chapitre a été publiée dans :
Saillenfest, Antoine & Dessalles, Jean-Louis (2012). Role of Kolmogorov Complexity on Interest in Moral Dilemma Stories, In N. Miyake, D. Peebles & R. Co-

oper (Eds.), Proceedings of the 34th Annual Conference of the Cognitive Science Society, 947-952. Austin, TX : Cognitive Science Society

Saillenfest, Antoine & Dessalles, Jean-Louis (2013). Using Unexpected Simplicity to Control Moral Judgments and Interest in Narratives, In Mark Finlayson, Bernhard Fisseni, Benedikt Löwe, Jan Christoph Meister (Eds.), 2013 Workshop on Computational Models of Narrative, 214-227. Dagstuhl, Germany : Schloss Dagstuhl - Leibniz-Zentrum für Informatik. OASICS 32

Saillenfest, Antoine & Dessalles, Jean-Louis (2014). Can Believable Characters Act Unexpectedly ?, Literary and Linguistic Computing (2014) 29(4) : 606-620.

7.1 Le dilemme intérêt-vraisemblance

L'auteur de fiction doit produire des récits intéressants dans lesquels les personnages doivent donner l'illusion de la rationalité. L'introduction d'actions inattendues (pour l'audience) peut avoir un effet négatif sur la reconnaissance des intentions des personnages qui réalisent ces actions. Nous reproduisons ici l'histoire de Mary et John introduite dans le chapitre 2 :

> « Mary et John sont très amoureux l'un de l'autre. Leur amour n'a cessé d'augmenter au fil du temps. Comme d'habitude, Mary déjeune avec John. Elle se lève, va dans la cuisine, saisit un couteau, revient à table et poignarde John dans le dos. »

L'action de Mary est surprenante mais le lecteur ne peut pas expliquer pourquoi elle a agi de la sorte. Le lecteur ne réussit pas à déterminer les intentions de Mary. La rationalité du personnage est donc menacée, un lecteur peut imaginer que Mary est folle. Un lecteur peut aussi imaginer qu'elle est espionne, que son amour pour John est feint et qu'elle le tue aujourd'hui pour prévenir l'occurrence d'un attentat qu'il s'apprêtait à commettre. Du point de vue du lecteur, le personnage doit rester vraisemblable, et en particulier le lecteur va continuer à le supposer rationnel pendant un moment.

Un lecteur accepte d'être placé temporairement dans un état d'incertitude, mais

l'auteur doit veiller à donner des éléments d'explication qui vont progressivement lever cette incertitude ; c'est un cas particulier de postdictibilité (Kintsch, 1980, p.89). Dans le cas présent, la postdictibilité consiste à donner les éléments qui vont permettre de reconnaître les intentions de Mary et de restaurer sa crédibilité.

L'auteur est donc confronté à un dilemme, que nous avons appelé dilemme vraisemblance - intérêt (Saillenfest and Dessalles, 2014a) :

Dilemme vraisemblance - intérêt : l'auteur est tiraillé entre deux objectifs parfois contraires que constituent le maintien de la vraisemblance du monde fictif (et des personnages qui le peuplent) et la nécessité de créer de la tension narrative tout au long de la narration.

Jusqu'ici, nous nous sommes principalement intéressés aux effets des incertitudes sur la réponse émotionnelle de l'audience d'un récit. Nous avons montré que l'inattendu associé à des situations passées, présentes et anticipées est source de tension narrative chez l'audience. Nous allons donc maintenant étudier les effets de ces incertitudes, et donc le rôle de la complexité, dans le maintien de la vraisemblance du monde fictif (et en particulier des personnages fictifs) par un auteur de récit.

Nous verrons que ce modèle peut nous permettre de définir des critères de pertinence narrative à respecter lors du processus de génération d'un récit. En particulier, nous réinterpréterons le dilemme intérêt-vraisemblance à l'aune de ces modèles.

7.2 Conflits cognitifs et nécessité des actions

Dans le chapitre 5 nous avons introduit les conflits cognitifs et montré qu'ils sont source de l'action des personnages. Lorsqu'un être rationnel fait face à un conflit cognitif entre un désir et un élément de se représentation interne du monde (une croyance), il est amené à calculer un plan, c'est-à-dire un ensemble d'événements ordonnés (ou partiellement ordonnés), dont la réalisation permettrait de sortir de la situation d'inconfort (cognitif) créé par le conflit. La réalisation de ce plan va donc entraîner la réalisation d'actions qui sont intentionnelles du point de vue de

celui qui agit (cet acteur est supposé rationnel).

Une action est un événement qui est le fait d'un acteur. Nous faisons l'hypothèse qu'un acteur rationnel est capable d'anticiper un certain nombre de conséquences d'une action donnée dans un contexte donné. Dans la suite, nous adoptons la notation $S_a^{\mathcal{A}}$ pour désigner l'ensemble des conséquences anticipées d'une action a par l'acteur $\mathcal{A}$.

Une conséquence $s \in S_a^{\mathcal{A}}$ est, en tant que conséquence possible de a, moins inattendue une fois l'action réalisée. Ainsi, on a :

$$\forall s \in S^a; U(s||\neg a) - U(s||a) \geq 0 \tag{7.1}$$

où $U(s||\neg a)$ désigne l'inattendu d'un état de fait s lié à la non-réalisation de l'action a. Si l'on se place d'un point de vue chronologique, $U(s||\neg a)$ correspond à la valeur de $U(s)$ calculée avant la réalisation de l'action a (il peut être utile de préciser ici que a, en tant qu'action n'est pas une situation, ou un état de fait. a n'a pas d'étendue temporelle, a disparaît dès sa réalisation terminée).

La réalisation d'actions par un agent a pour but de résoudre des conflits cognitifs auxquels il est confronté. La réalisation d'une action a donc pour but de faire baisser l'inattendu de l'occurrence d'un état de fait désiré ou de prévenir l'occurrence d'un état de fait non-désiré.

Soit s une situation désirée source de conflit et a une action réalisée dans le but de résoudre le conflit associé à s. La nécessité de s avant la réalisation de a vaut $\nu(s||\neg a) = E(s) - U(s||\neg a)$. Et on a $U(s||\neg a) - U(s||a) \geq 0$.

Comme nous l'avons fait précédemment pour les états de faits (voir chapitre 5), nous allons attribuer aux actions des nécessités. On notera $\nu_s(a)$ la nécessité de l'action a héritée de la conséquence particulière s de a. Une action est d'autant plus nécessaire qu'elle favorise la réalisation d'un état de fait désiré. De manière symétrique, une action est d'autant moins nécessaire qu'elle favorise la réalisation d'un état de fait non-désiré.

D'après ce qui précède, pour une situation désirée s source de conflit et a une action réalisée dans le but de résoudre le conflit associé à s, on obtient : $\nu_s(a) = E_h(s) + U(s) - U(s||a) = \nu(s||\neg a) + U(s||\neg a) - U(s||a)$.

Plus généralement, nous définissons la nécessité d'une action a héritée d'une

conséquence s anticipée particulière de valence $\epsilon(s)$ ainsi :

$$\begin{gathered}\nu_s(a) = \epsilon(s)\,|E_h(s) + U(s) - U(s||a)| \\ (\epsilon(s) = 1 \text{ si } s \text{ est désirée, } \epsilon(s) = -1 \text{ sinon})\end{gathered} \tag{7.2}$$

Cette notion de nécessité est à la base de la définition de ce que nous définirons comme la reconnaissance des intentions attribuées à un acteur (voir section 7.3.1). Un agent rationnel est supposé agir de manière à favoriser la réalisation d'états qu'il désire. Ainsi, un agent rationnel est censé favoriser les actions qui, de son point de vue, sont nécessaires.

La nécessité étant une notion subjective, certains états de faits désirés par certains agents sont non-désirés par d'autres agents. Nous montrerons que ces différences individuelles sont source de jugements qui peuvent être très différents de la part de plusieurs entités rationnelles qui observent une même situation. Nous nous intéresserons dans la partie 7.4 à ces jugements, et en particulier nous donnerons une définition de la notion de responsabilité attribuée à un acteur.

Dans le chapitre 5, la nécessité des états de faits nous a permis de définir la notion de conflit cognitif. Un conflit cognitif lié à une situation se produit lorsque celle-ci reçoit des évaluations de signes opposés. Une action peut avoir à la fois des conséquences désirées, sources de nécessités positives, et non désirées, sources de nécessités négatives. Une action peut donc avoir une nécessité positive et une nécessité négative en même temps, ce qui est à l'origine de dilemmes d'ordre moraux. Nous aborderons la question du jugement moral dans la partie 7.4.

7.3 Reconnaissance des intentions d'un acteur

7.3.1 Modélisation cognitive des intentions

L'intentionnalité peut être définie comme le principe par lequel les actions des personnages et leur comportement apparaît naturel (et rationnel) pour un observateur externe. La causalité joue ici un rôle primordial, une action réalisée de manière intentionnelle pour atteindre un objectif désiré sera reconnue comme telle si la réalisation de cette action apparaît comme nécessaire au regard de sa conséquence

en prenant le point de vue du personnage qui agit. Ainsi, pour un observateur $\mathcal{O}$ qui se place du point de vue de $\mathcal{A}$, nous définissons l'intention attribuée à un acteur $\mathcal{A}$ (qui réalise une action a) et héritée de la conséquence anticipée s par la formule (Saillenfest and Dessalles, 2014a; Saillenfest and Dessalles, 2013) :

$$Int^{\mathcal{O}\to\mathcal{A}}(a,s) = |\nu_s^{\mathcal{O}\to\mathcal{A}}(a)| \tag{7.3}$$

Ce qui, d'après l'équation 7.2, donne :

$$Int^{\mathcal{O}\to\mathcal{A}}(a,s) = E^{\mathcal{O}\to\mathcal{A}}(s) - U^{\mathcal{O}\to\mathcal{A}}(s||a) \tag{7.4}$$

Rappelons que la notation $\mathcal{O} \to \mathcal{A}$ signifie que l'observateur $\mathcal{O}$ prend le point de vue de $\mathcal{A}$ pour calculer la valeur de la fonction (voir la partie 6.3).

Lorsque l'action décrit totalement la conséquence, ce qui signifie que l'acteur anticipe un scénario causal de taille (la taille en *bits* des paramètres) $C_w^{\mathcal{O}\to\mathcal{A}}(s||a)$ entre son action et la conséquence s totalement déterminée $C_d^{\mathcal{O}\to\mathcal{A}}(s|a) = 0$. Dans la plupart des cas l'inattendu est un inattendu causal, $U^{\mathcal{O}\to\mathcal{A}}(s||a) = C_w^{\mathcal{O}\to\mathcal{A}}(s||a)$, ainsi :

$$Int^{\mathcal{O}\to\mathcal{A}}(a,s) = E^{\mathcal{O}\to\mathcal{A}}(s) - C_w^{\mathcal{O}\to\mathcal{A}}(s||a) \tag{7.5}$$

La nécessité d'une action héritée de conséquences non-désirées est négative. Or l'intention associée à ces conséquences non-désirées est positive d'après l'équation 7.3. Cela signifie donc qu'un acteur peut réaliser des actions avec l'intention de produire des conséquences qu'il ne désire pas. On peut distinguer trois cas dans lesquelles des situations non-désirées peuvent être produites suite à l'action d'un agent-acteur :

— l'action est réalisée dans un autre but par l'acteur. L'acteur est conscient des conséquences non-désirées mais réalise l'action pour produire une (ou plusieurs) autres conséquences désirées.
— la situation non-désirée n'a pas été anticipée.
— l'action a été réalisée par inadvertance

Prenons un exemple pour illustrer le premier de ces cas : un aventurier se trouve piégé dans une salle au trésor. Pour sortir de la salle, un seul et unique moyen

existe, l'aventurier peut activer un levier : il sait que cela provoquera l'apparition immédiate d'une issue de secours mais il sait que son action provoquera aussi la disparition immédiate du trésor. L'aventurier décide d'actionner le levier.

La formule 7.3 indique que, pour l'aventurier, l'intention de perdre son trésor en tirant sur le levier vaut $E^{aventurier}(\text{“}tresor_perdu\text{”})$ et que l'intention de sortir de la salle au trésor en tirant sur le levier vaut $E^{aventurier}(\text{“}etre_libre\text{”})$.

Dire que l'aventurier a l'intention de perdre le trésor en tirant sur le levier peut sembler faux. Or cette conséquence a bien été anticipée par l'aventurier, donc il a bien perdu le trésor intentionnellement. Classiquement, on dira que cette conséquence est 'un dommage collatéral', que l'action a été réalisée 'en connaissance de cause'. Cette conséquence négative est 'secondaire', le personnage est rationnel parce qu'une conséquence principale de 'plus grande valeur' justifie *a posteriori* la réalisation de l'action malgré une conséquence négative anticipée. En d'autres termes, l'action est bien jugée *a posteriori* comme étant intentionnelle si l'observateur est en mesure de reconnaître que les gains anticipés par l'acteur suite à la réalisation de cette action sont supérieurs aux pertes anticipées par cet acteur.

Pour évaluer l'intentionnalité d'une action, nous proposons donc de considérer la somme (pondérée uniquement par la valence des désirs des conséquences anticipées) des intentions particulières associées à chacune des conséquences anticipées. Autrement dit, nous proposons de prendre en compte la contribution de chacune des conséquences anticipées. Nous définissons le niveau d'intentionnalité d'une action a attribué par un observateur $\mathcal{O}$ à l'acteur $\mathcal{A}$ supposé avoir anticipé un ensemble de conséquences $S_a^{\mathcal{O}\to\mathcal{A}}$ ainsi (Saillenfest and Dessalles, 2013) :

$$\begin{aligned} Int^{\mathcal{O}\to\mathcal{A}}(a) &= \sum_{s\in S_a^{\mathcal{O}\to\mathcal{A}}} \nu_s^{\mathcal{O}\to\mathcal{A}}(a) \\ &= \sum_{s\in S_a^{\mathcal{O}\to\mathcal{A}}} \epsilon^{\mathcal{O}\to\mathcal{A}}(s) Int^{\mathcal{O}\to\mathcal{A}}(a,s) \qquad (7.6) \\ \epsilon^{\mathcal{O}\to\mathcal{A}}(s) &= +1 \text{ si } \mathcal{O} \text{ considère que } s \text{ est désirée par } \mathcal{A}\,,\ -1 \text{ sinon} \end{aligned}$$

Ce qui peut être réécrit sous la forme :

$$Int^{\mathcal{O}\to\mathcal{A}}(a) = \sum_{s\in S_a^{\mathcal{O}\to\mathcal{A}}} \epsilon^{\mathcal{O}\to\mathcal{A}}(s)\left(E^{\mathcal{O}\to\mathcal{A}}(s) - U^{\mathcal{O}\to\mathcal{A}}(s||a)\right) \tag{7.7}$$

$\epsilon^{\mathcal{O}\to\mathcal{A}}(s) = +1$ si $\mathcal{O}$ considère que s est désirée par $\mathcal{A}$, -1 sinon

Ainsi, un personnage qui agit de manière intentionnelle sera reconnu comme tel par un observateur si cet observateur est capable de déterminer un ensemble de conséquences anticipées par ce personnage qui permet d'attribuer un niveau d'intentionnalité positif aux actions de ce personnage. Ce qui permet de dire que, si on note $A_{\mathcal{A}}$ l'ensemble des actions d'un acteur $\mathcal{A}$, on a :

$$\mathcal{A} \text{ est rationel pour un observateur } \mathcal{O} \Rightarrow \forall a \in A_{\mathcal{A}}, Int^{\mathcal{O}\to\mathcal{A}}(a) \geq 0 \tag{7.8}$$

Pour toute conséquence non-anticipée d'une action on a :

$$\forall s \notin S_a^{\mathcal{O}\to\mathcal{A}}, Int^{\mathcal{O}\to\mathcal{A}}(a, s) = 0 \tag{7.9}$$

Dans le cas où l'action est réalisée par inadvertance, aucune des conséquences n'est anticipée. La seule nécessité associée à l'action est la nécessité $-U(a)$ (l'action est donc un événement surprenant), ce qui rend toute action réalisée par inadvertance non intentionnelle d'après les formules 7.9 et 7.6 ; si l'action est réalisée par inadvertance : $Int^{\mathcal{O}\to\mathcal{A}}(a) = 0$.

Enfin, nous attirons l'attention sur le fait que, dans les modèles introduits dans cette partie, les intentions calculées sont attribuées à des acteurs pour des actions réalisées par ces acteurs. Les actions, comme les conséquences anticipées, dépendent donc des acteurs concernés. Un observateur n'attribue pas d'intention à un acteur pour une action qu'il ne réalise pas.

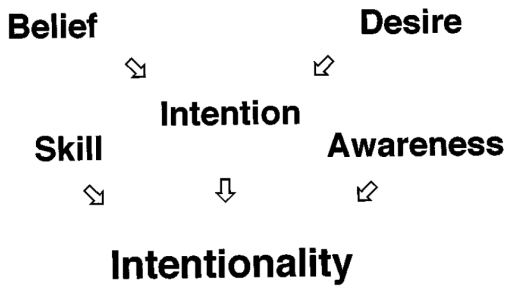
7.3.2 Interprétation et prédictions du modèle

Le sens commun du concept d'intention

L'équation 7.3 rend compte du concept d'intention dans son acception la plus courante.

> "*For an intention to be ascribed, a desire (for an outcome) and a belief (about the action-outcome link) must be present. For an action to be seen as performed intentionally, however, skill and awareness have to be present as well. Thus people distinguish between intention as a mental state and intentionality as the quality of an action*" (Malle, 2010)

FIGURE 7.1 – Modèle de l'intentionalité - Malle et Knobe (1997)



D'après ce modèle (voir aussi fig 7.1), on ne peut attribuer d'intentions que lorsqu'une situation est désirée et qu'il existe un lien (causal) entre l'action et cette situation (et si l'acteur a les capacités de réaliser l'action, ce que l'on peut inclure dans le lien causal entre ce qu'il fait et ce qu'il pense voir se produire). Dans notre modèle, cela est caractérisé par l'existence d'un ensemble $S^{\mathcal{O}\to\mathcal{A}}$ de conséquences qu'un observateur $\mathcal{O}$ estime anticipées par $\mathcal{A}$. Certaines de ces conséquences ne sont pas désirées, dans le sens où $\mathcal{A}$ est supposé ne pas souhaiter qu'elles se produisent. La notion de 'désir' dans la définition de Malle est ambiguë, sa définition ne fait pas la distinction entre les conséquences particulières (un acteur peut avoir l'intention de provoquer la réalisation de situations qui sont, de son point de vue, de valence négative, ce dont rend compte l'équation 7.4) et la somme globale pondérée des contributions des conséquences particulières (qui, elle, doit être positive pour que

l'action soit reconnue comme intentionnelle (voir l'équation 7.7 et la proposition 7.8)).

Dans les équations de l'intention, nous ne prenons en compte que les contributions des conséquences anticipées (voir équation 7.7). Dans le cas où une conséquence non-anticipée se produit, alors elle ne contribue pas à renforcer ou affaiblir (selon sa valence) l'intentionnalité de l'action. La réalisation d'une conséquence non-anticipée s d'une action a est inattendue, sa nécessité est donc $-U(s)$; cette conséquence est un événement inattendu (*a minima* pour l'acteur) et n'affecte pas l'intentionnalité de son action si l'observateur est conscient de la non-anticipation de cette conséquence par l'acteur.

En résumé, l'évaluation de l'intentionnalité d'une action par un observateur ne dépend que de la contribution pondérée des conséquences supposées anticipées par l'acteur, et non des conséquences réelles. Illustrons ceci par un exemple : une personne donne un médicament à un enfant pour le soigner, mais ne sait pas que cet enfant est allergique à ce médicament ; un observateur conscient de cet état de connaissance de l'acteur ne jugera pas que l'action a été réalisée intentionnellement dans le but de blesser l'enfant (ce qui est la conséquence réelle de l'action et non-désirée par l'acteur) mais qu'elle a été réalisée de manière intentionnelle dans le but de soigner l'enfant (ce qui est la conséquence anticipée de l'action et désirée par l'acteur). De même si un acteur veut tuer une personne en la poignardant dans son son sommeil, alors l'équation 7.4 prédit : l'acteur à l'intention de tuer la personne en la poignardant (y compris si la personne en question est déjà morte et que l'observateur qui juge pense que l'acteur ne le sait pas) ; d'après l'équation 7.7, $Int^{\mathcal{O}\to\mathcal{A}}(a) = E^{\mathcal{O}\to\mathcal{A}}(\text{la victime est morte}) > 0$, l'observateur estime que le tueur pense que la victime va mourir suite à son acte et a donc l'intention de tuer la victime en la poignardant.

Dans le modèle de l'intention de Malle (figure 7.1), une action n'est intentionnelle que si l'acteur est conscient de l'action. Si l'acteur n'est pas conscient de l'action, il n'est pas non plus conscient de ses conséquences. Comme nous l'avons montré précédemment, cela revient à une évaluation nulle de l'intentionnalité associée à cette action car les conséquences d'une action réalisée par inadvertance sont non-anticipées (voir les équations 7.4 et 7.9).

Jugement et hiérarchie des conséquences

Définir l'intention à la fois comme une notion graduelle ainsi que comme la somme pondérée d'intentions particulières héritées de conséquences anticipées de nos actes permet, entre autres, d'établir une classification desdites conséquences. Ainsi, lorsqu'il s'agit de juger de l'intention d'un acteur qui réalise une action donnée, on va distinguer :

— les conséquences principales, c'est-à-dire celle qui contribuent majoritairement au calcul de l'intention de réaliser l'action.

— les effets anticipés collatéraux (ou conséquences anticipées collatérales) de moindre importance (*side effects*), c'est-à-dire l'ensemble des bénéfices ou des dommages collatéraux anticipés qui viennent renforcer ou affaiblir l'intention de réaliser l'action.

Analysons l'exemple classique suivant introduit par Knobe (2003) (il existe deux versions de l'histoire, les éléments en gras indiquent les différences entre ces versions)

> "*The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm* ***(resp. help)*** *the environment.' The chairman of the board answered, 'I don't care at all about harming* ***(resp. helping)*** *the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was harmed* ***(resp. help)****.* " (Knobe, 2003)

La question posée par Knobe est de savoir si le chef d'entreprise sauve ou détruit l'environnement de manière intentionnelle en décidant de lancer le programme. Selon la version du texte présentée aux personnes interrogées, les avis divergent. Dans la version dans laquelle l'environnement est sauvé, une majorité de répondants déclare que le chef d'entreprise ne sauve pas l'environnement de manière intentionnelle. Dans la version dans laquelle l'environnement est détruit, une majorité de répondants considère que le chef d'entreprise a détruit l'environnement de manière intentionnelle. Cependant la majorité des personnes interrogées préfèrent dire que la sauvegarde ou la destruction de l'environnement sont des

conséquences (secondaires) d'une action réalisée en connaissance de cause (*knowingly*) si cette option leur est proposée (Guglielmo and Malle, 2010). Ceci est cohérent avec le modèle de l'équation 7.7. Si l'on prend le point de vue du chef d'entreprise qui n'accorde que très peu d'importance à tout ce qui touche l'environnement, l'émotion suscitée par le bénéfice financier $E^{\mathcal{O}\to Ch}(benefice)$ est bien plus grande que celle suscitée par la destruction ($E^{\mathcal{O}\to Ch}(dest_env)$) ou la préservation ($E^{\mathcal{O}\to Ch}(sauv_env)$) de l'environnement qui sont donc, de son point vue, des effets collatéraux d'importance moindre (voire négligeable) anticipés, et donc son action est réalisée en connaissance de cause, ce sont des bénéfices (ou des dommages) collatéraux. Si on suppose que le gain financier et, selon l'histoire, la destruction et la sauvegarde de l'environnement sont des conséquences directes et immédiates de la décision du *chairman*, les intentions (du chef d'entreprise, en prenant la décision a de lancer le programme) de faire un bénéfice, de détruire l'environnement et de préserver l'environnement valent, en appliquant l'équation 7.4 :

$$Int^{\mathcal{O}\to Ch}(a, benefice) = E^{\mathcal{O}\to Ch}(benefice)$$

$$Int^{\mathcal{O}\to Ch}(a, dest_env) = E^{\mathcal{O}\to Ch}(dest_env)$$

$$Int^{\mathcal{O}\to Ch}(a, sauv_env) = E^{\mathcal{O}\to Ch}(sauv_env)$$

L'intentionnalité de la réalisation du programme vient donc principalement de la contribution du bénéfice financier anticipé, la préservation (resp. la destruction) de l'environnement vient (légèrement) renforcer (resp. atténuer) cette valeur d'intentionnalité. Formellement, en utilisant l'équation 7.7, dans le cas où le lancement du programme conduit à la destruction de l'environnement :

$$Int^{\mathcal{O}\to Ch}(a) = E^{\mathcal{O}\to Ch}(benefice) - E^{\mathcal{O}\to Ch}(dest_env)(\approx E^{\mathcal{O}\to Ch}(benefice))$$

Dans le cas où le lancement du programme conduit à la préservation de l'environnement :

$$Int^{\mathcal{O}\to Ch}(a) = E^{\mathcal{O}\to Ch}(benefice) + E^{\mathcal{O}\to Ch}(sauv_env)(\approx E^{\mathcal{O}\to Ch}(benefice))$$

Subjectivité et différences de jugement

Il est important de préciser pour chacune des notions qui calcule quoi en prenant le point de vue de qui. La valence des conséquences anticipées peut, par exemple, ne pas être la même pour l'acteur et l'observateur qui prend le point de vue de l'acteur. Un observateur peut penser que l'acteur désire (resp. ne désire pas) un état de fait alors que l'acteur ne le désire pas (resp. le désire). Dans l'exemple de Mary et John reproduit en début de chapitre, l'auteur donne au lecteur une information sur les sentiments de Mary à l'égard de John "Mary aime beaucoup John"). Ainsi, lorsque Mary poignarde John, ce qui a pour conséquence de blesser John, conséquence de valence négative pour le lecteur qui se place du point de vue de Mary, alors le calcul de l'intention est négatif (c'est la seule conséquence anticipée par le lecteur prenant le point de vue de Mary à ce stade de l'histoire). Cela signifie donc que, pour le lecteur, l'acte de Mary qui consiste à poignarder John n'est pas intentionnel (à ce stade). Si, en réalité, Mary déteste John, ce qui aurait donc été caché au lecteur, alors l'action de Mary est bien intentionnelle de son propre point de vue (elle a le désir de tuer John, les capacités de le faire, sait que le coup de couteau sera certainement fatal et est consciente de son acte). Nous serions donc face à un cas où une action intentionnelle n'est pas reconnue comme telle par un observateur.

Il peut également exister un écart entre les conséquences anticipées par un observateur qui prend le point de vue d'un acteur et les conséquences 'réellement' anticipées par cet acteur. Si on reprend le cas d'une personne qui donne un médicament à un enfant, un observateur peut penser que la personne veut soigner l'enfant alors que l'acteur sait que l'enfant est allergique au médicament et veut le blesser. L'observateur considérera donc que l'acteur avait l'intention de soigner l'enfant, alors que l'acteur avait en réalité l'intention de le blesser.

Un accord entre l'observateur et l'acteur sur l'ensemble des conséquences anticipées par ce dernier ne garantit pas que l'un et l'autre vont attribuer une égale intentionnalité à une action. Chacun ayant sa propre connaissance, l'observateur

et l'acteur ne vont pas nécessairement s'accorder sur l'évaluation de l'émotion attachée à un état de fait donné ou sur l'évaluation de l'inattendu d'une situation donnée. Par exemple, si la réalisation d'une action donnée a pour effet de sauver cinq personnes (qui sont en danger de mort) mais également de tuer une autre personne, alors un observateur pourrait ne pas comprendre qu'un acteur ne réalise pas l'action s'il en a les moyens (sauver la vie de cinq personnes a objectivement 'plus de valeur' que la mort d'une personne). Or si la personne à sacrifier est un proche de l'acteur, la valeur associée par l'acteur à la mort de cette personne peut surpasser le gain des cinq vies sauvées, ce qui, du point de vue de l'acteur en capacité d'agir, rend l'inaction préférable à l'action (dans l'équation 7.5, la simplicité de la victime fait baisser la complexité $C_d^{\mathcal{O}\to\mathcal{A}}(s)$ de la conséquence positive "ne pas tuer la personne seule", et donc la contribution de cette conséquence positive pour l'acteur dépasse la contribution négative de la mort des cinq personnes, et sa non-action sera jugée comme étant intentionnelle *a posteriori*).

Limites du modèle des intentions

Le modèle proposé précédemment a ses limites. Le calcul de l'intentionnalité d'une action est une somme valuée des conséquences anticipées par l'acteur de son action. Cette somme correspond à l'occurrence d'une 'conséquence globale' (ce que nous appelons S_a) qui est une conjonction de conséquences particulières (la somme porte sur l'ensemble des situations s composant S_a). Parfois, la conséquence d'une action peut être une disjonction de conséquences, ce qu'illustre l'exemple suivant :

Un individu doit choisir entre deux loteries, l'une peut lui rapporter 1000€ avec 90% de chances, l'autre peut lui rapporter 1000€ avec 30% de chances. Il choisit la seconde loterie (1000€ avec 30% de chances).

L'individu a-t-il l'intention de gagner 1000€ en choisissant la seconde loterie ? Le modèle précédent ne répond pas immédiatement à la question. En effet, le choix de la seconde loterie (on notera ce choix de la loterie 2, qui est une action, a_2. On notera a_1 le choix de la loterie 1) engendre une disjonction d'alternatives-conséquences : la première alternative alt_1 est un gain de 1000€, la seconde, alt_2, est une alternative dont on ne connaît pas la nature. On sait juste que la se-

conde alternative exclut la première ; ainsi on peut dire que $alt_2 = \neg alt_1$. On peut évaluer l'influence du choix a_2 sur l'évolution de l'inattendu de alt_1. En effet, si on suppose que le choix n'est pas inattendu pour l'acteur, alors avant le choix on a $U^{\mathcal{O}\to\mathcal{A}}(alt_1) = -log_2(0.9)bits$; pour l'observateur, l'acteur sait qu'il est dans une situation où le gain de 1000€ est *potentiellement* très peu inattendu. Puis il réalise a_2 et on a donc $U^{\mathcal{O}\to\mathcal{A}}(alt_1||a_2) = -log(0.3)bits$. On est donc dans une situation où $U^{\mathcal{O}\to\mathcal{A}}(alt_1) - U^{\mathcal{O}\to\mathcal{A}}(alt_1||a_2) < 0$, autrement dit le choix a défavorisé l'occurrence de l'alternative alt_1. Cela crée une sorte de conflit avec l'hypothèse où cette alternative serait désirée. Ce genre de conflit n'apparaît pas si on fait la supposition qu'à l'inverse $alt_2(= \neg alt_1)$ est désirée puisque $U^{\mathcal{O}\to\mathcal{A}}(alt_2) - U^{\mathcal{O}\to\mathcal{A}}(alt_2||a_2) = 0$ (l'alternative 2 peut *potentiellement* se réaliser avec 70% de chances avant le choix de l'acteur et peut toujours potentiellement se réaliser avec 70% de chances après le choix a_2). Ainsi supposer que alt_2 est désiré conduit à un calcul d'intentionnalité positif.

L'alternative 1 est-elle pour autant non-désirée ? Pas forcément, mais si elle est désirée, alors l'intention d'obtenir cette conséquence est moindre que celle d'obtenir l'occurrence de l'alternative 2. Intuitivement, on peut penser qu'un observateur répondra que l'acteur n'a pas l'intention d'obtenir les 1000€ en choisissant la seconde loterie. Mais on peut également penser que si on dévoile la nature de l'alternative 2 (par exemple si il s'agit de gagner une voiture) alors un observateur pourrait être tenté de dire que l'acteur à moins l'intention de gagner les 1000€ que la voiture.

Le cas des conséquences de type disjonction introduit un degré supplémentaire dans le calcul de l'intention des acteurs et nous pensons qu'il nécessite la réalisation par l'observateur d'un calcul séquentiel visant à établir quelles situations conduisent à des situations de conflit autour de l'action et quelles conséquences permettent de maintenir la rationalité de l'acteur.

7.4 Responsabilité et jugement moral

Dans cette partie nous allons prolonger notre étude au jugement moral, à l'attribution de responsabilité à un acteur. Nous parlerons de niveau de responsabilité attribuée pour désigner le niveau d'approbation (*level of praise*) ou de

désapprobation (*level of blame*, intensité du reproche).

7.4.1 Modélisation cognitive de l'attribution de responsabilité

Comme pour la notion d'intention, la capacité à prédire les conséquences d'une action, et donc les connaissances et croyances des acteurs que l'on juge responsables ou non de l'occurrence d'une situation, semblent jouer un rôle important pour déterminer le niveau de responsabilité. Cependant, bien que ces notions soient proches, il semble que l'attribution des intentions et de la responsabilité diffèrent au niveau de la sévérité des conséquences telles qu'elles sont évaluées par un observateur. Bien qu'il puisse y avoir accord entre plusieurs observateurs sur le niveau d'intentionnalité à attribuer à une action réalisée par un acteur donné, ces mêmes observateurs peuvent individuellement attribuer un niveau de responsabilité différent. Une majorité d'observateurs jugent un acteur plus responsable de l'occurrence d'une situation dans laquelle cinq personnes meurent que de l'occurrence d'une situation similaire dans laquelle il y a une seule victime... sauf si ladite victime est proche (meilleur ami, parent, etc.) de l'observateur (Saillenfest and Dessalles, 2012) (voir aussi A.3 en annexe).

Cependant, dans l'attribution de la responsabilité, la capacité supposée d'anticipation de l'acteur ne doit pas être négligée.

> "*Consider a doctor who treats a patient with a particular drug resulting in the patient's death. The doctor's treatment is a cause of the patient's death ; indeed, the doctor may well bear degree of responsibility 1 for the death. However, if the doctor had no idea that the treatment had adverse side effects for people with high blood pressure, he should perhaps not be held to blame for the death. Actually, in legal arguments, it may not be so relevant what the doctor actually did or did not know, but what he should have known. Thus, rather than considering the doctor's actual epistemic state, it may be more important to consider what his epistemic state should have been. But, in any case, if we are trying to determine whether the doctor is to blame for the patient's death, we must take into account the doctor's epistemic state.*" (Chockler and Halpern, 2004)

Ainsi, il semble que l'attribution de responsabilité dépend, non pas de la réponse émotionnelle de l'acteur, mais de celle de l'observateur, tout en prenant en compte la supposée anticipation de l'acteur.

Nous définissons le degré de responsabilité $Res^{\mathcal{O}}(\mathcal{A}, a, s)$ qu'un observateur $\mathcal{O}$ attribue à un acteur $\mathcal{A}$ pour une conséquence spécifique s d'une action a par l'équation suivante (Saillenfest and Dessalles, 2013) :

$$Res^{\mathcal{O}}(\mathcal{A}, a, s) = E^{\mathcal{O}}(s) - U^{\mathcal{O}\to\mathcal{A}}(s||a) \text{ si positif} \tag{7.10}$$

Nous parlons aussi de niveau d'approbation (ou de désapprobation) selon la valence de la conséquence considérée (ce qui peut devenir une intensité de 'félicitation' ou de reproche associée à l'occurrence possible de ladite conséquence).

Comme précédemment, une action peut avoir plusieurs conséquences. Dans le cas de la responsabilité, les conséquences prises en compte ne sont pas uniquement celles qui sont supposées avoir été anticipées par l'acteur, mais également celles constatées par l'observateur. Dans l'équation précédente, si une conséquence constatée n'a pas été anticipée, alors $U^{\mathcal{O}\to\mathcal{A}}(s||a) = U^{\mathcal{O}\to\mathcal{A}}(s)$ et nous pouvons alors supposer que $U^{\mathcal{O}\to\mathcal{A}}(s) = U^{\mathcal{O}}(s)$. Ainsi, si s est une conséquence constatée par $\mathcal{O}$ mais supposée non anticipée (parfois parce que supposée non-anticipable) par $\mathcal{A}$, alors la responsabilité de $\mathcal{A}$ attribuée par $\mathcal{O}$ pour s conséquence de a vaut, en utilisant l'équation 4.5 :

$$Res^{\mathcal{O}}(\mathcal{A}, a, s) = E_h^{\mathcal{O}}(s) \tag{7.11}$$

Le terme $U^{\mathcal{O}\to\mathcal{A}}(s||a)$ peut également être assimilé à $C_w^{\mathcal{O}\to\mathcal{A}}(s||a)$ dans les cas ou les conséquences sont anticipées et supposées être le résultat d'un scénario causal anticipé (la conséquence est supposée totalement décrite $C_d^{\mathcal{O}\to\mathcal{A}}(s|a) = 0$). Ainsi, dans ce cas, l'équation 7.10 devient :

$$Res^{\mathcal{O}}(\mathcal{A}, a, s) = E^{\mathcal{O}}(s) - C_w^{\mathcal{O}\to\mathcal{A}}(s||a) \text{ si positif} \tag{7.12}$$

7.4.2 Modèle du jugement moral

La somme des responsabilités de l'acteur pour chacune des conséquences individuelles d'une action donnée est ce que nous appellerons jugement moral de l'observateur à l'encontre de l'acteur pour une action. Le jugement moral vise à évaluer ce que l'observateur considère comme bien ou mal, c'est-à-dire ce qu'il approuve ou non.

Nous définissons le jugement moral MJ, ou degré d'approbation, d'un observateur $\mathcal{O}$ lié à une action a réalisée par un acteur $\mathcal{A}$ dont les conséquences observées (qui forment l'ensemble $S_a^{\mathcal{O}}$) et celles que l'observateur suppose avoir été anticipées par l'acteur (qui forment l'ensemble $S_a^{\mathcal{O}\to\mathcal{A}}$) forment un ensemble $S_a^{\mathcal{O}} \cup S_a^{\mathcal{O}\to\mathcal{A}}$ par la formule (Saillenfest and Dessalles, 2013; Saillenfest and Dessalles, 2012) :

$$\begin{aligned} MJ^{\mathcal{O}}(\mathcal{A}, a) &= \sum_{s \in S_a^{\mathcal{O}} \cup S_a^{\mathcal{O}\to\mathcal{A}}} \epsilon^{\mathcal{O}}(s) Res^{\mathcal{O}}(\mathcal{A}, a, s) \\ \epsilon^{\mathcal{O}}(s) &= +1 \text{ si } \mathcal{O} \text{ désire } s, \ -1 \text{ sinon} \end{aligned} \tag{7.13}$$

7.4.3 Interprétation et prédictions du modèle

Le modèle introduit précédemment entend rendre compte du sens commun de responsabilité. Ce dont rend compte la formule précédente est l'attribution d'une responsabilité d'ordre moral. Il ne s'agit pas, en particulier, d'une responsabilité d'ordre causal. Un père qui empoisonne par accident son enfant en lui donnant un médicament dans le but de le soigner est très responsable du point de vue causal (son action a pour conséquence directe l'empoisonnement de l'enfant), mais bien moins responsable du point de vue moral (il ne savait pas, et ne pouvait pas savoir ou n'était pas censé savoir, que son action provoquerait cette conséquence).

Dans le modèle précédent, nous considérons que lorsqu'il s'agit d'attribuer une responsabilité morale à un acteur, un observateur utilise ses propres valeurs pour évaluer les conséquences (ce que traduit la présence du terme $E^{\mathcal{O}}$ dans l'équation 7.10) tout en prenant en compte les capacités d'anticipation de l'agent (ce que traduit la présence du terme $U^{\mathcal{O}\to\mathcal{A}}(s||a)$ dans l'équation 7.10).

La présence des termes $E^{\mathcal{O}}$ et $\epsilon^{\mathcal{O}}$ dans l'équation 7.13 permet de faire quelques prédictions. Le niveau de responsabilité attribué dépend des désirs de l'observateur

(et donc de leur valence et leur intensité) et non des motivations de l'agent. Ainsi, un observateur désapprouvera l'action d'un agent si la conséquence de ses actes produit une situation que cet observateur ne désire pas (et ce même si l'acteur désire cette situation). Un ami de la victime d'un tueur va blâmer le tueur même si ce dernier désirait la mort de la victime. Notons ici qu'un juge, dans un cadre législatif, ne prend pas (et ne dois pas prendre) en compte ses propres émotions. Le juge dans un cadre législatif prend le point de vue d'un observateur plus abstrait, un humain moyen, afin que le jugement ne soit pas affecté par les effets de simplicité, et soit donc le moins subjectif possible.

Par ailleurs, l'intensité émotionnelle dépend de la simplicité (pour l'observateur) de la situation conséquence des actes de l'agent. Ainsi, dans l'exemple du paragraphe précédent, la proximité avec la victime du tueur aura une influence sur le niveau de responsabilité attribuée. L'équation 7.10 indique en particulier qu'un individu pris au hasard attribuera moins de responsabilité au tueur qu'un proche de la victime. Autrement dit, le reproche envers le tueur sera plus fort de la part d'un proche de la victime (voir l'expérience en annexe A.2).

Dans l'équation 7.10, le terme $U^{\mathcal{O}\to\mathcal{A}}(s||a)$ rend compte du rôle important des capacités d'anticipation supposées de l'agent dans l'attribution de responsabilité. Dans l'exemple d'un empoisonnement accidentel d'un enfant par son père qui lui a donné à manger pour la première fois un aliment auquel l'enfant est allergique (ce que le père ne pouvait pas savoir *a priori*), la plupart des observateurs vont considérer que l'empoisonnement de l'enfant ne pouvait pas être anticipé par le père. Le lien causal entre l'action de donner à manger à l'enfant et sa conséquence directe (l'empoisonnement de l'enfant) est donc très inattendu du point de vue du père. Pour un observateur qui se place du point de vue du père, l'empoisonnement de l'enfant une fois qu'il a avalé l'aliment allergène reste extrêmement complexe à produire. Cela peut être assimilé à une forme d''empoisonnement spontané', l'événement est donc aussi inattendu que s'il s'était produit spontanément. Ainsi $U^{\mathcal{O}}(s) = U^{\mathcal{O}\to\mathcal{A}}(s)$, ce qui donne $Res^{\mathcal{O}}(\mathcal{A}, a, s) = E_h^{\mathcal{O}}(s)$ (voir l'équation 7.11). Cette valeur correspond donc à l'anticipation émotionnelle d'un empoisonnement spontané de l'enfant, cette valeur est faible. D'un point de vue moral, le père de l'enfant sera jugé peu responsable (même si, là encore, d'un point de vue causal, cela ne se serait pas produit s'il n'avait pas donné à manger à l'enfant). D'une

manière générale, lorsqu'il s'agit d'un accident, et que l'acteur $\mathcal{A}$ n'avait pas les moyens ou n'était pas censé savoir que son action a pouvait mener à la conséquence s, la contribution au calcul global de responsabilité attribuée par un observateur $\mathcal{O}$ est $Res^{\mathcal{O}}(\mathcal{A}, a, s) = E_h^{\mathcal{O}}(s)$.

La présence du terme $U^{\mathcal{O}\to\mathcal{A}}(s||a)$ dans l'équation 7.10 explique également pourquoi les enfants ou les personnes 'folles' (ainsi que les animaux) sont, la plupart du temps, tenues pour peu responsables de leurs actions. Cela signifie que les observateurs-juges les considèrent comme étant incapables de réaliser les calculs de complexité qui permettent de distinguer $U^{\mathcal{O}\to\mathcal{A}}(s||a)$ de $U^{\mathcal{O}\to\mathcal{A}}(s)$.

7.5 Analyse empirique et validation des modèles

Nous avons réalisé deux expériences (présentées en annexe A.2 (il s'agit d'une expérience préliminaire) et en annexe A.3 (publiée dans (Saillenfest and Dessalles, 2012))) afin de tester certaines prédictions des modèles introduits précédemment concernant l'évaluation des intentions (voir équation 7.4), de la responsabilité (voir équation 7.10) et du niveau d'approbation moral (voir équation 7.13) liés à l'occurrence d'une action par un agent.

Nous avons testé l'influence des relations interpersonnelles entre une victime d'un accident et un observateur qui juge le niveau d'intentionnalité et de responsabilité d'un acteur (voir expérience A.2). L'expérience montre que l'attribution des intentions n'est globalement pas affectée par les liens d'amitié ou de parenté (lien frère-frère ou frère-sœur) qui unissent la victime de l'accident et l'observateur qui juge. Cela tend à confirmer que celui qui juge des intentions d'un acteur ne tient pas compte de sa propre réponse émotionnelle liée à l'occurrence possible d'une conséquence (de l'action jugée) anticipée par l'acteur. Cela est cohérent avec la formule 7.4 qui prédit une variation de l'attribution des intentions en fonction de l'état mental supposé de l'acteur, et non de celui de l'observateur.

En revanche, l'intensité des jugements de responsabilité portés par un observateur à l'encontre d'un acteur sont affectés par la simplicité (pour l'observateur) de la conséquence réelle (et observée) des actes de l'acteur. En effet, l'expérience en annexe A.2 montre que, lorsque la victime d'un accident est simple pour l'observateur (lien de sang, lien d'amitié ou relation entre une personne célèbre et un fan

de cette personne), la responsabilité attribuée par cet observateur est significativement plus élevée que lorsqu'il n'existe aucun lien entre la victime et l'observateur (et donc lorsque la conséquence n'est pas anormalement simple pour ce dernier).

Par ailleurs, dans l'expérience en annexe A.3, nous pouvons constater que l'évaluation morale de l'agent par un observateur est également affectée par la simplicité des conséquences anticipées par l'observateur. En effet, l'acteur est jugé moralement plus responsable si l'une des victimes a un lien de parenté ou d'amitié avec cet acteur, ce qui rend donc cette victime anormalement simple pour le lecteur (voir phase 3 de l'expérience en annexe A.3). Ceci est donc cohérent avec les prédictions des équations 7.10 et 7.13.

L'anticipation du niveau de complexité causale suite à la réalisation de l'action (*ie.* $C_w^{\mathcal{O}\to\mathcal{A}}(s||a)$) semble effectivement prise en compte par l'observateur qui juge (le niveau d'approbation des actions diminue à mesure que la complexité anticipée pour produire la conséquence positive augmente (voir expérience en annexe A.3, phase 2) ou que la complexité anticipée par l'agent pour produire la conséquence négative diminue (voir expérience en annexe A.3, phase 1)). Cela est cohérent avec les équations 7.10 et 7.13, l'observateur évalue son jugement moral en utilisant ses propres valeurs émotionnelles mais se place dans l'état mental de l'acteur pour évaluer le niveau d'anticipation des conséquences (ce qui correspond à la présence du terme $C_w^{\mathcal{O}\to\mathcal{A}}(s||a)$ dans l'équation 7.12).

Dans l'expérience en annexe A.2, nous avons aussi souhaité évaluer l'influence des informations supplémentaires et des facteurs atténuants sur l'attribution de responsabilité. Lorsque l'on informe un observateur d'une conséquence positive (jusque-là cachée) que l'acteur cherche à atteindre en réalisant son action (on suppose la conséquence positive du point de vue de l'acteur et de l'observateur), les participants à l'expérience A.2 indiquent majoritairement qu'ils pensent que cette conséquence positive va atténuer le niveau de responsabilité attribué à l'acteur. En d'autres termes, le niveau de désapprobation de l'observateur à l'égard des actes de l'acteur diminue. Ceci est cohérent avec les prédictions de l'équation 7.13 sur l'additivité des sources de responsabilité pour l'évaluation du jugement moral, et en particulier la prise en compte des conséquences anticipées par l'acteur. Par ailleurs, la divulgation d'informations tendant à rendre l'accident plus inattendu (du point de vue de l'acteur), est considérée comme ayant un pouvoir atténuant au

niveau de l'attribution de responsabilité par un observateur (ce qui est cohérent avec l'équation 7.10, cela fait augmenter la valeur de $U^{\mathcal{O}\to\mathcal{A}}(s||a)$).

7.6 Résoudre le dilemme intérêt vraisemblance

La reconnaissance des intentions d'un personnage dans un récit est l'un des principaux ingrédients de sa vraisemblance au yeux d'un lecteur (Riedl and Young, 2004). Or parfois, les personnages peuvent agir de manière inattendue et donc, pour un temps au moins, paraître non rationnels. Lorsque les intentions d'un personnage ne sont pas reconnues à un instant donné de l'histoire, la rationalité du personnage est mise en suspens.

Plaçons-nous dans un cas simple : un observateur $\mathcal{O}$ (le lecteur d'un roman par exemple) observe une scène dans laquelle un acteur $\mathcal{A}$ réalise une action a dont la conséquence (supposée unique, anticipée et résultant d'un processus causal déterministe pour l'acteur) est s. Le modèle de l'équation 7.4 prédit que le niveau d'intentionnalité attribué par $\mathcal{O}$ à $\mathcal{A}$ vaut $Int^{\mathcal{O}\to\mathcal{A}}(a,s) = \epsilon^{\mathcal{O}\to\mathcal{A}}(s)|E^{\mathcal{O}\to\mathcal{A}}(s) - U^{\mathcal{O}\to\mathcal{A}}(s||a)|$.

Pour l'observateur, la situation peut être intéressante car inattendue ($E^{\mathcal{O}} = E_h^{\mathcal{O}} + U^{\mathcal{O}}$) ou, par procuration, intéressante parce qu'elle l'est pour l'acteur ($E^{\mathcal{O}\to\mathcal{A}} = E_h^{\mathcal{O}\to\mathcal{A}} + U^{\mathcal{O}\to\mathcal{A}}$).

Ainsi, au moment où l'observateur $\mathcal{O}$ observe une scène dans laquelle un acteur $\mathcal{A}$ réalise une action a dont la conséquence (supposée unique, anticipée et résultant d'un processus causal déterministe pour l'acteur) est s, le dilemme Intérêt-Vraisemblance (ou dans ce cas Intérêt-Crédibilité d'où la notation IC) peut être formalisé ainsi :

$$IC^{\mathcal{O}}(\mathcal{A},a,s) : \begin{cases} I^{\mathcal{O}}(s) = [E_h(s) + U(s)]^{\mathcal{O} \text{ ou } \mathcal{O}\to\mathcal{A}} \\ Int^{\mathcal{O}\to\mathcal{A}}(a) = \epsilon^{\mathcal{O}\to\mathcal{A}}(s)|E^{\mathcal{O}\to\mathcal{A}}(s) - U^{\mathcal{O}\to\mathcal{A}}(s||a)| \end{cases} \tag{7.14}$$

où I désigne l'intérêt.

Dans l'introduction de cette partie, nous avons reproduit l'exemple de l'histoire du meurtre de John par Mary. Du point de vue du lecteur, la mort de John est surprenante car inattendue. Or si ce lecteur prend le point de vue de Mary pour

calculer les intentions de cette dernière, alors la valence de la mort de John (on note †John cette situation) est négative (car le lecteur pense que Mary aime John) et, l'acte de Mary (noté a) n'étant *a priori* pas réalisé par inadvertance, Mary étant capable d'anticiper les conséquences de son acte, la mort de John est très peu inattendue pour Mary suite à l'acte. Ainsi, le système d'équations Intérêt-Crédibilité appliqué à l'histoire de Mary et John (et calculé pour un lecteur Lec) devient :

$$IC^{Lec}(Mary, a, \dagger\text{John}) : \begin{cases} I^{Lec}(\dagger\text{John}) = E_h^{Lec}(\dagger\text{John}) + U^{Lec}(\dagger\text{John})(>> 0) \\ Int^{Lec \to Mary}(a) = -E^{Lec \to Mary}(\dagger\text{John})(<< 0) \end{cases} \tag{7.15}$$

Cette situation est un cas classique de situation surprenante mais dans laquelle l'un des personnages semble agir de manière irrationnelle, l'intentionnalité de son acte étant négative. C'est l'exemple général d'une situation surprenante qui vient remettre en cause la rationalité (et donc la crédibilité) d'un personnage. Or, comme cela crée un conflit chez le lecteur, la rationalité d'un personnage supposé rationnel est remise en cause.

Le dilemme intérêt-crédibilité est donc assimilable à la formation (voulue par l'auteur) d'un (méta-)conflit cognitif dans l'esprit du lecteur autour de la crédibilité de l'un des protagonistes du récit dans le but de produire de l'intérêt. Cette suspension de la crédibilité d'un personnage peut être acceptée par le lecteur, mais l'auteur doit, avant la fin du récit, présenter au lecteur les éléments de résolution de ce conflit, et donc restaurer la crédibilité du personnage (ou alors désigner ce personnage comme fou, irrationnel, ayant agit par inadvertance ou n'ayant pas anticipé les conséquences de son acte).

Le conflit autour de la rationalité d'un personnage est désigné comme étant un méta-conflit car ce conflit est un ensemble de conflits potentiels. Dans l'histoire de Mary, ce conflit peut porter directement sur la rationalité de Mary (Est-elle folle ? Est-elle sous l'emprise d'une personne ou d'une substance qui affecte sa conscience ?), sur la valence de la conséquence (Est-ce qu'elle désire la mort de John ? Est-ce bien John qu'elle a tué ?) ou encore sur l'existence de conséquences cachées au lecteur (Est-ce que Marie tue John sous la menace d'un autre individu

(et donc elle veut sauver sa vie) ?).

La restauration de la rationalité d'un personnage peut prendre des formes variées. D'après le modèle des intentions présenté précédemment, à la fin d'un récit, l'un des quatre types de situations suivantes doit être atteint :

- L'action qui, dans un premier temps, semble non-intentionnelle est intentionnelle.
- L'action a été réalisée par inadvertance (le calcul de l'intentionnalité n'est donc pas valable)
- La conséquence n'a pas été anticipée (le calcul de l'intentionnalité n'est donc pas valable)
- Le personnage n'est pas rationnel au moment de l'acte (folie passagère, manipulation, etc.). Le personnage n'a donc pas conscience de l'acte qu'il réalise et de ses conséquences. Le niveau d'intention est donc nul.

Comme cela est dit plus haut, la notion de postdictibilité est essentielle dans l'écriture d'un scénario ; Kintsch (1980, p.89) la définit ainsi : *The text as a whole must hang together and make sense to the reader, so that he is able to construct a coherent macrostructure in which each text unit has its place and is meaningfully related to other sections of the text. Weird and unpredictable statements in a text are interesting only in so far as they are well motivated within the text as a whole, at least by hindsight.* Ainsi, une action dont la conséquence peut paraître surprenante à un moment donné du texte, ce qui remet en cause la rationalité du personnage-acteur à ce moment, est postdictible si, à la fin du récit, de manière rétrospective, elle est jugée intentionnelle. La révélation des éléments de restauration de la rationalité du personnage-acteur fournissent une explication qui rend, rétrospectivement, la conséquence de ses actes non-inattendue.

Notons $A_{\mathcal{A}}$ l'ensemble des actions d'un acteur $\mathcal{A}$ et Act l'ensemble des acteurs d'une séquence narrative. D'après la définition du critère de postdictibilité, on peut écrire :

$$\text{Le critère de postdictabilité est respecté} \Rightarrow$$
$$\text{À la fin du récit, pour } \mathcal{O}, \forall\, \mathcal{A} \in Act, \forall a \in A_{\mathcal{A}}, \forall s \in S_a^{\mathcal{O}\to\mathcal{A}}, Int^{\mathcal{O}\to\mathcal{A}}(a) \geq 0 \tag{7.16}$$

ce qui signifie que toute action d'un acteur rationnel du récit doit rétrospectivement, à la fin du récit, être jugée comme étant intentionnelle par l'audience $\mathcal{O}$. C'est un critère de pertinence structurelle.

Par ailleurs, en notant $Recit*$ l'ensemble des situations s du récit (sans l'événement perturbateur initial), nous pouvons écrire :

$$\begin{array}{c}\text{Le critère de postdictabilité est respecté} \Rightarrow \\ \text{À la fin du récit, pour } \mathcal{O}, \forall s \in Recit*, U^{\mathcal{O}}(s) = 0\end{array} \tag{7.17}$$

ce qui signifie que, rétrospectivement, toute situation du récit qui n'est pas l'événement perturbateur doit être non-inattendue pour l'audience $\mathcal{O}$. Autrement dit, toute situation doit avoir été intégrée à la représentation qu'à l'audience du monde fictif et pouvoir être expliquée *a posteriori* (sauf l'événement perturbateur qui peut rester inexpliqué, c'est un prétexte pour raconter l'histoire). C'est un critère de pertinence émotionnelle.

Chapitre 8

Complexité et analyse de scénario

Nous avons jusqu'ici proposé des modèles quantitatifs de l'intérêt et de la reconnaissance des intentions basés sur la notion de complexité. Dans ce chapitre, notre objectif est de montrer que ces modèles peuvent servir d'outils d'analyse et d'évaluation de scénarios existants.

Nous allons donc analyser une séquence narrative, nous allons détailler les variations de complexité qui interviennent dans cette séquence et montrer comment ils peuvent être interprétés à l'aune des modèles introduits précédemment.

Nous utiliserons en particulier les équations définissant l'intérêt narratif (équations 4.4 et 4.5) ainsi que les équations qui définissent la reconnaissance des intentions (équations 7.4 et 7.6).

Voici l'extrait du scénario du film "*Indiana Jones and the Last Crusade*" (Lucas and Meyjes, 1989) que nous analysons dans cette partie :

At that moment DONOVAN and ELSA appear. DONOVAN turns to face the display of cups.

DONOVAN : *Which one is it ?*

KNIGHT : *You must choose. But choose wisely. For as the True Grail will bring you life, the False Grail will take it from you.*

DONOVAN : *I'm not a historian. I have no idea what it looks like. Which one is it ?*

ELSA : *Let me choose.*

DONOVAN : *Thank you, Doctor.*

Elsa and Indy exchange looks. He thinks he is seeing her in her true colors.

ELSA chooses a cup—a solid gold, emerald encrusted goblet. DONOVAN instantly takes it from her.

DONOVAN : *Oh, yes. It's more beautiful than I'd ever imagined. This certainly is the cup of the King of Kings.*

DONOVAN rushes to the well and fills the goblet with water.

DONOVAN : *Eternal life !*

DONOVAN drinks from the goblet. Then, DONOVAN'S entire body starts to convulse. His face contorts in agony. He grabs his stomach and turns toward ELSA. [. . .] He starts to age—fast ! [. . .] As he falls his body breaks into flames, then shatters against the wall.

KNIGHT : *He chose . . . poorly.*

INDY studies the array of chalices.

ELSA : *It would not be made out of gold.*

INDY picks up another cup—a simple earthenware jug.

INDY : *That's the cup of a carpenter.*

He and ELSA exchange a look.

INDY : *There's only one way to find out.*

INDY goes to the well and fills the earthenware jug with water, then pauses. INDY brings the jug to his lips and takes several large swallows. A strange sensation overcomes him, a feeling of peace and contentment . . . and we see his wounds begin to heal.

Note préliminaire 1 : Le visionnage du film montre qu'il y a une trentaine de coupes sur l'autel, nous considérerons qu'il y en a 32 ($5 = \log_2(32)$, ce qui

permettra d'avoir des valeurs entières de complexité).

Note préliminaire 2 : Cette analyse a été publiée dans :
Saillenfest, Antoine & Dessalles, Jean-Louis (2014). Can Believable Characters Act Unexpectedly ?, Literary and Linguistic Computing (2014) 29(4) : 606-620.

8.1 Analyse en complexité

Étape 1 - Présentation de la loterie et des enjeux

At that moment DONOVAN and ELSA appear. DONOVAN turns to face the display of cups.

DONOVAN : Which one is it ?

KNIGHT : You must choose. But choose wisely. For as the True Grail will bring you life, the False Grail will take it from you.

DONOVAN : I'm not a historian. I have no idea what it looks like. Which one is it ?

Ce passage introduit une loterie. Les personnages se trouvent face à un autel sur lequel se trouvent les 32 coupes ; le choix d'une coupe peut conduire un acteur à deux conséquences possibles, une mort immédiate ou la vie éternelle. Seule l'une des coupes est le Graal, c'est la seule coupe qui permet d'accéder à l'immortalité. Les 31 autres coupes sont des leurres, de faux Graals, qui provoquent la mort immédiate de celles et ceux qui les utilisent.

Les conséquences sont telles qu'un lecteur attribuera un modèle standard (une attitude standard) des protagonistes à l'égard des conséquences. Ainsi, on peut proposer que le lecteur se représente ainsi l'attitude des personnages :

— chaque protagoniste désire avoir la vie éternelle et ne désire pas mourir immédiatement.
— un protagoniste désire qu'un allié ait la vie éternelle et ne désire pas le voir mourir.
— un protagoniste désire qu'un ennemi n'ait pas la vie éternelle et désire qu'il meure immédiatement.

Cela signifie en particulier que :

— Donovan veut la vie éternelle et ne veut pas mourir.
$\epsilon^{\mathcal{O}\to Donovan}(s_{Donovan_a_la_vie_eternelle}) = +1$
$\epsilon^{\mathcal{O}\to Donovan}(s_{Donovan_est_mort}) = -1$

— Indiana (Indy) Jones veut la vie éternelle et ne veut pas mourir.
$\epsilon^{\mathcal{O}\to Indy}(s_{Indy_a_la_vie_eternelle}) = +1$
$\epsilon^{\mathcal{O}\to Indy}(s_{Indy_est_mort}) = -1$

— Elsa veut que Donovan ait la vie éternelle et ne veut pas qu'il meure.
$\epsilon^{\mathcal{O}\to Elsa}(s_{Donovan_a_la_vie_eternelle}) = +1$
$\epsilon^{\mathcal{O}\to Elsa}(s_{Donovan_est_mort}) = -1$

Le spectateur, à ce stade, n'a pas d'information lui permettant de déterminer laquelle des coupes est le Graal. Le lecteur est informé que Donovan n'a pas non plus la connaissance nécessaire pour réaliser une action autre qu'un choix aléatoire parmi les coupes. Or, si Donovan réalise un choix aléatoire a_{rand}, l'inattendu de la situation $s_{Donovan_a_le_Graal}$ "Donovan a le Graal" vaut :

$U^{\mathcal{O}\to Donovan}(s_{Donovan_a_le_Graal}||a_{rand}) = 5bits$

En effet, le Graal étant unique, sa description ne consomme pas de bits. Par ailleurs, avec un choix aléatoire, l'inattendu de la situation $s_{Donovan_a_Un_Faux_Graal}$ "Donovan a un faux Graal" vaut :

$U^{\mathcal{O}\to Donovan}(s_{Donovan_a_Un_Faux_Graal}||a_{rand}) = -\log_2(31/32)bits \approx 0$

À ce stade de l'histoire, l'inattendu, pour Donovan, d'avoir la vie éternelle est équivalent à celui de trouver le Graal (soit de l'ordre de 5 *bits*), ce qui est source de nécessité négative. Or ce dernier désire vraiment la vie éternelle, ce qui est source de nécessité positive. Ainsi, il y a conflit cognitif dans l'esprit de Donovan, ce qui le pousse à agir pour résoudre le conflit.

Les chances de tomber sur le mauvais Graal sont élevées du point de vue de Donovan, celui-ci cherche donc des éléments pour calculer un scénario causal alternatif moins risqué lui permettant d'accéder plus certainement à la vie éternelle.

Mettre une trentaine de coupes sur l'autel plutôt que 3 n'est donc pas "un hasard", c'est un choix délibéré des auteurs. La réaction de Donovan, qui refuse de choisir une coupe et espère obtenir des informations supplémentaires qui permettraient de réduire les risques de mourir sont cohérentes avec l'attitude d'un personnage rationnel. Dans le cas présent, Donovan n'accorde pas à la perspective

d'avoir la vie éternelle une valeur suffisamment grande (par rapport à la perspective de mourir immédiatement) et donc refuse le choix aléatoire d'une coupe sur l'autel.

Étape 2 - Information partagée sur la connaissance d'Elsa

ELSA : Let me choose. DONOVAN : Thank you, Doctor.

Analyse en complexité Rappeler ici qu'Elsa est Docteur n'est pas anodin. Ce qui empêche Donovan de choisir aléatoirement une coupe est l'inattendu trop élevé de tomber sur le Graal en choisissant une coupe aléatoirement. Préciser qu'Elsa est docteur revient à indiquer au lecteur qu'elle est experte, et que Donovan en est conscient. Du point de vue de Donovan, Elsa est capable de réaliser une action (un choix) qui permet d'obtenir de manière quasi-certaine (avec un inattendu très faible) le Graal.

Ainsi, pour le spectateur qui prend le point de vue de Donovan, qui lui-même prend le point de vue d'Elsa, cette dernière est en mesure de réaliser une action a telle que :

$U^{\mathcal{O}\to Donovan\to Elsa}(s_{Elsa_a_le_Graal}||a) \approx 0$

Note : la notation $\mathcal{O} \to Donovan \to Elsa$ signifie que le lecteur pense que Donovan pense qu'Elsa (va très certainement choisir le bon Graal). Cette notation indique que le lecteur prend le point de vue de Donovan qui prendrait lui-même le point de vue d'Elsa.

Si on suppose que le spectateur lui-même a confiance en Elsa, alors on a également :

$U^{\mathcal{O}\to Elsa}(s_{Elsa_a_le_Graal}||a) \approx 0$

Étape 3 - Information partagée sur l'attitude d'Elsa envers Donovan

ELSA and INDY exchange looks. He thinks he is seeing her in her true colors.

Le spectateur et Indy sont informés de l'attitude d'Elsa envers Donovan. Cependant, plusieurs interprétations sont possibles.

Première interprétation : Alors que le lecteur pensait jusqu'ici qu'Elsa et Donovan étaient alliés, cette information (surprenante) vient modifier la connaissance de l'état mental d'Elsa pour le lecteur. En particulier, on a désormais :

$\epsilon^{\mathcal{O}\to Elsa}(s_{Donovan_a_la_vie_eternelle}) = -1$

$\epsilon^{\mathcal{O}\to Elsa}(s_{Donovan_est_mort}) = +1$

Le lecteur pense donc qu'Elsa veut voir mourir Donovan et ne veut pas qu'il ait la vie éternelle.

Le lecteur, surpris de découvrir qu'Elsa va trahir Donovan, va à partir de ce moment (et ça ne sera pas résolu à la fin de la séquence analysée ici) être curieux de savoir pourquoi Elsa agit ainsi.

Deuxième interprétation : Indiana Jones se rend compte qu'Elsa est bien alliée à Donovan et qu'il l'avait vraiment mal jugée, ce qui ne modifie pas l'état mental supposé d'Elsa. Dans le film, cet échange de regards est très furtif et le spectateur conserve l'idée qu'Elsa est alliée à Donovan.

Étape 4 - Action d'Elsa

ELSA chooses a cup—a solid gold, emerald encrusted goblet.

Parmi toutes les coupes présentes sur l'autel, Elsa choisit la plus belle (action a_{Elsa}) et l'a donc en sa possession (conséquence $s_{Elsa_a_la_plus_belle_coupe}$).

Le choix de la plus belle coupe n'est pas anodin, l'objet est unique grâce à une caractéristique simple, ce qui le rend donc très simple à décrire d'après l'équation 4.13 ($C_d \approx 0$). Le choix de la plus belle coupe provoque un saut de complexité du type $C_w - C_d \approx 5\ bits$ pour le spectateur, ce choix est surprenant car inattendu. En dépit d'informations sur les motivations du choix, cela revient à un choix aléatoire d'un objet très simple parmi 32 objets.

Ce choix surprenant affecte négativement la crédibilité d'Elsa tant que les motivations de son choix ne sont pas connues. Cette situation illustre le dilemme crédibilité-intérêt : Elsa réalise une action surprenante (pour le spectateur, In-

diana Jones et Donovan), ce qui crée de l'intérêt, mais ses intentions ne sont pas reconnues par le spectateur.

Ici, plus précisément, le spectateur ne parvient pas à déterminer les conséquences anticipées de l'acte d'Elsa. Si le spectateur suppose qu'elle n'est pas alliée à Donovan, alors son action est intentionnelle car l'inattendu de la mort de Donovan est approximativement égal à l'inattendu d'avoir choisi un faux Graal (qui est très faible)

$U^{\mathcal{O}\to Elsa}(s_{Donovan_est_mort}||a_{Elsa}) \approx 0$

Et donc :

$Int^{\mathcal{O}\to Elsa}(s_{Donovan_est_mort}, a_{Elsa}) \approx +E^{\mathcal{O}\to Elsa}(s_{Donovan_est_mort})$

L'action d'Elsa est donc bien reconnue comme étant intentionnelle par le lecteur si on la suppose non-alliée à Donovan.

Dans le film, il n'y a pas véritablement de raison de penser qu'Elsa n'est pas alliée à Donovan. Donc dans ce cas, le choix d'Elsa reste inattendu au regard de ses motivations supposées et l'intentionnalité de son choix est négative.

Étape 5 : Levée de l'incertitude de Donovan et action de Donovan

DONOVAN instantly takes it from her.

DONOVAN : *Oh, yes. It's more beautiful than I'd ever imagined. This certainly is the cup of the King of Kings.*

DONOVAN rushes to the well and fills the goblet with water.

DONOVAN : *Eternal life !*

DONOVAN drinks from the goblet.

En affirmant que le Graal est (avec certitude) la plus belle des coupes car c'est celle du "roi des rois", Donovan donne au spectateur (et aux autres protagonistes) une information sur son état mental et en particulier sur le fait qu'un saut de complexité de génération s'opère dans son esprit. Il nous informe que $C_w(s_{Elsa_a_le_Graal}||a_{Elsa}) = 0$, il est maintenant certain qu'Elsa a choisi le vrai Graal et donc qu'il va avoir la vie éternelle. Ce saut de complexité est de l'ordre

de 5 *bits*, entre la complexité *a priori* de trouver le bon Graal au hasard et la complexité nulle après le choix du plus beau. Ce saut de complexité de génération $\Delta C_w = 5\ bits$ est un exemple de 'simplication' (voir équation 4.27), une variation brutale de complexité de génération qui est source d'intérêt. C'est une baisse brutale de la curiosité (de Donovan et du spectateur) postérieure au choix surprenant (pour Donovan et le spectateur) de la plus belle coupe.

Par ailleurs, cela vient aussi faire baisser la complexité du choix d'Elsa pour Donovan. Rétrospectivement, le choix d'Elsa n'est plus inattendu pour Donovan :

$$U^{\mathcal{O}\to Donovan\to Elsa}(s_{Elsa_a_le_Graal}||a_{Elsa}) = 0$$

Et comme choisir le bon Graal conduit à donner la vie éternelle à Donovan :

$$U^{\mathcal{O}\to Donovan\to Elsa}(s_{Elsa_a_le_Graal}||a_{Elsa}) = U^{\mathcal{O}\to Donovan\to Elsa}(s_{Donovan_a_la_vie_eternelle}||a_{Elsa}) \approx 0$$

Ainsi :

$$Int^{\mathcal{O}\to Donovan\to Elsa}(s_{Donovan_a_la_vie_eternelle}, a_{Elsa}) \approx + E^{\mathcal{O}\to Donovan\to Elsa}(s_{Donovan_a_la_vie_eternelle})$$

Le lecteur pense donc bien, rétrospectivement, que Donovan pense qu'Elsa à l'intention de lui apporter la vie éternelle en choisissant la plus belle coupe.

L'action $a_{Donovan}$ de Donovan consiste à prendre le (supposé) Graal des mains d'Elsa, courir jusqu'au puits, remplir la coupe et boire son contenu (on considère que ceci constitue une seule et même action). Donovan étant certain qu'Elsa a choisi le vrai Graal, on a :

$$U^{\mathcal{O}\to Donovan}(s_{Donovan_a_la_vie_eternelle}||a_{Donovan}) = 0$$

Ainsi, l'action de Donovan est bien jugée intentionnelle par un observateur car :

$$Int^{\mathcal{O}\to Donovan}(s_{Donovan_a_la_vie_eternelle}, a_{Donovan}) = \\ + E^{\mathcal{O}\to Donovan}(s_{Donovan_a_la_vie_eternelle})$$

L'équation précédente signifie que l'action de Donovan est bien jugée comme étant intentionnelle et qu'elle est réalisée dans le but d'avoir, avec certitude, la vie éternelle.

Elle est aussi jugée intentionnelle par les autres protagonistes qui partagent la même connaissance de l'état mental de Donovan que le lecteur.

Étape 6 : Surprise et mort de Donovan

Then, DONOVAN'S entire body starts to convulse. His face contorts in agony. He grabs his stomach and turns toward ELSA. [. . .] He starts to age—fast! [. . .] As he falls his body breaks into flames, then shatters against the wall.

KNIGHT : *He chose . . . poorly.*

Du point de vue du lecteur qui pense qu'Elsa est une alliée de Donovan, la mort de Donovan est surprenante, le lecteur était certain qu'Elsa avait choisi le vrai Graal en prenant la plus belle coupe, et donc que Donovan aurait la vie éternelle. La surprise est d'autant plus grande que Donovan est un personnage très simple, c'est le principal ennemi d'Indiana Jones dans l'histoire.

$U^{\mathcal{O}}(s_{Donovan_est_mort}) >> 0$

L'échange de regard de l'étape 8.1 entre Indiana Jones et Elsa ré-intervient après la mort de Donovan. Cet échange de regard intervient comme facteur explicatif de la mort de Donovan *a posteriori*, il préparait une baisse rapide de la complexité de génération, ce que nous appelons une 'simplication'. *A posteriori*, la mort de Donovan n'est plus inattendue :

$U^{\mathcal{O}}(s_{Donovan_est_mort}) = 0$

En revanche, l'échange de regard (qui informe qu'Elsa n'est pas alliée de Donovan) explique la mort de Donovan mais n'explique pas l'attitude d'Elsa à l'égard de Donovan. Ainsi, l'attitude d'Elsa est surprenante et, ensuite source de curiosité (le fait qu'Elsa ne soit pas alliée de Donovan reçoit une nécessité $\nu^{\mathcal{O}}(s_{Elsa_non_alliee_de_Donovan}) >$

0, ce qui est source d'un conflit et de curiosité (voir partie 6.2.1 et l'équation 6.2). La curiosité à ce sujet demeure jusqu'à la fin de l'extrait étudié dans ce chapitre.

Étape 7 : Indiana Jones face au choix

INDY studies the array of chalices.

Indiana Jones se retrouve (à peu près) face à la même alternative que Donovan au début de l'extrait analysé. Un choix aléatoire n'est toujours pas une bonne solution pour résoudre le conflit auquel il est confronté.

Étape 8 : Partage de connaissance d'Elsa et première action d'Indiana Jones

ELSA : *It would not be made out of gold.*

INDY picks up another cup—a simple earthenware jug.

Pour la même raison que précédemment, le choix d'une coupe simple (action a_{1_Indy}), car elle a une caractéristique simple qui la rend unique (la coupe la moins belle), provoque chez le lecteur un saut de complexité maximal car la complexité de génération explose aussi (c'était le choix le moins attendu). Le lecteur est donc très surpris. La phrase d'Elsa laisse penser qu'elle a, en tant qu'experte, une explication (une complexité de génération faible) qui, de son point de vue, rend le choix d'une coupe qui n'est pas en or peu inattendu. On ne sait pas si Indiana Jones partage cette connaissance.

Calculer l'intentionnalité de l'action a_{1_Indy} est difficile à ce stade. Si on suppose qu'Indiana Jones réalise un choix aléatoire, alors la réalisation de l'action favorise beaucoup l'occurrence de la conséquence non désirée "Indiana Jones meurt". C'est une source de suspense, car la complexité de génération de la situation désirée ("Indiana Jones a la vie éternelle") est très élevée (voir équation du suspense 6.3).

Si on suppose qu'il choisit cette coupe de manière rationnelle, autrement dit qu'il a bien l'intention d'avoir la vie éternelle en la choisissant, alors le lecteur manque d'information pour expliquer le choix de la coupe la moins belle (et faire baisser l'inattendu lié à ce choix). Ce choix est donc source de curiosité car il y a

une nécessité très négative associée au choix (voir équation de la curiosité 6.2).

Quoi que l'on suppose, cette action crée un conflit d'ordre cognitif pour le lecteur qui prend le point de vue d'Indiana Jones autour de sa survie, qui reçoit deux nécessités de signes opposés (Indiana Jones est supposé vouloir vivre mais sa survie est inattendue car il est inattendu qu'il ait choisi le bon Graal en choisissant la coupe la moins belle).

Étape 9 : Explication du choix d'Indiana Jones - Saut de complexité

INDY : *That's the cup of a carpenter.*

Cette phrase vient résoudre le conflit précédent. La phrase d'Elsa dans l'étape précédente avait préparé cette explication. L'action d'Indiana Jones est bien intentionnelle, il a bien l'intention d'avoir la vie éternelle. On a :

$U^{\mathcal{O}\to Indy}(s_{Indy_a_la_vie_eternelle}||a_{1_Indy}) = 0$

Et donc :

$Int^{\mathcal{O}\to Indy}(s_{Indy_a_la_vie_eternelle}, a_{1_Indy}) = E^{\mathcal{O}\to Indy}(s_{Indy_a_la_vie_eternelle})$

Par ailleurs, ce saut de complexité (une baisse rapide) est à nouveau une 'simplication' qui, rétrospectivement, rend le choix de la coupe la moins belle non inattendu. Un premier saut de complexité de génération d'une situation simple à décrire provoque un effet de surprise, la révélation ultérieure de l'explication causale ayant conduit à la production de cette situation provoque un second saut (une baisse rapide) de complexité (de génération) et rend *a posteriori* la situation non inattendue. La 'simplication' rend le choix de la coupe la moins belle postdictible.

Étape 10 : Deuxième action d'Indiana Jones

He and ELSA exchange a look.

INDY : *There's only one way to find out.*

INDY goes to the well and fills the earthenware jug with water, then pauses.

L'indication *"[Indiana Jones] pauses"* n'est pas fortuite. Elle indique que, malgré un choix rationnel de la coupe, l'issue de l'action a_{2_Indy} reste incertaine

pour Indiana Jones. Ainsi, à ce stade :

$U^{\mathcal{O}\rightarrow Indy}(s_{Indy_a_la_vie_eternelle}||a_{2_Indy}) > 0$

Pour le spectateur, qui partage cette incertitude, cela crée un suspense (voir équation 6.3), l'issue souhaitée est inattendue.

Étape 11 : Troisième et dernière action d'Indiana Jones

Cette dernière action a_{3_Indy} lève l'incertitude qui créait du suspense, et vient clore le conflit cognitif autour de la recherche du Graal. L'occurrence de l'issue étant légèrement incertaine, cela provoque un saut d'inattendu (et donc de l'intérêt) lorsque ladite issue se produit.

$U^{\mathcal{O}}(s_{Indy_a_la_vie_eternelle}) > 0$

À la fin de l'action, il ne reste, du point de vue du spectateur (et d'Indiana Jones) qu'un conflit cognitif non résolu, l'attitude d'Elsa envers Donovan (ce n'est pas, en réalité, le seul conflit non-résolu, mais nous avons ici limité notre étude au conflits introduits dans la partie du scénario étudiée, en omettant ceux qui pré-existent au début de la scène).

8.2 Bilan de l'analyse

L'analyse précédente vise à proposer un nouveau mode d'analyse des scénarios basé sur la notion de complexité, et en particulier montre comment les modèles introduits dans les chapitres antérieurs permettent de réaliser une analyse en complexité d'un scénario existant. Dans le tableau 8.1 nous avons repris les éléments de pertinence de l'analyse précédente.

Nous pouvons voir que cet extrait a été optimisé du point de vue de la pertinence. Les alternatives proposées aux personnages sont extrêmes, du point de vue émotionnel. Les personnages doivent choisir entre la vie éternelle et la mort immédiate. Par ailleurs, tout élément de cet extrait contribue à la création d'un effet en termes de pertinence, que cela soit pour créer de l'intérêt ou pour maintenir la crédibilité des personnages. Ainsi, cet extrait est un bon exemple de pertinence totale, c'est-à-dire un exemple dans lequel chaque élément présenté au spectateur, chaque mot prononcé par un personnage, chaque action, chaque événement

TABLE 8.1 – Critères de pertinence narrative dans l'extrait d'Indiana Jones et la Dernière Croisade

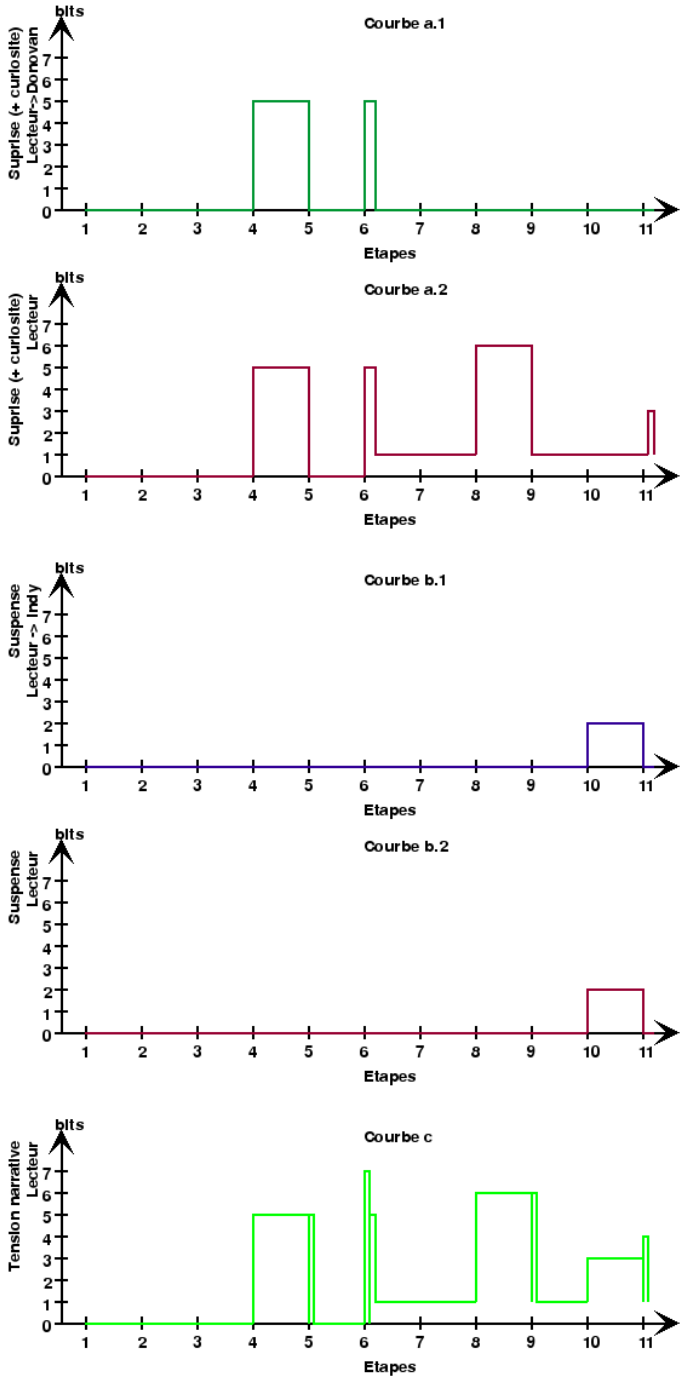
Critère de pertinence	***Commentaire***
Les valeurs extrêmes de l'émotion	Les enjeux extrêmes augmentent l'intensité émotionnelle donc l'intérêt (équation 4.5).
La simplicité des personnages et des coupes	Les trois personnages sont extrêmement simples. Indiana Jones est le personnage principal, Donovan l'ennemi principal et Elsa le personnage féminin principal. Les coupes choisies sont la plus belle et la moins belle. Cela fait baisser la complexité de description et augmente l'intérêt (voir partie 4.3.4 et équation 4.4)
Le nombre de coupes sur l'autel	Le fait qu'il y ait une trentaine de coupes sur l'autel n'est pas fortuit, la complexité du choix aléatoire est de 5 *bits*. Cela crée un conflit chez Donovan (lié au risque de mourir) et contribue à créer de la surprise quand Elsa ou Indiana Jones choisissent les coupes les plus simples (voir équation 6.1).
Surprise, conflit et curiosité	Les choix des coupes et la mort de Donovan sont des situations surprenantes et sources de conflits car très peu nécessaires *a priori* et nécessaires *a poteriori*. Ces nécessités nouvelles qui créent les conflits sont sources de curiosité pour le spectateur (voir équation 6.2)
Les personnages sont crédibles	Les intentions des personnages peuvent être évaluées à chaque étape (voir équations 7.4 et 7.7). Quand Elsa ou Indiana Jones choisissent des coupes, il y a mise en danger de leur rationalité pour créer de l'intérêt (ce sont des situations de dilemme intérêt-crédibilité (voir le système 7.14))
Simplication post-surprise	Cet extrait contient trois exemples de simplication qui font chuter brutalement la complexité de génération des situations surprenantes (voir 4.27). Ces exemples de simplication suivent tous le même schéma : un élément qui prépare la simplication est présenté avant une situation surprenante, ensuite la situation surprenante est présentée et enfin la simplication a lieu en utilisant l'élément présenté *a posteriori*.
Hésitation d'Indiana Jones et suspense	Il y a un suspense lorsqu'Indiana Jones hésite à utiliser la coupe. Cette hésitation rend la perspective d'une issue favorable à Indiana Jones plus incertaine (l'augmentation de l'inattendu *a priori* crée du suspense d'après l'équation 6.3)
Quasi-totale postdictibilité de la scène	Les situations inattendues de cette extrait sont postdictibles suite aux simplications réalisées (baisse brutale de complexité de génération suite à l'occurrence de situations surprenantes et révélation de l'issue "Indiana Jones a la vie éternelle" qui met fin au suspense). L'attitude d'Elsa envers Donovan n'est cependant pas postdictible. Une explication est présentée ultérieurement dans le récit.

contribue à rendre le récit pertinent.

L'analyse en complexité nous permet par ailleurs de tracer des courbes de variation de complexité du point de vue des personnages et donc de l'intérêt narratif pour le lecteur (voir figure 8.1). Comme on peut le voir dans cet extrait, l'intérêt est ressenti de manière directe ou par transitivité (quand un personnage du récit est surpris par exemple). Nous considérons que la surprise d'un personnage est celle ressentie par le spectateur car l'ensemble des personnages présents est bien défini et chacun est simple pour le lecteur (ce sont les personnages principaux du récit).

Ceci nous permet donc de tracer une courbe d'évolution globale de l'intérêt narratif. La scène ne se termine pas avec une résolution totale des conflits (du point de vue du lecteur). L'attitude d'Elsa envers Donovan est toujours source de conflit à la fin de cet extrait, aucune explication n'a été fournie qui permet d'expliquer le comportement d'Elsa.

FIGURE 8.1 – *Courbes, en fonction des étapes définies dans la partie 8.1, des variations de surprise (et l'intérêt argumentatif créé par les conflits non résolus) dans l'esprit de Donovan (courbe a.1) et celui du lecteur (courbe a.2), des variations de suspense du point de vue d'Indiana Jones (courbe b.1) et du spectateur (courbe b.2) et des variations de l'intérêt cumulé (suspense + surprise + curiosité post-surprise + émotion standard) pour le spectateur (courbe c)*



Chapitre 9

Approche cognitive de la génération automatique des récits

Dans les précédents chapitres, nous avons présenté une théorie cognitive de l'intérêt narratif, la théorie de la Simplicité. Nous avons montré comment définir, à partir des notions introduites dans ce modèle la notion de conflit cognitif. Nous avons montré que le conflit cognitif est source d'intérêt, à la fois pour ceux qui y sont confrontés mais aussi pour un observateur extérieur conscient de la présence d'un conflit interne à un individu. Nous avons par ailleurs montré que le conflit est source d'action, les individus qui font face à des conflits calculent des plans, c'est-à-dire des actions à réaliser, dans le but de résoudre lesdits conflits.

Dans le chapitre 5, nous avons montré que la notion de conflit, dans le cadre narratif, pouvait être assimilée à un ensemble de conflits, d'ordre cognitif, internes aux acteurs rationnels d'un récit. Nous avons par ailleurs émis l'idée qu'un récit est équivalent à une suite ordonnée de conflits narratifs présentés à l'audience dans le but de produire un effet (*ie.* créer une tension narrative) sur cette dernière.

Par ailleurs, dans le chapitre 6, nous avons présenté un modèle cognitif de la surprise basé sur la notion de conflits. Ce modèle vise à rendre compte de l'attitude d'un individu suite à l'occurrence d'un événement surprenant.

Dans cette partie, nous nous intéressons à l'application des modèles cognitifs introduits dans les chapitres précédents, dans le but de générer automatiquement des scénarios en utilisant une approche cognitivement plausible (nous détaillerons

des conditions de cette approche dans la partie 9.2). Le conflit étant un élément central de la structure d'un récit, nous allons proposer un modèle de génération basé sur la formation, l'identification et la résolution de conflits internes à un personnage. Nous allons en particulier montrer que ce modèle peut être utilisé dans le cadre du développement d'un générateur de scénario pour produire automatiquement des séquences narratives configurées par la surprise.

Ce modèle étant basé sur la résolution de conflits, c'est-à-dire en particulier à la formation de plans par les personnages rationnels du récit qui font face aux conflits, nous introduirons un modèle de résolution de conflit basé sur la notion d'abduction que nous utiliserons comme système de planification à l'échelle d'un acteur rationnel. Nous montrerons qu'un tel modèle permet de générer des plans qui, du point de vue du personnage, sont intentionnels (au sens de l'équation 7.6).

Enfin, nous proposerons une analyse de notre approche. En particulier, nous monterons que les séquences narratives générées respectent les critères de pertinence émotionnelle et structurelle introduits précédemment.

Note préliminaire : Les conditions d'une approche cognitive et l'utilisation de la procédure CAN pour la génération de plans intentionnels ont été publiées : *Saillenfest, Antoine & Dessalles, Jean-Louis (2014). A Cognitive Approach to Narrative Planning with Believable Characters, In Mark A. Finlayson, Jan Christoph Meister, Emile G. Bruneau (Eds), 2014 Workshop on Computational Models of Narrative, OASIcs 41,177-181, Leibniz-Zentrum für Informatik, Dagstuhl, Germany : Schloss Dagstuhl.*

9.1 Modèle de la surprise et génération de récits surprenants

9.1.1 Un modèle de génération récursif de séquences narratives surprenantes

Dans le chapitre 6, nous avons introduit un modèle de l'attitude d'un individu suite à l'observation de l'occurrence d'une situation surprenante (ce que nous

avons appelé modèle de la surprise configuré par les conflits, voir 6.4). Nous faisons l'hypothèse qu'un personnage rationnel dans la fiction peut être traité comme un humain. Ainsi, nous pouvons utiliser notre modèle de l'attitude suite à l'observation d'une situation surprenante aux personnages dans une fiction.

Nous faisons également l'hypothèse que l'audience d'un récit partage l'émotion des personnages du récit (voir la partie 6.3 pour une discussion sur les émotions ressenties par l'audience d'un récit par procuration). Ainsi, nous supposons que la surprise ressentie par un personnage l'est aussi par l'audience d'un récit.

Dans le but de générer des séquences narratives intéressantes, nous souhaitons en particulier générer des séquences configurées par la surprise. D'après ce qui précède, nous allons en particulier nous intéresser à la génération de séquences dans lesquelles un personnage est surpris (ce qui peut être présenté à l'audience pour que celle-ci partage l'émotion) et appliquer à ce personnage le modèle de la surprise présenté dans le chapitre 6 pour simuler sa réaction suite à l'occurrence de l'événement surprenant.

Ainsi, dans les séquences narratives que nous souhaitons générer, le personnage sera confronté à l'occurrence d'un événement surprenant, cela provoquera une mise à jour de sa connaissance et *a minima* un conflit cognitif interne à sa connaissance (il va être amené à calculer une hypothèse pour expliquer ce qu'il a observé). *Via* le partage des émotions, l'audience est, elle aussi, surprise et cherche donc aussi, par application du modèle de la surprise, à expliquer l'occurrence de l'événement. Comme l'indique le critère de pertinence émotionnelle lié à la postdictibilité (critère 7.17), l'audience doit pouvoir expliquer rétrospectivement tous les événements du récit. Ce qui n'est pas nécessairement le cas du personnage. Le personnage peut, à la fin de l'histoire racontée, être encore dans un état d'incertitude concernant l'occurrence de certains événements. Ainsi, nous souhaitons générer des séquences narratives dans lesquelles :

— un personnage est surpris et l'audience, qui doit avoir à sa disposition suffisamment d'éléments pour reconnaître que le personnage est surpris, partage sa surprise.
— une explication (pour intégrer la situation surprenante) est fournie à l'audience (pas nécessairement au personnage).

Nous souhaitons que les séquences narratives configurées par la surprise fassent

'avancer le récit', nous souhaitons que l'événement surprenant soit source d'action pour le personnage. Ainsi, nous souhaitons que les événements surprenants rencontrés par ce personnage le mettent dans une situation d'inconfort (en éloignant la perspective d'un état désiré ou en le rapprochant d'un état non-désiré du monde).

Or si la situation surprenante vient empêcher la réalisation d'un plan destiné à résoudre un conflit, alors le personnage fait à nouveau face à ce conflit une fois la situation surprenante observée, et donc la surprise sera bien source d'action pour tenter, à nouveau, de résoudre le conflit.

D'après l'ensemble des remarques précédentes, nous proposons un modèle récursif de génération de scénario basé sur la génération récursive de séquences configurées par la surprise ressentie par un personnage et qui l'empêchent de résoudre un conflit (principal), déclencheur initial de son action. Ce modèle est présenté dans la figure 9.1.

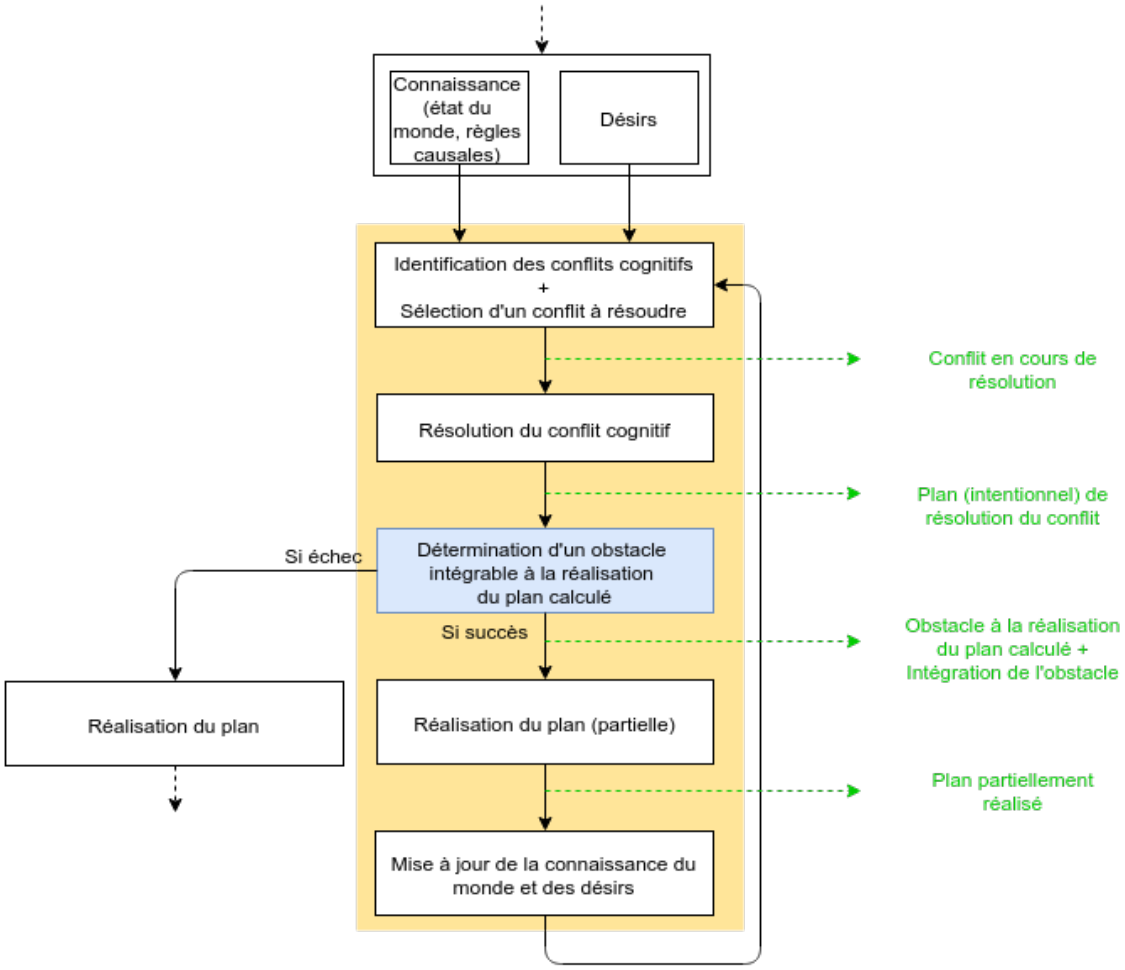
9.1.2 Les structures événementielles enrichies CRRS (Conflit-Résolution-Réalisation-Surprise)

Parce que nous devons pouvoir justifier les actions du personnage, nous proposons de conserver la mémoire de ses intentions, c'est-à-dire des plans calculés pour résoudre les conflits auxquels ce personnage est confronté. Ainsi, le modèle de génération précédent permet de construire des séquences événementielles enrichies CRRS (conflit-résolution-réalisation-surprise) qui contiennent :

- un plan calculé par le personnage-acteur qui résout un conflit
- la sous-partie du plan qui est (effectivement) réalisée par le personnage-acteur
- une situation inattendue qui empêche la réalisation totale du plan (surprenante pour le personnage-acteur)
- l'intégration (une 'explication') de la situation inattendue qui perturbe la réalisation du plan (pour l'audience)

Le modèle de génération décrit plus haut consiste à générer récursivement des structures événementielles CRRS jusqu'à ce que les plans calculés pour résoudre les conflits ne puissent plus être perturbés (une perturbation est un élément qui empêche la réalisation d'un plan calculé ; ici il s'agit de la non-validité d'une

FIGURE 9.1 – Modèle de génération d'une séquence événementielle de résolution de conflit par perturbations successives de la réalisation des plans calculés par l'acteur. Le modèle génère des séquences CRRS (Conflit-Résolution-Réalisation-Surprise) successives, jusqu'à ce que le système ne puisse plus perturber la réalisation d'un plan calculé, et donc laisse le conflit être résolu, ce qui marque la fin de l'histoire.



précondition qui empêche la réalisation d'au moins une action du plan). Ainsi après la génération de la dernière structure CRRS, le personnage est censé pouvoir calculer un plan pour résoudre les conflits et le réaliser dans son intégralité (s'il ne peut pas ultimement résoudre les conflits, nous avons affaire à un problème

lié au non-respect du critère de postdictibilité).

Les structures CRRS peuvent servir comme outil de décomposition et d'analyse des histoires configurées par la surprise (du point de vue d'un personnage). Prenons l'exemple suivant :

> « John veut récupérer le trésor qui se trouve dans le château, pour cela il décide de traverser le pont, ce qui lui permettra de prendre le chemin qui mène au château et d'aller chercher le trésor. En arrivant près du pont, il constate que celui-ci s'est effondré. C'était un pont en bois et le bois a pourri. Il ne peut pas le réparer donc il décide de prendre un chemin plus long à travers la forêt. En prenant ce chemin, il réussit à rejoindre le château et à récupérer le trésor. »

Cette histoire peut être décomposée en une structure CRRS et une structure CRR (conflit-résolution-réalisation), autrement dit une structure finale du processus de génération présenté précédemment. L'histoire peut donc être décomposée ainsi :

CRRS :

Conflit : avoir le trésor

Plan calculé : aller jusqu'au pont, le traverser, prendre le chemin qui mène au chateau, entrer dans le château, prendre le trésor

Plan réalisé : aller jusqu'au pont

Intégration d'élément surprenant : Le pont a pourri, il s'est effondré

CRR :

Conflit : avoir le trésor

Plan calculé : Prendre un chemin plus long, prendre le chemin qui mène au château, entrer dans le château, prendre le trésor

Plan réalisé : Prendre un chemin plus long, prendre le chemin qui mène au château, entrer dans le château, prendre le trésor

*(Intégration d'élément surprenant : *void*)*

L'action de traverser le pont ne sera jamais réalisée mais elle est importante pour comprendre les intentions initiales de John. Le plan calculé dans la structure CRR tient compte de la mise à jour de la connaissance (le pont est cassé, ce qui n'était pas supposé *True* (au sens logique) précédemment). Ainsi, l'action de traverser le pont est considérée comme non-réalisable (elle est affectée d'une

nécessité négative très forte).

Le mécanisme de génération présenté plus haut consiste en l'identification de conflits dans l'esprit d'un personnage et le calcul par ce personnage de plans pour les résoudre. Après avoir présenté des conditions d'une approche cognitive du processus de génération de scénarios, nous présenterons la procédure CAN, une procédure de raisonnement basée sur la résolution de conflits *via* un mécanisme d'abduction (Dessalles, 2008b). Nous allons voir que la procédure CAN est une procédure cognitivement plausible de raisonnement que nous pouvons utiliser pour le calcul de plans intentionnels pour la résolution de conflits cognitifs.

9.2 Conditions d'une approche cognitive de la génération de scénarios

Le problème classique dans le domaine de la génération de scénario est celui qui consiste à générer la *fabula*, c'est-à-dire la génération d'une séquence ordonnée temporellement d'événements entre le début et la fin (au sens temporel de ces notions) de l'histoire (Bal, 1997). Ce processus de génération doit respecter certains impératifs pour former un récit acceptable. Les événements doivent respecter les règles causales du monde fictif et l'audience doit être capable de déterminer les intentions des personnages au cours du récit (Bates, 1994; Graesser et al., 1994) (voir aussi la partie 7.3.1).

Les actions réalisées par les personnages ne doivent pas affecter négativement la suspension consentie de l'incrédulité de l'audience (Bates, 1994). L'un des aspects essentiels de la crédibilité des personnages est leur intentionnalité, c'est-à-dire la manière avec laquelle ils choisissent d'agir ou de se comporter afin que leurs actes apparaissent comme naturels (voire rationnels) pour un observateur donné (voir la partie 7.3.1 pour une définition de l'intentionnalité).

Nous allons détailler dans cette partie les conditions d'une approche cognitive de la génération de récits (Saillenfest and Dessalles, 2014b).

Dans le domaine de la génération de scénario, il a été montré que le comportement des personnages doit être perçu comme étant *goal oriented* (motivé par la réalisation d'un objectif) pour que lesdits personnages apparaissent comme

crédibles (Riedl and Young, 2010). Cependant, tous les systèmes *goal-oriented* ne sont pas cognitivement plausibles.

Plusieurs systèmes utilisent une approche basée sur la manipulation de scripts (*script-based models*) (Gervàs, 2009). Ces modèles ne font pas partie des approches cognitivement plausibles. L'être humain utilise des scripts (les scripts sont des structures qui contiennent un ensemble d'événements - dont des actions - ordonnés pour réaliser une tâche donnée dans un certain contexte) pour réaliser un grand nombre de tâches quotidiennement, par exemple pour boire ou aller à la boulangerie. Cependant, un système de génération de scénarios doit être capable de modifier des scripts à la volée et, en l'absence de script approprié, de produire un plan inédit en utilisant un ensemble de connaissances sur le monde fictif (par exemple des états et des règles causales).

Précédemment, nous avons mentionné la nécessité d'avoir un système *goal-oriented* afin de satisfaire à la contrainte de crédibilité des personnages. Cependant, du point de vue cognitif, il n'est pas raisonnable d'imaginer que l'être humain possède, de manière permanente, un ensemble possiblement infini de buts prêts à être 'satisfaits'. Un modèle cognitivement plausible de génération de plans (*ie.* un ensemble ordonné d'actions à réaliser) au niveau de l'agent devrait plutôt générer les plans, non pas à partir d'une liste de buts pré-existants, mais à la volée, en se basant sur les éléments qui créent des conflits cognitifs (voir chapitre 5 pour une définition des conflits cognitifs) internes à cet agent.

Un modèle cognitivement plausible pour la génération de plans à l'échelle d'un personnage ne doit pas faire appel à un oracle externe qui lui apporterait des éléments d'information sur la valeur de vérité de telle ou telle proposition. Les plans sont les résultats de processus de raisonnement internes à l'agent calculés du point de vue de cet agent, en utilisant uniquement sa propre connaissance du monde et ses préférences.

Un modèle cognitivement plausible de calcul des plans doit réaliser un calcul séquentiel des plans. Plusieurs études en psychologie suggère que le processus de planification chez l'être humain apparentée à un processus de planification à ordre-partiel (Rattermann et al., 2001). Les systèmes de planification à ordre-partiel construisent des plans en manipulant des plans partiels et en révoquant (si nécessaire) seulement certaines parties du plan global. Les performances de l'être

humain en termes de temps de calcul ou de nombre d'opérations réalisées sont proches de celles des systèmes de planification à ordre-partiel.

Enfin, un modèle cognitivement plausible de planification n'est pas censé générer un plan optimal qui serait celui qui reçoit la meilleure évaluation (selon un ensemble variés d'objectifs). Nous attendons d'un système cognitivement plausible qu'il génère des plans qui soient (*a minima*) acceptables.

Dans la suite, nous allons introduire la notion d'abduction. Nous allons définir l'abduction et nous montrerons les liens entre cette méthode de raisonnement et le processus de création littéraire, en particulier la génération de scénarios. Ensuite, nous présenterons un modèle de raisonnement délibératif pour la résolution de conflits cognitifs basé sur l'abduction. Puis nous montrerons comment utiliser ce modèle pour générer des séquences d'actions intentionnelles à l'échelle d'un personnage.

9.3 CAN : une procédure cognitivement plausible de résolution des conflits

9.3.1 L'abduction, un processus de raisonnement délibératif et de création littéraire

Dans Éthique à Nicomaque, Aristote décrit ainsi le processus de calcul des hypothèses (ou des plans) pour parvenir à atteindre un but :

> « [...] Nous ne délibérons pas des fins, mais des moyens pour y parvenir. [...] une fois qu'on a posé la fin, on regarde la question de savoir et par quels moyens on peut l'atteindre et si plusieurs moyens paraissent en mesure de l'atteindre, on examine quel est le plus facile et le plus beau. Mais s'il n'y en a qu'un seul pour arriver à cette fin, on cherche comment il permet d'y arriver et aussi par quel moyen ce moyen lui-même peut être atteint, jusqu'à pouvoir parvenir au premier moyen dans la chaîne causale, lequel, dans la recherche, est l'ultime chose à découvrir. » (Aristote, nc, p.146-147 (1112b))

Le processus décrit par Aristote est similaire à l'abduction. L'abduction, telle

qu'elle est définie par Peirce (CP 5.189) consiste à générer une hypothèse qui explique une observation. Peirce décrit ainsi le fonctionnement de l'abduction :

> "*The surprising fact, C, is observed ; But if A were true, C would be a matter of course, Hence, there is reason to suspect that A is true.*" (CP 5.189)

Autrement dit, l'abduction est un raisonnement qui prend comme point de départ des conséquences observées et tente de remonter aux causes. Pour reprendre les mots de Sylvie Catellin :

> « Le point de départ de l'abduction est un fait ou un phénomène perçu comme surprenant, inattendu, ou s'inscrivant contre ce qui était jusqu'alors tenu pour acquis. [...] l'abduction est une procédure de normalisation d'un fait aberrant, un effort de raisonnement que l'on entreprend lorsqu'il y a rupture de notre système d'attentes, un raisonnement imaginatif qui fait appel à nos connaissances implicites. » (Catellin, 2014, p.73-74)

L'abduction est centrale dans l'intelligence humaine (Hobbs et al., 1993), en particulier dans les contextes de la résolution de problèmes ou de diagnostic (Magnani, 2001; Ganascia and Debru, 2007; Habib and Ganascia, 2009). Elle joue également un rôle très important dans la création littéraire. En particulier, c'est le type de raisonnement utilisé pour résoudre une enquête à partir d'indices. Zadig, qui à partir d'indices est capable de produire des hypothèses sur tel ou tel animal qu'il n'a pas vu, utilise l'abduction. Un autre exemple célèbre d'utilisation de raisonnement abductif en littérature est le personnage de Sherlock Holmes. La sagacité de Sherlock Holmes, qui lui permet de résoudre les enquêtes, est le résultat d'un raisonnement de type abduction créative (Eco, 1992, p.278). D'une manière générale, l'abduction est le modèle de raisonnement à l'œuvre dans les récits d'enquête, policière ou médicale. Nous pouvons citer les séries télévisées *House M.D.* (Dr House) ou *CSI : Crime Scene Investigation* (Les Experts) comme exemples de récits basés sur l'abduction.

9.3.2 La procédure Conflit-Abduction-Négation

La procédure CAN (Conflit-Abduction-Négation) (Saillenfest and Dessalles, 2014b; Dessalles, 2008a) est basée sur la résolution de conflits *via* l'abduction. Cette procédure n'est pas *goal-oriented* mais *problem-oriented* : les objectifs sont générés à la volée, en fonction des conflits cognitifs internes (à un agent), au moment où ils sont détectés. Le résultat de la procédure CAN, qui consiste en un ensemble d'actions et de justifications, peut conduire un observateur à percevoir le comportement de l'agent comme étant *goal-oriented*, tout en ignorant comment les objectifs (*goals*) ont été générés.

Dans la partie 5, nous avons introduit la notion de conflit cognitif. L'origine des conflits cognitifs vient d'évaluations - les nécessités - opposées d'une même situation (ou d'un même prédicat), par exemple lorsqu'une situation non désirée (qui reçoit donc une évaluation de valence négative) se produit (elle reçoit également une évaluation de valence positive et de grande intensité car l'inattendu de sa non-occurrence est élevé). En particulier, nous avons montré que les événements surprenants étaient sources de conflits cognitifs (voir chapitre 6).

L'abduction a pour but de "résoudre" les conflits cognitifs, ce qui peut consister à expliquer les conflits, apporter une justification, ou encore à déterminer les actions à réaliser pour faire baisser l'intensité du ou des conflits identifiés. Ce qui est parfois appelé processus de "normalisation d'un fait aberrant" (Catellin, 2014) ou encore intégration d'un fait surprenant suite à l'occurrence d'une situation surprenante est en fait, du point de vue de l'agent confronté à cette surprise, un processus de détermination d'une hypothèse explicative qui permet de faire baisser l'intensité d'un conflit cognitif.

Voici une description du fonctionnement de la procédure CAN (voir également la figure 9.2) :

——————————La procédure Conflit-Abduction-Négation——————————

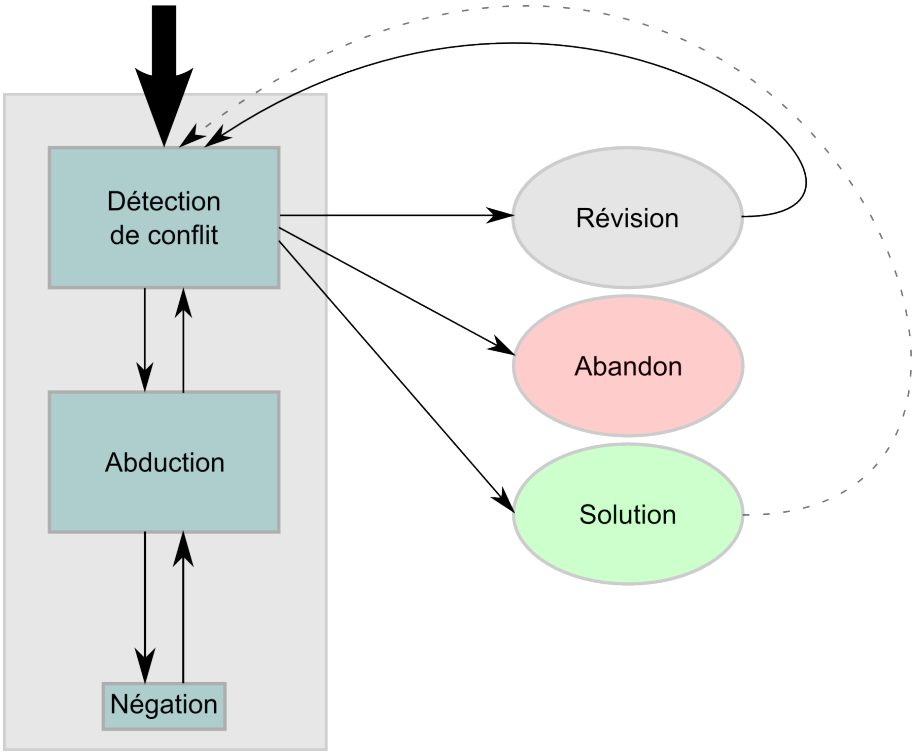
— Étape 1 : Un conflit cognitif est détecté, c'est-à-dire qu'une situation s reçoit deux évaluations de valences opposées $v_1 < 0$ et $v_2 > 0$. La procédure de résolution de ce conflit commence
— Étape 2 : La procédure réalise une abduction à partir de s, elle cherche une cause "faible" qui fait baisser l'intensité du conflit. Si l'évaluation de

la cause est plus faible que $|v_1|$, alors le conflit porte désormais sur cette cause.

— Étape 3 : Si la phase d'abduction échoue, v_2 est remplacée par $|v_1|$. La procédure est lancée à nouveau à partir de la négation de s (la négation de s est aussi une source de conflit).

— Étape 4 : Si la procédure ne parvient pas à calculer une solution qui résout le conflit, soit une évaluation doit être révisée (et la procédure recommence), soit la procédure se termine sans fournir de solution.

FIGURE 9.2 – Procédure Conflit-Négation-Abduction (Dessalles, 2008a)



La procédure CAN satisfait aux conditions de plausibilité cognitive énoncées précédemment. Les objectifs d'un agent sont générés à la volée à partir des conflits cognitifs détectés. Les plans pour résoudre ces conflits n'utilisent que les valeurs internes (la connaissance interne) de l'agent qui les calcule. Par ailleurs, la procédure

CAN est assimilable à un système de planification à ordre-partiel. Les plans ne sont pas recalculés intégralement lorsqu'une contradiction est détectée.

Présentons un exemple de raisonnement (réalisé par un personnage, John) pour trouver un trésor qui utilise la procédure CAN (on trouvera après cet exemple un détail, pour chacune des possibilités, des différentes étapes de la procédure)

John veut le trésor, mais ne l'a pas encore trouvé.
John sait que le trésor se trouve dans le château. John doit aller dans le château qui se trouve de l'autre côté du pont.
John doit traverser le pont et John se trouve juste avant le pont.
Possibilité numéro 1 : ... [Le pont n'est pas cassé] John décide de traverser le pont, d'aller au château et de prendre le trésor.
Possibilité numéro 2 : ... [Le pont est cassé] John ne peut rien faire pour changer cet état de fait.
John ne peut pas traverser le pont.
John peut atteindre l'autre côté du pont en empruntant un chemin plus long.
John décide de prendre le chemin le plus long pour aller de l'autre côté du pont, puis au château pour prendre le trésor.

——Analyse du calcul du plan (possibilité numéro 1) ——-
Conflit : "J'ai le trésor" désiré mais non-vrai
Abduction : "action : Prendre le trésor"
–**Propagation sur** "Je suis dans le château"
–**Abduction :** "action : Aller dans le château"
– –**Propagation sur** "Je suis de l'autre côté du pont
– –**Abduction :** "action : Traverser le pont"
– – –**Propagation sur** "Je suis près du pont" ET "Le pont n'est pas cassé"
– – –**Fin de la procédure : succès**
– –**Fin de la procédure : succès**
–**Fin de la procédure : succès**
Fin de la procédure : succès
Plan calculé : Traverser le pont, Aller dans le château, Prendre le trésor

——Analyse du calcul du plan (possibilité numéro 2) ——-
Conflit : "J'ai le trésor" désiré mais non-vrai
Abduction : "action : Prendre le trésor"
–**Propagation sur** "Je suis dans le château"
–**Abduction :** "action : Aller dans le château"
– –**Propagation sur** "Je suis de l'autre côté du pont
– –**Abduction :** "action : Traverser le pont"
– – –**Propagation sur** "Je suis près du pont" ET "Le pont n'est pas cassé"
– – –**Abduction impossible** à partir de "Le pont n'est pas cassé"
– – –**Négation :** ¬"Le pont n'est pas cassé"
– – –**Abduction impossible** à partir de ¬"Le pont n'est pas cassé"
– – **Fin de la procédure : échec** "action : Traverser le pont" impossible
– –**Abduction :** "action : Prendre le long chemin"
– – –**Propagation sur** "Je suis près du pont"
– – –**Fin de la procédure : succès**
– –**Fin de la procédure : succès**
–**Fin de la procédure : succès**
Fin de la procédure : succès
Plan calculé : Prendre le long chemin, Aller dans le château, Prendre le trésor

9.3.3 La procédure conflit-abduction : un modèle argumentatif pour la planification d'actions

≪ Et si d'aventure, on se heurte à une impossibilité, on renonce à agir, comme par exemple si l'on a besoin d'argent mais qu'on n'est pas en mesure de s'en procurer. Mais si une possibilité apparaît, on se met à l'action. Or sont possibles les choses qu'on peut faire par nous-mêmes. Celles qui se font en effet par l'intermédiaire des amis sont faites d'une certaine façon par nous-mêmes, puisque leur point de départ est en nous. Par ailleurs, quand on a quelque chose à faire, on recherche tantôt les outils, tantôt la façon de les utiliser. Or il en va de même dans toutes les autres recherches : tantôt on cherche par quel moyen atteindre une fin, tantôt de quelle façon ou par quel autre moyen celui-ci peut être mis en œuvre. ≫ (Aristote, nc, p.147 (1112b))

Il y a dans cet extrait une distinction faite entre le raisonnement qui permet de résoudre un conflit et la mise en œuvre effective de la résolution de ce

conflit. Comme le montre l'exemple précédent, l'utilisation de la procédure CAN sert à diagnostiquer une situation, c'est-à-dire à imaginer ce qui peut permettre de résoudre le conflit, mais fournit également un plan, c'est-à-dire un ensemble d'actions, à réaliser pour résoudre effectivement le conflit. Aussi, dans ce qui suit, nous allons utiliser la procédure CAN comme outil pour la planification (Saillenfest and Dessalles, 2014b). En proposant à un personnage de diagnostiquer les origines d'un conflit en utilisant une base d'actions qu'il peut réaliser dans le monde fictif (si les préconditions de ses actions sont valides, *True* au sens logique), cela nous fournira un ensemble d'actions que le personnage pourra tenter de réaliser ensuite pour résoudre le conflit, c'est-à-dire passer de la résolution mentale (il imagine le plan à réaliser) à la résolution 'concrète' (il tente de réaliser le plan calculé).

La procédure conflit-abduction, comme système de planification d'actions, permet en outre de générer des plans intentionnels (du point de vue du personnage). En effet, si un conflit porte sur une situation désirée s (qui reçoit une nécessité $E(s)$ qui quantifie l'intensité du désir) mais inattendue (cette situation reçoit donc également une nécessité $-U(s)$), la phase d'abduction cherche une cause "faible" qui fait baisser l'intensité du conflit, c'est-à-dire une action telle que $U(s||a) < U(s)$, et donc, d'après l'équation 7.4 nous avons $Int^{\mathcal{A}}(s, a) > 0$ (l'action est intentionnelle du point de vue de l'acteur). Puis le conflit est propagé sur la précondition s_a de l'action a, avec une valeur $U(s) - U(s||a)$, et la procédure - récursive - va déterminer une action a' à réaliser pour faire baisser le conflit propagé sur cette précondition, qui sera donc intentionnelle du point de vue de l'acteur ($Int^{\mathcal{A}}(s_a, a') > 0$). Ainsi, pour le plan résultat de la procédure conflit-abduction, composé de n actions $a_1 * a_2 * ... * a_n$, dont les préconditions sont $s_1, s_2, ..., s_n$, nous avons (nous supposons que $s_{n+1} = s$) :

$$\forall j \in [2, n], U(s_j||a_{j-1}) < U(s_{j+1}) - U(s_{j+1}||a_j)$$

Et donc nous avons :

$$U(s) - \sum_{1}^{n} U(s_{j+1}||a_j) > 0$$

Et enfin :

$$U(s) - U(s||a_1 * a_2 * ... * a_n) > 0$$

Le plan, résultat de la procédure conflit-abduction, est donc bien intentionnel du point de vue du personnage-acteur.

Dans la suite, nous allons montrer comment mettre en œuvre la procédure CAN. Pour cela, nous allons présenter un modèle de représentation mentale du monde fictif par un personnage, puis mettre en œuvre la détection de conflits cognitifs en utilisant cette représentation et montrer comment calculer des plans pour les résoudre.

9.4 Mise en œuvre de la procédure conflit-abduction

L'ensemble de la mise en œuvre est réalisée en Python.

9.4.1 Représentation d'un monde fictif

Dans le but de générer des scénarios, nous devons choisir un modèle de représentation du monde fictif, c'est-à-dire de son état et des règles causales qui permettent de faire évoluer cet état. Pour cela nous choisissons de représenter le monde sous la forme de prédicats et des règles d'évolution de type opérateurs dans STRIPS (Fikes and Nilsson, 1972). Ainsi, un événement peut se produire si un ensemble de préconditions sont vérifiées et a pour conséquence de changer l'état de certains éléments du monde (dans les opérateurs de STRIPS, des états sont ajoutés et d'autres sont retirés suite à la mise en œuvre d'un opérateur. Nous nous contenterons de faire varier l'état logique des prédicats). Nous représentons les préconditions et les conséquences sous la forme de conjonctions de prédicats.

Nous allons considérer des événements de type actions (qui sont réalisées par un personnage). Nous traiterons l'influence de *world events via* l'utilisation de scripts, c'est pourquoi nous ne les prenons pas en compte dans la connaissance. Un personnage ne pourra raisonner que sur les actions qu'il peut réaliser.

L'état du monde et les règles causales qui le font évoluer ne sont pas objectives. Ainsi, ce que nous appelons représentation de l'état du monde est une

représentation subjective de l'état du monde fictif qu'à un personnage donné, pas une représentation objective.

Outre leur connaissance sur l'état du monde et l'état mental des autres protagonistes, les personnages ont un ensemble de désirs. Nous représentons un désir sous la forme d'un prédicat, affecté d'une valeur logique et d'une intensité. La valeur logique d'un désir représente l'état logique de cet état qui est désiré par le personnage.

Implémentation

La figure 9.3 donne des détails de la représentation du monde fictif (dans l'esprit d'un personnage).

L'état du monde est représenté par un ensemble de prédicats (classe `Predicate`) dont les attributs sont un nom et un booléen (le statut du prédicat). Un autre attribut conserve la valeur du statut dit 'simulé', lorsque nous analysons l'influence de certaines actions sur l'évolution du monde.

Une action, implémentée *via* la classe `Action` qui hérite des attributs de la classe `Event` a pour attributs un nom, une précondition et une conséquence.

Une précondition est une instance de la classe `Precondition` qui a comme attributs une liste de prédicats et un statut (booléen) qui indique l'état logique nécessaire du prédicat pour rendre la précondition valide (la validité d'une précondition est testée *via* une méthode `is_valid` (ou `simul_is_valid` si nous testons la validité de la précondition sur les états simulés)). L'attribut `possible` d'une instance de la classe `Action` évolue en fonction d'un test de validité de la précondition de cette instance.

Une conséquence est une instance de la classe `Consequence` qui a comme attribut une liste de prédicats et un statut (booléen) qui indique l'état logique du prédicat après réalisation de l'événement. Une méthode `is_in` de la classe `Consequence` teste l'appartenance d'un prédicat (qui aurait comme état logique un booléen donné) à la liste des conséquences.

Un désir est représenté comme une instance de la classe `Desire`, qui a comme attributs un nom, un statut (désiré) et une intensité. Le nom du désir est aussi un nom de prédicat par commodité, nous verrons dans la suite comment un désir est

transformé en préférence.

La classe `World` a pour attributs une liste de prédicats (qui rendent compte) de l'état du monde à un instant donné, une liste des événements connus, une liste des actions connues et une liste de désirs. L'initialisation de la classe `World` est réalisée *via* la lecture d'un fichier contenant un état initial du monde fictif (voir la partie gauche de la figure 9.5) et pré-traitée *via* une méthode d'une classe `Utils` (des méthodes utilitaires).

Le fichier contenant l'état initial du monde se compose de 3 parties principales :

— un état initial, défini comme une liste de prédicats (par exemple `est_avant` sur la figure 9.5) affectés d'un état logique (+ si l'état logique initial est $True$, – sinon).

— un ensemble d'événements (uniquement des actions dans le cas de la figure 9.5). Ces actions ont un nom, les préconditions (et les conséquences) sont représentées comme des conjonctions de prédicats (opérateur `&`) affectés d'un état logique.

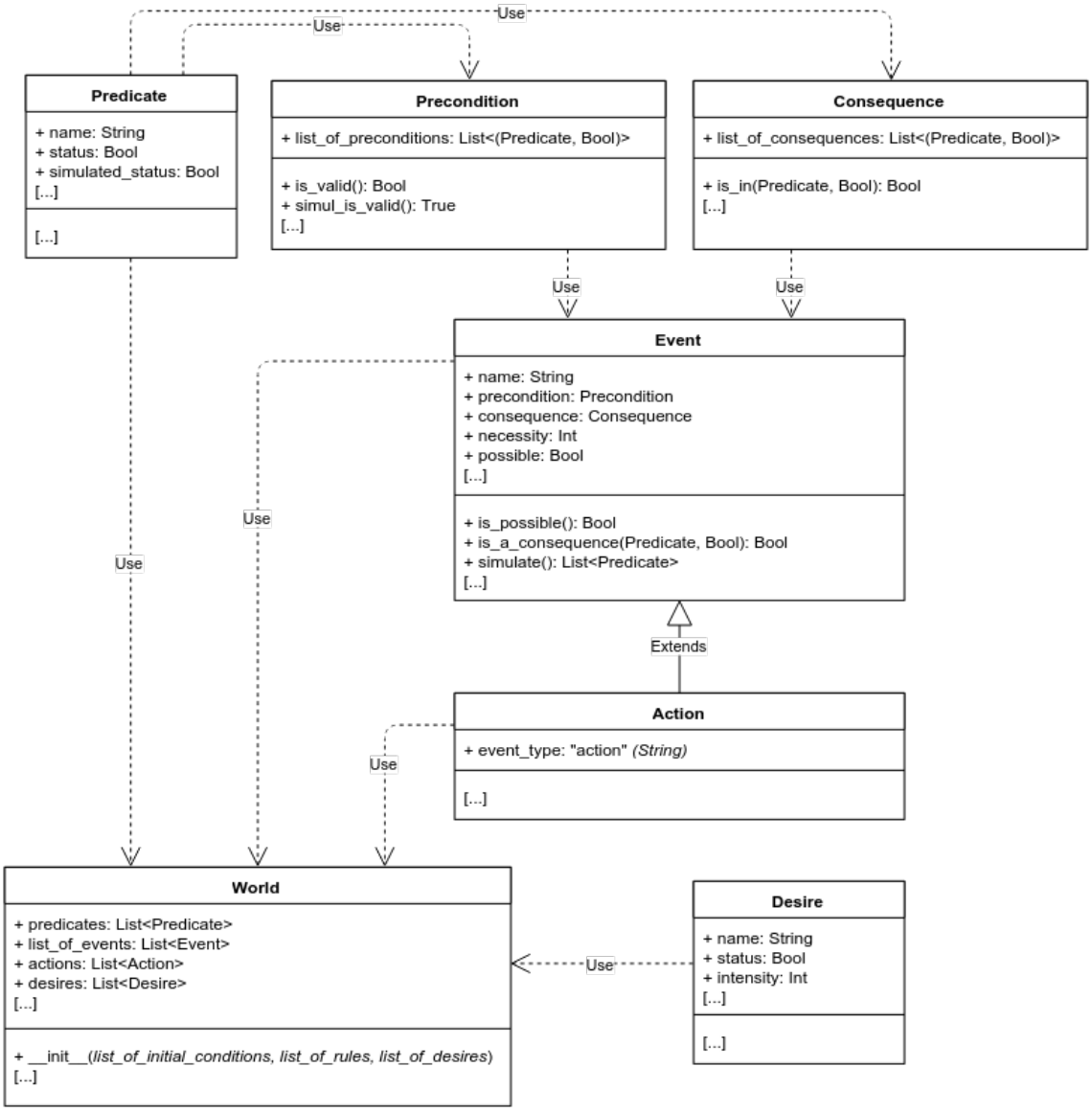
— un ensemble de désirs, qui ont un nom, un état logique désiré et une intensité (dans la figure 9.5, un désir nommé `a_le_tresor` est désiré $True$ avec une intensité `50`).

9.4.2 Détection des conflits cognitifs

Comme indiqué dans le chapitre 5, le conflit est à l'origine de l'action des personnages, mais également une source d'intérêt dans le récit. Le conflit est un conflit entre deux nécessités de valences opposées. Un type de conflit particulier est celui entre une croyance est un désir. Dans le modèle de représentation du monde interne à un personnage que nous avons introduit ci-dessus, un conflit désir-croyance est caractérisé par un désir (avec une intensité positive donnée) portant sur la valeur logique d'un prédicat donné et la croyance en l'état logique opposé du prédicat. Si un prédicat p est désiré $True$ avec une intensité n (et donc reçoit une nécessité $n > 0$) mais supposé (ou connu comme étant) $False$ (et donc reçoit une nécessité $-U(p) << 0$), il y a bien conflit.

Lors des phases de propagation, de nouveaux conflits apparaissent. En effet, l'intensité du désir pour un état donné est propagé sur les causes des actions ou

FIGURE 9.3 – Représentation du monde



événements lors de la phase d'abduction, ces causes devenant alors sources de conflits potentiels. La génération de conflits à la volée en fonction des objectifs que le personnage cherche à atteindre est une condition nécessaire de l'approche cognitive. Ces conflits ne sont pas permanents et disparaissent lorsque la procédure CAN renvoie un résultat ou échoue. Il peuvent être re-générés lors d'un appel

ultérieur de la procédure CAN.

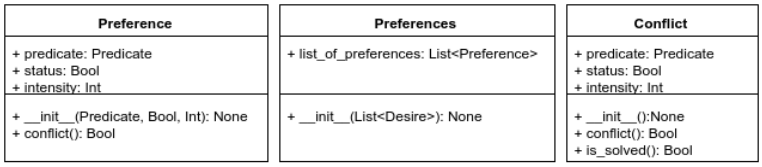
Implémentation

Une classe `Preference` a pour attributs un prédicat, un état logique qui correspond à l'état désiré de ce prédicat et une intensité de la préférence. Une méthode de la classe `Preference` permet de détecter si une préférence est source de conflit, c'est-à-dire si l'état logique (actuel), connu ou supposé, du prédicat en attribut n'est pas l'état logique souhaité pour ce prédicat.

Une classe `Preferences`, initialisée à partir des désirs du personnage, maintient à jour la mémoire de la liste des préférences pendant l'exécution du programme.

Une classe `Conflict` est initialisée avec les attributs d'une préférence détectée comme étant source de conflit (méthode `detect`). La classe `Conflict` possède en outre une méthode `is_solved` qui permet de détecter l'éventuelle résolution du conflit.

FIGURE 9.4 – Classes `Preference`, `Preferences` et `Conflict`



9.4.3 Le processus d'abduction

La procédure d'abduction détermine, parmi l'ensemble des actions possibles connues par un personnage donné, une action dont tout ou partie de la conséquence est source de conflit et qui n'a pas été préalablement écartée.

```
def possible(self, predicate, status, intensity, list_of_events):
#renvoie True si on trouve une abduction possible et
#stocke l'evenement qui permet l'abduction
#(ainsi que le predicat et le statut)
    for event in list_of_events:
```

```
    if event.is_a_consequence(predicate, status) and (event.necessity < -1*intensity) :
                self.event = event
                self.predicate = predicate
                self.status = status
                return True
    return False
```

La classe `Abduction` contient une méthode qui parcourt la liste des événements connus par le personnage afin de déterminer si une abduction est possible, c'est-à-dire si un événement a pour conséquence une source d'un conflit en cours de résolution. Une autre condition est que l'événement ne doit pas être affecté d'une nécessité négative bloquante. Les événements peuvent être réalisés si l'ensemble de leurs causes sont réalisables. Certaines abductions échouent, ce qui rend l'exécution de certains événements précédemment sélectionnés lors des phases successives d'abduction impossibles à réaliser. Ainsi, ces événements héritent de nécessités, par propagation de nécessités négatives affectées à leurs causes, bloquant leur exécution. Dans le cas précédemment détaillé de John désirant avoir le trésor qui se trouve dans le château, lorsque l'analyse des actions possibles de John ne lui permet pas de résoudre le conflit associé au prédicat `-pont_casse`, l'action `traverser_le_pont` hérite de cette nécessité négative et son exécution est bloquée, autrement dit lors d'une prochaine tentative d'abduction à partir du prédicat `est_apres_le_pont`, l'action `traverser_le_pont` ne sera pas retenue comme candidate potentielle. L'abduction réalise une analyse séquentielle des actions connues par le personnage, aussi l'ordre dans lequel ces actions sont données importe. Dans le cas où nous voudrions reproduire l'histoire de John et du pont cassé, il faut que le système explore l'action `traverser_le_pont` avant l'action `Prendre_le_long_chemin`. Sinon l'action `Prendre_le_long_chemin` sera sélectionnée comme bonne candidate pour l'abduction sans que le système explore la possibilité de traverser le pont. De ce point de vue, la procédure d'abduction a deux limitations évidentes : elle n'assure pas l'optimalité du plan calculé en terme de complexité de génération et elle est susceptible de tomber dans une boucle infinie.

La première limitation n'est pas problématique sur le plan cognitif, l'optimalité des plans calculés n'étant ni nécessaire ni recherchée. Nous nous attendons uniquement à ce que la procédure renvoie un plan acceptable, c'est-à-dire un plan qui fait baisser suffisamment l'intensité d'un conflit pour que celui-ci soit considéré comme résolu.

La seconde limitation n'est pas non plus problématique sur le plan de la plausibilité cognitive. En effet, il n'est pas impossible que la résolution d'un problème nous amène à entrer dans une boucle d'action (ou de simulation d'actions). Cependant, si nous affectons une complexité de génération à chaque action, la boucle infinie devient rapidement une solution inefficace pour résoudre le conflit, elle devient trop complexe ce qui nous pousse à la rejeter et à sortir de la procédure CAN.

9.4.4 La procédure conflit-abduction

La classe CAN met en œuvre la procédure CAN définie précédemment. Son exécution repose sur une connaissance du monde fictif à un instant donné et un jeu de préférences (initialisé à l'aide des désirs du personnage à cet instant). Elle garde en mémoire à tout instant le conflit en cours de résolution et une liste ordonnée des conflits à résoudre. Enfin elle conserve à chaque instant la mémoire du plan en cours de calcul et le dernier plan calculé. À travers une méthode de classe `solve_conflict` elle résout entièrement un conflit donné en appelant de manière récursive une méthode `solve` qui propage par abduction (via une méthode `propagate`) le conflit actuel sur les causes d'une action tout en maintenant à jour l'évolution du plan calculé. La méthode de classe `solve` affecte également des nécessités négatives aux actions lorsque la phase d'abduction échoue.

Nous pouvons voir dans la figure 9.5 un exemple de trace de l'exécution de la procédure CAN et la connaissance utilisée. Nous pouvons voir en particulier qu'une nécessité négative est affectée à l'action de traverser le pont ce qui rend donc sa réalisation impossible et oblige le personnage à recommencer la procédure CAN, le système ayant gardé en mémoire cette nécessité négative. Lors du deuxième appel de la procédure CAN, l'action qui consiste à traverser le pont n'est pas prise en

compte et la procédure décide d'explorer directement la possibilité d'emprunter le long chemin (ce qui est possible). Ainsi, cela permet de résoudre les conflits successifs.

Nous pouvons voir l'affichage du plan calculé sous forme d'arbre qui se lit en profondeur d'abord (En bas à droite de la figure 9.5 - les lignes se lisent par 2, la première correspond à l'action à réaliser, la seconde correspond aux prédicats qui changent d'état). Ainsi, ce plan se lit ainsi :

> Je ('je' correspond à John) dois prendre le long chemin (ce qui va changer l'état logique de `est_apres` et `est_avant`)
> Je dois aller au château (ce qui va changer l'état logique de `est_apres` et `est_au_chateau`)
> Je dois prendre le trésor (ce qui va changer l'état logique de `a_le_tresor`)

9.5 Génération récursive de séquences narratives

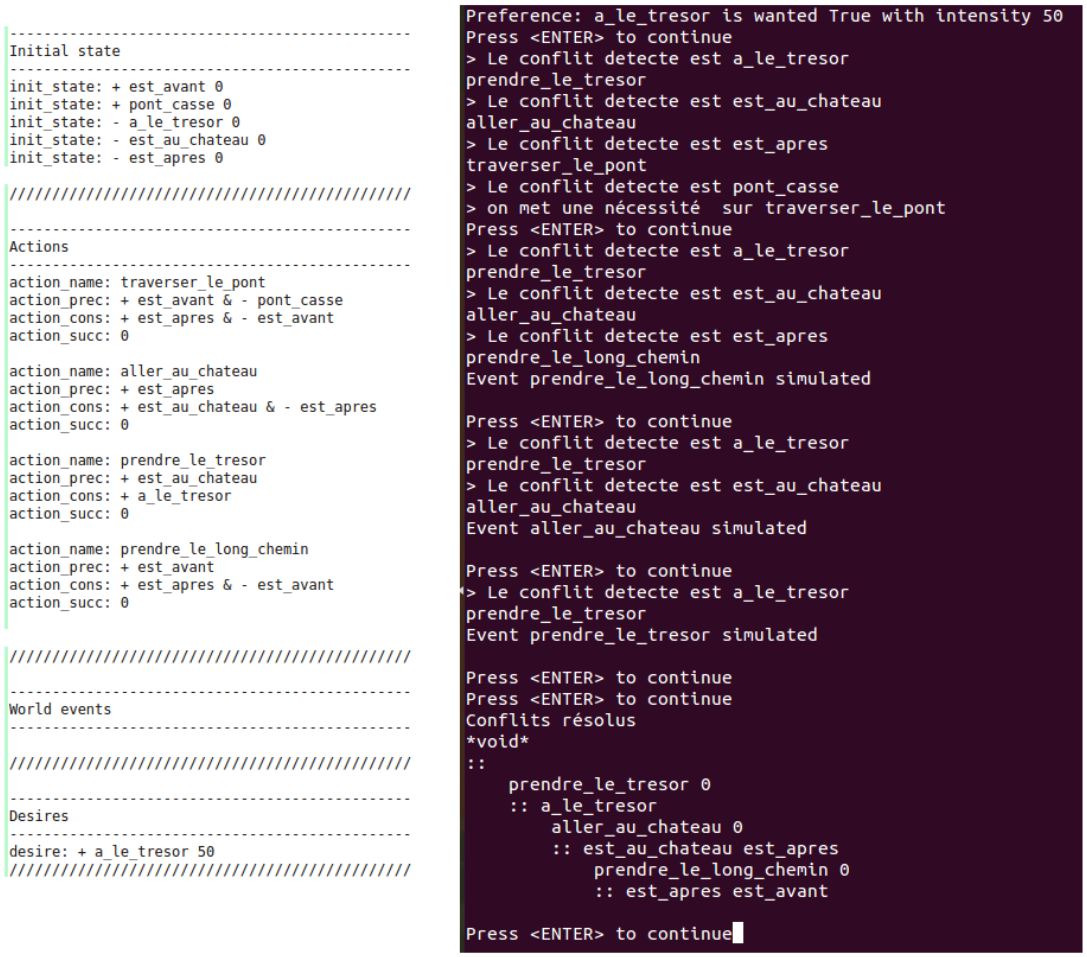
9.5.1 Génération et intégration de conflits inattendus

Le modèle de génération introduit précédemment (voir page 195) est basé sur la génération récursive de séquences CRRS, c'est-à-dire sur la génération successive de plans et d'obstacles à la réalisation de ces plans. Ces obstacles sont des conflits, ils sont inattendus du point de vue du personnage qui voit la réalisation du plan qu'il a calculé compromise.

Les plans sont des ensembles d'actions qui, à chaque étape du plan, modifient l'état du monde fictif. Plusieurs facteurs peuvent faire obstacle à la réalisation d'un plan. Tout d'abord, une action du plan ne produit pas les conséquences attendues. Cela qui peut avoir pour cause une connaissance fausse (ou incertaine) de l'état du monde fictif, et en particulier des préconditions de cette action. L'action en elle-même peut produire des résultats incertains. Enfin, l'action peut ne pas produire les conséquences attendues car la connaissance du monde, et de cette action en particulier est fausse (en d'autres termes, l'action n'est pas une règle causale du monde fictif).

Un autre facteur qui peut faire obstacle à la réalisation d'un plan est une intervention extérieure non anticipée qui vient affecter les préconditions d'une ou

FIGURE 9.5 – Trace de l'exécution de la procédure CAN avec la connaissance initiale définie.



```
-----------------------------------------------
Initial state
-----------------------------------------------
init_state: + est_avant 0
init_state: + pont_casse 0
init_state: - a_le_tresor 0
init_state: - est_au_chateau 0
init_state: - est_apres 0

////////////////////////////////////////////////

-----------------------------------------------
Actions
-----------------------------------------------
action_name: traverser_le_pont
action_prec: + est_avant & - pont_casse
action_cons: + est_apres & - est_avant
action_succ: 0

action_name: aller_au_chateau
action_prec: + est_apres
action_cons: + est_au_chateau & - est_apres
action_succ: 0

action_name: prendre_le_tresor
action_prec: + est_au_chateau
action_cons: + a_le_tresor
action_succ: 0

action_name: prendre_le_long_chemin
action_prec: + est_avant
action_cons: + est_apres & - est_avant
action_succ: 0

////////////////////////////////////////////////

-----------------------------------------------
World events
-----------------------------------------------

////////////////////////////////////////////////

-----------------------------------------------
Desires
-----------------------------------------------
desire: + a_le_tresor 50
////////////////////////////////////////////////
```

```
Preference: a_le_tresor is wanted True with intensity 50
Press <ENTER> to continue
> Le conflit detecte est a_le_tresor
prendre_le_tresor
> Le conflit detecte est est_au_chateau
aller_au_chateau
> Le conflit detecte est est_apres
traverser_le_pont
> Le conflit detecte est pont_casse
> on met une nécessité  sur traverser_le_pont
Press <ENTER> to continue
> Le conflit detecte est a_le_tresor
prendre_le_tresor
> Le conflit detecte est est_au_chateau
aller_au_chateau
> Le conflit detecte est est_apres
prendre_le_long_chemin
Event prendre_le_long_chemin simulated

Press <ENTER> to continue
> Le conflit detecte est a_le_tresor
prendre_le_tresor
> Le conflit detecte est est_au_chateau
aller_au_chateau
Event aller_au_chateau simulated

Press <ENTER> to continue
> Le conflit detecte est a_le_tresor
prendre_le_tresor
Event prendre_le_tresor simulated

Press <ENTER> to continue
Press <ENTER> to continue
Conflits résolus
*void*
::
    prendre_le_tresor 0
    :: a_le_tresor
        aller_au_chateau 0
        :: est_au_chateau est_apres
            prendre_le_long_chemin 0
            :: est_apres est_avant

Press <ENTER> to continue
```

plusieurs étapes du plan, ce qui rend sa réalisation impossible. Cela peut être dû à l'intervention d'un autre personnage ou un *world event* par exemple.

Dans la suite, nous allons considérer qu'un obstacle à la réalisation d'un plan calculé vient du constat, par un personnage, de la non-validité de l'état de sa connaissance concernant une précondition d'une action du plan. Par ailleurs, nous ne nous intéresserons qu'aux préconditions qui ne sont pas la conséquence d'une action passée du plan en cours de réalisation (nous considérons que la connaissance des règles causales (les actions, leurs préconditions et leurs conséquences)

du monde par le personnage est valide et que la réalisation d'une action produit la conséquence attendue (aucune incertitude quant à la réalisation d'une action dont les préconditions sont valides)). Ainsi, si un plan consiste à aller, en partant d'un village, jusqu'à un pont puis, une fois devant le pont, et si celui-ci n'est pas cassé, à le traverser, les seules préconditions qui seront analysées comme possibles obstacles sont la présence du personnage dans le village et la non-destruction effective du pont (le fait d'être devant le pont, précondition de l'action consistant à traverser le pont est conséquence de l'action consistant à aller jusqu'au pont)

Lorsqu'un obstacle potentiel est détecté, il faut pouvoir l'intégrer pour que l'obstacle soit valide dans le cadre de la génération d'un récit. Intégrer l'obstacle signifie fournir une explication au lecteur/spectateur qui permet de faire passer le prédicat de surprenant à non-surprenant (pour l'audience). Cette intégration est nécessaire afin de satisfaire le critère de postdictibilité. Comme cela est dit plus haut, l'intervention d'un autre personnage ou un *world event* peut modifier l'état du monde sans qu'un personnage le sache *a priori*. La présentation au lecteur de ces interventions externes constitue une manière d'intégrer l'événement surprenant. Si l'erreur est interne au personnage, c'est-à-dire si sa connaissance est fausse à cause d'une erreur d'interprétation ou de raisonnement, l'intégration peut aussi consister à présenter au lecteur/spectateur les raisons qui ont amené le personnage gêné dans la réalisation de son plan à faire une hypothèse fausse sur l'état du monde.

Dans la suite, nous nous limiterons à l'intégration d'obstacles *via* des causes externes.

Implémentation

Les plans sont représentés sous forme d'arbres d'événements et de modifications de prédicats qui servent à décrire l'état du monde. Le parcours de ces arbres permet de dresser la liste des préconditions nécessaires à la réalisation des différentes actions du plan. La mémoire des changements d'état du monde permet d'éliminer de cette liste un ensemble de préconditions qui sont conséquences d'actions précédemment réalisées (ou dont la réalisation est prévue si on se place avant la tentative de réalisation du plan). Cela permet d'obtenir une liste de prédicats, qui composent les préconditions, qui peuvent faire obstacle à la réalisation du plan

si leur état 'réel' est différent de leur état supposé.

Parmi ces prédicats, qui sont à ce stade des obstacles potentiels, il nous faut déterminer un obstacle effectif, c'est-à-dire un prédicat que l'on peut intégrer. Pour déterminer ceux qui peuvent être intégrés, nous allons fournir au système une base de prédicats qui peuvent être intégrés et les intégrations associées sous forme de scripts (voir 9.6). Ainsi, nous supposons que le système, qui agit comme un auteur virtuel, possède un ensemble de scripts prêts à l'emploi qui lui permettent d'affecter, si besoin, l'état logique d'un certain nombre de prédicats qui composent l'état du monde fictif. Ainsi, si un plan repose sur l'hypothèse d'un pont non-détruit, l'auteur dispose éventuellement d'un script qui lui permet de détruire le pont de manière opportune et qui fournit, en sus, une explication à cet état de fait pour le lecteur.

FIGURE 9.6 – Extrait de la base d'intégrations - Un prédicat `pont_casse` affecté de la valeur logique *True* (+) peut être intégré *via* l'explication `faire_pourir_le_pont`.

```
-----------------------------------------------
Integrations
-----------------------------------------------

[...]

integr_name: faire_pourir_le_pont
integr_cons: + pont_casse
```

Ainsi, à partir d'un plan, on peut obtenir une liste de prédicats nécessaires à la réalisation de ce plan et dont la négation est intégrable. Nous pouvons ainsi décider lequel de ces prédicats va être présenté au personnage dans son état nié au cours de la réalisation du plan calculé précédemment et ainsi surprendre le personnage (et l'audience). Puis nous pouvons proposer à l'audience une explication, *via* la base d'intégrations donnée au système.

9.5.2 Génération d'une structure CRRS

La figure 9.1 montre les différentes étapes de génération d'une brique structurelle CRRS (Conflit - Résolution - Réalisation - Surprise). D'après ce qui précède, nous possédons les éléments nécessaires pour générer ce type de structures à partir

de la connaissance qu'a un personnage de l'état du monde fictif et d'un ensemble de désirs de ce personnage. Nous savons :

- Détecter les conflits potentiels et sélectionner ceux qui doivent être résolus.
- Résoudre les conflits, c'est-à-dire calculer un plan (un arbre d'actions) à réaliser *via* l'utilisation d'une procédure de type conflit-abduction.
- Déterminer d'éventuels obstacles à la réalisation d'un plan donné.
- Déterminer les obstacles qui peuvent être intégrés *via* la consultation d'un ensemble donné de scripts d'intégration.
- Modifier l'état du monde suite à la réalisation d'un plan

Pour générer totalement une brique séquentielle CRRS, il reste à effectuer une mise à jour de la connaissance du personnage (en particulier il doit modifier sa connaissance de l'état constaté du prédicat-obstacle et évaluer les conséquences éventuelles de ce cette modification de la connaissance).

Le modèle de génération introduit plus haut (voir figure 9.1) est basé sur la génération récursive de séquences CRRS. Cette génération prend fin lorsque qu'aucun obstacle à la réalisation du plan, et qui est intégrable, ne peut être trouvé. Autrement dit, le processus de génération se termine lorsque l'auteur virtuel n'a plus de scripts pour provoquer des situations conflictuelles pour le personnage et l'empêcher de réaliser ses objectifs (c'est-à-dire réaliser le dernier plan de résolution des conflits calculé). Ainsi, nous générons un ensemble de séquences CRRS et la génération se termine avec la génération d'une séquence CRR (conflit-résolution-réalisation) (voir figure 9.7).

Implémentation

Le processus de génération d'une cellule CRRS a été implémenté *via* la classe `CRRS`, dotée d'une méthode `stabilize` qui réalise successivement les opérations de détection de conflit et résolution des conflits (ces deux opérations sont réalisées *via* les méthodes de la classe conflit-abduction `CAN`), puis cherche un obstacle intégrable à la réalisation du plan (*via* la méthode `find_surprise`) et, si un tel obstacle est identifié, provoque la réalisation partielle du plan et met à jour la connaissance du personnage.

Nous n'avons pas créé de classe CRR (conflit-résolution-réalisation). La méthode

FIGURE 9.7 – Structure des histoires générées via le modèle de la figure 9.1

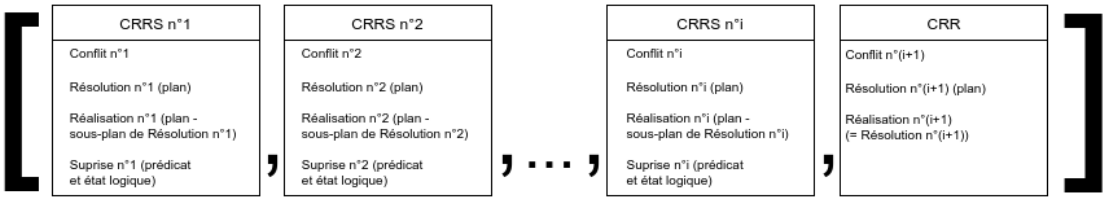
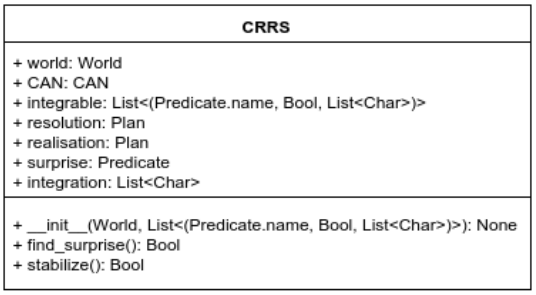


FIGURE 9.8 – Classe CRRS



`stabilize` de la classe `CRRS` renvoie la valeur booléenne *False* si un obstacle intégrable a été trouvé, et donc si le plan calculé pour résoudre les conflits initiaux ne peut pas être mené à son terme. Si cette méthode renvoie *True*, aucun obstacle satisfaisant n'a pu être trouvé, et donc le plan de résolution des conflits et le plan réalisé sont identiques, ce qui correspond au cas où nous avons une séquence CRR (voir figure 9.7). La valeur renvoyée par la méthode `stabilize` détermine donc la fin de la génération de la structure événementielle de notre scénario.

9.5.3 Génération récursive de la structure événementielle augmentée d'une séquence narrative localement surprenante

La génération d'une séquence de séquences CRRS (et d'une séquence CRR finale) est réalisée *via* une méthode `generate` de la classe `OneCharacterStory`. Voici cette méthode :

```
def generate(self):
    _CRRS = CRRS(self.world, self.integrable)
    if not _CRRS.stabilize():
        #Tant que l'on peut generer une sequence CRRS
      self.story.append(_CRRS)
      #on ajoute la sequence CRRS a l'histoire
      self.generate()
      #on repete l'operation
    else:
    #On ne peut plus generer de sequence CRRS
    #Une sequence CR a ete generee
      self.story.append(_CRRS)
      #On ajoute la sequence CR a l'histoire
    return True
```

Le contenu de chacune des séquences, présenté tel quel, ne constitue pas en tant que tel une histoire que l'on peut lire (et comprendre) facilement. Outre la génération de l'histoire, la classe `OnePersonStory` contient des méthodes pour présenter l'histoire. La question de la présentation de l'histoire est abordée dans la partie suivante.

9.5.4 Quelques considérations sur la génération du discours

Dans ce qui précède, la question de la présentation de l'histoire a été peu abordée. Nous nous sommes principalement intéressés à la génération d'événements

qui composent un scénario surprenant dans lequel les personnages agissent de manière intentionnelle.

Nous avons montré que le spectateur doit reconnaître que le personnage agit de manière intentionnelle. Nous décidons de présenter le plan calculé par le personnage pour résoudre les conflits cognitifs auxquels il est confronté ainsi que les règles causales et la connaissance utilisée par le personnage pour calculer ce plan afin de pouvoir reconnaître que ce plan est bien intentionnel. Il nous faut enfin présenter, pour que le plan soit reconnu comme étant intentionnel, les motivations du personnage (nous allons donc présenter les désirs du personnage).

Ensuite, nous devons présenter le plan effectivement réalisé, qui fait évoluer le monde et donc l'histoire.

Puis nous devons introduire (s'il existe) l'obstacle à la réalisation du plan. Comme nous l'avons dit au début de cette partie, nous considérons que ce qui surprend un personnage est intéressant pour une audience. Ainsi, nous allons présenter l'événement surprenant du point de vue du personnage.

Les événements surprenants doivent être intégrés et leur intégration présentée à l'audience. Ainsi, nous présenterons à l'audience l'intégration de l'événement surprenant après avoir présenté ce dernier.

Ce processus peut être répété pour chaque séquence CRRS (et dans une version abrégée à la séquence CRR finale).

Tout élément du récit doit être pertinent. Ainsi, tout élément de description du monde qui n'intervient pas dans le développement événementiel du scénario n'a pas à être présenté (c'est-à-dire tout élément de description du monde qui n'est ni une précondition ni une conséquence d'un événement du scénario augmenté et tout événement qui n'est pas un événement du scénario augmenté).

Par ailleurs, dans la mesure où les règles causales sont présentées (pour permettre à l'audience de reconnaître les intentions du personnage), la part de la connaissance du personnage qui est présentée à l'audience est la connaissance minimale nécessaire à la réalisation du plan. Si le personnage se trouve dans un village et doit prendre un chemin pour aller près d'un pont puis, si le pont n'est pas cassé, doit traverser le pont pour aller au château, alors, une fois ces règles causales introduites (et le plan calculé dans ce cas), la connaissance minimale du personnage à introduire est qu'il se trouve dans le village et qu'il pense que le

pont n'est pas détruit. Le fait d'être près du pont ou au château sont des états intermédiaires que l'audience peut déduire de la réalisation du plan car les règles causales ont été présentées *in extenso*.

Implémentation

Afin de présenter à une audience un texte 'lisible', chacun des prédicats et des événements du monde fictif contient des attributs qui permettent de le décrire (des chaînes de caractères qui forment des phrases ou morceaux de phrases pour décrire le prédicat ou l'action).

La méthode `write_story` de la classe `OnePersonStory` fait appel à un objet `Storyteller` qui possède une méthode `write` qui présente l'histoire sous la forme suivante :

- Les motivations du personnage sont présentées
- Pour chaque cellule CRRS :
  - La connaissance utile et pertinente du personnage est présentée
  - Le plan que le personnage à l'intention de réaliser est présenté
  - Le plan effectivement réalisé est présenté
  - La surprise est présentée du point de vue du personnage
  - L'intégration de la surprise est présentée à l'audience (sauf pour la dernière séquence où il n'y a pas de surprise)

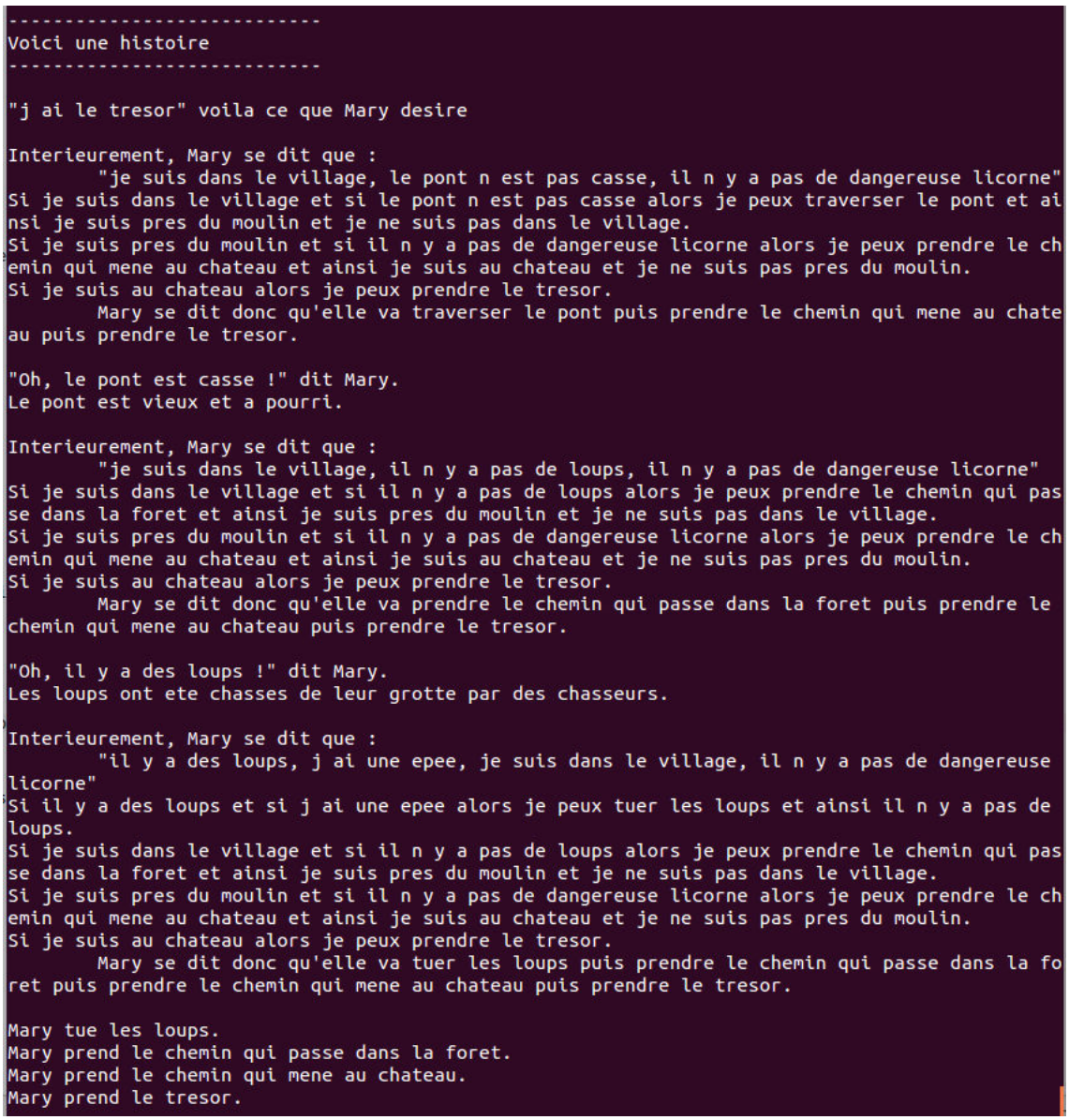
Un exemple d'histoire générée

La figure 9.9 montre un exemple d'histoire présentée de manière à être 'lisible'. Une nouvelle opération d'optimisation pourrait être réalisée en supprimant les parties de plans qui sont répétées (ou les prédicats répétés) de la présentation d'une structure CRRS à l'autre, ce qui allégerait considérablement l'ensemble.

9.6 Évaluation du modèle et perspectives

Dans cette section nous avons proposé un modèle de génération de séquences basé sur la détection (et la résolution) récursive de conflits d'ordre cognitif auxquels est confronté un personnage-acteur *via* l'introduction d'obstacles à la réalisation

FIGURE 9.9 – Un exemple d'histoire présentée à un lecteur



```
---------------------------
Voici une histoire
---------------------------

"j ai le tresor" voila ce que Mary desire

Interieurement, Mary se dit que :
        "je suis dans le village, le pont n est pas casse, il n y a pas de dangereuse licorne"
Si je suis dans le village et si le pont n est pas casse alors je peux traverser le pont et ai
nsi je suis pres du moulin et je ne suis pas dans le village.
Si je suis pres du moulin et si il n y a pas de dangereuse licorne alors je peux prendre le ch
emin qui mene au chateau et ainsi je suis au chateau et je ne suis pas pres du moulin.
Si je suis au chateau alors je peux prendre le tresor.
        Mary se dit donc qu'elle va traverser le pont puis prendre le chemin qui mene au chate
au puis prendre le tresor.

"Oh, le pont est casse !" dit Mary.
Le pont est vieux et a pourri.

Interieurement, Mary se dit que :
        "je suis dans le village, il n y a pas de loups, il n y a pas de dangereuse licorne"
Si je suis dans le village et si il n y a pas de loups alors je peux prendre le chemin qui pas
se dans la foret et ainsi je suis pres du moulin et je ne suis pas dans le village.
Si je suis pres du moulin et si il n y a pas de dangereuse licorne alors je peux prendre le ch
emin qui mene au chateau et ainsi je suis au chateau et je ne suis pas pres du moulin.
Si je suis au chateau alors je peux prendre le tresor.
        Mary se dit donc qu'elle va prendre le chemin qui passe dans la foret puis prendre le
chemin qui mene au chateau puis prendre le tresor.

"Oh, il y a des loups !" dit Mary.
Les loups ont ete chasses de leur grotte par des chasseurs.

Interieurement, Mary se dit que :
        "il y a des loups, j ai une epee, je suis dans le village, il n y a pas de dangereuse
licorne"
Si il y a des loups et si j ai une epee alors je peux tuer les loups et ainsi il n y a pas de
loups.
Si je suis dans le village et si il n y a pas de loups alors je peux prendre le chemin qui pas
se dans la foret et ainsi je suis pres du moulin et je ne suis pas dans le village.
Si je suis pres du moulin et si il n y a pas de dangereuse licorne alors je peux prendre le ch
emin qui mene au chateau et ainsi je suis au chateau et je ne suis pas pres du moulin.
Si je suis au chateau alors je peux prendre le tresor.
        Mary se dit donc qu'elle va tuer les loups puis prendre le chemin qui passe dans la fo
ret puis prendre le chemin qui mene au chateau puis prendre le tresor.

Mary tue les loups.
Mary prend le chemin qui passe dans la foret.
Mary prend le chemin qui mene au chateau.
Mary prend le tresor.
```

des plans calculés par ce personnage. Le conflit narratif, que l'on peut ramener à un ensemble de conflits d'ordre cognitifs à l'échelle des personnage, est à la fois la source d'intérêt narratif et le moteur de l'action des personnages dans les récits, ce qui nous a conduit à faire du conflit l'objet central de notre générateur de récits.

Notre approche de la génération est hybride, à la fois *author based* et *character based*. La détection de conflits cognitifs, la résolution de ces conflits et la réalisation des plans calculés représentent une simulation de la réaction d'un agent virtuel suite à l'observation d'un élément surprenant (pour cet agent) dans un monde fictif virtuel. L'insertion d'éléments surprenants, et leur intégration, relève cependant plutôt d'une forme de modélisation de la volonté d'un auteur virtuel de créer une tension narrative dans le récit. La structure CRRS est assimilable à une méthode qu'un auteur utiliserait (parmi une liste de méthodes possibles connues par cet auteur) pour créer une histoire.

Le système produit une séquence narrative localement cohérente. Il y a une unité de la séquence narrative car les obstacles empêchent la résolution d'un même conflit initial qui est ré-activé après l'observation de chaque événement surprenant. Ainsi, il y a localement une unité dans la séquence narrative, l'objectif initial du personnage est maintenu tout au long de la séquence.

Les actions du personnage sont perçues comme étant réalisées de manière intentionnelle. Le personnage connaît des actions (qui sont des éléments de la causalité du monde) qu'il réalise suite à un calcul de type abduction qui fait baisser l'intensité des conflits que ce personnage essaye de résoudre. Comme nous présentons à l'audience les règles causales connues par le personnage et nécessaires à la compréhension des plans que celui-ci tente de réaliser ultérieurement pour atteindre un objectif lui-même présenté à l'audience au début de la séquence narrative, l'audience perçoit les actions du personnages comme étant intentionnelles. Ainsi, le personnage semble crédible. Cette crédibilité des personnages n'est pas assurée *a posteriori*, le processus de génération et de résolution de conflits assure que les actions des personnages est bien intentionnelle au moment de la formation des actions qui vont, *in fine*, composer le récit.

L'intégration des situations surprenantes pour le personnage étant réalisée *via* la présentation d'une explication scriptée au lecteur, celle-ci respecte la causalité du monde fictif, les scripts étant indépendants des actions du personnage. Ainsi, le monde fictif est cohérent du point de vue de la crédibilité des personnages et au niveau du respect de la causalité qui fait évoluer l'état de ce monde.

Le modèle présenté dans ce chapitre produit une séquence narrative qui est une succession de séquences configurées par la surprise d'un personnage fictif. Cela crée

des séquences très répétitives, la surprise créant une tension brève ponctuellement. Les séquences événementielles générées contiennent les composants nécessaires à la production d'une séquence narrative localement surprenante, l'intérêt n'est pas intégré au récit *a posteriori*.

Il est facile d'imaginer des scripts plus évolués pour intégrer les situations surprenantes qui feraient donc baisser la curiosité suite à l'observation de la situation surprenante de manière graduelle, et donc permettraient de provoquer une baisse plus graduelle de la tension liée à cette curiosité. Par ailleurs, il est également possible d'imaginer un processus d'extension des parties de plans effectivement réalisées par le personnage afin de créer du suspense. Pour cela, il faudrait introduire une forme de granularité dans la représentation causale du monde afin de pouvoir retarder (artificiellement) l'occurrence de l'événement suivant.

Le système produit des séquences narratives dans lesquelles le critère de postdictibilité est respecté car tout événement surprenant est expliqué au lecteur. Si ultimement le personnage n'arrive pas à résoudre les conflits auxquels il est confronté, ce n'est pas nécessairement un échec du point de vue de la postdictibilité. Le personnage constate son impossibilité à résoudre les conflits et les abandonne progressivement, ce qui lui fait retrouver une forme de stabilité. Dans l'exemple donné plus haut de narration produite sur le système, si Mary ne peut pas atteindre le trésor à la fin de l'histoire, elle peut acter que le trésor est (pour le moment) inaccessible et abandonner sa quête, ce qui clôt l'histoire.

Le modèle de génération proposé est déterministe. En cela sa *replay value* est faible. Deux exécutions du programme avec la même base de connaissance donnée en argument fournissent le même résultat. Les calculs de plans sont réalisés de manière séquentielle, ainsi que le choix des obstacles à la réalisation des plans. Pour obtenir un modèle de génération qui produirait un ensemble de possibles narrations, il faudrait un modèle d'auteur virtuel plus évolué qui, selon ses objectifs propres, orienterait l'action des personnages en manipulant leurs désirs, les actions à leur disposition ou interdirait des séquences jugées peu intéressantes.

L'originalité de notre approche est de conserver en mémoire les intentions du personnages, ce qui permet de les présenter afin de maintenir la crédibilité du personnage au cours du récit. Ainsi, on ne peut pas vraiment parler de la génération d'une *fabula* mais d'un ensemble plus riche composés d'intentions (éventuellement

non réalisées), de plans réalisés (qui participent à l'évolution du monde fictif et du personnage au cours du récit) et de scripts qui viennent interférer dans les actions du personnage (et font aussi évoluer le mode fictif ou la connaissance que le personnage a de ce monde).

Chapitre 10

Conclusion et perspectives

Nous nous demandions en introduction comment 'programmer le singe savant', c'est-à-dire comment créer un programme capable de générer de bons récits en un temps acceptable. Notre travail de modélisation cognitive donne des clés pour créer de tels programmes.

Notre travail de modélisation cognitive se situe au carrefour de la psychologie, des études littéraires, de l'analyse quantitative des récits et de l'intelligence artificielle (en particulier le sous-domaine de la génération automatique de récits). À partir d'un modèle mathématique et cognitivement plausible de l'intérêt narratif, nous avons modélisé divers aspects de la pertinence narrative. Nous avons proposé des modèles cognitif de la réponse émotionnelle, d'attitudes ou en encore de jugements d'individus rationnels. Ces modèles théoriques, d'inspiration cognitive, peuvent être utilisés comme guides pour la génération de récits et, plus généralement, pour le développement d'agents virtuels artificiellement intelligents.

Cette approche peut être développée suivant de nombreux axes et ouvre par ailleurs la voie à de multiples applications. De nombreux modèles théoriques peuvent être testés empiriquement. Les modèles d'évaluation des actions *a posteriori* ouvrent la voie à l'exploration du rôle de la complexité dans le domaine de la prise de décision, c'est-à-dire à l'étude du rôle de la complexité *a priori*. Enfin, nous proposons également de poursuivre l'étude de la résolution des conflits cognitifs interne à un agent, en particulier l'étude des mécanismes d'argumentation et d'optimisation pour le calcul de l'explication la plus satisfaisante.

Bilan de l'approche

L'un des premiers objets de cette thèse est l'introduction et le développement d'une théorie alternative aux modèles probabilistes classiques basée sur les calculs de complexité. La complexité cognitive, dérivée de la notion de complexité de Kolmogorov, joue un rôle important dans la cognition, en particulier dans le domaine de la perception. Souvent, la complexité est délaissée au profit des probabilités lorsqu'il s'agit d'analyser des cas concrets. Nous avons montré que les calculs de complexité permettent de mieux rendre compte de ces situations.

Nous avons introduit la notion de conflit cognitif comme conflit entre deux évaluations d'un même état au sein de la connaissance d'un individu. Ces évaluations, les nécessités, reposent sur des calculs de complexité. L'évaluation d'une situation donnée peut provenir de sources différentes, d'une observation, d'une connaissance liée à un état supposé standard du monde ou encore de désirs propres à chaque individu. Une même situation peut recevoir en particulier deux évaluations de signes opposés, ce qui crée un conflit de nécessité (ce que nous appelons conflit cognitif interne à un individu). Nous avons évoqué la résolution desdits conflits par les personnages. Cette résolution des conflits peut être de nature différente, selon les sources d'évaluation qui sont à l'origine du conflit. L'explication ou la planification sont des modes de résolution différents, qui amènent le personnage à raisonner sur des règles différentes.

Nous avons proposé un modèle cognitif du conflit narratif, défini comme ensemble de conflits cognitifs, c'est-à-dire de conflits internes aux personnages du récit (traités comme des humains réels).

Par la suite, nous avons proposé des modèles quantitatifs pour l'évaluation de modalités classiques de l'intérêt narratif (la surprise, la curiosité et le suspense) dans une séquence narrative. Par ailleurs, nous avons élaboré un modèle de l'attitude d'un individu, possiblement un personnage dans un récit, postérieure à l'occurrence d'un événement surprenant. Ce modèle décrit le processus cognitif de formation de conflits suite à l'observation d'une situation surprenante et leur résolution éventuelle.

L'action des personnages fait évoluer le récit, mais ces actions doivent être perçues comme intentionnelles pour ne pas affecter négativement la crédibilité des

personnages-acteurs. Nous avons proposé des modèles pour quantifier l'intentionnalité, et, ces notions étant proches, nous avons également étendu notre étude à l'évaluation de la responsabilité attribuée aux personnages dans le récit (jugement moral). Ces modèles visent à évaluer, *a posteriori* la crédibilité d'un personnage suite à la réalisation d'une action et, ultimement, à vérifier que chacun des actes de chacun des personnages apparaît bien comme étant intentionnel pour l'audience.

L'auteur d'un récit étant constamment tiraillé entre la nécessité de créer des situations intéressantes et le maintien de la crédibilité des personnages, ces modèles, et ceux de l'intérêt introduits précédemment, permettent de contrôler à la fois localement, ainsi que globalement, le respect de ces deux contraintes.

Ces modèles peuvent servir à analyser les récits. Nous montrons, sur un exemple particulier que l'on peut utiliser les modèles introduits dans nos travaux pour suivre l'évolution de la reconnaissance des intentions des personnages et de la tension narrative. Par ailleurs, ces modèles nous permettent, globalement, de déterminer si la séquence analysée respecte ou non certains critères généraux de pertinence introduits.

Ces modèles peuvent également servir de guide en vue de la génération de séquences narratives pertinentes. Nous montrons, dans la dernière partie, comment le modèle de la surprise introduit précédemment peut servir de guide pour la création et le développement d'un modèle de génération de séquences narratives pertinentes, configurées par la surprise. Les modèles introduits dans ce manuscrit reposent sur des évaluations subjectives des états du monde par des personnages-acteurs, notre modèle repose donc essentiellement sur la simulation d'un comportement rationnel de personnages-agents.

Principes et contributions

Dans cette thèse, nous avons montré l'importance d'adopter une démarche qui est guidée par les principes. Cela permet une approche plus globale du processus de génération. La plupart des systèmes de génération de récit actuels mettent en place des méthodes particulières qui ne permettent de générer que certains types de récits particuliers (séquence narrative configurée par le suspense, par la surprise, séquence de type 'dilemme à résoudre', etc.). En utilisant des principes généraux,

et en définissant dans un même espace des modèles pour quantifier à la fois des notions relatives à la création de tension narrative et des notions qui renvoient à la vraisemblance des récits, nous ouvrons la voie au développement de modèles de génération plus généraux, qui permettent de développer des générateurs de récits qui produisent une plus grande variété d'effets sur l'audience.

Les principes théoriques qui ont guidés ces travaux sont :

- **La simplicité :** les individus réalisent effectivement des calculs de complexité pour évaluer (subjectivement) l'occurrence de situations, y compris dans un cadre narratif. La complexité ne doit pas être abandonnée au profit de calculs probabilistes ou le développement de modèles *ad hoc* (probabilistes ou non). Les effets de simplicité permettent d'aborder le processus de génération globalement, et de définir dans un même espace des notions relatives à la création de tension narrative (des critères de pertinence émotionnelle) et des notions relatives à la structure événementielle des récits (des critères de pertinence structurelle)
- **La procédure conflit-abduction :** c'est un modèle simple d'argumentation qui permet d'intégrer (de trouver une explication) une situation conflictuelle (ou une perspective de conflit). Ce modèle réalise des abductions successives dans le but de faire baisser l'intensité du conflit à résoudre, c'est-à-dire faire baisser l'intensité d'une nécessité source du conflit.
- **L'utilisation d'un modèle unique de résolution des conflits pour le calcul d'une explication et le calcul d'un plan d'actions :** un même modèle de résolution de conflit peut être utilisé à la fois pour expliquer l'occurrence d'un événement ainsi que pour calculer un plan d'actions à réaliser (le modèle de résolution raisonne sur des règles différentes selon le type de résolution)
- **L'utilisation du modèle de résolution des conflits à plusieurs niveaux dans le cadre de la génération de récits :** un modèle unique de résolution de conflits peut être utilisé à plusieurs niveaux dans le cadre de la génération de scénarios. Il peut à la fois être utilisé pour simuler le comportement d'agents rationnels confrontés à des conflits d'ordre cognitif, mais aussi pour résoudre des conflits au niveau de l'auteur (qui sont d'ordre cognitif internes à l'auteur, par exemple intégrer dans le monde fictif un

obstacle à la réalisation d'un plan d'un personnage). Là encore, le modèle raisonne sur des règles différentes, dans des espaces de représentation de la connaissance différents.

— **L'utilisation du modèle de résolution des conflits à plusieurs échelles dans le cadre de la génération de récits :** un modèle unique de résolution de conflits peut être utilisé pour résoudre des conflits à différentes échelles. Ainsi le modèle doit permettre de générer les 'grandes étapes' de la résolution d'un conflit narratif principal, mais aussi de générer les étapes intermédiaires qui permettent de lier entre elles ces 'grandes étapes' (ce qui peut être représenté par des conflits au niveau de l'auteur - il désire que les étapes soient liées mais elles ne le sont pas - et transformé en conflit à l'échelle des personnages), etc. L'algorithme de résolution doit pouvoir être utilisé à des niveaux de granularité divers de la représentation structurelle des récits.

Ce travail sur les principes nous a amené à modéliser cognitivement dans un même espace diverses notion liées à la pertinence des récits. En particulier :

— **Le conflit narratif :** nous l'avons défini comme équivalent à un ensemble de conflits cognitifs à l'échelle des personnages, c'est-à-dire comme un conflit entre des nécessités d'états (des évaluations basées sur des calculs de complexité) au sein de la représentation de la connaissance des personnages (voir chapitre 5).

— **La surprise :** nous avons défini le niveau de surprise comme étant égal à l'inattendu (défini lui-même comme un écart de complexité) d'une situation (voir équation 6.1). Par ailleurs, nous avons présenté un modèle de la surprise comme source de conflits d'ordre cognitif internes aux personnages qui observent l'occurrence d'une situation surprenante (voir partie 6.1). Ces personnages sont amenés à calculer une explication pour intégrer l'événement surprenant à leur représentation du monde (ce qui les amène à mettre à jour cette représentation du monde). Par ailleurs, l'occurrence de situations surprenantes peut faire apparaître des perspectives de conflits qui amènent les personnages à calculer des plans d'action.

— **La curiosité :** nous avons défini le niveau de curiosité associé à une situation comme l'intensité d'une nécessité qui a fait apparaître un conflit

cognitif autour de cette situation (voir équation 6.2). Le niveau de curiosité est donc tantôt égal à la taille de la meilleure explication d'une situation surprenante, tantôt égal à l'intensité émotionnelle perspective d'une situation dont l'occurrence anticipée est source de conflit.

— **Le suspense :** nous avons défini l'intensité du suspense comme l'évaluation *a priori* de l'intérêt d'une issue désirée, c'est-à-dire comme l'intensité émotionnelle *a priori* de cet événement. Cette intensité augmente avec la taille de la meilleure hypothèse causale qui permettrait de générer l'issue désirée (voir équation 6.3). La taille de cette hypothèse augmente à mesure que le temps passe et que l'on s'approche de la limite préalablement fixée de la phase de suspense (par exemple la fin d'un compte à rebours qui indique le moment où une bombe va exploser), ce qui fait augmenter le suspense.

— **La 'simplication' d'une situation :** nous avons appelé 'simplication' d'une situation une baisse de complexité de génération associée à cette situation. La 'simplication' *a posteriori* d'une situation contribue à la rendre postdictible. La 'simplication', en tant que variation de complexité cognitive, est source d'intérêt (voir page 105 et en particulier l'équation 4.27). Nous avons mis en évidence l'utilisation de 'simplications' brutales de situations surprenantes dans les récits dans le but de produire de l'intérêt.

— **La définition cognitive du critère de postdictibilité :** le critère de postdictibilité est respecté si toute situation du récit (sauf, éventuellement, l'élément perturbateur source du conflit initial) est non-inattendue pour l'audience de manière rétrospective, lorsque la fin du récit a été atteinte (voir la proposition 7.17).

— **La modélisation de la reconnaissance des intentions :** nous avons défini l'évaluation d'une intention particulière d'un acteur liée à la réalisation d'une action dans le but de produire une conséquence particulière anticipée (par cet acteur) comme l'intensité de la nécessité de l'action héritée de cette conséquence particulière calculée en prenant le point de vue de l'acteur (voir équation 7.4). L'intentionnalité est définie comme la somme (pondérée uniquement par les valences des conséquences) des contributions des conséquences supposées anticipées par l'acteur (voir équation 7.7). Ce modèle des intentions permet de calculer, dans le cadre d'une analyse

quantitative des récits, les intentions des personnages dans les récits et de déterminer les possibles mises en danger de la rationalité des personnages. Dans le cadre génératif, le modèle de résolution de conflit par abduction (dont le but est de faire baisser l'intensité du conflit à résoudre) est une possible implémentation de notre modèle cognitif des intentions (voir partie 9.3).

— **La modélisation de l'attribution de responsabilité et du jugement moral :** nous avons défini l'attribution de responsabilité comme une évaluation basée sur les conséquences observées d'une action, pour laquelle l'observateur utilise ses propres valeurs pour évaluer ces conséquences, mais se projette dans l'esprit de l'acteur pour évaluer la complexité anticipée par ce dernier pour produire la situation (voir équation 7.10). Autrement dit, la responsabilité d'un acteur est une évaluation de sa volonté de contribuer à l'occurrence d'un état de fait observé par un observateur-juge et évalué avec les propres valeurs de cet observateur-juge. Le jugement moral est la somme pondérée uniquement par les valences des conséquences observées des responsabilités particulières attribuées à chacune de ces conséquences (voir équation 7.13).

— **La définition d'un critère de crédibilité des personnages lié à la reconnaissance de leurs intentions :** ce critère de crédibilité des personnages correspond à la nécessité, pour l'audience, d'être capable rétrospectivement d'évaluer toutes les actions des personnages rationnels comme ayant été réalisées de manière intentionnelle (voir la proposition 7.16).

— **L'application du modèle de la surprise (*ie.* de l'attitude suite à l'observation d'une situation surprenante) à la génération de récits :** nous avons proposé un modèle de génération de récits (chapitre 9) qui consiste à générer récursivement des situations surprenantes pour un personnage, qui font obstacle à la réalisation des objectifs de ce personnage, et sont intégrées (par l'auteur virtuel) *via* des scripts. Les conflits sont représentés comme des conflits autour de l'état logique de prédicats qui font partie de la représentation du monde fictif d'un personnage (conflit entre deux nécessités calculées en fonction d'états logiques supposés et d'état logiques désirés de prédicats). La procédure de résolution des conflits par

abduction permet de calculer des plans d'actions intentionnels du point de vue des personnages (et présentés à l'audience pour qu'ils soient reconnus comme tels).

Les modèles existants de génération de scénarios implémentent, pour la plupart, *via* des méthodes *ad hoc*, uniquement certains aspects des modèles définis dans cette thèse. Nous pouvons analyser certaines de ces méthodes au regard des modèles présentés dans cette thèse.

Dans le modèle du conflit de Ware & Young (2010), le conflit narratif est un conflit qui apparaît lorsque deux plans valides (pour atteindre des buts) ne peuvent pas être exécutés tous les deux. C'est un cas particulier de conflit narratif au sens où nous l'avons défini. Les plans calculés qui ne peuvent pas être exécutés sont des résolutions de conflits cognitifs internes aux personnages impliqués dans le conflit narratif. L'impossibilité d'exécuter le plan est source d'une nécessité négative autour du but à atteindre, ce qui provoque un nouveau conflit et amène le ou les personnages à recalculer des plans pour atteindre leurs objectifs (résoudre leurs conflits internes). Le paradoxe narratif dans le système IDtension (Szilas, 2007; Szilas and Richle, 2013) est aussi un cas particulier de conflits dont les résolutions sont conflictuelles, ce qui amène le ou les protagonistes à re-calculer sans cesse un plan pour résoudre les conflits.

Le modèle Dramatis (O'Neill and Riedl, 2014) génère du suspense en introduisant la perspective d'occurrence d'un état non désiré pour un personnage et mesure le niveau de suspense en calculant le coût le moins élevé d'un opérateur qui permettrait à ce personnage de s'en sortir (c'est-à-dire de se trouver dans une situation désirée). C'est une application particulière de la définition cognitive du suspense (voir 6.3) dont la mesure augmente avec la taille de l'hypothèse causale la plus simple pour atteindre l'issue désirée. La simple introduction d'une issue qui crée une perspective de conflit est suffisante pour créer du suspense, car cela rend l'issue désirée plus incertaine et donc d'après l'équation 6.3 le suspense augmente. Le modèle de génération de récits de Miletitch et al. (2012) utilise cette méthode particulière pour créer du suspense en introduisant dans le cours du récit, *via* l'introduction d'une phrase d'accroche, une perspective d'issue négative. Le système Mexica (Pérez y Pérez and Sharples, 2001) utilise une approche similaire en contrôlant l'évolution de la tension narrative *via* la manipulation du "danger"

auquel font face les personnages.

Le modèle de création de surprise Prevoyant (Bae and Young, 2014) est une implémentation particulière du modèle de la surprise introduit dans la partie 6.1. Le système ôte de la structure événementielle des événements dont l'occurrence préalable à celle d'une situation donnée est nécessaire, ce qui crée une surprise lorsque la situation en question est présentée (la situation était non-nécessaire car des événements dont l'occurrence préalable est jugée nécessaire ont été omis). Cela crée un conflit autour de cette situation surprenante et donc de la curiosité (voir équation 6.2). Dans Prevoyant, les *flashbacks* provoquent une 'simplication' (voir équation 4.27) de l'événement surprenant (ils expliquent *a posteriori* l'occurrence de la situation surprenante). Cette 'simplication' a été préparée par le biais de 'présages' (*foreshadowing*). Elle est donc brutale, et donc particulièrement intéressante. Le mécanisme de prévoyant est le même mécanisme de 'simplication' observé à plusieurs reprise dans l'analyse de l'extrait du script d'Indiana Jones et la Dernière Croisade (voir chapitre 8).

Perspectives

Nous avons testé empiriquement certains aspects de ces modèles. Cependant, de nombreux autres aspects des modèles présentés mériteraient d'être testés, en particulier dans le cadre des jugements d'intentionnalité ou de responsabilité.

Les modèles développés dans nos travaux au sujet des jugements réalisés à propos de l'action rendent compte d'une appréciation *a posteriori* desdites actions. *A priori*, un individu peut être confronté à un ensemble de choix possibles, une disjonction d'actions possibles. Nous pensons en particulier que ce processus de décision est le résultat d'un processus dynamique d'évaluation et de comparaisons successives de complexités.

Nous avons émis l'hypothèse qu'un même modèle délibératif pouvaient être à l'œuvre pour résoudre ces conflits (c'est-à-dire pour calculer une explication ou planifier des actions à réaliser), ce qui doit être testé. Par ailleurs, le modèle d'argumentation basé sur la notion d'abduction que nous avons introduit en vue de générer des séquences narratives n'assure pas l'optimalité de l'hypothèse causale (une explication ou un plan) calculée, c'est-à-dire la plus simple localement (l'ob-

jectif étant d'assurer l'optimalité locale de l'intérêt) qui permet d'expliquer (ou de produire) au mieux une situation donnée.

Par ailleurs, lorsqu'un individu fait face à un objet inédit, l'explication qui permet d'intégrer cet objet modifie la représentation du monde de l'individu. Le système peut être amené à intégrer de nouvelles règles ou encore de nouveaux états. L'étude de la modification possible de l'espace de représentation du monde suite à un calcul de type argumentatif est un sujet à explorer. En particulier, il faudrait étudier les liens entre la modification de l'espace de représentation, qui est une forme de créativité, et les baisses de complexité au sein de la représentation du monde.

L'application des modèles aux domaines de l'analyse et la génération de scénarios doit être poursuivie. Dans le domaine de l'analyse, nous pouvons envisager la création d'un système d'évaluation automatique des récits, qui pourrait être une aide à la conception de scénarios. En utilisant une représentation qui permet de définir des modalités d'évaluation de la complexité, un tel système devrait pouvoir évaluer la pertinence des scénarios et pouvoir proposer des améliorations, au moins au niveau local, de ces scénarios.

Dans le domaine de la génération de scénarios, les modèles présentés peuvent servir de guide pour la création de modèles de génération de séquences narratives configurées par le suspense ou encore la curiosité. Par ailleurs, la génération de récits impliquant plusieurs personnages nécessite l'introduction et le maintien à jour de représentations du monde plus riches que celles que nous avons proposées. Un personnage donné doit, en particulier, être en mesure de se représenter au moins partiellement la connaissance de l'état du monde d'autres personnages (c'est la condition nécessaire pour réaliser une projection dans l'esprit des autres personnages) si l'un des objectifs du système est que ce personnage calcule des jugements moraux à l'égard de ces autres personnages.

Outre l'application des modèles au domaine de la génération de scénario, les modèles introduits précédemment peuvent être utilisés dans d'autres cadres. Les modèles de reconnaissance des intentions et de jugement moral peuvent servir à développer des agents virtuels de type 'juges' ou des systèmes experts pour l'aide au jugement et à la prise de décision.

Enfin, nos travaux portent sur la pertinence narrative. Chaque élément in-

troduit dans un récit doit servir à satisfaire au moins un critère de pertinence narrative. Nous avons détaillé dans nos travaux certains de ces critères, que nous avons répartis entre critères de pertinence structurelle et critères de pertinence émotionnelle. Un étude plus approfondie de cette notion peut permettre de mettre en avant une liste plus exhaustive de critères, et également de les modéliser. Par ailleurs, nous pensons qu'un récit doit respecter un critère de pertinence totale, c'est-à-dire que tout élément non pertinent n'a pas sa place dans le récit.

Dénouement

Contrairement aux approches précédentes de la génération de scénarios, nous avons traité les questions de pertinence structurelle et de pertinence émotionnelle dans un même espace de représentation. L'auteur qui produit un récit est sans cesse tiraillé entre le maintien de la cohérence du récit et la création d'intérêt narratif. L'intérêt ne vient pas d'une réorganisation postérieure des événements du récit ou de la construction d'un discours postérieure à la création de la structure événementielle. L'intérêt d'un récit est produit en même temps que sa structure événementielle. Nous avons donc fait l'hypothèse et montré que les mêmes mécanismes cognitifs, fondés sur des calculs de complexité cognitive, peuvent rendre compte de critères de pertinence liés au maintien de la vraisemblance ou la création d'intérêt. En cela, notre approche de l'analyse et la production de récits est originale. Si cette approche est poursuivie et étendue, elle peut permettre à la génération automatique de récits de ne plus être un rêve inaccessible mais une réalité. Ce jour-là, les machines donneront l'illusion d'avoir acquis une grande part d'humanité.

FIN

Annexe A

Annexe - Expériences

Dans cette section sont regroupées les annexes, en particulier les expériences évoquées dans ce manuscrit et leurs résultats.

A.1 Rôle de la complexité de description dans la reconnaissance d'un possible cas de plagiat

Cette expérience a été publiée dans :

Saillenfest, A. & Dessalles, J.-L. (2015). Some probability judgments may rely on complexity assessments. Proceedings of the 37th Annual Conference of the Cognitive Science Society, 2069-2074. Austin, TX : Cognitive Science Society.

Modèle théorique étudié

Cette expérience a pour but de tester les prédictions du modèle d'évaluation d'un possible cas de plagiat (voir partie 4.5).

Supposons qu'un individu se trouve face à un texte t_2 et trouve que ce texte ressemble fortement à un autre texte t_1 qu'il connaît (on note A l'algorithme, c'est-à-dire l'ensemble des transformations, qui transforme t_1 en t_2). Il peut s'agir d'une coïncidence ou d'un plagiat. La mesure de l'inattendu de la présence de t_2 si on fait l'hypothèse d'une coïncidence permet d'évaluer les chances de faire face à un cas de plagiat (plus la coïncidence est inattendue, plus le plagiat semble flagrant, voir l'équation 4.23). L'inattendu de la coïncidence vaut, d'après l'équation 4.23 :

$U(T_1 \& T_2) \geq C_w(T_1) + C_w(T_2) - C_d(t_1) - C_d(A) - C_d(t_2|T_1 \& A)$

Ce modèle permet de réaliser plusieurs prédictions (voir partie 4.5.2). Les prédictions testées sont les suivantes :

- Plus le texte soupçonné d'être le produit d'un plagiat est long, plus la coïncidence est perçue comme inattendue, et donc plus l'hypothèse de plagiat est privilégiée.
- La coïncidence est d'autant plus inattendue que les livres dans lesquels les textes se trouvent sont courts. Et donc l'hypothèse de plagiat devient une explication plus satisfaisante.
- Un plagiat sera plus évident si les textes impliqués se trouvent dans des endroit remarquables (au début par exemple) des livres dans lesquels on les trouve.
- Le plagiat est d'autant plus flagrant que l'algorithme qui transforme le texte plagié en une copie modifiée est simple à décrire.

Questionnaire original en français

Commencez par lire la mise en situation suivante. Veuillez ensuite répondre aux questions :

Mme Schmidt accuse M. Durand d'avoir plagié un passage que M. Durand aurait copié du recueil de nouvelles de Mme Schmidt.

1- Dans la proposition suivante, quelle valeur rendrait à vos yeux le plagiat le plus flagrant :

Le livre de Mme Schmidt fait 154 // 654 pages.

2- Dans la proposition suivante, quelle valeur rendrait à vos yeux le plagiat le plus flagrant :

Le roman de M. Durand fait 162 // 443 pages.

3- Dans la proposition suivante, quelle valeur rendrait à vos yeux le plagiat le plus flagrant :

Le passage mentionné par Mme Schmidt se trouve page 3 // 43 de son recueil.

4- Dans la proposition suivante, quelle valeur rendrait à vos yeux le plagiat le plus flagrant :

Le passage mentionné par Mme Schmidt fait 9 // 19 lignes.

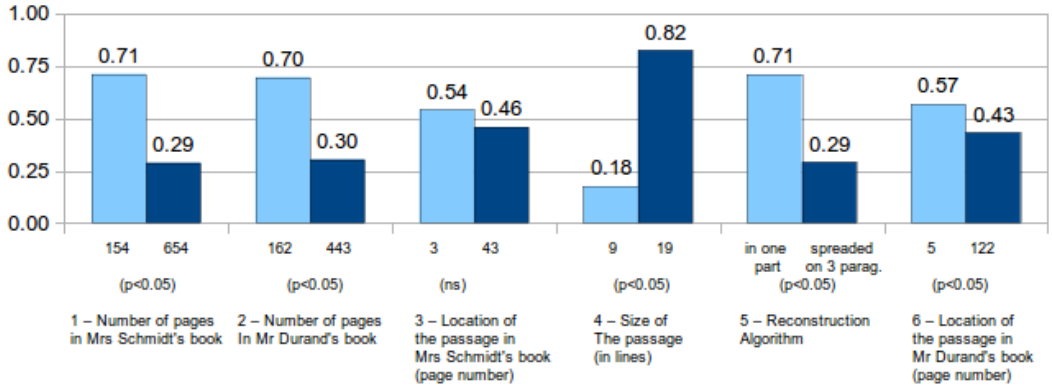
5- Dans la proposition suivante, quelle proposition rendrait à vos yeux le plagiat le plus flagrant :

On retrouve le passage mentionné par Mme Schmidt d'un seul tenant // réparti sur trois paragraphes dans le roman de M. Durand.

6- Dans la proposition suivante, quelle valeur rendrait à vos yeux le plagiat le plus flagrant :

On retrouve le passage mentionné par Mme Schmidt à la page 5 // 122 du roman de M. Durand.

FIGURE A.1 – Pourcentages de réponses pour chaque alternative des 6 propositions concernant un possible cas de plagiat



Méthode et Participants

352 individus (âgés de 16 à 63 ans, âge moyen de 28,64 ans (déviation standard 6,66), 276 individus de sexe féminin) ont participé au test en ligne. Les participants ont été recrutés via des messages sur les réseaux sociaux est une campagne d'affichage. Chacun devait lire l'une des 6 versions de l'histoire avec attention. Puis ils devaient répondre aux questions.

Un traitement manuel des résultats a permis d'écarter les fichiers de résultats incomplets ou ceux dont le temps de réponse était inférieur à 30 secondes.

Résultats et discussion

Les résultats (en pourcentage) des réponses aux questions est présenté dans la figure A.1. Un test binomial a permis de déterminer les écarts significatifs. On constate un effet significatif ($p < 0,05$) lié au nombre de pages des livres, à la longueur du passage, à la localisation du passage dans le livre de M. Durand ainsi qu'à la complexité de l'algorithme qui permet de reconstruire un texte à partir de l'autre. La localisation du passage dans le livre de Mme Schmidt n'a pas d'effet significatif ($p \approx 0,11$). Ces résultats sont en accord avec les prédictions ci-dessus.

A.2 Intentions et Responsabilité - L'accident de la route

Modèle théorique étudié

Cette expérience préliminaire avait pour but de tester les prédictions du modèle de l'intention (voir équation 7.4) et de la responsabilité (voir équations 7.13). Dans le cas présent, l'acteur s'appelle John, on a donc $\mathcal{A} = John$. L'observateur, qui varie selon la version de l'histoire présentée aux participants, est noté $\mathcal{O}$.

Si on note s la conséquence de l'acte de John (*ie.* l'accident), de valence négative pour John comme pour les observateurs, l'intention attribuée à John par un observateur vaut, d'après l'équation 7.4 :

$$Int^{\mathcal{O}}(John, a, s) = E^{\mathcal{O}\to John}(s) - C_w^{\mathcal{O}\to John}(s||a) \tag{A.1}$$

D'après l'équation 7.10, la responsabilité attribuée par un observateur à John vaut :

$$Res^{\mathcal{O}}(John, a, s) = |E^{\mathcal{O}}(s) - C_w^{\mathcal{O}\to John}(s||a)| \tag{A.2}$$

Par ailleurs, si on suppose que l'observateur accorde une certaine valeur à une conséquence atténuante s_1 (donc de valence positive pour cet observateur), un jugement moral (un degré d'approbation) égal à :

$$MJ^{\mathcal{O}}(John, a) = |Res^{\mathcal{O}}(John, a, s_1)| - |Res^{\mathcal{O}}(John, a, s)| \tag{A.3}$$

Ces modèles prédisent que la simplicité de la conséquence s pour l'observateur (liée aux liens entre cet observateur et la victime de l'accident) affectera l'attribution de la responsabilité par cet observateur, mais pas le niveau d'intentionnalité attribuée.

En outre, le modèle prédit que la révélation à l'observateur d'une circonstance atténuante ou d'éléments complémentaires pouvant avoir favorisé (sans que John le sache) la réalisation de la conséquence négative font croître le degré d'approbation (et donc font diminuer la responsabilité attribuée à John).

Questionnaire original en français

Test responsabilité - Janvier 2013

Voici un test court (environ 4 minutes) qui s'inscrit dans le cadre d'une série d'expériences sur l'étude de la responsabilité.

Ce test est anonyme.

Prenez le temps de lire attentivement le texte qui vous sera présenté. Ce texte ne varie pas d'une question à l'autre. Ensuite, veuillez s'il vous plaît répondre aux questions qui sont posées en-dessous.

Note : le texte présenté était l'un des 6 suivants, l'attribution du texte était aléatoire. En gras est indiqué la partie de l'histoire qui variait d'un participant à l'autre, cette mise en forme n'apparaît pas dans le texte présenté aux participants. Au début de chaque version de l'histoire est indiquée une référence qui sert à identifier chacune des 6 versions.

AF John est en déplacement professionnel. Il se sent fatigué cette nuit là. Il décide tout de même de rentrer chez lui en voiture au lieu de rester dormir à l'hôtel. Alors qu'il roule depuis seulement quelques minutes, il entend un choc sur sa voiture. Il sort de sa voiture pour voir ce qu'il a percuté. Il réalise alors qu'il a renversé un cycliste. Rapidement arrivés sur les lieux, les pompiers évacuent la victime qui souffre de graves lésions. Le lendemain, le journal Grand-Ouest annonce que la victime est **un acteur au sommet de sa carrière**. **Un fan de cet acteur** prend connaissance de ce récit en lisant le journal...

AR John est en déplacement professionnel. Il se sent fatigué cette nuit là. Il décide tout de même de rentrer chez lui en voiture au lieu de rester dormir à l'hôtel. Alors qu'il roule depuis seulement quelques minutes, il entend un choc sur sa voiture. Il sort de sa voiture pour voir ce qu'il a percuté. Il réalise alors qu'il a renversé un cycliste. Rapidement arrivés sur les lieux, les pompiers évacuent la victime qui souffre de graves lésions. Le lendemain, le journal Grand-Ouest annonce que la victime est **un acteur au sommet**

de sa carrière. **Un lecteur du journal** prend connaissance de ce récit en lisant le journal...

SF John est en déplacement professionnel. Il se sent fatigué cette nuit là. Il décide tout de même de rentrer chez lui en voiture au lieu de rester dormir à l'hôtel. Alors qu'il roule depuis seulement quelques minutes, il entend un choc sur sa voiture. Il sort de sa voiture pour voir ce qu'il a percuté. Il réalise alors qu'il a renversé un cycliste. Rapidement arrivés sur les lieux, les pompiers évacuent la victime qui souffre de graves lésions. Le lendemain, le journal Grand-Ouest annonce que la victime est **une personne qui habite à 500m du lieu de l'accident**. **Un ami de la victime** prend connaissance de ce récit en lisant le journal...

SB John est en déplacement professionnel. Il se sent fatigué cette nuit là. Il décide tout de même de rentrer chez lui en voiture au lieu de rester dormir à l'hôtel. Alors qu'il roule depuis seulement quelques minutes, il entend un choc sur sa voiture. Il sort de sa voiture pour voir ce qu'il a percuté. Il réalise alors qu'il a renversé un cycliste. Rapidement arrivés sur les lieux, les pompiers évacuent la victime qui souffre de graves lésions. Le lendemain, le journal Grand-Ouest annonce que la victime est **une personne qui habite à 500m du lieu de l'accident**. **Le frère de la victime** prend connaissance de ce récit en lisant le journal...

SR John est en déplacement professionnel. Il se sent fatigué cette nuit là. Il décide tout de même de rentrer chez lui en voiture au lieu de rester dormir à l'hôtel. Alors qu'il roule depuis seulement quelques minutes, il entend un choc sur sa voiture. Il sort de sa voiture pour voir ce qu'il a percuté. Il réalise alors qu'il a renversé un cycliste. Rapidement arrivés sur les lieux, les pompiers évacuent la victime qui souffre de graves lésions. Le lendemain, le journal Grand-Ouest annonce que la victime est **une personne qui habite à 500m du lieu de l'accident**. **Un lecteur du journal** prend connaissance de ce récit en lisant le journal...

PR John est en déplacement professionnel. Il se sent fatigué cette nuit là. Il décide tout de même de rentrer chez lui en voiture au lieu de rester dormir à l'hôtel. Alors qu'il roule depuis seulement quelques minutes, il entend un choc sur sa voiture. Il sort de sa voiture pour voir ce qu'il a percuté. Il réalise alors qu'il a renversé un cycliste. Rapidement arrivés sur les lieux, les pompiers évacuent la victime qui souffre de graves lésions. Le lendemain, le journal Grand-Ouest annonce que la victime est **une femme enceinte de 6 mois**. **Un lecteur du journal** prend connaissance de ce récit en lisant le journal...

Note : dans les questions ci-dessous, #Lecteur est à remplacer par l'identité du lecteur selon l'histoire présentée : ce fan de la victime, ce lecteur, cet ami de la victime, le frère de la victime

1. Selon vous, quel degré de responsabilité [#Lecteur] attribuera-t-il à John pour avoir blessé la victime ? (de 0 : Pas du tout responsable, à 5 : Extrêmement responsable)
2. Selon vous, [#Lecteur] jugera-t-il l'action de John intentionnelle et si oui à quel degré ? (de 0 : Pas du tout intentionnelle, à 5 : Extrêmement intentionnelle)
3. Selon vous, [#Lecteur] jugera-t-il John imprudent et si oui à quel degré ? (de 0 : Pas du tout imprudent, à 5 : Extrêmement imprudent)
4. "[#Lecteur] apprend que John a choisi de rentrer chez lui ce soir là pour réconforter sa femme qui avait appris la mort de son père le même jour." Cela va-t-il modifier le degré de responsabilité attribué par [#Lecteur] à John ? Si oui dans quelle mesure ? (Le degré de responsabilité baissera / ne changera pas / augmentera)
5. "L'enquête policière montre que l'éclairage de la route ne fonctionnait pas sur le lieu de l'accident." Cela va-t-il modifier le degré de responsabilité attribué par [#Lecteur] à John ? Si oui dans quelle mesure ? (Le degré de responsabilité baissera / ne changera pas / augmentera)

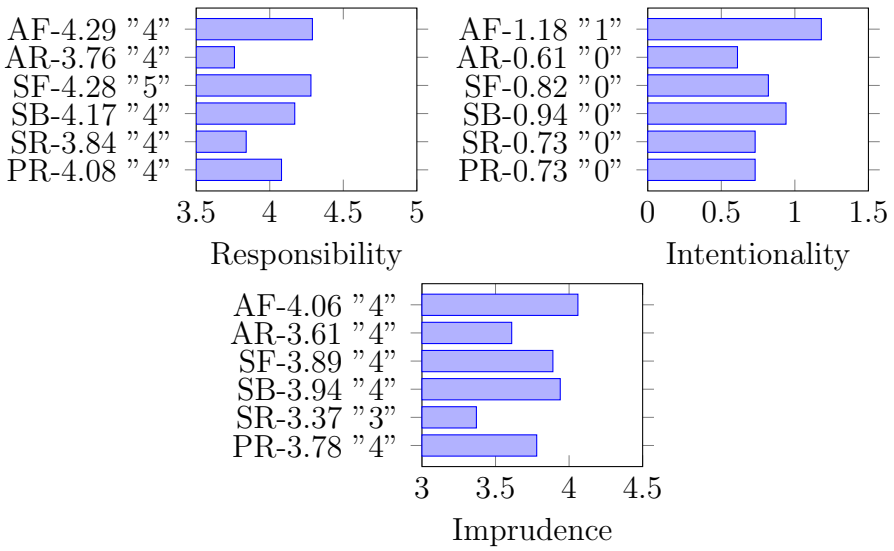


FIGURE A.2 – Moyenne "+ valeur médiane" des niveaux de responsabilité, d'intentionnalité et d'imprudence attribués pour les six versions de l'histoire testées (les références en ordonnée (AF, AR, SF, SB, SR et PR) correspondent aux références des six versions de l'histoire)

Méthode et Participants

454 individus (âgés de 15 à 76 ans, âge moyen de 29,52 ans (déviation standard 9,66), 290 individus de sexe féminin) ont participé au test en ligne. Les participants ont été recrutés *via* des messages sur les réseaux sociaux et une campagne d'affichage. Chacun devait lire l'une des 6 versions de l'histoire avec attention. Puis ils devaient répondre aux questions.

Un traitement manuel des résultats a permis d'écarter les fichiers de résultats incomplets ou qui contenaient des réponses aléatoires.

Résultats et discussion

Les taux de responsabilité attribués (et les taux d'attribution médians) sont présentés dans la Figure A.2. Les résultats ne suivent pas une distribution normale, nous avons donc utilisé un test de Mann-Whitney-Wilcoxon pour réaliser des comparaisons deux-à-deux entre les niveaux d'attribution de responsabilité,

d'intention et d'imprudence. (voir Tableau A.1)

TABLE A.1 – *p-values* venant du test de Mann-Whitney-Wilcoxon pour les niveaux de responsabilité, d'intention et d'imprudence attribués (chaque version de l'histoire est indiquée par sa référence (AF, AR, SF, SB, SR ou PR))

Versions appairées	Resp.	Intent.	Imprudence
AF / AR	$.007 **$	$.0119*$	$.00457 **$
SF / SB	ns	ns	ns
SF / SR	$< .001 ***$	ns	$.00266 **$
SB / SR	$.00826 **$	ns	$< .001 ***$
AR / SR	ns	ns	ns
AR / PR	ns	ns	ns
SR / PR	ns	ns	$0.0177*$

Le tableau A.2 regroupe les résultats des évaluations par les participants de l'évolution des niveaux de responsabilité attribués par l'observateur qui reçoit des informations complémentaires.

TABLE A.2 – Évaluation, par les participants, de l'effet des informations complémentaires (fournies au lecteur) sur le niveau de responsabilité attribué par ce lecteur (le niveau de responsabilité augmente (↗)/ diminue (↘)/ ne change pas (=)) (chaque version de l'histoire est indiquée par sa référence (AF, AR, SF, SB, SR ou PR))

	Information complémentaire					
	"réconfort de la femme"			"éclairage de la route éteint"		
	↗	↘	=	↗	↘	=
AF	3%	70%	27%	0%	61%	39%
AR	0%	77%	23%	0%	82%	18%
SF	7%	57%	36%	1%	66%	32%
SB	1%	56%	43%	2%	67%	31%
SR	3%	67%	30%	2%	79%	19%
PR	0%	57%	43%	0%	77%	23%

Les situations qui impliquent des amis, des membres de la famille ou n'importe quelle personne "simple" pour l'observateur entraînent une plus forte attribution de responsabilité de la part de cet observateur envers l'acteur (John). Ces résultats sont cohérents avec le modèle théorique (voir 7.10 et 7.13) qui prédit que l'attribution de responsabilité dépend de la réponse émotionnelle de l'observateur.

Les niveaux d'intentionnalité attribués ne sont globalement pas affectés par les relations entre la victime et l'observateur. Ceci est cohérent avec le modèle théorique (voir 7.7) qui prédit que l'attribution des intentions ne dépend que de l'état mental supposé de l'acteur (le niveau d'intentionnalité élevé dans l'histoire AF peut être lié à un supposé état mental "conspirateur" chez le fan).

Les participants considèrent majoritairement que les facteurs atténuants ("il veut réconforter sa femme") ou les informations supplémentaires ("l'éclairage de la route est éteint") tendent à faire baisser le niveau de responsabilité attribué par l'observateur. Cela est cohérent avec le modèle de l'équation 7.13. Le facteur atténuant ("il veut réconforter sa femme") introduit un terme positif dans l'équation du jugement moral, le modèle prédit donc que l'action sera plus approuvée (et John sera jugé moins responsable). De même, l'information sur l'éclairage de la route fait baisser la valeur de $C_w^{\mathcal{O}}(s)$ dans l'équation 7.10, et donc la responsabilité attribué à John (voir équation 7.10). Du point de vue de l'observateur, l'accident est moins inattendu une fois cette information dévoilée ; il est important de noter que la complexité de génération du point de vue de l'agent $C_w^{\mathcal{O}\to\mathcal{A}}(s||a)$ dans l'équation 7.10 reste inchangée après révélation de l'information sur l'éclairage, John ne pouvait pas anticiper ce problème.

Divers paramètres de l'histoire ont été manipulés pour tester les niveaux d'attribution d'intention et de responsabilité par un observateur spécifique. Ces résultats tendent à confirmer que distinguer les réponses émotionnelles de l'agent et de l'observateur est capital, et que cela est le principal point de divergence entre les notions d'intentionnalité et de responsabilité.

A.3 Le dilemme moral - La mine inondée

Cette expérience a été publiée dans :

Saillenfest, Antoine and Dessalles, Jean-Louis (2012). Role of Kolmogorov Complexity on Interest in Moral Dilemma Stories, In N. Miyake, D. Peebles & R. Cooper (Eds.), Proceedings of the 34th Annual Conference of the Cognitive Science Society, 947-952. Austin, TX : Cognitive Science Society

Modèle théorique étudié

Cette expérience avait pour but de tester les prédictions du modèle de l'intérêt (voir équation 4.4 et 4.5) et du jugement moral (équation 7.13). Dans le cas présent, l'acteur s'appelle Tom, on a donc $\mathcal{A} = Tom$ dans les équations précédentes. On conservera la notation $\mathcal{O}$ pour le lecteur observateur.

Dans les phases 1 à 3, l'issue est la même, il y a 5 survivants (notons cette conséquence désirée s_{+5}) et une victime (situation non désirée s_{-1}). Ainsi le système formé par les équations 4.4, 4.5 et 7.13 appliqué à l'histoire de la mine inondée, pour les phases 1 à 3 devient :

$$\begin{cases} I^{\mathcal{O}}(s) = E_h^{\mathcal{O}\to *}(s) + U^{\mathcal{O}\to *}(s) \\ MJ^{\mathcal{O}}(\mathcal{A}, a) = +|E^{\mathcal{O}}(s_{+5}) - C_w^{\mathcal{O}\to\mathcal{A}}(s_{+5}||a)| - |E^{\mathcal{O}}(s_{-1}) - C_w^{\mathcal{O}\to\mathcal{A}}(s_{-1}||a)| \end{cases} \tag{A.4}$$

Le modèle prédit donc que :

— Si on maintient la complexité de génération de la situation-conséquence constante (*ie.* $C_w^{\mathcal{O}\to\mathcal{A}}(s_{+5}||a) + C_w^{\mathcal{O}\to\mathcal{A}}(s_{-1}||a)$ constant), l'intérêt ne varie pas. En revanche, les actions dont les conséquences négatives sont les plus directes seront les moins approuvées (le jugement moral est une fonction croissante de $C_w^{\mathcal{O}\to\mathcal{A}}(s_{-1}||a)$). Ceci est testé dans l'expérience 1.

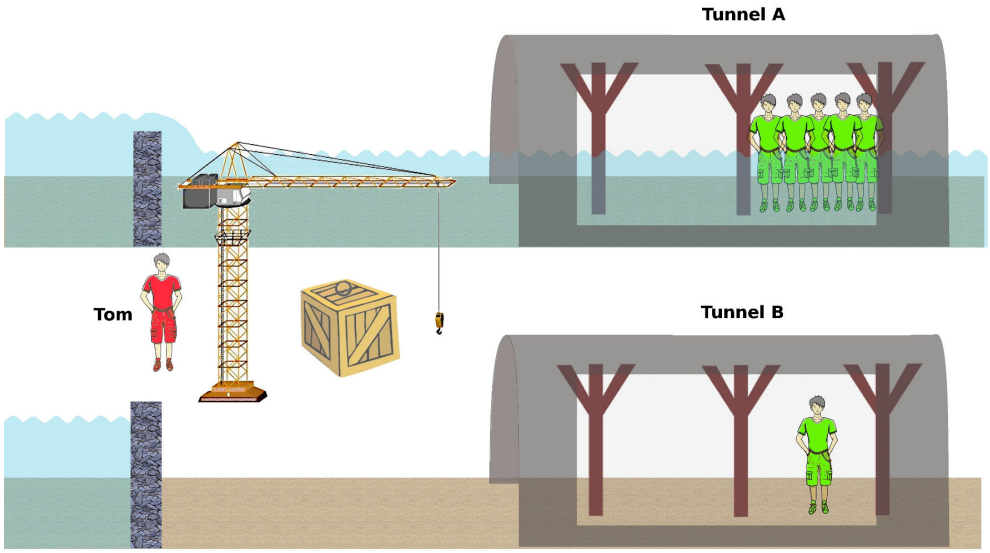
— Les actions dont les conséquences positives sont les plus directes seront les plus approuvées (le jugement moral est une fonction décroissante de $C_w^{\mathcal{O}\to\mathcal{A}}(s_{+5}||a)$) mais les moins intéressantes (l'intérêt est une fonction croissante de $C_w^{\mathcal{O}\to\mathcal{A}}(s_{+5}||a)$). Ceci est testé en phase 2.

— Les actions dont les conséquences négatives sont les plus simples pour le lecteur (ou dont l'émotion hypothétique est la plus forte pour le lecteur) seront les moins approuvées (le jugement moral croît avec $C_d^{\mathcal{O}}(s_{-1})$ et décroît avec $E_h(s_{-1})$) mais les plus intéressantes (l'intérêt décroît avec $C_d^{\mathcal{O}}(s_{-1})$ et croît avec $E_h(s_{-1})$). Ceci est testé en phase 3.

La phase 4 teste l'influence d'un déroulement inattendu des événements. Un déroulement inattendu des événements provoque une hausse de l'intérêt narratif. par ailleurs, le jugement moral dépendant à la fois des conséquences anticipées et réelles de l'action, la survenue d'un événement inattendu qui provoque une conséquence plus négative qu'attendue est censée faire baisser le niveau d'approbation moral de ladite action. Nous testons ces prédictions dans la phase 4.

Questionnaire original en français

FIGURE A.3 – Illustration du dilemme de la mine inondée



Vous trouverez ci-après le questionnaire présenté aux participants de l'expérience du dilemme moral (étude de variation de l'histoire de la mine inondée).

Questionnaire

Vous allez lire un récit fictif comportant des variantes. Nous avons besoin de votre jugement pour anticiper les réactions des lecteurs de l'histoire. **Le début du récit est identique pour les 4 phases du test.**

À chaque phase du test, plusieurs fins alternatives vous seront proposées. Vous devrez donner votre avis sur l'aspect moral de chaque alternative ainsi que leur intérêt via un système de notation.

Lisez bien le texte une fois. Dans chaque phase, lisez-bien les alternatives et les questions.

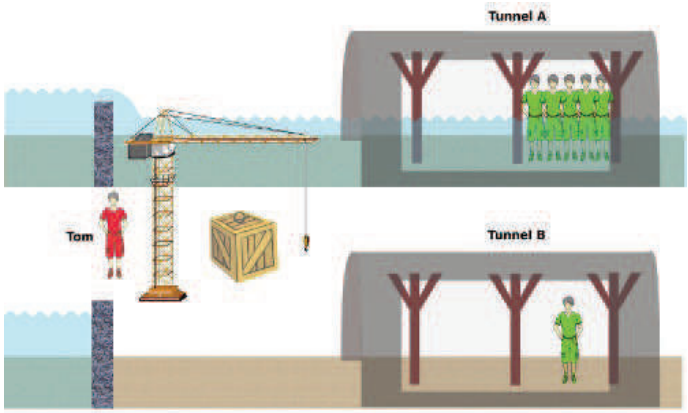
Pour chaque phase du test, vous pouvez, si vous le souhaitez, laisser un ou plusieurs commentaire(s) dans la zone prévue à cet effet.

Commencez par répondre aux deux questions suivantes :

O Homme O Femme

Votre âge : ____

Vous pouvez commencer le questionnaire !



Phase 1 :

Tom est un ouvrier mineur qui a quitté l'Écosse, son pays natal, pour aller travailler dans une mine en Argentine. Voici ce qui lui est arrivé il y a deux ans, alors qu'il surveillait l'amont du chantier.
La rivière était en crue soudaine. Elle était en train d'inonder l'un des deux tunnels de la mine. Tom savait que dans ce tunnel (le tunnel A) se trouvaient 5 personnes. Dans l'autre tunnel (le tunnel B) se trouvait une seule personne. L'eau montait rapidement, et le courant dans le tunnel A augmentait, les personnes bloquées allaient être noyées.
Tom sait que l'interruption du courant dans l'un des tunnels va de pair avec l'inondation du second et la mort de celui ou ceux qui s'y trouvent. Il est impossible d'interrompre le courant dans les deux tunnels en même temps.
Tom était le seul à pouvoir agir. Il se trouvait à l'entrée des deux tunnels, près d'une grue et d'une caisse lourde et volumineuse.

---

1. ***Tom poussa la caisse à l'entrée du tunnel A pour stopper le courant dans ce tunnel.***
   Cela interrompit le courant dans le tunnel A, le tunnel B fut inondé.
   Cinq personnes furent sauvées, une personne mourut noyée.
2. ***Tom monta dans la grue, saisit la caisse avec le crochet, amena la caisse au-dessus de l'entrée du tunnel A et la lâcha au milieu du courant pour stopper le courant dans ce tunnel.***
   Cela interrompit le courant dans le tunnel A, le tunnel B fut inondé.
   Cinq personnes furent sauvées, une personne mourut noyée.
3. ***Tom brisa le barrage du tunnel B pour inonder immédiatement ce tunnel.***
   Cela interrompit le courant dans le tunnel A, le tunnel B fut inondé.
   Cinq personnes furent sauvées, une personne mourut noyée.

D'après vous, les lecteurs de l'histoire vont-ils approuver les actions de Tom ?
(donner une note entre -5 (désapprouvent) et 5 (approuvent))

Pour l'alternative 1 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 2 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 3 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

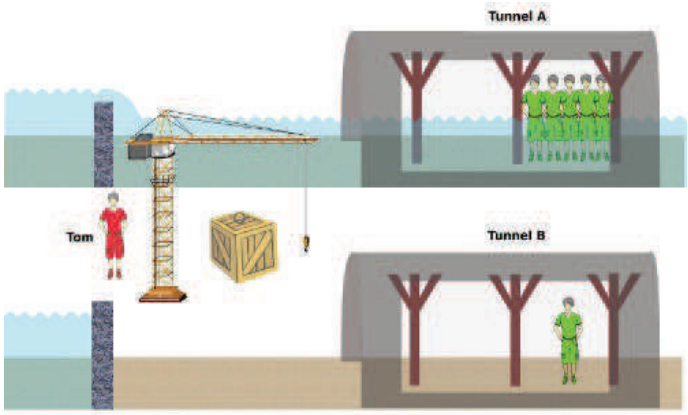
D'après vous, les lecteurs de l'histoire vont-ils trouver les alternatives intéressantes ?
(donner une note entre -5 (pas intéressante) et 5 (intéressante))

Pour l'alternative 1 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 2 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 3 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Commentaires :



Phase 2 :

Tom est un ouvrier mineur qui a quitté l'Écosse, son pays natal, pour aller travailler dans une mine en Argentine. Voici ce qui lui est arrivé il y a deux ans, alors qu'il surveillait l'amont du chantier.
La rivière était en crue soudaine. Elle était en train d'inonder l'un des deux tunnels de la mine. Tom savait que dans ce tunnel (le tunnel A) se trouvaient 5 personnes. Dans l'autre tunnel (le tunnel B) se trouvait une seule personne. L'eau montait rapidement, et le courant dans le tunnel A augmentait, les personnes bloquées allaient être noyées.
Tom sait que l'interruption du courant dans l'un des tunnels va de pair avec l'inondation du second et la mort de celui ou ceux qui s'y trouvent. Il est impossible d'interrompre le courant dans les deux tunnels en même temps.
Tom était le seul à pouvoir agir. Il se trouvait à l'entrée des deux tunnels, près d'une grue et d'une caisse lourde et volumineuse.

---

1. *Tom poussa la caisse devant l'entrée du tunnel A.*
   Elle fut emportée par le courant, heurta les étais (piliers) dans le tunnel et, comme l'espérait Tom, certains étais se brisèrent et une partie du toit à l'entrée du tunnel s'effondra.
   Cela interrompit le courant dans le tunnel A. Les cinq personnes furent sauvées, mais le tunnel B fut inondé et une personne mourut noyée.

2. *Tom poussa la caisse à l'entrée du tunnel A.*
   Comme il l'espérait, son poids la maintint sur place en travers du courant.
   Cela interrompit le courant dans ce tunnel. Les cinq personnes furent sauvées, mais le tunnel B fut inondé et une personne mourut noyée.

3. *Tom poussa la caisse devant l'entrée du tunnel A.*
   Elle fut emportée par le courant, heurta les étais (piliers) dans le tunnel, des poutres tombèrent du plafond, elles furent elles aussi emportées par le courant et se retrouvèrent bloquées par d'autres étais ce qui, comme l'espérait Tom, forma un nouveau barrage.
   Cela interrompit le courant dans le tunnel A. Les cinq personnes furent sauvées, mais le tunnel B fut inondé et une personne mourut noyée.

4. *Tom poussa la caisse devant l'entrée du tunnel A.*
   Elle fut emportée par le courant et, comme l'espérait Tom, se retrouva coincée dans les étais du tunnel.
   Cela interrompit le courant dans le tunnel A. Les cinq personnes furent sauvées, mais le tunnel B fut inondé et une personne mourut noyée.

D'après vous, les lecteurs de l'histoire vont-ils approuver les actions de Tom ?
(donner une note entre -5 (désapprouvent) et 5 (approuvent))

Pour l'alternative 1 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 2 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 3 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 4 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

D'après vous, les lecteurs de l'histoire vont-ils trouver les alternatives intéressantes ?
(donner une note entre -5 (pas intéressante) et 5 (intéressante))

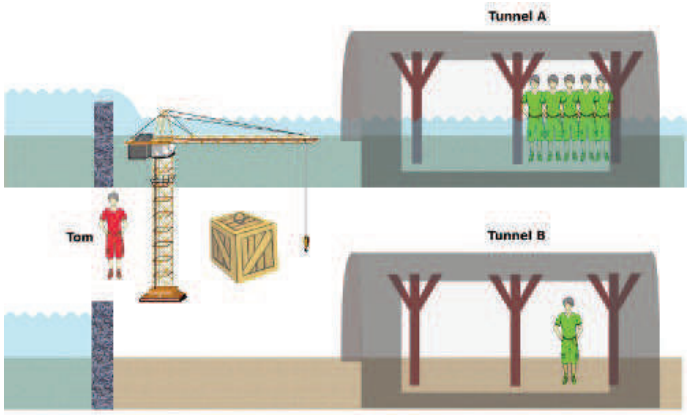
Pour l'alternative 1 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 2 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 3 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 4 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Commentaires :



Phase 3 :

Tom est un ouvrier mineur qui a quitté l'Écosse, son pays natal, pour aller travailler dans une mine en Argentine. Voici ce qui lui est arrivé il y a deux ans, alors qu'il surveillait l'amont du chantier.
La rivière était en crue soudaine. Elle était en train d'inonder l'un des deux tunnels de la mine. Tom savait que dans ce tunnel (le tunnel A) se trouvaient 5 personnes. Dans l'autre tunnel (le tunnel B) se trouvait une seule personne. L'eau montait rapidement, et le courant dans le tunnel A augmentait, les personnes bloquées allaient être noyées.
Tom sait que l'interruption du courant dans l'un des tunnels va de pair avec l'inondation du second et la mort de celui ou ceux qui s'y trouvent. Il est impossible d'interrompre le courant dans les deux tunnels en même temps.
Tom était le seul à pouvoir agir. Il se trouvait à l'entrée des deux tunnels, près d'une grue et d'une caisse lourde et volumineuse.

---

1. ***Tom savait que c'était un de ses amis qui se trouvait dans le tunnel B.***
   Tom poussa la caisse à l'entrée du tunnel A.
   Cela interrompit le courant dans le tunnel A, le tunnel B fut inondé.
   Cinq personnes furent sauvées et ***l'ami de Tom mourut noyé.***

2. ***Tom savait que c'était un enfant de 10 ans qui se trouvait dans le tunnel B.***
   Tom poussa la caisse à l'entrée du tunnel A.
   Cela interrompit le courant dans le tunnel A, le tunnel B fut inondé.
   Cinq personnes furent sauvées et ***l'enfant mourut noyé.***

3. *Tom poussa la caisse à l'entrée du tunnel A.*
   Cela interrompit le courant dans le tunnel A, le tunnel B fut inondé.
   Cinq personnes furent sauvées et ***une personne mourut noyée.***

4. ***Tom savait que c'était son propre cousin qui se trouvait dans le tunnel B.***
   Tom poussa la caisse à l'entrée du tunnel A.
   Cela interrompit le courant dans le tunnel A, le tunnel B fut inondé.
   Cinq personnes furent sauvées et ***le cousin de Tom mourut noyé.***

D'après vous, les lecteurs de l'histoire vont-ils approuver les actions de Tom ?
(donner une note entre -5 (désapprouvent) et 5 (approuvent))

Pour l'alternative 1 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 2 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 3 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 4 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

D'après vous, les lecteurs de l'histoire vont-ils trouver les alternatives intéressantes ?
(donner une note entre -5 (pas intéressante) et 5 (intéressante))

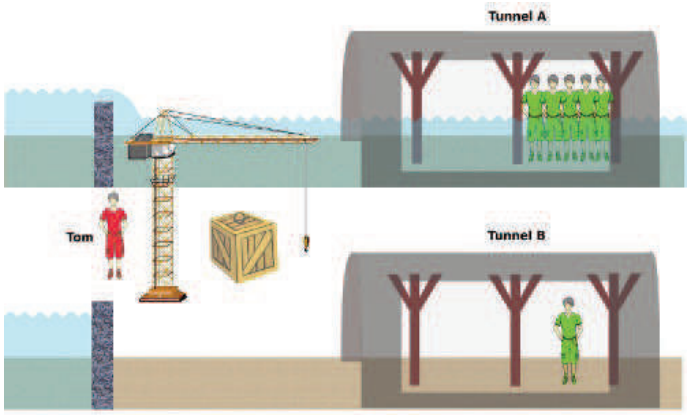
Pour l'alternative 1 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 2 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 3 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 4 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Commentaires :



Phase 4 :

Tom est un ouvrier mineur qui a quitté l'Écosse, son pays natal, pour aller travailler dans une mine en Argentine. Voici ce qui lui est arrivé il y a deux ans, alors qu'il surveillait l'amont du chantier.
La rivière était en crue soudaine. Elle était en train d'inonder l'un des deux tunnels de la mine. Tom savait que dans ce tunnel (le tunnel A) se trouvaient 5 personnes. Dans l'autre tunnel (le tunnel B) se trouvait une seule personne. L'eau montait rapidement, et le courant dans le tunnel A augmentait, les personnes bloquées allaient être noyées.
Tom sait que l'interruption du courant dans l'un des tunnels va de pair avec l'inondation du second et la mort de celui ou ceux qui s'y trouvent. Il est impossible d'interrompre le courant dans les deux tunnels en même temps.
Tom était le seul à pouvoir agir. Il se trouvait à l'entrée des deux tunnels, près d'une grue et d'une caisse lourde et volumineuse.

---

1. *Tom brisa le barrage du tunnel B pour inonder immédiatement ce tunnel.*
   Mais ça ne se passa pas comme prévu, les ruines du barrage formèrent un nouveau barrage à l'entrée du tunnel B, cela ne suffit pas à stopper le courant dans le tunnel A.
   Cinq personnes moururent, une personne survécut.

2. *Tom poussa la caisse à l'entrée du tunnel A pour stopper le courant dans ce tunnel.*
   Cela interrompit le courant dans ce tunnel A, le tunnel B fut inondé.
   Cinq personnes furent sauvées, une personne mourut noyée.

3. *Tom brisa le barrage du tunnel B pour inonder immédiatement ce tunnel.*
   Cela interrompit le courant dans ce tunnel A, le tunnel B fut inondé.
   Cinq personnes furent sauvées, une personne mourut noyée.

4. *Tom poussa la caisse à l'entrée du tunnel A pour stopper le courant dans ce tunnel.*
   Mais ça ne se passa pas comme prévu, la caisse ne suffit pas à stopper le courant dans le tunnel A.
   Cinq personnes moururent, une personne survécut.

D'après vous, les lecteurs de l'histoire vont-ils approuver les actions de Tom ?
(donner une note entre -5 (désapprouvent) et 5 (approuvent))

Pour l'alternative 1 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 2 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 3 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 4 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

D'après vous, les lecteurs de l'histoire vont-ils trouver les alternatives intéressantes ?
(donner une note entre -5 (pas intéressante) et 5 (intéressante))

Pour l'alternative 1 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 2 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 3 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Pour l'alternative 4 :
○ ○ ○ ○ ○ ○ ○ ○ ○ ○
-5 -4 -3 -2 -1 1 2 3 4 5

Commentaires :

Méthode et Participants

Un total de 64 participants (agés de 19 à 65 ans, moyenne de 26,11 ans (déviation standard : 7,72), 26 participants de genre féminin) a pris part au test. L'étude a été menée en ligne. Les participants, principalement des étudiants-ingénieurs, ont été recrutés *via* les réseaux sociaux et une campagne d'affichage.

Il leur a été demandé de lire l'histoire de la mine inondée avec attention. Pour chacune des phases du test, plusieurs fins alternatives étaient proposées, dans lesquelles les actions de Tom, les liens causaux ainsi que les conséquences variaient. Les alternatives ont été présentées dans le même ordre à tous les participants. En utilisant la numérotation définie ci-après, l'ordre de présentation des alternatives pour les phases 1 à 4 était 1-2-3, 3-1-4-2, 2-4-1-3 et 1-4-2-3.

Pour chacune des alternatives, les participants devaient répondre aux deux questions suivantes *via* une échelle composée de 10 points (voir le questionnaire précédent) :

— D'après vous, les lecteurs de l'histoire vont-ils approuver les actions de Tom ? (donner une note entre -5 (désapprouvent) et 5 (approuvent))

— D'après vous, les lecteurs de l'histoire vont-ils trouver les alternatives intéressantes ? (donner une note entre -5 (pas intéressante) et 5 (intéressante))

La note 0 n'était pas proposée au participants pour les forcer à choisir entre approbation et désapprobation (resp. intéressant et non-intéressant). Les temps de réponses (avec déviation standard) pour les 4 phases étaient 3min32s(1min12s), 3min12s(1min36s), 1min36s(1min17s) et 2min17s(1min31s). Les résultats ont été traités manuellement pour éliminer les résultats aléatoires ou incomplets.

Les différentes phases du test explorent successivement le rôle de la complexité (au sens de Kolmogorov) sur l'évaluation du jugement moral et de l'intérêt narratif.

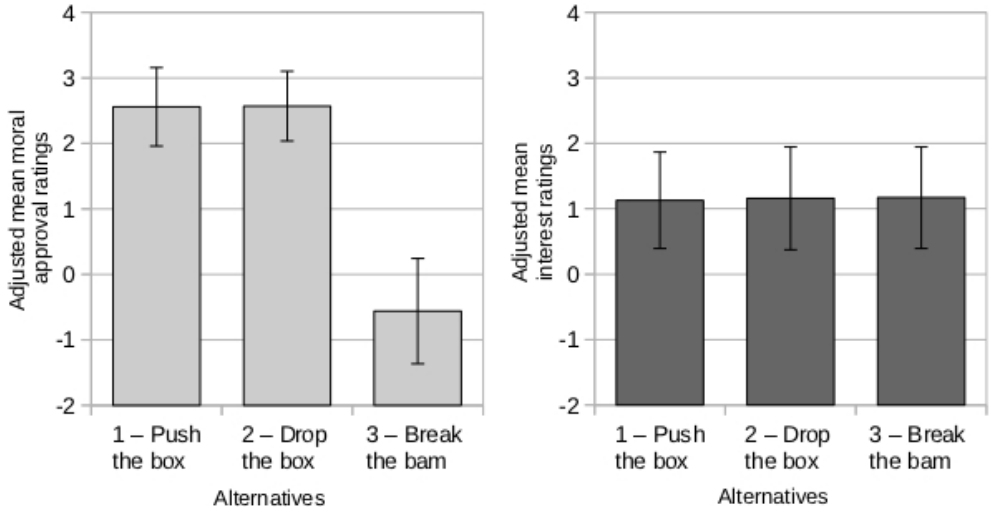
Résultats et discussions

Phase 1

Les études précédentes ont montré que les actions qui “”blessent une victime” ont tendant à être moralement approuvées si leurs conséquences négatives sont plus indirectes (Cushman et al., 2006). Dans la phase 1 de ce test, nous avons essayé de

reproduire ce résultat et d'explorer les effets des variation de la complexité causale sur l'intérêt narratif. Nous analysons également le rôle de la complexité de l'action.

FIGURE A.4 – Résultats de la phase 1 - influence de la distance causale entre l'action et sa conséquence négative : niveau moyen d'approbation (à gauche) et d'intérêt (à droite) pour chacune des trois alternatives testées. Les barres d'erreur indiquent l'intervalle de confiance à 95%



Il y a un effet significatif de l'intention de l'acteur (F-test : $F(1, 190) = 60,87$, $p < 0,0001$) sur le jugement moral mais pas d'effet significatif sur l'intérêt ($F(1; 190) = 0,0045$, $p = 0,95$). Une comparaison par paires (*pairwise comparison*) des deux alternatives impliquant la caisse révèle que la manière d'amener la caisse au milieu du courant n'a pas d'effet sur le jugement moral ($p = 1$) et sur l'évaluation de l'intérêt ($p = 1$) (voir figure A.4).

Dans cette phase, nous avons répliqué un résultat classique au sujet du problème du tramway : les personnes tendent à n'approuver les actions qui ont pour effet de provoquer la mort d'un unique individu que s'il s'agit d'un effet secondaire et non d'un moyen.

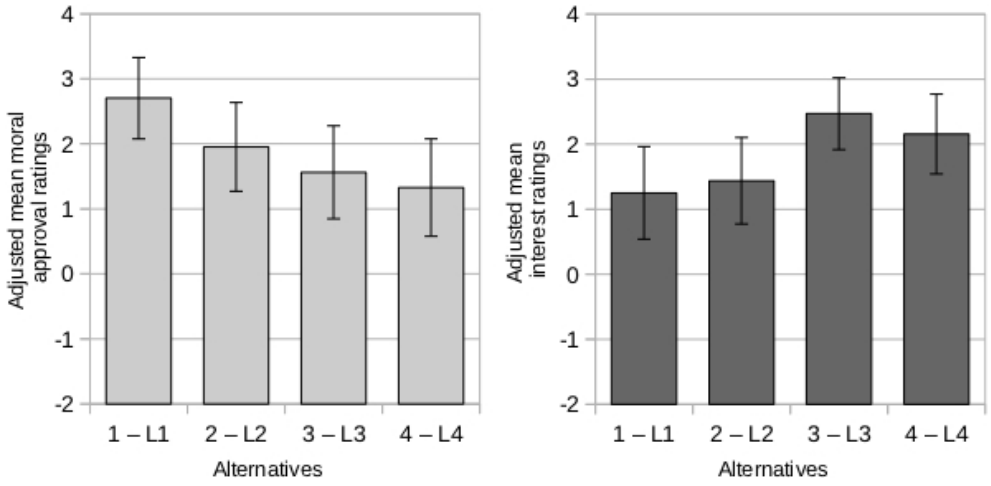
Le principal résultat de cette phase est que les situations qui sont le moins approuvées par les participants ne sont pas nécessairement les plus intéressantes. Le fait de blesser quelqu'un comme moyen ou comme effet secondaire n'a d'effet que sur le jugement moral, pas sur l'intérêt de l'histoire. Ceci est cohérent avec les modèles de l'intérêt (voir équation 4.4 et 4.5 et du jugement moral 7.13). En

effet, dans cette phase, l'intensité émotionnelle des conséquences anticipées n'est pas manipulée. Dans les alternatives 1 et 2, la chaîne causale entre l'action de Tom et ses conséquences négatives est plus complexe que dans l'alternative 3. Dans la mesure où toutes les conséquences sont aussi inattendues les unes que les autres, notre modèle prédit effectivement que l'intérêt narratif ne sera pas affecté. Notre modèle (voir équation 7.13) prédit en outre que les actions ont plus tendance à être approuvées si leurs conséquences désirées sont plus directes et leurs conséquences non-désirées plus indirectes.

Phase 2

La phase 2 explore l'influence de la longueur de la chaîne causale d'événements entre une action et ses conséquences sur l'évaluation morale et l'intérêt narratif.

FIGURE A.5 – Résultats de la phase 2 - influence de la distance causale entre l'action et sa conséquence négative : niveau moyen d'approbation (à gauche) et d'intérêt (à droite) pour chacune des trois alternatives testées. Les barres d'erreur indiquent l'intervalle de confiance à 95%



On constate un effet significatif de la longueur des chaînes causales d'événements, à la fois sur l'évaluation morale ($F(3; 252) = 3,01$, $p = 0.03$), mais aussi sur l'intérêt ($F(3; 252) = 3,29$, $p = 0,02$) (voir figure A.5).

Dans cette phase, seule la complexité causale de génération est manipulée. Dans la mesure où l'inattendu (et donc l'intérêt) est une fonction croissante de

cette complexité (voir équation 4.4), des chaînes déterministes d'événements plus longues rendent les conséquences plus inattendues. Par ailleurs, dans l'équation du jugement moral (voir équation 7.13), l'approbation de l'action est une fonction décroissante de cette complexité (pour les conséquences positives). Ainsi les prédictions de notre modèle sont correctes : plus la complexité de génération d'une conséquence positive augmente, plus cette conséquence sera intéressante mais l'action sera moins approuvée. Jones (1991) utilise l'expression *probability of effect* pour faire référence à la probabilité d'occurrence d'une conséquence négative. Nous pensons que la notion de probabilité est inadaptée ; dans beaucoup de cas la longueur de la chaîne d'événements est une notion plus pertinente, même si cette chaîne est déterministe. La notion d'inattendu définie comme différence de complexité capture correctement ce phénomène.

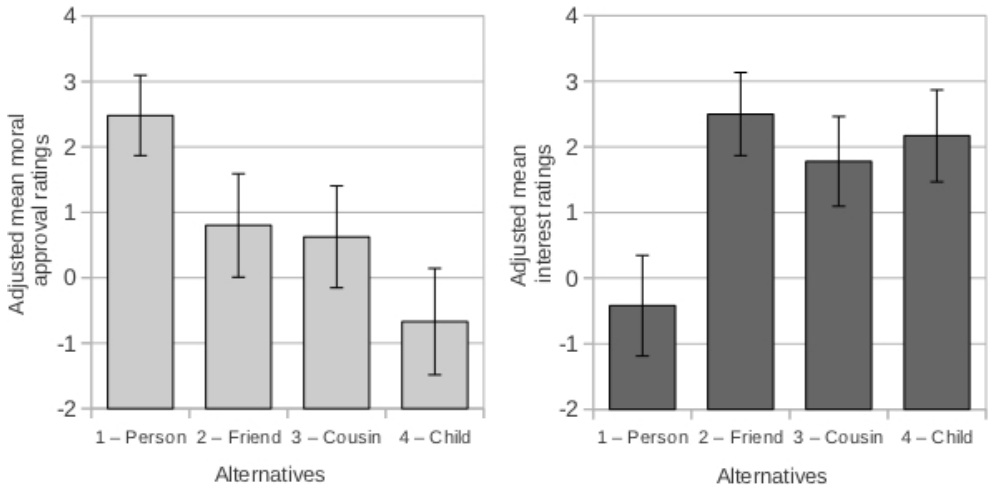
Phase 3

Les études précédentes sur la moralité (voir (Bleske-Rechek et al., 2010)) ont montré que l'identité des victimes influence notre jugement moral. Dans cette phase, nous étudions l'influence de la victime dans le tunnel B sur le jugement moral et l'intérêt narratif.

On constate un effet significatif de l'identité de la victime (une personne non-définie, un cousin de Tom, un ami de Tom ou un enfant de 10 ans) sur le jugement moral ($F(3, 252) = 11,79$, $p < 0,05$) et sur l'intérêt narratif ($F(3, 252) = 14,26$, $p < 0,05$). Une série de comparaisons par paires à permis de clarifier la nature de cette interaction. Le niveau d'approbation est significativement plus grand lorsque la victime est non-définie (personne non-définie *vs.* un ami de Tom : $F(1, 126) = 11,32$, $p = 0,001$, personne non-définie *vs.* un cousin de Tom : $F(1, 126) = 14,07$, $p = 0,0003$, personne non-définie *vs.* un enfant de 10 ans : $F(1, 126) = 38,25$, $p < 0,0001$) et les histoires sont jugées comme étant significativement moins intéressantes (personne non-définie *vs.* un ami de Tom : $F(1, 126) = 34,37$, $p < 0,0001$, personne non-définie *vs.* un cousin de Tom : $F(1, 126) = 18,32$, $p < 0,0001$, personne non-définie *vs.* un enfant de 10 ans : $F(1, 126) = 24,87$, $p < 0,0001$) (voir figure A.6).

Les alternatives dans lesquelles la victime est moins complexe à décrire (pour

FIGURE A.6 – Résultats de la phase 3 - influence de l'identité de la victime : niveau moyen d'approbation (à gauche) et d'intérêt (à droite) pour chacune des trois alternatives testées. Les barres d'erreur indiquent l'intervalle de confiance à 95%



Tom) (*ie.* celles impliquant un cousin de Tom ou un ami de Tom) sont plus intéressantes du point de vue narratif que celle dans laquelle la victime est non-définie. Comme le prédit notre modèle, l'inattendu augmente (voir équation 4.4) et l'approbation diminue (voir le terme $C_d^{\mathcal{O}\to\mathcal{A}}(s||a)$ dans l'équation 7.13) quand les conséquences non désirées sont plus simples. Ainsi, le modèle prédit que les alternatives impliquant un cousin ou un ami de Tom seront moins approuvées que celle impliquant une personne non-définie, mais qu'elles seront plus intéressantes.

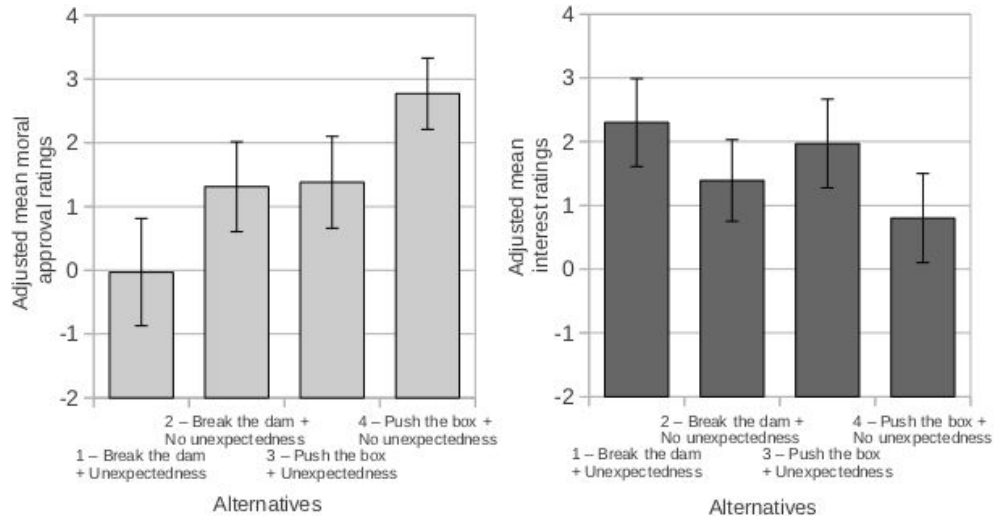
Lorsque la victime est un enfant, l'émotion hypothétique est plus grande (on accorde plus de valeur à la vie d'un enfant qu'à celle d'un adulte). Notre modèle fait une prédiction correcte lorsque la victime est un enfant ; cette alternative sera moins approuvée mais plus intéressante.

Notre modèle est en accord avec les résultats de (Bleske-Rechek et al., 2010) ainsi qu'avec certains composants de l'intensité morale de Jones tels que la proximité (définie comme le sentiment de proximité d'un agent moral vis-à-vis de la victime) ou encore le consensus social (*The social consensus of the moral issue is defined as the degree of social agreement that a proposed act is evil (or good)*(Jones, 1991)).

Phase 4

Deux éléments sont manipulés dans cette phase : l'action plus ou moins directe de Tom et la présence ou l'absence d'inattendu dans la chaîne d'événements.

FIGURE A.7 – Résultats de la phase 4 - influence de la distance causale entre l'action et sa conséquence négative et de la présence d'un déroulement inattendu ou non des événements : niveau moyen d'approbation (à gauche) et d'intérêt (à droite) pour chacune des trois alternatives testées. Les barres d'erreur indiquent l'intervalle de confiance à 95%



Ces résultats sont cohérents avec les observations des phases précédentes. Ils montrent que les variations de distance causale anticipées par Tom n'affectent que le jugement moral. Par ailleurs, ces résultats montrent une influence de l'augmentation combinée de l'inattendu de la conséquence et de son intensité émotionnelle hypothétique (il y a 5 morts au lieu d'un anticipé) sur le niveau d'intérêt narratif, qui augmente, mais également sur le niveau d'approbation moral, qui diminue (voir figure A.7).

Ces résultats ne font pas que confirmer les résultats des phases précédentes. Ils viennent aussi confirmer les prédictions du modèle sur l'alternative la moins approuvée (et la plus intéressante) "Tom casse la barrage + la conséquence est inattendue", ainsi que les prédictions sur l'alternative la plus approuvée (et la moins intéressante) "Tom pousse la caisse + un résultat déterministe". Cependant,

le modèle n'est pas assez précis pour prédire laquelle des alternatives 2 et 3 sera la plus approuvée ou la plus intéressante. Il faudrait pour cela avoir une estimation précise des valeurs de complexité, ce qui dépend de la connaissance de l'observateur mais également de son histoire personnelle.

Annexe B

Glossaire

Narration - Récit

Adressée à une audience, une narration est constituée d'un ensemble d'événements se rapportant à un sujet et constituant un tout, c'est-à-dire qu'elle a un commencement (une phase d'exposition durant laquelle sont généralement introduits le monde fictif et les personnages principaux et qui se termine par la présentation d'un élément perturbateur), un milieu (il s'agit d'une phase durant laquelle l'essentiel de l'action du récit se produit, les personnages font face à des problèmes qu'ils tentent de résoudre ; cette phase se termine par le *climax* - un pic de tension narrative - au cours duquel la plupart des problèmes auxquels sont confrontés les protagonistes sont résolus) et une fin (c'est une phase de dénouement, les protagonistes du récits résolvent les éventuels conflits narratifs non encore résolus, le monde fictif retrouve une forme de 'stabilité', la tension narrative diminue et le récit s'achève). Les événements qui composent une narration peuvent être le fait d'acteurs (ce sont alors des actions) ou non.

Pertinence narrative

C'est l'ensemble des critères qui contribuent à donner de la valeur, sur le plan narratif, à un récit. Cela regroupe à la fois des critères structurels, c'est-à-dire des critères qui contribuent à renforcer la qualité du récit du point de vue de la structure des événements qui le composent, ainsi que des critères de pertinence émotionnelle, c'est-à-dire des critères qui contribuent à rendre le récit plus

intéressant.

Vraisemblance

Le monde fictif doit être vraisemblable, c'est-à-dire cohérent et crédible. La cohérence renvoie directement au respect des règles causales (les événements qui font évoluer l'état du monde d'un état antérieur à un état futur) lorsqu'un événement (qu'il soit ou non le fait d'un acteur) est introduit dans le récit. Ces règles causales sont introduites par l'auteur ou peuvent provenir du genre de la narration. Tout en maintenant la cohérence du récit, l'auteur doit maintenir une forme de crédibilité, en particulier au niveau des personnages. La cohérence et la crédibilité permettent à l'audience du récit d'être absorbés dans le récit. En particulier, la crédibilité des personnages repose en grande partie sur la capacité d'une audience à reconnaître leurs intentions.

Intentionnalité

Un modèle classique en psychologie affirme qu'un personnage qui agit de manière intentionnelle doit nécessairement désirer la réalisation d'un certain état de fait (désir de résultat) et avoir un niveau de connaissance suffisant qui lui permet de déterminer comment son action peut mener à cet état de fait (Forguson, 1989). Ce modèle a été complété par Shaver (1985) qui a proposé une définition plus élaborée : une action est intentionnelle si l'agent (acteur) désire un résultat, est conscient des conséquences de son action ainsi que de sa capacité à pouvoir réaliser ladite action. Le modèle de Malle & Knobe (1997) détaille quant à lui cinq composants nécessaires qui permettent de déterminer si un résultat (ou un but) a été atteint de manière intentionnelle. L'agent doit désirer ce résultat, avoir une connaissance qui lui permet de déterminer que son action lui permet d'atteindre ce résultat, réaliser l'action volontairement, être conscient du fait qu'il réalise l'action au moment où il la réalise, et être apte à réaliser l'action de manière fiable.

Jugement moral et responsabilité

Nous appelons jugement moral d'un observateur à l'encontre d'un acteur pour une action la somme des responsabilités de l'acteur pour chacune des conséquences

individuelles de cette action. Le jugement moral vise à évaluer ce que l'observateur considère comme bien ou mal, c'est-à-dire ce qu'il approuve ou non.

Comme pour la notion d'intention, la capacité à prédire les conséquences d'une action, et donc les connaissances et croyances des acteurs que l'on juge responsables ou non de l'occurrence d'une situation, semblent jouer un rôle important pour déterminer le niveau de responsabilité. Cependant, bien que ces notions soient proches, il semble que l'attribution des intentions et de la responsabilité diffèrent au niveau de la sévérité des conséquences telles qu'elles sont évaluées par un observateur. Bien qu'il puisse y avoir accord entre plusieurs observateurs sur le niveau d'intentionnalité à attribuer à une action réalisée par un acteur donné, ces mêmes observateurs peuvent individuellement attribuer un niveau de responsabilité différent.

Intérêt - Tension narrative

L'intérêt d'un récit correspond à la réponse émotionnelle de l'audience confrontée à ce récit. Cette réponse émotionnelle suit généralement un 'profil de tension narrative' et augmente graduellement dans la phase intermédiaire du récit, entre l'apparition de l'élément perturbateur et le climax. La réponse émotionnelle face à un récit peut provenir de la référence à de grands 'thèmes' tels que la mort, le sexe ou encore la politique (Kintsch, 1980). Cette réponse émotionnelle vient également du caractère inattendu des situations présentées à l'audience (Kintsch, 1980). Parmi les modalités classiques de l'intérêt dans le cadre narratif, on peut citer la surprise, la curiosité ou encore le suspense (Baroni, 2007).

Conflit narratif

Le conflit narratif est défini comme l'incompatibilité entre les objectifs de deux personnages ou forces. Au cœur de la notion de conflit narratif se trouve l'idée qu'un ensemble d'actions intentionnelles de personnages (planifiées dans le but d'être réalisées) pour atteindre un ou plusieurs objectifs se trouvent contrecarrées par des événements inattendus (des obstacles). Ces événements peuvent être des actions d'autres personnages ou non. Il est classique de distinguer plusieurs sortes de conflits en fonction des forces en présence. Classiquement, on distingue trois

formes de conflits narratifs :

1. le conflit entre personnages, c'est-à-dire un conflit qui oppose au moins deux protagonistes dont les objectifs sont contradictoires. Un exemple typique est le conflit "héros-*nemesis*" (Batman contre le Joker ou encore l'affrontement final entre les trois personnages principaux dans "Le Bon, la Brute et le Truand").
2. le conflit entre un personnage et la nature (c'est-à-dire entre un personnage et des éléments non contrôlés par une entité rationnelle appartenant au monde fictif). Dans ce cas, ce sont des *world events* qui viennent perturber la réalisation des plans d'un ou plusieurs protagonistes du récit (un exemple typique est l'arrivée d'une tornade dans le film Twister).
3. le conflit interne à un personnage. Dans ce type de conflit, le personnage est opposé à lui-même, il doit faire face à des incohérences internes. Un exemple classique de conflit est le dilemme moral. On peut citer comme exemple ce dilemme moral classique : un protagoniste doit décider s'il se met en danger pour sauver un autre protagoniste.

Les récits sont généralement organisés autour d'un conflit narratif principal dont la présentation marque la fin de la première phase du récit et dont la résolution marque le passage à la troisième phase du récit (c'est une façon de redéfinir le critère d'unité, l'ensemble du récit est organisé autour d'un conflit narratif principal). Ainsi la structure en trois actes peut être résumée ainsi :

avant le conflit – du conflit à sa résolution – après la résolution du conflit

Génération automatique de scénario

Le domaine de la génération automatique de scénarios est une branche de l'intelligence artificielle dont le but est de développer, à partir d'un contexte initial donné, des algorithmes qui produisent des narrations. Le mot narration, dans ce contexte, a souvent été réduit aux aspects structurels définis plus haut. Ainsi, comme le souligne Pablo Gervàs (Gervàs, 2013) "*The underlying concept of "story" in SGAs (Story Generator Algorithms) is functional and does not imply any aesthetic notion. This is important because it sets the context for evaluation of generated*

stories, for which having a surface realization as a readable and appealing text is not necessarily a core issue".

L'objectif principal de la génération de scénarios a, jusqu'ici, été la génération de *fabula* structurellement pertinentes, c'est-à-dire des suites d'événements cohérents (au niveau causal et au niveau global de l'intrigue) et crédibles. Pour y parvenir, diverses approches ont été explorées, qui peuvent être classées en trois catégories :

— les approches centrées sur la fonction de l'auteur
— les approches centrées sur la personnalité des personnages
— les approches hybrides.

Les aspects esthétiques de la narration, en particulier la génération d'histoires intéressantes, ont été souvent négligés. Il y a néanmoins eu des tentatives de contrôler la réponse émotionnelle. Nous verrons que plusieurs générateurs implémentent des méthodes *ad hoc* pour générer des récits intéressants.

Dilemme vraisemblance - intérêt

L'auteur d'un récit est confronté à un dilemme que nous avons appelé dilemme vraisemblance - intérêt (Saillenfest and Dessalles, 2014a) :

Dilemme vraisemblance - intérêt : l'auteur est tiraillé entre deux objectifs parfois contraires que constituent le maintien de la vraisemblance du monde fictif (et des personnages qui le peuplent) et la nécessité de créer de la tension narrative tout au long de la narration.

Émotions ressenties par procuration

Nous pensons que l'audience est capable de prendre le point de vue d'un protagoniste du récit (*via* une projection dans l'esprit du protagoniste, une forme primaire de théorie de l'esprit) ce qui permet à l'audience d'utiliser l'évaluation de l'état du monde de ce protagoniste et ainsi de ressentir les émotions ressenties par ce protagoniste.

Postdictibilité

Kintsch (1980) définit ainsi la postdictibilité :

"*The text as a whole must hang together and make sense to the reader, so that he is able to construct a coherent macrostructure in which each text unit has its place and is meaningfully related to other sections of the text. Weird and unpredictable statements in a text are interesting only in so far as they are well motivated within the text as a whole, at least by hindsight.*" (Kintsch, 1980, p.89)

Ce critère est est un critère de restauration de la vraisemblance du monde fictif. C'est l'une des différences importantes qui peut exister entre le récit de faits réels et une fiction. Dans le monde réel, un événement non-expliqué nous met dans une situation d'inconfort mais est intéressant à raconter. Dans le cadre d'une fiction, le critère de postdictibilité oblige l'auteur à fournir une explication à toute situation présentée à l'audience avant la clôture du récit afin de ne pas laisser l'audience dans une situation d'inconfort à la fin du récit.

Inattendu

Dans un sens général, sont inattendus les événements perçus comme rares (ou de faible probabilité d'occurrence), les attentes non-confirmées ou les situations non-anticipées. L'inattendu est une notion subjective, la probabilité d'occurrence d'un événement est nécessairement une probabilité subjective d'occurrence, c'est-à-dire qu'un événement perçu comme rare par un individu peut ne pas être perçu comme tel par un autre individu. Maguire et al. (2011) considèrent que l'inattendu d'une situation (l'intensité de la surprise associée à l'occurrence d'une situation) dépend du 'niveau de difficulté' pour intégrer cette situation dans une représentation existante du monde.

La théorie de la Simplicité (Dessalles, 2006) propose un modèle formel de calcul de l'inattendu comme différence de complexité cognitive entre une complexité attendue de génération (qui correspond à la taille (exprimée en nombre de *bits*) de la meilleure hypothèse causale pour générer la situation compte-tenu de notre connaissance sur l'état du monde et des règles causales qui régissent son évolution) et une complexité observée (qui correspond à la taille (exprimée en nombre de *bits*)

de la meilleure description de la situation compte-tenu des connaissances de l'observateur). Cette théorie réalise, entre autres, de bonnes prédictions quant à l'attitude des individus confrontés à certains types d'événements (ou de situations) jugées inattendues : les coïncidences, les tirages de loterie remarquables, les situations rares, les situations atypiques , les rencontres fortuites, etc.

Complexité de Kolmogorov

La complexité de Kolmogorov d'un objet peut être définie ainsi : la complexité d'un objet est la taille du plus petit programme qui permet de le produire.

Formellement : soit X un objet observé en sortie d'une machine de Turing M_U. On note $l(p)$ la taille d'un programme p, $M_U(p)$ la réalisation du programme p au travers de la machine M_U et $K(X)$ la complexité de Kolmogorov de l'objet X.

$$K(X) = min_p(l(p)|M_U(p) = X)$$

L'un des résultats fondamentaux relatif à la complexité de Kolmogorov est le théorème d'invariance qui dit que la taille de la description minimale d'un objet est invariante (à une constante près) entre les différents langages universels (Li and Vitányi, 2008, p.104-105). On peut donc parler de la complexité de Kolmogorov d'un objet indépendamment du langage utilisé pour le décrire.

Un autre résultat classique concernant la complexité de Kolmogorov est sa non-calculabilité.

Complexité cognitive

La complexité joue un rôle important en cognition, en particulier au niveau des processus perceptifs (dans ce cadre on parle de complexité cognitive, c'est une notion dérivée de la notion de complexité de Kolmogorov). L'être humain a une propension à considérer le objets les plus simples à décrire (*ie.* les moins complexes) comme les plus probables.

La non-calculabilité de la complexité de Kolmogorov est souvent utilisée pour affirmer que l'être humain n'a pas accès à la complexité des objets auxquels il est confronté. Cette conclusion hâtive a poussé certains chercheurs à abandonner la notion de complexité au profit de l'inférence statistique (Griffiths and Tenenbaum,

2003) ou se sont tournés vers des mesures alternatives et calculables de complexité telles que la reconnaissance de motifs ou la complexité booléenne (Simon, 1972; Feldman, 2004).

Cependant, nous pouvons conserver la notion de complexité en considérant que l'humain à des ressources limitées. L'être humain possède une puissance de calcul qui lui permet, entre autres, d'identifier la répétition de motifs, ce qui lui permet ainsi de réaliser des opérations de compression face aux situations rencontrées. D'une manière plus générale, toute forme de reconnaissance de structures permet de réaliser une compression. Dès lors que nous disposons d'un modèle cognitif qui représente ces mécanismes, nous sommes en mesure de calculer la taille de la meilleure compression que peut réaliser un individu en un temps donné. On peut ainsi conserver la notion de complexité, en lui donnant ce sens de taille de la meilleure compression réalisée en un temps fini et en utilisant un modèle cognitif donné. En lui donnant ce sens, la complexité devient une notion calculable.

Dans le cadre de la théorie de la Simplicité (Dessalles, 2006), l'inattendu d'une situation est défini comme une différence entre une complexité attendue (dite de génération) et une complexité de description.

La complexité de génération correspond à la complexité de la situation, considérée comme résultat de l'exécution d'un programme 'causal' à travers une machine-monde (*W-machine*), c'est-à-dire à la taille de la plus courte hypothèse (causale) permettant de générer la situation.

La complexité de description d'une situation correspond à la complexité de la situation, considérée comme étant le résultat de l'exécution d'un programme 'descriptif' à travers une machine-observation (*O-machine*), c'est-à-dire à la taille de la plus courte description de la situation.

Nécessité d'un état

La nécessité d'un état est un modèle d'évaluation des états basé sur les attitudes que nous avons envers ces états, c'est-à-dire l'ensemble des croyances et des désirs associés à ces états (et leurs correspondants négatifs). Chaque nécessité a une intensité et une valence.

L'évaluation des nécessités peut provenir de sources diverses (observation di-

recte, croyance antérieure, inférence, désir, etc.), ce qui peut donner naissance à des conflits d'ordre cognitifs.

Conflit cognitif

Un conflit cognitif est un conflit entre (au moins) deux nécessités de signes opposés d'un même état de fait (Saillenfest and Dessalles, 2014b).

On peut citer les cas de conflits courants (ou classiques) suivants :

1. un état désiré mais dont la non-occurrence est crue (ou inférée)
2. un état non-désiré mais dont l'occurrence est crue (ou inférée)
3. un état de fait dont l'occurrence ainsi que la non-occurrence sont crues (ou inférées).

Nous proposons dans ce manuscrit qu'un conflit narratif peut être assimilé à un ensemble de conflits d'ordre cognitifs à l'échelle des protagonistes rationnels qui sont impliqués dans ledit conflit narratif.

Abduction

L'abduction, telle qu'elle est définie par Peirce (CP 5.189) consiste à générer une hypothèse qui explique une observation. Peirce décrit ainsi le fonctionnement de l'abduction :

> "*The surprising fact, C, is observed ; But if A were true, C would be a matter of course, Hence, there is reason to suspect that A is true.*" (CP 5.189)

Autrement dit, l'abduction est un raisonnement qui prend comme point de départ des conséquences observées et tente de remonter aux causes. Pour reprendre les mots de Sylvie Catellin :

> « Le point de départ de l'abduction est un fait ou un phénomène perçu comme surprenant, inattendu, ou s'inscrivant contre ce qui était jusqu'alors tenu pour acquis. [...] l'abduction est une procédure de normalisation d'un fait aberrant, un effort de raisonnement que l'on entreprend lorsqu'il y a rupture de notre système d'attentes, un raisonnement imaginatif qui fait appel à nos connaissances implicites. » (Catellin, 2014, p.73-74)

La procédure CAN (Conflit-Abduction-Négation) (Saillenfest and Dessalles, 2014b; Dessalles, 2008a) est une procédure de résolution de conflits cognitifs *via* l'abduction. C'est une procédure de détermination d'une hypothèse explicative qui permet de faire baisser l'intensité d'un conflit cognitif.

Bibliographie

Aristote (b335). *Poétique*. Le Livre de Poche. (env. 335 avant JC, traduction par Michel Magnien de 1990).

Aristote (nc). *Éthique à Nicomaque*. Flammarion. (traduction et présentation par Richard Bodéüs, 2004).

Bae, B.-C. and Young, R. M. (2014). A computational model of narrative generation for surprise arousal. *Computational Intelligence and AI in Games, IEEE Transactions on*, 6(2) :131–143.

Bailey, P. (1999). Searching for storiness : Story-generation from a reader's perspective. In *Working notes of the Narrative Intelligence Symposium*, pages 157–164.

Bal, M. (1997). *Narratology : Introduction to the theory of narrative*. University of Toronto Press.

Baron-Cohen, S. (1999). *The descent of mind : Psychological perspectives on hominid evolution.*, chapter The evolution of a theory of mind., pages 261–277. New York, NY, US : Oxford University Press.

Baroni, R. (2007). *La tension narrative : suspense, curiosité et suprise*. Seuil.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, 37(7) :122–125.

Beaugrande, R. and Colby, B. N. (1979). Narrative models of action and interaction. *Cognitive Science*, 3(1) :43–66.

Bleske-Rechek, A., Nelson, L., Baker, J., and Brandt, M. R. S. (2010). Evolution and the trolley problem : people save five over one unless the one is young, genetically related, or a romantic partner. *Journal of Social, Evolutionary, and Cultural Psychology*, 4(3) :115–127.

Borel, É. (1913). La mécanique statique et l'irréversibilité. *Journal de Physique Théorique et Appliquée*, 3(1) :189–196.

Bortolussi, M. and Dixon, P. (2003). *Psychonarratology, Foundations for the empirical study of literary response*. Cambridge University Press.

Boulton, M. (1975). *Anatomy of the Novel*. Routledge & Kegan Paul.

Bremond, C. (1966). La logique des possibles narratifs. *Communications*, 8(1) :60–76.

Brewer, W. F. (1996). *Suspense : Conceptualizations, theoretical analyses, and empirical explorations*, chapter The nature of narrative suspense and the problem of rereading, pages 107–127. Routledge.

Catellin, S. (2014). *Sérendipité : du conte au concept*. Seuil.

Chater, N. (1999). The search for simplicity : A fundamental cognitive principle ? *The Quarterly Journal of Experimental Psychology*, 52A(2) :273–302.

Chater, N. and Brown, G. D. (2008). From universal laws of cognition to specific cognitive models. *Cognitive Science*, 32(1) :36–67.

Chater, N. and Vitányi, P. M. B. (2003). Simplicity : a unifying principle in cognitive science ? *TRENDS in Cognitive Sciences*, 7(1) :19–22.

Chatman, S. B. (1993). *Reading Narrative Fiction*. Macmillan Publishing Company, New York.

Cheong, Y.-G. and Young, R. M. (2015). Suspenser : A story generation system for suspense. *Computational Intelligence and AI in Games, IEEE Transactions on*, 7(1) :39–52.

Chion, M. (1985). *Écrire un scénario*. Cahiers du cinéma - I.N.A.

Chockler, H. and Halpern, J. Y. (2004). Responsibility and blame : a structural-model approach. *Journal of Artificial Intelligence Research*, 22(1) :93–115.

Cohen, J. (2001). Defining identification : A theoretical look at the identification of audiences with media characters. *Mass Communication & Society*, 4(3) :245–264.

Cushman, F., Young, L., and Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment. *Psychological Science*, 17(12) :1082–1089.

Dehn, N. (1981). Story generation after tale-spin. In *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 1*, IJCAI'81, pages 16–18, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Dessalles, J.-L. (2006). A structural model of intuitive probability. In D. Fum, F. D. M. and Stocco, A., editors, *Proceedings of the seventh International Conference on Cognitive Modelling*, pages 86–91, Trieste, IT. Edizioni Goliardiche.

Dessalles, J.-L. (2008a). A computational model of argumentation in everyday conversation : A problem-centred approach. In *Proceedings of the 2008 Conference on Computational Models of Argument (COMMA 2008)*, pages 128–133, Amsterdam, The Netherlands, The Netherlands. IOS Press.

Dessalles, J.-L. (2008b). *La pertinence et ses origines cognitives : nouvelles théories*. Lavoisier.

Eco, U. (1992). *Les limites de l'interprétation*. Grasset.

Feldman, J. (2004). How surprising is a simple pattern ? quantifying "eureka !". *Cognition*, 93(3) :199–224.

Fikes, R. E. and Nilsson, N. J. (1972). Strips : A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3) :189–208.

Fish, S. (1989). Being interdisciplinary is so very hard to do. *Profession*, pages 15–22.

Forguson, L. (1989). *Common sense*. Cambridge Univ Press.

Freytag, G. (1894). *Freytag's technique of the drama : an exposition of dramatic composition and art*. Scott, Foresman.

Ganascia, J.-G. and Debru, C. (2007). Cybernard : A computational reconstruction of claude bernard's scientific discoveries. In *Model-Based Reasoning in Science, Technology, and Medicine*, pages 497–510. Springer.

Gerrig, R. J. (1989). Reexperiencing fiction and non-fiction. *Journal of Aesthetics and Art Criticism*, 47(3) :277–280.

Gerrig, R. J. (1993). *Experiencing narrative worlds : On the psychological activities of reading*. Yale University Press.

Gervàs, P. (2009). Computational approaches to storytelling and creativity. *The AI magazine*, 30(3) :49–62.

Gervàs, P. (2013). *the living handbook of narratology*, chapter Story Generator Algorithms. Hamburg : Hamburg University.

Graesser, A., Singer, M., and Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3) :371–375.

Graesser, A. C., Lang, K. L., and Roberts, R. M. (1991). Question answering in the context of stories. *Journal of Experimental Psychology : General*, 120(3) :254–277.

Green, M. C., Brock, T. C., and Kaufman, G. F. (2004). Understanding media enjoyment : The role of transportation into narrative worlds. *Communication Theory*, 14(4) :311–327.

Griffiths, T. L. and Tenenbaum, J. B. (2003). Probability, algorithmic complexity, and subjective randomness. In *Proceedings of the 25th annual conference of the cognitive science society*, pages 480–485.

Guglielmo, S. and Malle, B. F. (2010). Can unintended side effects be intentional ? resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin*, 36(12) :1635–1647.

Habib, B. and Ganascia, J.-G. (2009). The reasoning process underlying claude bernard's scientific discoveries. In *Proceedings of the IJCAI-09 Workshop on Abductive and Inductive Knowledge Development (AIAI09) of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*, pages 48–53. Pasadena, California, USA.

Hart, H. L. A. and Honoré, T. (1985). *Causation in the law*. Clarendon Press ; Oxford University Press, Oxford ; New York.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational research*, 60(4) :549–571.

Hobbs, J. R., Stickel, M. E., Appelt, D. E., and Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63(1–2) :69 – 142.

Hochman, B. (1985). *Character in literature*. Cornell University Press Ithaca.

Johnson, J. T. and Drobny, J. (1985). Proximity biases in the attribution of civil liability. *Journal of Personality and Social Psychology*, 48(2) :283–296.

Jones, T. M. (1991). Ethical decision making by individuals in organizations : An issue-contingent model. *The Academy of Management Review*, 16(2) :366–395.

Kintsch, W. (1980). Learning from text, levels of comprehension, or : Why anyone would read a story anyway. *Poetics*, 9(1-3) :87 – 98.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(279) :190–194.

Lagnado, D. A. and Channon, S. (2008). Judgments of cause and blame : The effects of intentionality and foreseeability. *Cognition*, 108(3) :754 – 770.

Lavandier, Y. (2004). *La dramaturgie*. Le Clown & l'enfant.

Lebowitz, M. (1985). Story-telling as planning and learning. *Poetics*, 14(6) :483–502.

Li, M. and Vitányi, P. (2008). *An Introduction to Kolmogorov Complexity and its Applications*. Springer-Verlag, 3rd edition.

Lipsky, E. (1956). *The Mystery Writer's Handbook*, chapter Suspense, pages 103–112. New York : Harper.

Lucas, G. and Meyjes, M. (1989). Indiana jones and the last crusade.

Macedo, L., Cardoso, A., Reisenzein, R., Lorini, E., and Castelfranchi, C. (2009). Artificial surprise. *Handbook of research on synthetic emotions and sociable robotics : New applications in affective computing and artificial intelligence*, pages 267–291.

Magnani, L. (2001). *Abduction, Reason, and Science*. Kluwer Academic/Plenum Publishers.

Maguire, P., Moser, P., Maguire, R., and Keane, M. (2013). A computational theory of subjective probability [featuring a proof that the conjunction effect is not a fallacy]. In *Proceedings of the Thirty-Fifth Annual Conference of the Cognitive Science Society*, pages 960–965.

Maguire, R., Maguire, P., and Keane, M. T. (2011). Making sense of surprise : an investigation of the factors influencing surprise judgments. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 37(1) :176–186.

Malle, B. F. (2010). Intentional action in folk psychology. *Blackwell companion to the philosophy of action. Oxford, UK : Blackwell.*

Malle, B. F. and Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33(2) :101–121.

Malle, B. F. and Nelson, S. E. (2003). Judging mens rea : the tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences & the Law*, 21(5) :563–580.

Mateas, M. and Sengers, P. (1999). Narrative intelligence. In *Proceedings AAAI Fall Symposium on Narrative Intelligence*, pages 1–10.

Mateas, M. and Stern, A. (2003). Façade : An experiment in building a fully-realized interactive drama. In *Game Developers Conference*, volume 2.

Meehan, J. R. (1977). Tale-spin, an interactive program that writes stories. In *Proceedings of the 5th international joint conference on Artificial intelligence - Volume 1*, IJCAI'77, pages 91–98, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Meyer, W.-U., Reisenzein, R., and Schützwohl, A. (1997). Toward a process analysis of emotions : The case of surprise. *Motivation and Emotion*, 21(3) :251–274.

Miletitch, R., Sabouret, N., and Ochs, M. (2012). Susciter l'émotion dans la narration automatique. *Technique et Science Informatiques*, 31(4) :477–501.

Miller, D. T. and Gunasegaram, S. (1990). Temporal order and the perceived mutability of events : Implications for blame assignment. *Journal of Personality and Social Psychology*, 59(6) :1111–1118.

N'gbala, A. and Branscombe, N. R. (1995). Mental simulation and causal attribution : When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology*, 31(2) :139–162.

Norrick, N. R. (2000). *Conversational Narrative : Storytelling in everyday talk*. John Benjamins Publishing Company.

Ochs, M., Sabouret, N., and Corruble, V. (2009). Simulation of the dynamics of nonplayer characters' emotions and social relations in games. *Computational Intelligence and AI in Games, IEEE Transactions on*, 1(4) :281–297.

Ortony, A. and Partridge, D. (1987). Surprisingness and expectation failure : what's the difference ? In *Proceedings of the 10th international joint confe-*

rence on Artificial intelligence-Volume 1, pages 106–108. Morgan Kaufmann Publishers Inc.

O'Neill, B. and Riedl, M. (2014). Dramatis : A computational model of suspense. In *Proceedings of the 28th AAAI conference on artificial intelligence*, pages 944–950.

Peirce, C. S. (1974). *Collected papers of charles sanders peirce*. Harvard University Press.

Pérez y Pérez, R. and Sharples, M. (2001). Mexica : A computer model of a cognitive account of creative writing. *Journal of Experimental & Theoretical Artificial Intelligence*, 13(2) :119–139.

Pérez y Pérez, R. and Sharples, M. (2004). Three computer-based models of storytelling : Brutus, minstrel and mexica. *Knowledge-Based Systems*, 17(1) :15 – 29.

Phares, E. J. and Wilson, K. G. (1972). Responsibility attribution : Role of outcome severity, situational ambiguity, and internal-external control. *Journal of Personality*, 40(3) :392–406.

Pizarro, D. A., Uhlmann, E., and Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39(6) :653 – 660.

Prince, G. (1987). *A Dictionary of Narratology*. University of Nebraska Press.

Rattermann, M. J., Spector, L., Grafman, J., Levin, H., and Harward, H. (2001). Partial and total-order planning : evidence from normal and prefrontally damaged populations. *Cognitive Science*, 25(6) :941–975.

Reisenzein, R. (2000a). Exploring the strength of association between the components of emotion syndromes : The case of surprise. *Cognition & Emotion*, 14(1) :1–38.

Reisenzein, R. (2000b). The subjective experience of surprise. *The message within : The role of subjective experience in social cognition and behavior*, pages 262–279.

Riedl, M. and Young, R. (2004). An intent-driven planner for multi-agent story generation. In *Proceedings of the Third International Joint Conference on*

Autonomous Agents and Multiagent Systems-Volume 1, pages 186–193. IEEE Computer Society.

Riedl, M. O. and Young, R. M. (2005). An objective character believability evaluation procedure for multi-agent story generation systems. In Panayiotopoulos, T., Gratch, J., Aylett, R., Ballin, D., Olivier, P., and Rist, T., editors, *Intelligent Virtual Agents*, volume 3661 of *Lecture Notes in Computer Science*, pages 278–291. Springer Berlin Heidelberg.

Riedl, M. O. and Young, R. M. (2010). Narrative planning : Balancing plot and character. *Journal of Artificial Intelligence Research*, 39(1) :217–268.

Rumelhart, D. E. (1975). Notes on a schema for stories. *Representation and understanding : Studies in cognitive science*, pages 211–236.

Rumelhart, D. E. (1980). *Theoretical Issues in Reading Comprehension*, chapter Schemata : The building blocks of cognition., pages 33–58. Hillsdale, NJ : Lawrence Erlbaum Associates.

Saillenfest, A. and Dessalles, J.-L. (2012). Role of kolmogorov complexity on interest in moral dilemma stories. In Miyake N., P. D. and R., C., editors, *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pages 947–952. Austin, TX : Cognitive Science Society.

Saillenfest, A. and Dessalles, J.-L. (2013). Using unexpected simplicity to control moral judgments and interest in narratives. In Finlayson, M. A., Fisseni, B., Löwe, B., and Meister, J. C., editors, *2013 Workshop on Computational Models of Narrative*, volume 32 of *OpenAccess Series in Informatics (OASIcs)*, pages 214–227, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Saillenfest, A. and Dessalles, J.-L. (2014a). Can believable characters act unexpectedly ? *Literary and Linguistic Computing*, 29(4) :606–620.

Saillenfest, A. and Dessalles, J.-L. (2014b). A cognitive approach to narrative planning with believable characters. In Mark A. Finlayson, Jan Christoph Meister, E. G. B., editor, *2014 Workshop on Computational Models of Narrative*, volume 41 of *OpenAccess Series in Informatics (OASIcs)*, pages 177–181, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Saillenfest, A. and Dessalles, J.-L. (2015). Some probability judgments may rely on complexity assessments. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., and Maglio, P. P., editors, *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, pages 2069–2074. Austin, TX : Cognitive Science Society.

Schank, R. C. and Abelson, R. P. (1977). *Scripts, plans, goals and understanding : an inquiry into human knowledge structures. Hillsdale, NJ : L.* Erlbaum.

Schlenker, B., Britt, T., Pennington, J., Murphy, R., and Doherty, K. (1994). The triangle model of responsibility. *Psychological Review*, 101(4) :632–652.

Sharples, M. (1999). *How we write : Writing as creative design*. Psychology Press.

Shaver, K. G. (1985). *The attribution of blame : causality, responsibility, and blameworthiness*. Springer-Verlag, New York.

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological review*, 79(5) :369–382.

Smith, T. C. and Witten, I. H. (1991). A planning mechanism for generating story text. *Literary and Linguistic Computing*, 6(2) :119–126.

Sternberg, M. (1992). Telling in time (ii) : Chronology, teleology, narrativity. *Poetics Today*, 13(3) :463–541.

Szilas, N. (2007). A computational model of an intelligent narrator for interactive narratives. *Applied Artificial Intelligence*, 21(8) :753–801.

Szilas, N. and Richle, U. (2013). Towards a computational model of dramatic tension. In Finlayson, M. A., Fisseni, B., Löwe, B., and Meister, J. C., editors, *2013 Workshop on Computational Models of Narrative*, volume 32 of *OpenAccess Series in Informatics (OASIcs)*, pages 257–276, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Teigen, K. H. and Keren, G. (2003). Surprises : low probabilities or high contrasts ? *Cognition*, 87(2) :55 – 71.

Terrell, D. (1994). A test of the gambler's fallacy : Evidence from pari-mutuel games. *Journal of risk and uncertainty*, 8(3) :309–317.

Theune, M., Faas, S., Nijholt, A., and Heylen, D. (2003). The virtual storyteller : Story creation by intelligent agents. In *Proceeding of the Technologies for In-*

teractive Digital Storytelling and Entertainment (TIDSE) Conference, pages 204–215. Fraunhofer IRB.

Thiriot, S. and Kant, J.-D. (2008). Using associative networks to represent adopters' beliefs in a multiagent model of innovation diffusion. *Advances in Complex Systems (ACS)*, 11(02) :261–272.

Toolan, M. (2012). *Narrative : A critical linguistic introduction*. Routledge, New York.

Trabasso, T. and Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and language*, 24(5) :595–611.

Truffaut, F. and Hitchcock, A. (1975). *Le cinéma selon Hitchcock*. Seghers.

Turner, S. R. (1993). *Minstrel : a computer model of creativity and storytelling*. PhD thesis, University of California at Los Angeles.

Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning : the conjunction fallacy in probability judgment. *Psychological review*, 90(4) :293–315.

van den Broek, P. (1988). The effects of causal relations and hierarchical position on the importance of story statements. *Journal of Memory and Language*, 27(1) :1–22.

Vorderer, P. (1996). *Suspense : Conceptualizations, theoretical analyses, and empirical explorations*, chapter Toward a psychological theory of suspense, pages 233–254. Routledge.

Ware, S. G. and Young, R. M. (2010). Modeling narrative conflict to generate interesting stories. In *Proceedings of the Sixth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, pages 210–215.

Zillmann, D. (1995). Mechanisms of emotional involvement with drama. *Poetics*, 23(1) :33–51.

Modélisation Cognitive de la Pertinence Narrative en vue de l'Évaluation et de la Génération de Récits

Antoine SAILLENFEST

RÉSUMÉ : Une part importante de l'activité de communication humaine est dédiée au récit d'événements (fictifs ou non). Ces récits doivent être cohérents et intéressants pour être pertinents. Dans le domaine de la génération automatique de récits, la question de l'intérêt a souvent été négligée, ou traitée *via* l'utilisation de méthodes *ad hoc*, au profit de la cohérence des structures narratives produites. Nous proposons d'aborder le processus de création des récits sous l'angle de la modélisation quantitative de critères de pertinence narrative *via* l'application d'un modèle cognitif de l'intérêt événementiel. Nous montrerons que cet effort de modélisation peut servir de guide pour concevoir un modèle cognitivement plausible de génération de narrations.

MOTS-CLEFS : Pertinence narrative, complexité de Kolmogorov, complexité cognitive, inattendu, intérêt, crédibilité, postdictibilité, conflit cognitif, surprise, suspense, intention, responsabilité, jugement moral, génération automatique de récits, probabilité subjective, abduction

Cognitive Modeling of Narrative Relevance : Towards the Evaluation and the Generation of Stories

Antoine SAILLENFEST

ABSTRACT : *Humans devote a considerable amount of time to producing narratives. Whatever a story is used for (whether to entertain or to teach), it must be relevant. Relevant stories must be believable and interesting. The field of computational generation of narratives has explored many ways of generating narratives, especially well-formed and understandable ones. The question of what makes a story interesting has however been largely ignored or barely addressed. Only some specific aspects of narrative interest have been considered. No general theoretical framework that would serve as guidance for the generation of interesting and believable narratives has been provided. The aim of this thesis is to introduce a cognitive model of situational interest and use it to offer formal criteria to decide to what extent a story is relevant. Such criteria could guide the development of a cognitively plausible model of story generation.*

KEY-WORDS : Narrative relevance, Kolmogorov complexity, cognitive complexity, unexpectedness, interest, believability, postdictability, cognitive conflict, surprise, suspense, intentionality, responsibility, moral judgment, automatic story generation, subjective probability, abduction